Brightness Analysis in Finite Geometries:
Probing Protein Interactions in Cellular, Cell-free and Aqueous Environments

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

Patrick John Macdonald

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

Joachim D. Mueller, Advisor

December 2012

Acknowledgements

Graduate school is a marvelous and terrible experience, and I'd like to thank my fellow lab members for their help along the road. Lindsey Hillesheim, Joe Skinner and Bin Wu got me started, teaching me the ways of the lab and providing endless advice when things went wrong. They created a fun atmosphere, including music recommendations and lots of jokes that still fostered my development as a scientist.

My particular thanks go to Elizabeth Smith and Jolene Johnson. My close peers, we've provided each other with support in the low times, admiration in the high times, and nobody ever complained if you spent too much time talking to yourself. Whether as a source of research advice, a soundboard for complaints, or an opponent in an exchange of sarcastic witticisms, Elizabeth and Jolene have been crucial to my sanity as well as becoming my very good friends. My deepest gratitude to them both and my best wishes.

At the head of the lab is the dynamic combination of our advisor, Jochen Mueller and Yan Chen, premier experimentalist and research associate. I, and the rest of the lab, have been very fortunate to have such strong scientists with (usually extremely) different research approaches. Their energy, their drive, and their mentorship have shaped and refined my problem solving abilities, and have given me the preparation and experience to take on whatever comes next.

And always, at the core of things, is the support my family. Curiosity about the way things work and a drive to learn is something of a tradition in my family. I'll be a fourth generation engineer in my father's line and my mother is also a biomedical engineer. Her family has several members who "seem to know something about everything," a state of knowledge I've joked about achieving for years. Nevertheless, the kernel of that ambition is quite real, and I have been blessed to grow up in an environment where books are loved, words are revered, and it's never a bad idea to think about out how a given doohickey triggers the whatsit. And so, I affirm my deep love and respect for my parents, Frank and Aileen Macdonald, and grudgingly offer the same to my sister and brother, Jane and Kevin, for teaching me much about perseverance, drive, attention to detail, and never (literally never) passing up the opportunity to prevent my ego from getting overinflated (via squashing).

Last, but never least, are my close friends from high school and college: Caitlin, Justin and Ben; Desmond, Stephen, Matthew, Beth and Jennie. Though we are scattered about the country now, they've also been there for me. Also fortunately (from my perspective), they have also undergone painfully long careers in higher education, so that I never felt alone, and despite our struggles, there was always laugher to share.

My thanks to you all. I really couldn't have done it without you.

This work has been supported by the National Science Foundation (PHY-0346782) and the National Institutes of Health (GM64589, GM091743).

Dedication

For my grandfathers, who inspired me to be curious, to tinker, and to learn.

Abstract

Fluorescence fluctuation spectroscopy (FFS) is a powerful technique for quantitatively analyzing protein interactions. Using brightness analysis methods, we are uniquely able to measure the stoichiometry of protein complexes. FFS is particularly valuable because it allows measurements within living cells. This thesis demonstrates that measuring in very small volumes, such as *E. coli* cells, introduces a bias into the measured brightness. We show that this bias is a result of accumulative sample loss, or photodepletion, and that we can account for this effect and recover correct brightness values. Similarly, very thin samples, such as cell cytoplasm, introduce a bias due to the sample being shorter along the vertical axis than the volume of the excitation light. We introduce z-scan FFS and theory to identify and model thin samples and to recover unbiased data. Although measuring in cells is a primary strength of the FFS technique, some studies require the greater degree of experimental control afforded by solution measurements. Thus, we characterize cell-free expression solution for FFS measurements, an environment that offers increased control but permits genetic fluorescent labeling. We take advantage of this system to perform chromophore maturation experiments as a function of temperature on three common fluorescent proteins: EGFP, EYFP and mCherry. Our results prove that EGFP has fast maturation and is a good reporter for fluorescence experiments. Finally, we apply FFS and brightness analysis to the enzyme, APOBEC3G. We reveal that APOBEC3G interactions with RNA and single-stranded DNA are sequence dependent, which has important implications for the mechanism by which APOBEC3G packages itself into HIV-1 viral particles and restricts the virus to prevent infection.

Table of Contents

List of Tables vi

List of Figures vii

List of Abbreviations viii

1. Introduction 1
   1.1 Quantifying protein interactions 1
   1.2 Fluorescence fluctuation spectroscopy 1
   1.3 An historical reflection: GFP and biological fluorescence 5
   1.4 Extending brightness analysis: New sample environments 6

2. FFS 10
   2.1 Concepts of FFS 10
   2.2 Instrumentation 10
   2.3 Brightness and stoichiometry 14
   2.4 Analysis 17
      2.4.1 Photon count moments 17
      2.4.2 Photon counting histogram 17
      2.4.3 Autocorrelation analysis 18
      2.4.4 Deadtime, afterpulsing and background 19
   2.5 Apparent brightness 21
   2.6 Brightness titration 22
   2.7 Droplet sample protocol 24

3. Photodepletion 27
   3.1 Introduction 27
   3.2 Theory 28
      3.2.1 Brightness analysis 28
      3.2.2 The effect of photodepletion on brightness for a stationary signal 30
      3.2.3 The effect of photodepletion on brightness for a non-stationary signal 35
   3.3 Materials and methods 37
   3.4 Results 39
   3.5 Discussion 50

4. Z-scan FFS 53
   4.1 Introduction 53
   4.2 Demonstration of brightness bias 54
   4.3 Formulation of z-scan FFS 55

4.4 PSF model 60
4.5 Oil- and water-immersion objectives 64
4.6 Materials and methods 70
4.7 Z-scan fluorescence intensity curves of cells 71
4.8 Gamma factor 73
4.9 Brightness by z-scan FFS 76
4.10 Discussion 79

5. Chromophore Maturation and Cell-Free Expression 84
5.1 Introduction 84
5.2 Materials and methods 86
5.3 Cell-free synthesis of EGFP 88
5.4 Kinetics of chromophore maturation 93
5.5 Brightness and titration studies 97
5.6 Discussion 101

6. APOBEC 106
6.1 Introduction 106
6.2 A3G in cells 107
6.3 Materials 109
6.4 A3G in solution 112
6.5 A3G interactions with RNA 112
6.6 A3G interactions with ssDNA 119
6.7 FFS diffusion studies 123
6.8 Oligonucleotide brightness 131
6.9 Discussion 134

7. Summary 139

References 142

Appendices 153
Appendix A. APOBEC and reducing environments 153
Appendix B. Temperature-induced A3G oligomerization 154
Appendix C. Bright and Dark A3G studies 159
Appendix D. A3G dimer dissociation in cytosol analogs 165
Appendix E. Dual-color FFS of A3G-RNA 168

List of Tables

Table 5.1 EGFP brightness in different environments. 92
Table 6.1 RNA/DNA copy number in APOBEC complexes. 133
Table B.1 Labeled and mostly unlabeled A3G show temperature aggregation. 157
Table C.1 Bright & Dark A3G mixing study. 162
Table D.1 A3G dimer dissociation in cell extract. 166

List of Figures

Figure 1.1 A diffusing fluorophore gives rise to fluctuations. 2
Figure 1.2 The concept of brightness. 4
Figure 2.1 The information in fluorescence fluctuations. 12
Figure 2.2 Instrument setup. 13
Figure 2.3 Single-species and apparent brightness. 15
Figure 2.4 Autocorrelation curve. 16
Figure 2.5 Monomer-Dimer Titration 23
Figure 2.6 Protein Interaction Droplet Protocol 26
Figure 3.1 FFS data in the presence of photodepletion. 32
Figure 3.2 Conceptual picture of monomer/dimer sample photodepletion effects. 33
Figure 3.3 Brightness bias as a function of time and lost intensity fraction. 40
Figure 3.4 Brightness bias at two different fluorophore concentrations. 41
Figure 3.5 EGFP brightness of cells in the presence of photodepletion. 45
Figure 3.6 Dimeric EGFP brightness of cells in the presence of photodepletion. 46
Figure 3.7 NTF2 Oligomerization in *E. coli*. 49
Figure 4.1 Thin samples bias brightness in conventional FFS data. 56
Figure 4.2 Z-scan approach to finite sample geometries of the living cell. 57
Figure 4.3 PSF models. 63
Figure 4.4 Focal depth dependence of oil-immersion objective. 68
Figure 4.5 Focal depth dependence of water-immersion. 69
Figure 4.6 Intensity z-scans fit with mGL PSF model. 72
Figure 4.7 Gamma ratio: experiment and theory. 75
Figure 4.8 Z-scan corrected brightness analysis. 78
Figure 5.1 Cell-free protein synthesis. 91
Figure 5.2 Chromophore maturation time. 94
Figure 5.3 Temperature dependence of maturation rate coefficient *k*. 95
Figure 5.4 Droplet dilutions. 98
Figure 5.5 Cell-free NTF2. 100
Figure 6.1 A3G and A3A in cells. 108
Figure 6.2 A3G in solution. 111
Figure 6.3 Sequence dependent RNA interactions with A3G. 115
Figure 6.4 EGFP-labeled APOBEC and Test+ RNA droplet titration. 117
Figure 6.5 EGFP-labeled APOBEC and CA RNA droplet titration. 118
Figure 6.6 Sequence dependent ssDNA interactions with A3G. 121
Figure 6.7 GU ssDNA versus GT ssDNA A3G oligomerization. 122
Figure 6.8 Autocorrelation curves for A3G and RNA. 124
Figure 6.9 Diffusion studies: APOBEC-oligonucleotide interaction 128
Figure 6.10 Diffusion study binding of A3G and Test+ RNA. 129
Figure 6.11 Diffusion study binding of A3G and CA RNA. 130
Figure 6.12 CA ssDNA dissociates A3G oligomers. 138
Figure A.1 A3A and A3G in a reducing environment. 153
Figure B.1 Temperature-induced A3G aggregation. 154
Figure B.2 37°C A3G runaway aggregation. 155
Figure B.3 Temperature threshold of A3G oligomerization. 156
Figure C.1 A3G sample photobleaches as a dimer. 160
Figure C.2 Bright & Dark dimeric protein mixing. 161
Figure C.3 Compiled Bright & Dark A3G mixing data. 163
Figure C.4 Bright & Dark A3G mixtures complex with Test+ RNA. 164
Figure D.1 Low concentration cellular A3G. 165
Figure E.1 Average A3G and RNA copy numbers in A3G-RNA complex. 169

List of Abbreviations

3DG	three-dimensional Gaussian
A3G	APOBEC 3G
ACF	autocorrelation function
APD	avalanche photodiode
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
EGFP	enhanced green fluorescent protein
$EGFP_2$	dimeric EGFP
EYFP	enhanced yellow fluorescent protein
FCS	fluorescence correlation spectroscopy
FFS	fluorescence fluctuation spectroscopy
FRET	Förster resonance energy transfer
GL	Gaussian-Lorentzian
HMM	high molecular mass
LMM	low molecular mass
mGL	modified Gaussian-Lorenztian
NA	numerical aperture
NC	nucleocapsid
NTF2	nuclear transport factor 2
OPE	one-photon excitation
PCH	photon counting histogram
PSF	point spread function
SNR	signal-to-noise ratio
TPE	two-photon excitation
Vif	viral infectivity factor

1. Introduction

1.1 Quantifying protein interactions

The human being is an amazingly complex organism. As cellular and microbiological studies advance, it has become clear that cells are not simply miniature test tubes filled with "$A + B \rightleftharpoons AB$" reactions. The developing picture indicates that living organisms, at the core, are networks of biochemical assembly lines composed of large protein machines, molecular motors, and powered valves.[1] With molecular complexes of 10 or more proteins performing many critical functions, the quantitative study of protein interactions is both much more important and much more difficult. Proteins are truly the building blocks of cellular and biological function, and fluorescence microscopy has proved to be one of the most valuable tools in penetrating down to ground level to parse the complexity into comprehensible and individual protein mechanisms. At the heart of the approach is this: fluorescently-labeled proteins can be "seen." It's a way to tag and locate molecules in a discrete fashion. Fluorescently-labeled proteins can be imaged and localized as a population distributed through the cell, or tracked individually as they perform their functions. Protein interactions can be studied in real-time and their strength, frequency, and protein ingredients can be identified. The ability both to discover and to *quantify* this information makes fluorescence microscopy one of the most powerful tools for studying biological function.

1.2 Fluorescence fluctuation spectroscopy

Fluorescence fluctuation spectroscopy (FFS) is a powerful method for examining protein interactions. The technique is based on the idea that focusing laser light into a sample creates only a very small volume in which fluorescently-labeled particles are excited and observed. Fluorescent molecules diffusing through this volume give rise to signal fluctuations (Figure 1.1).[2,3]

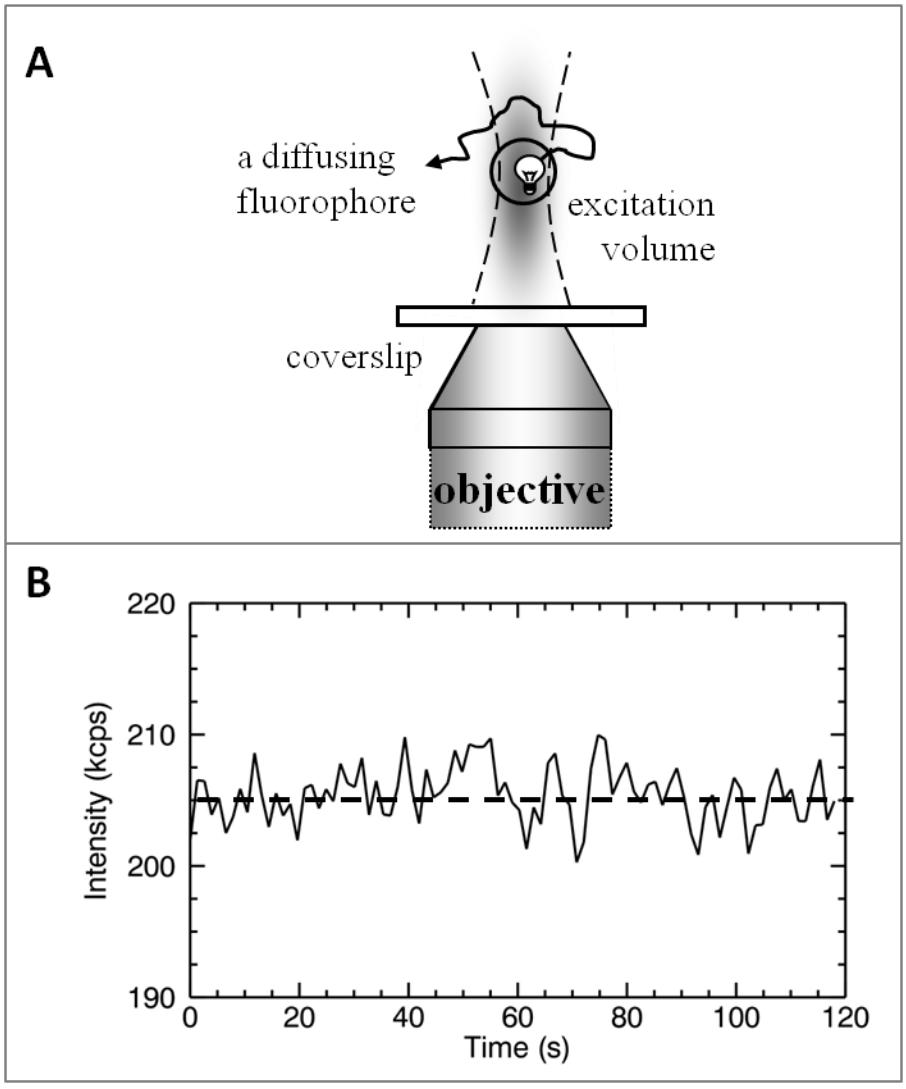

Figure 1.1 **A diffusing fluorophore gives rise to fluctuations.**

(A) Focused laser light creates a small excitation volume through which fluorescent particles diffuse. (B) An example of an intensity trace showing how the fluorescence intensity fluctuates around an average value as particles move into and out of the excitation volume. The fluctuating signal contains information about the sample that can be extracted through statistical analysis.

The averaged fluorescence intensity is stable, but as individual molecules move into and out of the excitation volume, they generate small fluctuations about that average. FFS characterizes static and dynamic properties of the sample by exploiting the information embedded in the fluctuating signal.[4,5] We will discuss some of the various statistical methods for extracting that information in the next chapter.

One valuable parameter determined from an FFS measurement is brightness. Brightness is a measure of the fluorescence per particle, generally expressed in units of photon counts per second per molecule (cpsm). Because brightness is a property of the molecule—as opposed to intensity which is a property of the ensemble—it is a powerful tool for studying molecular interactions. Consider the fluorescent label by itself (Figure 1.2A). For a given fluorescence instrument and under a given excitation power, that label has a brightness λ. The average fluorescence intensity $\langle F \rangle$ depends on the average number of particles present in the excitation volume. This label may then be attached to a protein of interest (Figure 1.2B). If the protein exists as a monomer, then the brightness of the sample remains equal to λ. However, if the protein forms dimers, then the brightness will be equal to 2λ, because the protein complex carries two fluorophores which produce, on average, twice the signal.

Consider another example where protein X and protein Y are both tagged with the same fluorescent label. If, when performing a series of experiments, protein X alone has brightness λ, protein Y has brightness λ, and mixture of X and Y has brightness 2λ, we have direct evidence of interaction between X and Y. This experiment employed uniform, identical labels. By labeling only one protein at a time, or using labels of more than one color it becomes possible to identify more complicated stoichiometries as, for instance, X_3Y. Brightness is a straightforward parameter that lends itself extremely well to the study of protein complexes.

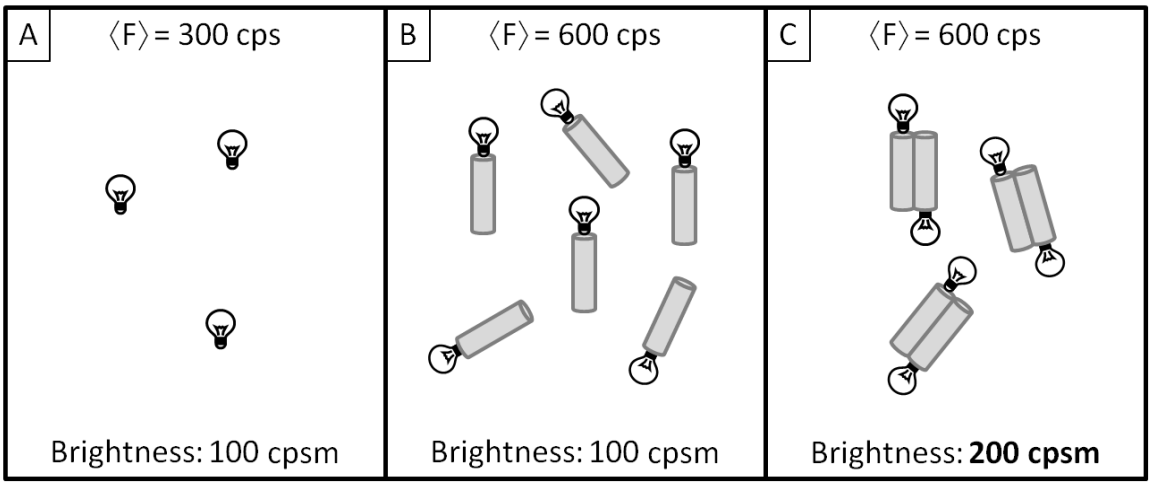

Figure 1.2 **The concept of brightness.**

(A) A depiction of three fluorescent labels in the observation volume, each with a brightness of 100 counts per second per molecule and a total average intensity of 300 counts per second. This fluorescent label can be attached to a protein of interest in order to study its behavior. Note that while the average intensity is complicated by protein concentration, brightness easily identifies (B) a monomeric and (C) a dimeric protein sample. Brightness is a straightforward and powerful tool for quantifying protein interactions.

Because of its ability to characterize protein-protein interactions, brightness analysis forms the main drive behind this thesis. One spectacular aspect of FFS and brightness analysis is that this approach can probe directly into living organisms, detecting and measuring the association of proteins in intact cells to identify interactions and stoichiometries. This information contributes to the development of a molecular understanding of the protein machinery and its role in the biological organism.[1] Brightness analysis of protein complexes in cells has been successfully demonstrated for both homocomplexes, where all proteins carry the same fluorophore,[6] and developed into dual-color analysis to study heterocomplexes.[7,8] This thesis, however, focuses almost exclusively on some of the ways single-color brightness analysis can be used to study protein interactions.

1.3 An historical reflection: GFP and biological fluorescence

In many ways it was the jellyfish that set brightness analysis on the road to success. In December of 2008, the Nobel Prize in Chemistry was awarded jointly to Osamu Shimomura, Martin Chalfie and Roger Y. Tsien "for the discovery and development of the green fluorescent protein, GFP." Shimomura identified and purified GFP and later elucidated the structure of the GFP chromophore.[9,10] Martin Chalfie applied it as a marker to study gene expression in *C. elegans*,[11] demonstrating the power of GFP for scientific study in living organisms. Finally, Roger Tsien showed that oxygen is the only external reagent necessary for GFP fluorescence, developed the brighter and more photostable enhanced green fluorescent protein (EGFP), and went on to generate a variety of fluorescent proteins of various colors, stabilities, and folding speeds, including CFP (cyan) and EYFP (yellow), as well as mCherry (red) and a set of other red-shifted fluorescent proteins termed the mFruits.[12–15]

However, it was Douglas Prasher, working at a car dealership at the time of the Nobel prize awarding, who initially had the vision of GFP as a fluorescent marker,[16] though even he may not have foreseen the full scope of its effect on biological fluorescence. After failing to obtain a new grant, he was forced to leave the field and go

in other directions, although not before inspiring Chalfie and Tsien to push forward on GFP and supplying them with the protein vector. An unlucky series of canceled projects and downsizing drove him from job to job and further away from science, the expenses of family and children forcing him to take an immediately available job. The full tragedy culminated in his position as the courtesy van driver at a Toyota dealership when he was invited to the 2008 Nobel prize ceremony—as a guest.[17] Prasher eventually returned to science, but perhaps without the full honors due him for the initial insight and the eventual wonder of the green fluorescent protein.

Wild-type GFP, EGFP, and their many variations effectively revolutionized the field of biological fluorescence. For the most part, they are monomeric labels that can be genetically-tagged to proteins in a huge variety of organisms because they don't require any additional proteins or chemicals to generate fluorescence, merely oxygen. All in all, the palette of fluorescent proteins available has proven, and continues to prove, immensely useful for cellular applications. EGFP, for instance, is an amazingly robust fluorescent protein. Shaped as a barrel with the fluorescence-generating chromophore tucked inside, EGFP is largely insensitive to changes in the local environment, provided that the pH remains stable. 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.[18] It is this property that makes EGFP such a valuable tool, and offers the scope to push brightness experiments in new directions.

1.4 Extending brightness analysis: New sample environments

We first explore the challenges of applying FFS in small sample compartments. Photodepletion, or sample attrition from excitation light photodamage, is a non-trivial problem for small sample volumes. In a related vein, we explore what happens if the sample compartment being measured is smaller than the focus of the excitation light. We provide methods and corrections to surmount both of these sample compartment challenges. The next sections move the focus away from cellular compartments to show the value of FFS for *in vitro* experiments. We demonstrate that FFS can be used in combination with cell-free expression systems to study the maturation of fluorescent

proteins, by controlling the start and stop of the expression reaction. Finally, we demonstrate the utility of FFS and brightness analysis when applied to protein-nucleotide interactions in aqueous solution. The interactions of APOBEC proteins with DNA and RNA have important biological significance, but cannot be studied directly in cells because of the lack of control over fluorescent labeling, specific DNA/RNA sequences, and other reaction parameters inside the cell.

On a broader level, developing FFS into a tool that probes protein behavior both in cell and aqueous solution environments can help us to gain insight into the differences between these two sample environments. For example, a developing issue in protein studies, is the effect of macromolecular crowding.[19] The insides of cells are dense with an thousands of assorted types of proteins, as well as DNA and RNA. This crowding has been shown to have various functional effects from changes in diffusion to altered protein binding affinity and specificity.[20–22] For example, the GroEL chaperonin protein assists protein folding in *E. coli* in a cyclic fashion, triggered by the binding and release of its cofactor GroES. It has been demonstrated that a release of partially-folded peptides between cycles does not occur in the presence of macromolecular crowding,[23] indicating that GroEL is designed for a crowded environment, and may function abnormally in simple aqueous solution. This result suggests that earlier, purely *in vitro* studies of GroEL drew some invalid conclusions regarding the immature peptide release. A vast number of protein assays are performed *in vitro*, so it is important that we be able to carefully study how the environment shapes protein behavior in order to properly relate *in vivo* and *in vitro* experiments. FFS is uniquely placed to apply the same technique to both cellular and solution environments, and to recover quantitative data on protein binding and concentration. This thesis extends FFS and develops a flexible measurement platform from which such issues as macromolecular crowding can be investigated.

Whenever a technique is pushed in new directions or into new territories, a fresh set of challenges are encountered. Here, we present a more detailed introduction for the first part of this thesis and a closer look at the challenges associated with small sample geometries. Many biophysical fields of study have a somewhat non-standard mission statement: Think small! This approach is growing in popularity as new technologies

provide access to nanoparticles, microfluidic devices and microdroplets, as well as the more naturally occurring microorganisms. Chapter 3 addresses one of the complications encountered when trying to apply FFS to such small sample volumes. Fluorescence techniques are plagued by the phenomenon of photobleaching, wherein the absorption of photons has a finite probability of destroying fluorescence instead of exciting it. Most researchers simply decrease the excitation power to a point where the effect of photobleaching is negliglible, although the need for good signal statistics imposes a lower limit on this approach. However, in very small sample volumes (~ 1 picoliter), even weak photobleaching accumulates in a process we term photodepletion, and because of the limited sample reservoir, causes the destruction of a statistically significant fraction of the sample population. FFS analysis requires that the average fluorescence intensity be stable, a condition that is not met in a gradually shrinking sample population. We adapt theory and technique to incorporate the effects of photodepletion into FFS brightness analysis and extract accurate quantitative information.

We explore a related question when the sample is small in height rather than in volume. Chapter 4 deals with thin layer sample geometries, like cell cytoplasm. When the focal volume of the excitation light is taller than the sample, the measured brightness is biased. To avoid this artifact, we introduce z-scan FFS which consists of scanning the laser along the vertical axis through the sample, followed by a stationary FFS measurement. We fit the z-scan intensity trace to a model to determine the geometry of the sample, which information can then be used to calculate an accurate brightness from the FFS data.

Beginning in chapter 5, we will return to the importance of the sample environment. Proteins operate in cells, making cellular measurements most suited to forming a true understanding of protein function. However, because precise experimental control of the cell environment is typically not feasible, it is still necessary to investigate many open questions through direct measurements in aqueous solution. In this chapter, we develop a compromise between these two approaches by employing cell-free protein expression. Cell-free expression systems essentially strip the cytoplasm and ribosomal machinery out of the cell, and so provide the researcher with an *in vitro* system that can

still transform DNA plasmids into proteins. Such a system allows direct experimental control just as with *in vitro* experiments, while maintaining the elegance and efficiency of genetic fluorescent protein labeling. We take advantage of this combination to study the chromophore maturation time, which is the time required for a fluorescent protein to go from an unfolded state to a fluorescence-generating molecule. Addressing the issues inherent in developing a new system, chapter 5 characterizes the application of FFS in cell-free expression solutions. We demonstrate that all EGFP proteins fully mature in minutes in cell-free solution, that brightness is a robust parameter specifying stoichiometry, and conclude that combining cell-free expression and FFS provides a powerful technique for quick, quantitative study of chromophore maturation and protein-protein interaction.

The final chapter uses FFS and brightness analysis to study APOBEC3G, a DNA cytosine deaminase enzyme with innate immune activity against retroviruses, notably HIV-1. We were surprised to discover through brightness analysis that the oligomerization of APOBEC3G is more complex than anticipated. The homo-interactions of the enzyme with itself are influenced by several environmental factors including temperature and solution environment, although these phenomena require further study. Most excitingly, we identify differences between APOBEC3G interactions with DNA versus RNA, as well as a sequence dependence to these interactions. These results may help to explain how the enzyme accommodates packaging into the virus and fulfills its role in restricting HIV-1.

2. FFS

In this chapter, we provide some background information about the techniques and instrumentation used throughout this work.

2.1 Concepts of FFS

FFS utilizes the intensity fluctuations of fluorescent molecules passing through a very small optical observation volume to determine concentration, diffusion coefficient, and brightness of the molecules.[24] Consider the burst of photons that is detected each time a fluorescent molecule crosses the observation volume. The frequency of the bursts provides a measure of the particle density or concentration within the observation volume. The duration of a burst correlates with the time it takes the particle to cross the observation volume, characterizing the diffusion. Lastly, the amplitude of a burst yields information about the number of fluorescent particles clustered together in a single molecule (Figure 2.1, *left panel*). This burst model is a simplified picture. In general, the number of particles present in the observation volume does not permit the identification of individual bursts (Figure 2.1, *right panel*), but statistical analysis can extract the same information. Statistical methods tools such as fluorescence correlation spectroscopy (FCS)[4,25], photon counting histogram (PCH)[26,27], and many others, are available to analyze the signal fluctuations.[28] FCS uses the correlation function to capture the temporal information of the physical process, while PCH uses the amplitude distribution of the fluctuations to characterize the concentration and brightness of each fluorescent species. Moment analysis was originally developed in the late 80s and early 90s[29–31] and is an alternative technique for studying the brightness of fluorophores.[32,33] We will introduce these techniques in more detail below.

2.2 Instrumentation

Our lab focuses primarily on two-photon excitation wherein two photons, each carrying half the necessary energy, must reach the fluorophore simultaneously in order to excite fluorescence.[34,35] Here, simultaneously means within the Heisenberg energy-time

uncertainty. To achieve this excitation, there must be a large number of photons focused both in space and in time. We use a Zeiss Axiovert 200 microscope (Thornwood, NY) and a mode-locked Ti:sapphire laser (Tsunami or MaiTai, Spectra Physics, Mountain View, CA). The laser has a pulse width of ~100 fs and a frequency of 80 MHz, and the wavelength can be tuned from 700-1000 nm. The laser is directed through a beam expander before passing into the microscope. It is then reflected up through the objective by a dichroic mirror (#740DCSPXR, Chroma, Rockingham, VT). A Zeiss 63x C-Apochromat water immersion objective (NA = 1.2) focuses the excitation light onto the sample. The resulting fluorescence light is collected by the same objective and passed through the dichroic mirror which transmits from 400-750 nm. An additional bandpass filter (FF01-750/SP-25) removes scattered laser light. In the detection compartment of the instrument, an optional dichroic mirror allows the fluorescence signal to be split by emission color into two detection channels. Each channel is equipped with an avalanche photodiode (APD) (SPCM-AQ-14, Perkin-Elmer, Dumberry, Québec) that records photon counts. The TTL pulse signal generated by the APD is connected directly to a data acquisition card (FLEX02, Correlator.com, Bridgewater, NJ or ISS, Champagne, IL). The photon counts are stored and later analyzed using programs written in IDL, version 5.4 or later (Exelis Visual Information Solutions, Boulder, CO).

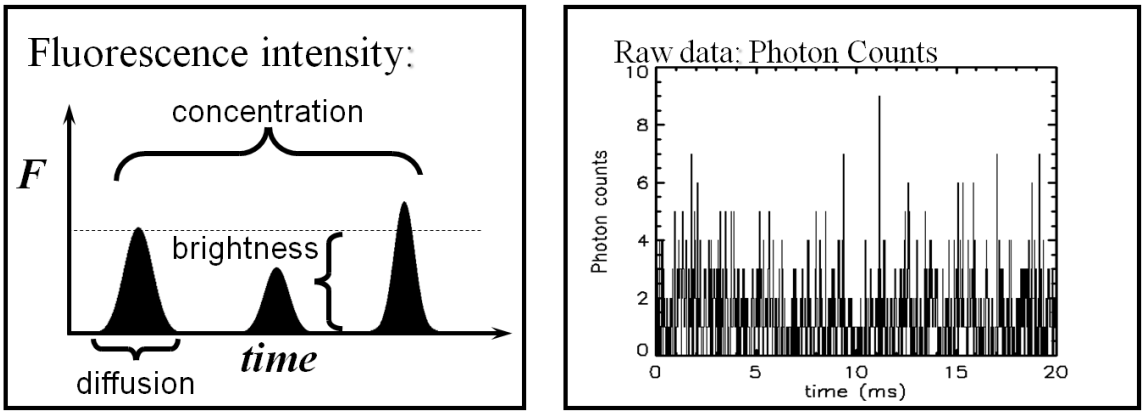

Figure 2.1 **The information in fluorescence fluctuations**.

The left panel shows a cartoon of fluorescence bursts from molecules passing through the observation volume. The frequency of the bursts provides information about concentration, the duration is correlated with the diffusion coefficient, and the amplitude provides a measure of the brightness. As the right panel shows, the cartoon of individual bursts is a simplified picture. Most FFS data have many overlapping bursts, but the same information can be extracted through statistical analysis.

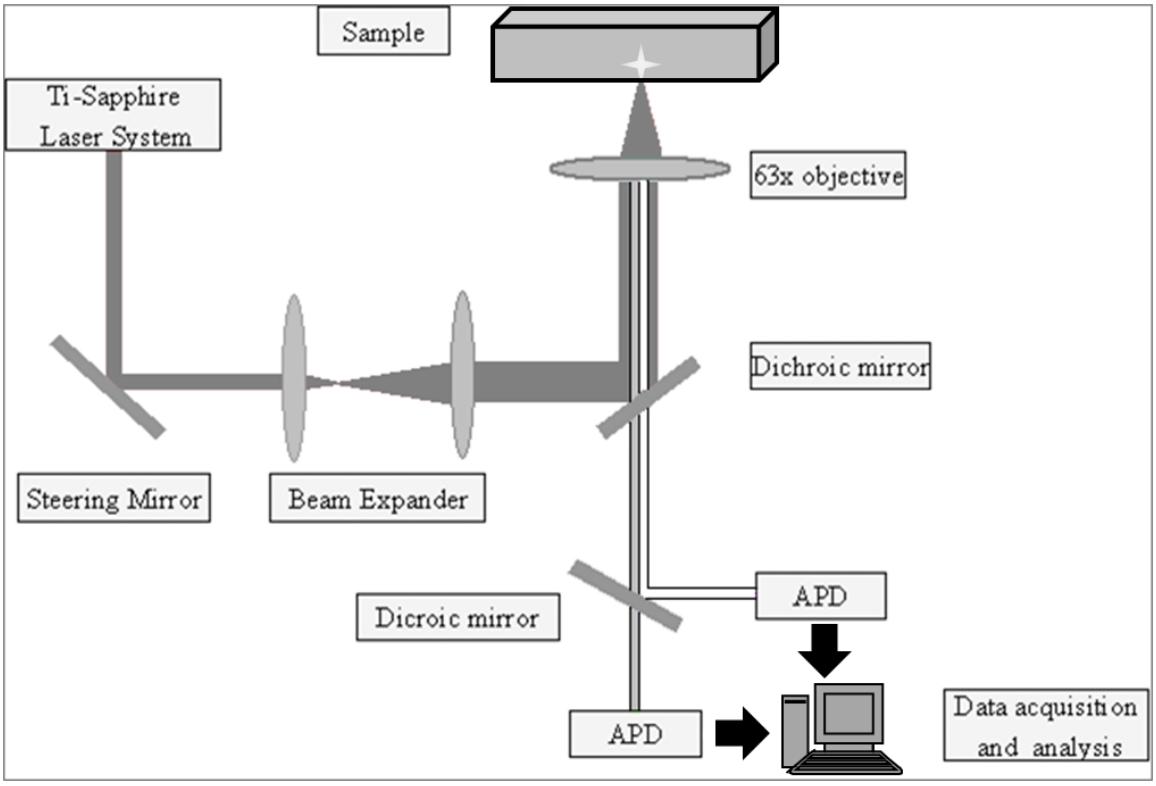

Figure 2.2 **Instrument setup.**

Pulsed laser light is steered through a beam expander onto a dichroic mirror which reflects the excitation light up through an objective onto the sample. The same objective collects the fluorescence light, which passes through the dichroic mirror down to the detection pathway. Here, a second (optional) dichroic mirror permits the color separation of the fluorescence light onto avalanche photodiode detectors. The data from the APDs is recorded to computer memory for further analysis and storage.

2.3 Brightness and stoichiometry

Brightness is a unique FFS parameter that specifies the average rate of photon counts per fluorescent particle. We focus in this section on brightness, because brightness experiments are central to this thesis. 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*. Thus, FFS is a vital tool in providing quantitative data that help clarify the molecular interactions that drive cellular processes. The relationship between brightness and stoichiometry is very straightforward. To emphasize this point and to simplify the discussion, we introduce the concept of dimensionless brightness (also referred to as normalized brightness),

$$b = \frac{\lambda_{sample}}{\lambda_{label}} \tag{2.1}$$

where λ_{sample} is the brightness of the sample and λ_{label} is the brightness of the label alone (Figure 2.3A). Note that λ in this thesis has units of photon counts per second. We will also employ the "ε" brightness notation which has units of photon counts per data bin, and is thus dependent on sampling frequency f. The two values are related by $\lambda = \varepsilon \cdot f$ in the absence of undersampling.[33]

By either definition, normalized brightness provides information about the average number of proteins in a complex. The best fluorophores for brightness experiments, e.g. EGFP, are resistant to quenching, and so are unaffected by proximity to other proteins. Thus an EGFP-labeled protein sample that forms monomers also returns $b = 1$ (Figure 2.3B). For a dimeric protein sample, $b = 2$ (Figure 2.3C), as each protein complex now contains two EGFP labels, and so on and so forth for higher order oligomers (Figure 2.3D). Nor is the approach limited to small complexes; other work from our lab uses an identical formulation to investigate self-assembling viral proteins and calculates b-values or "copy numbers" over 1000.[36] The concept of brightness analysis is very simple and very powerful, and has been successfully used to quantify biologically relevant stoichiometry in both cells and solution.[37,38]

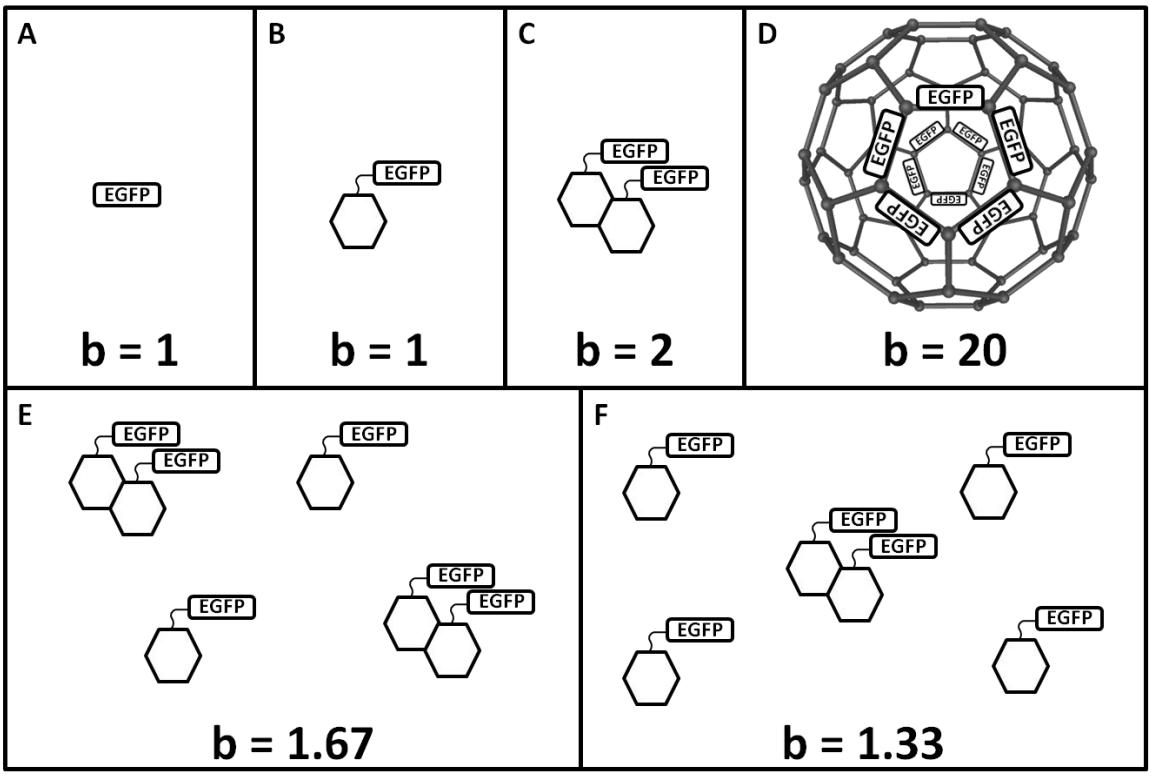

Figure 2.3 **Single-species and apparent brightness**.

Experimental brightness is normalized by establishing the inherent brightness of the monomeric fluorescent label (A). Then, the label is attached to a protein of interest and the measured normalized brightness will return (A) a value of 1 for a monomeric sample, (B) 2 for a dimeric sample, (C) etc. In the case of a mixed brightness species sample, the situation is not quite as clear cut. The simplest—and sometimes only—brightness measurement available is the average or apparent brightness. As shown in Eq. 2.10, apparent brightness returns a fractional intensity weighted average of the mixture (E & F).

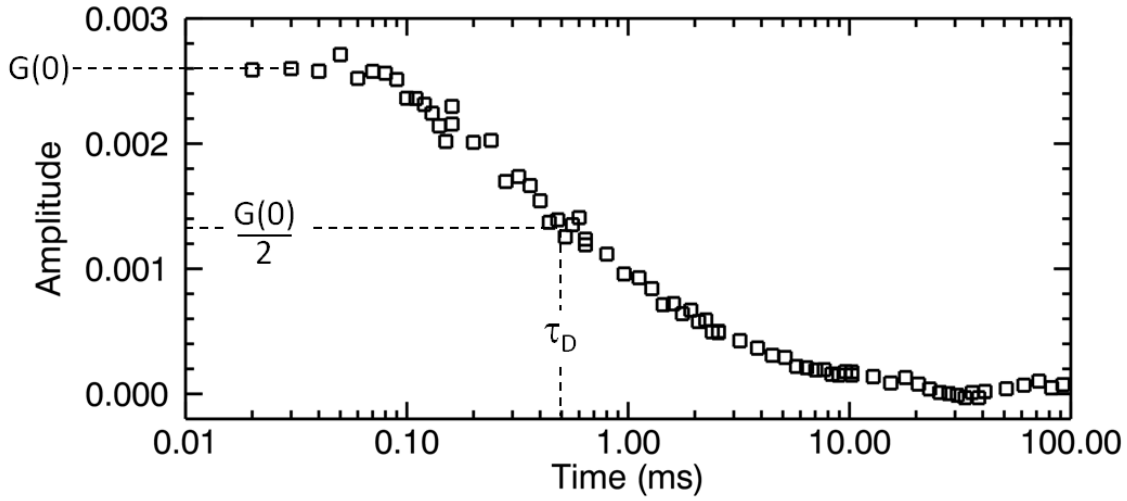

Figure 2.4 **Autocorrelation curve.**

The plot displays a sample autocorrelation curve. On short time scales, most of the same particles are still in the excitation volume and so correlations are strong and the amplitude is large. Over time, as the starting particle population diffuses away, the amplitude decays to zero. From the extrapolated $G(0)$ value, we get information about concentration and so can calculate a brightness. From the decay, we gain information about diffusion of the particles. Diffusion time τ_D, the time it takes for the amplitude to halve, is defined as the characteristic time that a particle stays in the excitation volume.

2.4 Analysis

We briefly present some common analysis methods used to calculate brightness from FFS data, as well as a few sources of error common to all brightness experiments and the corrections thereof, as used in this thesis.

2.4.1 Photon count moments

Access to the raw photon counts of an FFS experiment allows a direct calculation of the photon count moments. Each data point represents the number of photons k detected during a time interval T. The first two moments of the photon counts determine Mandel's Q-parameter

$$Q_T = \frac{\left\langle \Delta k^2 \right\rangle_T - \left\langle k \right\rangle_T}{\left\langle k \right\rangle_T} = \gamma_2 \tilde{\lambda} T \tag{2.2}$$

which in turn can be related to the brightness $\tilde{\lambda}\ \tilde{\lambda}\ \tilde{\lambda}$. The subscript T is included to highlight that moments were calculated from photon counts sampled with a time interval T. Note that unlike Mandel's Q-parameter the apparent brightness is model dependent, because the value of the shape factor γ_2 is determined by the point spread function (PSF) model, as will be the case in the other analysis methods presented. Three different PSF models are widely used. The three-dimensional Gaussian PSF with $\gamma_2 = 1/\left(2\sqrt{2}\right)$, the two-dimensional Gaussian PSF with $\gamma_2 = 1/2$, and the squared Gaussian-Lorentzian PSF with $\gamma_2 = 3/16$. Additional modifications to these PSF models or any other PSF model require an explicit calculation of the shape factor γ_2 for consistent interpretation of brightness values. This thesis uses the squared Gaussian-Lorentzian PSF unless otherwise indicated.

2.4.2 Photon counting histogram

The origins of brightness analysis in our lab stem from an analysis method called the photon counting histogram (PCH).[26] As the name implies, the raw photon count data is plotted as a histogram and then fit to a theoretical model. PCH assumes that the

particles are effectively stationary during the acquisition of each data point, thus PCH data must be taken with a fast sampling frequency. Importantly, the technique can distinguish species by concentration and brightness, making it complementary to autocorrelation analysis which may distinguish species by diffusion time.[4,39] Since brightness is a more sensitive parameter than diffusion time, PCH is frequently a better technique for resolving species. However, as most of the work in this thesis only requires measurement of the average brightness, our analysis mostly consists of the straightforward and more easily implemented autocorrelation analysis.

2.4.3 Autocorrelation analysis

The time-dependent fluorescence intensity $F(t)$ of the FFS measurement is used to determine its autocorrelation function (ACF),

$$G(\tau) = \frac{\langle \delta F(t+\tau)\delta F(t) \rangle}{\langle F(t) \rangle^2} \tag{2.3}$$

where $\delta F(t) = F(t) - \langle F \rangle$ is the fluctuation in fluorescence intensity about the average. The ACF measures self-similarity of a times series and emphasizes the characteristic time constants of underlying processes.[40,41] Thus the decay of the ACF provides information about molecular diffusion. Specifically, the diffusion time τ_D, defined as the time required for the autocorrelation amplitude to fall to half its initial value (shown in Figure 2.4), characterizes the average duration a particle remains in the excitation volume. A fit of the autocorrelation function to a simple diffusion model determines τ_D and the autocorrelation amplitude $G(0)$, which is related to the average number of molecules N_{PSF} present in the excitation volume,

$$G(0) = \frac{\gamma_2}{N_{PSF}} \tag{2.4}$$

with γ_2 representing the shape factor of the FFS observation volume.[39] We convert N_{PSF} into a concentration by $c = N_{PSF}/V_{PSF}$. The value of V_{PSF} is determined from an FFS measurement of a dye sample with known concentration.[6] The brightness is calculated from the fluctuation amplitude and the average fluorescence intensity,

$$\lambda = \frac{G(0) \cdot \langle F \rangle}{\gamma_2} \tag{2.5}$$

and reflects the average fluorescence intensity received per particle. The brightness of a monomeric fluorescent protein, such as λ_{EGFP}, serves as reference value and is determined by a calibration measurement of a sample containing only the fluorescent protein. Recall that the normalized brightness *b* provides an experimental measure of the stoichiometry of the protein complex, e.g., $b = 2$ indicates a dimeric protein.

2.4.4 Deadtime, afterpulsing and background

Although brightness analysis is straightforward, there are a few challenges. First and foremost, correct identification of the fluctuation amplitude requires proper treatment of deadtime and afterpulsing effects from the detector when fitting the autocorrelation function. Second, both the cellular environment and the solution environment can give rise to a background signal that biases the data, particularly at low intensities.

Photodetectors used in photon counting experiments are never ideal.[33,42,43] Here we will focus specifically on avalanche photodiodes (APDs) because they are the detector most commonly used in fluorescence fluctuation experiments. All brightness experiments that employ APDs must contend with two primary artifacts, deadtime and afterpulsing. Deadtime is a fixed period after the collection of a photon during which the detector cannot register any other events. This temporary "blindness" leads to a decrease in the number of photons detected and is of particular concern in high count rate experiments. Deadtime has the effect of narrowing the photon count distribution. At high count rates more events are lost since there is a greater probability of additional photons arriving while the detector is blind to new events. Typical APDs have deadtimes of 50 ns, although newer models have deadtimes as low as 20 ns. For an average intensity of 850 kilocounts/second and a detector with 50-ns deadtime, the uncorrected brightness will be 15% lower than the true brightness. Thus in high count rate experiments, the effect of deadtime on brightness is not negligible and will result in lower brightness if

there is no correction applied. For the autocorrelation brightness values in this thesis, we employ the following correction:[44]

$$\lambda^{corr} = \frac{\lambda^{\dagger}}{1 - 3\langle F \rangle T^{\dagger}} \tag{2.6}$$

where $\lambda^{\dagger}$ is the deadtime-effected brightness and $T^{\dagger}$ is the deadtime of the detector. Simply calculating the brightness directly from the raw data will result in $\lambda^{\dagger}$, a brightness that includes the effects of deadtime. However, we have previously characterized the detector to identify the deadtime $T^{\dagger}$ and knowing the average fluorescence $\langle F \rangle$, it is simple to use the above formula to calculate the corrected brightness value.

Afterpulses are spurious events that occur with a small probability after the detection of an actual photon. This leads to an increase in the number of apparent photons detected. Afterpulses will have the effect of broadening the photon count distribution because the afterpulses will artificially increase the counts in the higher channels, which leads to an artificially high brightness. For newer APD detectors the afterpulse probability can be sufficiently low that its effect on many experiments is negligible. For instance, APDs used in our lab typically have an afterpulse probability of 0.2% to 0.5%.

It is important to be aware of the background signal in any brightness measurement, especially with a low brightness or low concentration sample. The background signal, defined as any low brightness signal of non-sample origin, is nearly impossible to eliminate and comes from multiple sources. Detectors have a certain level of dark counts. There is an additional level of room light counts, since it is generally difficult to shield all other sources of light around an experimental setup. Finally, there is the background autofluorescence from the sample itself. Aqueous buffers have low but non-zero background counts. Cells are notoriously autofluorescent although this can be greatly reduced by carefully selecting cell lines, cellular location and excitation wavelength. Thus, particularly when beginning a new series of experiments, FFS measurements should be taken of the buffer alone or of an untransfected cell. On a known system these measurement are less crucial unless the experiments are conducted at a low concentration that is close to the background level, in which case it is again very important to have parallel and freshly-prepared "blank" controls. When the fluorescence

signal exceeds the background by a factor of 10, background effects will be less than 10% and may be considered negligible as they fall within the expected noise of brightness experiments. At higher percentages of background, the recovered brightness will be artificially reduced and, in this thesis, is corrected by a first order approximation,

$$\lambda^{corr} = \frac{\lambda^* \cdot F}{F - F_{background}} \tag{2.7}$$

where λ^* is the background effected brightness, and $\langle F_{background} \rangle$ is the total background signal.

2.5 Apparent brightness

Most experiments are conducted on samples representing a mixture of brightness species. The signal statistics from a single FFS measurement are frequently insufficient to directly resolve a mixture of two or more brightness species, particularly in cellular FFS experiments. Given these constraints the best approach is to forgo direct resolution of brightness species in favor of calculating the apparent brightness $\tilde{\lambda}$ of the sample. The apparent brightness for a mixture of m species with brightness λ_i and number of molecules N_i for the *i*-th species is defined by

$$\tilde{\lambda} = \frac{\sum_{i=1}^{m} \lambda_i^2 N_i}{\sum_{i=1}^{m} \lambda_i N_i} \tag{2.8}$$

Because this equation is quadratic in the brightness λ_i, it is useful to rewrite it using the fractional intensity f_i of species *i*,

$$f_i = \frac{\langle F_i \rangle}{\langle F \rangle} = \frac{\lambda_i N_i}{\sum_{i=1}^{N} \lambda_i N_i}, \tag{2.9}$$

where $\langle F_i \rangle$ is the fluorescence intensity of species *i* and $\langle F \rangle$ is the total fluorescence of the sample. Combining these two equations demonstrates that the apparent brightness reflects a fractional intensity weighted brightness of the sample,

$$\tilde{\lambda} = \sum_{i=1}^{N} f_i \lambda_i \ . \qquad (2.10)$$

The final panels (E & F) of Figure 2.3 display the normalized apparent brightness for two mixed samples of monomer and dimer species, and illustrate how the ratio of the subpopulations affects the brightness value. For the remainder of the thesis we will consider brightness synonymous with apparent brightness.

2.6 Brightness titration

Because apparent brightness reduces the complexity of the sample to a single numerical value, it is prudent to couple it with another observable that leads to predictable changes in apparent brightness. In this thesis, we are interested in protein association. Because protein association is concentration dependent, a binding curve of the association process can be constructed by systematic variation of the protein concentration and simultaneous brightness measurements of the sample.[6] Since cells vary in the level of protein expressed, we take advantage of the wide concentration range by selecting cells from low, medium, and high expression levels. For each cell, the normalized apparent brightness b is graphed as a function of the fluorescently-labeled protein concentration. Such a plot characterizes the binding curve of a cellular protein in its native environment and is referred to as brightness titration.[45] In a hypothetical monomer-dimer protein equilibrium (Figure 2.5), the normalized brightness is close to 1 at low concentrations, which indicates a monomeric protein state. Because higher concentrations favor dimerization, the normalized brightness climbs with increasing concentration and saturates at $b = 2$, indicating the presence of a dimeric protein population. The brightness values at intermediate concentrations are between 1 and 2, which indicates a sample containing a mixture of monomeric and dimeric proteins (as in Figure 2.3E&F). Thus 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.[46]

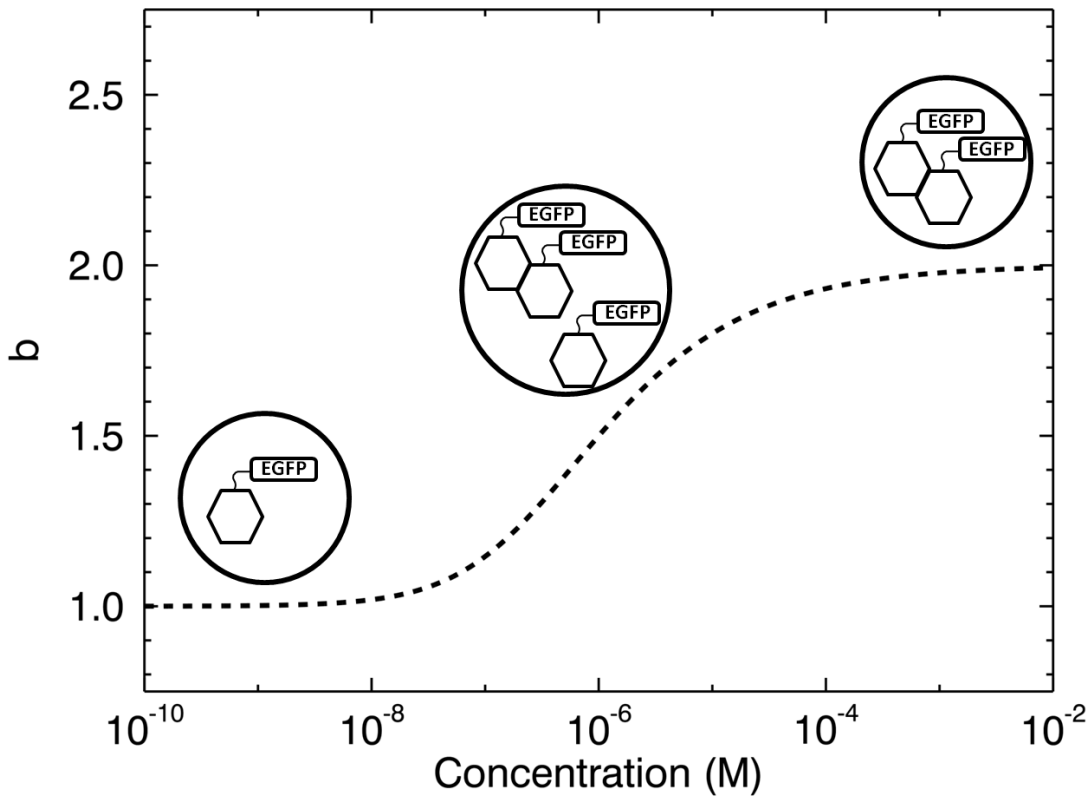

Figure 2.5 **Monomer-Dimer Titration**

A conceptual brightness titration in cells of an EGFP-labeled protein with a dissociation constant $K_D \sim 1\ \mu M$. At low concentrations, the sample is predominantly monomeric and has an apparent brightness close to 1. At high concentrations the equilibrium favors the dimeric form and $b = 2$. By measuring multiple cells expressing different levels of the protein, we can generate a curve that provides a more reliable picture of the protein behavior and identifies the affinity.

2.7 Droplet sample protocol

Most of the solution experiments in the later chapters are conducted in droplet samples. We present here a protocol for free-standing droplets placed on a glass surface (Figure 2.6A). Due to the difficulty of purification or the expense of cell-free solution, small sample volumes represent important savings in time and expense. In the case of the cell-free samples, small droplets also permit efficient oxygen exchange with the sample which is important for fluorophore maturation, as will be discussed in chapter 5. We developed a droplet protocol for FFS experiments that represents the best compromise between small sample volumes and practical considerations such as sample manipulation and evaporation.

We typically perform FFS measurements in 8-well chambered coverglass slides. By covering the top opening of the 8-well with a microscope slide, sealed down with vacuum grease, we create small, isolated chambers that exhibit minimal evaporation over a period of ~6 hours. Within each well, we first silanize the bottom surface to prevent sample adhesion. A volume (~150 µL) of SigmaCote sufficient to cover the surface is added to a well and left to react for 30 seconds. We then tilt the 8-well and remove the SigmaCote from one corner, endeavoring to prevent excess SigmaCote from evaporating off the surface of glass. Such evaporation can leave behind a residue that may interact with the sample. After the SigmaCote is removed, we rinse the well to remove byproducts of the silanization reaction.

We have found 4 µL samples to be small enough to save material while being large enough to prevent strong evaporation. In samples of 1 µL or less, evaporation is very difficult to prevent and the sample becomes difficult to mix, dilute or otherwise manipulate with standard pipettes. To further assist with sample manipulation, we make small measurement wells on the glass surface by drawing 3-4 mm diameter circles with a vacuum grease filled syringe. Such circles may also be generated through a stamp approach, but the syringe method creates slightly higher “walls” that prevent spillover when moving the 8-well or when diluting a sample.

We also perform titration experiments in the droplet form, but because of the small volumes used, a typical titration is not feasible. Instead a series of droplets are created at different interaction mixtures and are then measured individually as in the cell titration. Figure 2.6B depicts such an experiment to study the interaction of sample X and sample Y. Sample X is fluorescently labeled and a single stock solution is prepared. Sample X will be held at a fixed concentration in all the droplets in order to maintain the same fluorescence intensity. A series of Y dilutions are generated before being mixed 1:1 with X in their respective droplets. Brightness measurements in each droplet create a titration curve of the concentration-dependent stoichiometry of the labeled sample X as it interacts with Y.

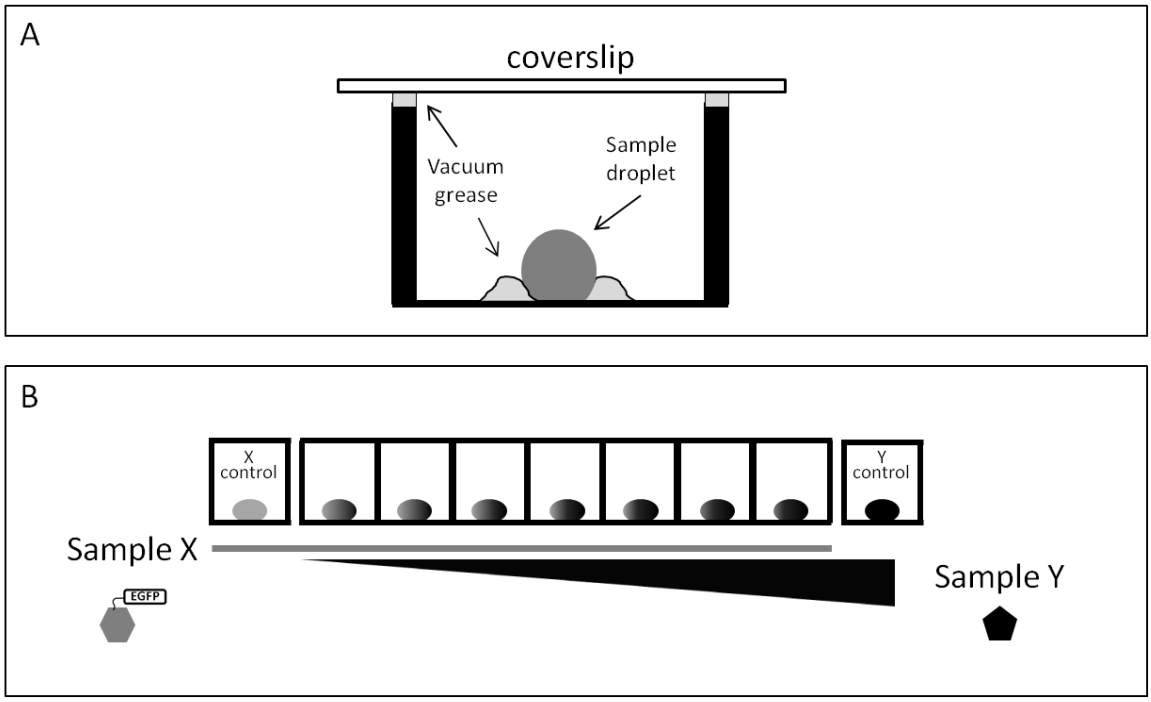

Figure 2.6 **Protein Interaction Droplet Protocol**

(A) The cartoon displays the droplet sample preparation. A circle of vacuum grease prevents the droplet from moving or spreading. A coverslip and some additional vacuum grease seal the top surface to prevent evaporation. The free standing droplet provides good oxygen exchange for chemical reactions as well as allowing easy access for further mixing or dilution. (B) In chapter 6, we perform several protein interaction titrations in a droplet assay. Because we work with very small sample volumes to save material, a standard titration is not feasible. Consider the interaction of sample X with sample Y. Sample X is fluorescently labeled and held at a fixed concentration. Sample Y is not labeled and is mixed into different droplets at different concentrations. A brightness measurement of all the droplets creates a titration curve that identifies the stoichiometry of X in interactions with Y. The labeling situation can be reversed for additional information.

3. Photodepletion

In order to extend FFS into other environments, we must be able to deal with the constraints imposed by those environments. Small sample environments have a finite reservoir of fluorescence material. When that reservoir is on the order of picoliters, mild photobleaching leads to non-trivial sample depletion. This depletion biases FFS measurements because FFS analysis relies on a stationary average fluorescence signal. For instance, any FFS experiments performed in *E. coli*—with a volume on the order of a femtoliter—will exhibit unavoidable photodepletion and a significant bias in brightness. To combat this, we develop protocols and theory that allow us to account for the presence of the bias and to extract the corrected brightness value. This chapter describes the theoretical model and its validation both in solution droplets and in mammalian cell nuclei. I collaborated with my colleague Kwang Ho Hur, who applied the photodepletion model to *E. coli* cells. I include one figure showing preliminary *E. coli* results from his work.

3.1 Introduction

Although photobleaching is not noticeable as a direct effect in most FFS measurements, it nevertheless occurs with a small probability and reduces the number of fluorophores in the sample. *In vitro* FFS experiments are typically conducted with a sample volume that exceeds the optical volume by many orders of magnitude, thus providing a large enough reservoir so that the concentration of fluorophores in the sample is virtually constant. In contrast, a biological cell may be viewed as a picoliter cuvette with a finite amount of fluorophores. The cumulative reduction of the fluorophore concentration by photobleaching will be referred to as photodepletion throughout the rest of the paper.

Photodepletion results in a decreasing fluorescence intensity which biases the interpretation of FFS data. We develop and verify a simple model that takes the influence of photodepletion on brightness into account. Some FFS experiments in cells require long data acquisition times to gather the necessary statistics for quantifying protein interactions. Photodepletion effects that are completely negligible for short data

acquisition times will become apparent once the measurement time exceeds a certain limit.[47] As a model system, we use Alexa 488 in microdroplets created from a water-oil emulsion. We then perform photodepletion experiments on EGFP in cells to demonstrate that the model is sufficiently robust for *in vivo* applications. Finally, we apply this approach to a biological system.

In *E. coli* cells, short data acquisition times can no longer be used to prevent photodepletion effects. The *E. coli* sample volumes are small enough that the photodepletion effects are very strong. If the data acquisition time is reduced to avoid the photodepletion bias, there is no longer sufficient statistics for good brightness calculation. In this regime, there is an estimation bias[48] inherent in correlation data that prevents the extraction of accurate brightness data. Thus, an approach that accounts for photodepletion is required for quantitative brightness analysis of FFS experiments in *E. coli*.

Throughout this thesis we will employ nuclear transport factor 2 (NTF2) as a model system of a functional dimer. In later chapters, the oligomerization state of NTF2 will have biological relevance, but as it is not an *E. coli* protein, it is simply a test for brightness doubling in this chapter. We apply our photodepletion model to identify the brightness of NTF2 and EGFP in *E. coli* cells.

3.2 Theory

We briefly review brightness analysis using the "ε" brightness notation, recalling that ε is defined as photon counts per data acquisition bin, rather than per second. This notation is simpler for this application. We subsequently introduce FFS theory to account for photodepletion in the cases of stationary and non-stationary fluorescence signals.

3.2.1 Brightness analysis

We only consider samples containing a single type of fluorophore, i.e., the green fluorescent protein EGFP. In single-channel FFS each species is specified by its brightness ε and its occupation number N. Moment analysis determines the brightness

and number of molecules in the observation volume from the first two moments of the photon counts, $\langle k \rangle$ and $\langle k^2 \rangle$. For a single species the average photon count is

$$\langle k \rangle = \varepsilon N . \tag{3.1}$$

The brightness of the sample is determined by the Q-factor[49]

$$Q = \gamma\varepsilon = \frac{\langle k^2 \rangle - \langle k \rangle^2 - \langle k \rangle}{\langle k \rangle}, \tag{3.2}$$

where γ is a shape factor that depends on the form of the point spread function (PSF).[39] These two equations uniquely determine the brightness and occupation number. The brightness ε is a function of the sampling frequency *f*, and is related to the absolute brightness by $\lambda = \varepsilon \cdot f$.

In the presence of multiple species, each with its own brightness ε_i and occupation number N_i, the first two moments determine an apparent brightness and occupation number via

$$\langle k \rangle = \varepsilon_{app} N_{app} , \tag{3.3}$$

$$Q = \gamma\varepsilon_{app} . \tag{3.4}$$

The apparent parameters are determined by the brightness and occupation number of all species,[29,39]

$$\varepsilon_{app} = \frac{\sum_i \varepsilon_i^2 N_i}{\sum_i \varepsilon_i N_i}, \tag{3.5}$$

$$N_{app} = \frac{\left(\sum_i \varepsilon_i N_i \right)^2}{\sum_i \varepsilon_i^2 N_i} . \tag{3.6}$$

Here we define the normalized brightness as $b = \varepsilon_{app} / \varepsilon$, where the measured brightness ε_{app} is divided by the brightness ε of a single EGFP. This definition is exactly equivalent to that defined earlier, i.e., the normalized brightness of an *n*-mer is $b_{n-\mathrm{mer}} = n$.

In the following sections, we consider how photodepletion affects brightness analysis in cells.

3.2.2 The effect of photodepletion on brightness for a stationary signal

For simplicity we will assume a two-state model to describe bleaching, $A \xrightarrow{k} D$. The brightness of the molecule in the fluorescent state *A* is ε, while it is zero in the bleached state *D*. The reaction $A \to D$ is described by a rate coefficient *k*, which depends on the fluorophore, the excitation power, and other experimental parameters. While it is possible to evoke more complex bleaching models, the simple two-state process is sufficient to describe the data presented in this paper.

Photodepletion leads to a decay in the average fluorescence intensity $I = \langle k \rangle f$ where *f* is the sampling frequency, as demonstrated in Figure 3.1A for measurements of EGFP inside the nucleus of a cell. Thus, data acquired in the presence of photodepletion is non-stationary. Because FFS analysis requires a stationary signal, we consider a sufficiently short time interval of duration *T* so that the fluorescence signal may be regarded as stationary over that time interval. The fluorescence moments at time *t* are defined by a time average over the interval *T*,

$$\left\langle k(t)^r \right\rangle = \frac{1}{T} \int_{t-T/2}^{t+T/2} k(t')^r dt' . \tag{3.7}$$

This approach allows us to determine the mean $\langle k(t) \rangle$ and $Q(t)$ for each segment of the measured fluorescence signal. If the signal is approximately stationary over the data segment, the brightness $\varepsilon(t)$ and the occupation number $N(t)$ are determined by Eqs. 3.3 and 3.4. Because photodepletion reduces the concentration of fluorophores, the occupation number of EGFP molecules $N(t)$ decreases with time (Figure 3.1B). The

brightness $\varepsilon(t)$ is time independent in the case where a single, monomeric fluorescent species, such as EGFP, is present (Figure 3.1C).

Photobleaching of individual chromophores is statistically independent. From the viewpoint of a single chromophore, photodepletion is described by a probability $p(t)$ to remain fluorescent at time *t*, and a probability $1-p(t)$ to be. If a single rate coefficient *k* describes bleaching then

$$p(t) = e^{-kt}. \tag{3.8}$$

The fluorescence intensity at time *t* is described by $I(t) = \langle k \rangle f = \varepsilon p(t) N_0 f$ with N_0 denoting the total number of chromophores in the observation volume at time 0. The remaining intensity fraction is defined by

$$r_I = \frac{I(t)}{I(0)} = p(t). \tag{3.9}$$

We also define the lost intensity fraction $1-r_I$, and will refer to these two quantities throughout the remainder of the work.

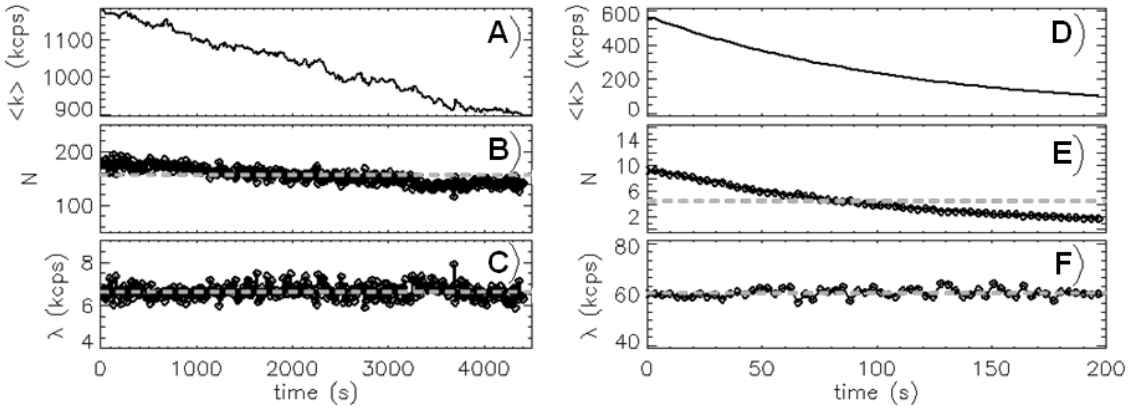

Figure 3.1 **FFS data in the presence of photodepletion.**

EGFP is transiently transfected into COS cells and a prolonged measurement is taken in the nucleus using two-photon excitation (left panels). Fluorescence intensity (A), occupation number (B), and brightness (C) are plotted as a function of time, after having divided the data into segments and analyzing each segment independently to extract the aforementioned parameters. Note that while the overall sample population is being depleted, the brightness remains stable. Panels (D), (E), and (F) show the fluorescence intensity, occupation number, and brightness of Alexa 488 in a microdroplet in the presence of severe photodepletion.

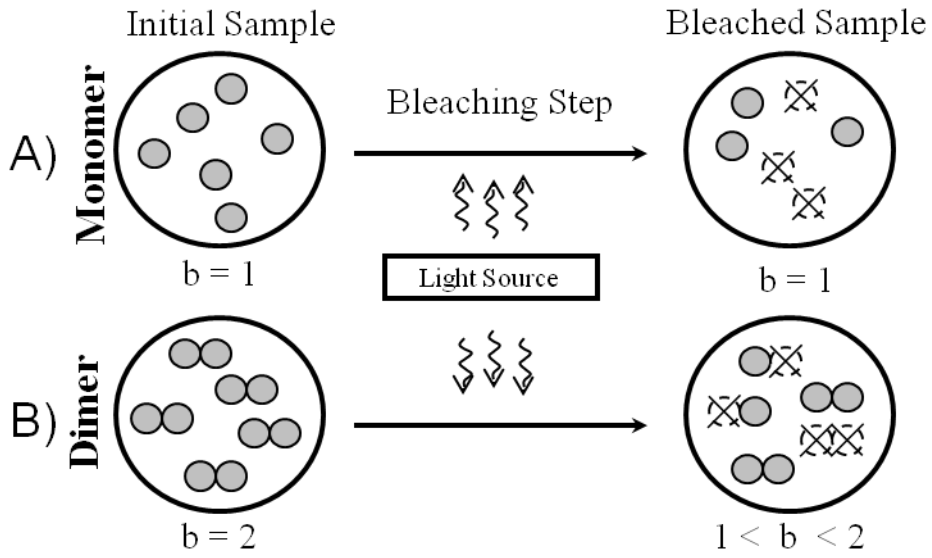

Figure 3.2 **Conceptual picture of monomer/dimer sample photodepletion effects.**

(A) Bleaching of a monomer population has no effect on brightness, because bleached molecules are not contributing to the fluorescence signal. The normalized brightness remains $b = 1$. (B) In the case of a dimer population, bleaching introduces a population with one of the two fluorophores surviving, thereby creating a mixed brightness population with an apparent brightness smaller than the original.

In contrast to fluorescence intensity, the influence of photodepletion on brightness is more complex. Consider first a monomeric fluorescent protein A with brightness ε. Bleaching will lead to two populations, A and D. The brightness of the remaining population in state A is ε, while the brightness of the bleached state D is zero. Because the population D does not contribute to the fluorescence signal, the brightness of the sample remains unchanged (see Figure 3.2A). A population of dimers A_2, on the other hand, has a brightness of 2ε (Figure 3.2B). Bleaching of one or two chromophores of A_2 leads to the states AD and D_2. Because bleaching is a stochastic event, a statistical mix of proteins in each of the three states is attained. The brightness of AD is half that of the dimer A_2, which leads, according to Eq. 3.5, to an apparent brightness for the sample of less than 2ε (Figure 3.2B). Thus, the brightness of the dimer decreases if photodepletion is present.

Let us explicitly treat the case of a pure *n*-mer species A_n. Bleaching of $n - s$ chromophores leads to the species A_sD_{n-s} with brightness $s\varepsilon$. Its fractional population is determined by the probability

$$p_{n,s}(t) = \binom{n}{s} p^s (1-p)^{n-s}. \tag{3.10}$$

According to Eq. 3.5, all species contribute to the normalized brightness of the sample. Evaluation of Eq. 3.5 for fractional populations given by Eq. 3.10 reveals a simple linear dependence of the brightness of the *n*-mer on the remaining intensity fraction r_I,

$$b_{n-\mathrm{mer}}(r_I) = \frac{1}{\varepsilon} \frac{\sum_{s=0}^{n} (s\varepsilon)^2 p_{n,s}}{\sum_{s=0}^{n} s\varepsilon\, p_{n,s}} = 1 + (n-1)p = 1 + (n-1)r_I. \tag{3.11}$$

The normalized brightness starts at a value of *n* in the absence of photodepletion ($r_I = 1$) and decreases linearly to a value of one as the remaining intensity fraction approaches zero. This result reflects that the last surviving fluorescent population of an *n*-mer A_n is A_1D_{n-1} with exactly one functioning fluorophore, which implies a normalized brightness

of one. As a consequence, the normalized brightness of a monomeric protein species (n = 1) is independent of photodepletion as shown for EGFP in Figure 3.1.

In cells, however, one frequently encounters the simultaneous presence of protein species in different oligomeric states. We now extend our result to an arbitrary mixture of oligomeric states. Consider a sample of A_n , A_{n-1} , … , A_0 with a fractional population of α_i for the oligomer A_i. By definition, $\sum_i \alpha_i = 1$. Depletion of the oligomer A_i due to bleaching leads to the species A_sD_{i-s} with a normalized brightness of s and a fractional population $p'_{i,s}(t) = \alpha_i p_{i,s}(t)$. The sum over all fractional populations equals one at all times. The apparent brightness of the mixture is calculated analogously to the case of the n-mer A_n (Eq. 3.11) by summing up the contributions from all oligomers in all of their bleached states,

$$b_{app}(r_I) = \frac{1}{\varepsilon} \frac{\sum_{r=0}^{n} \sum_{s=0}^{n-r} (s\varepsilon)^2 p'_{n-r,s}}{\sum_{r=0}^{n} \sum_{s=0}^{n-r} s\varepsilon p'_{n-r,s}} = 1 + (b_{app,0} - 1) r_I , \tag{3.12}$$

where $b_{app,0} = b_{app}(r_I = 1)$ is the apparent brightness of the mixture in the absence of photobleaching,

$$b_{app,0} = \frac{\sum_{r=0}^{n} \alpha_r r^2}{\sum_{r=0}^{n} \alpha_r r} . \tag{3.13}$$

Again, the slope plus one of the normalized brightness versus r_I curve is a measure of the normalized brightness $b_{app,0}$ of the sample in the absence of photodepletion.

3.2.3 The effect of photodepletion on brightness for a non-stationary signal

The theory presented so far assumes a stationary signal. In other words, the data segment chosen must be short enough that the fluorescence intensity is approximately constant. Choosing a longer data segment is often advantageous from an experimental point of view as it increases the signal-to-noise ratio (SNR). But if the intensity

decreases over the duration of the segment, the condition of stationarity is violated, which leads to a bias in the parameters obtained from Eqs. 3.3 and 3.4. We now describe how to account for the bias introduced by the non-stationary signal.

The photon count moments for a data segment of duration T are obtained by

$$\overline{\left\langle K^r \right\rangle^s} = \frac{1}{T}\int_{t'=0}^{T} \left\langle k^r(t') \right\rangle^s dt' . \tag{3.14}$$

The biased apparent brightness $\bar{\varepsilon}_{app}$ and occupation number $\bar{N}_{app}$ are defined by

$$\overline{\left\langle K \right\rangle} = \bar{\varepsilon}_{app}\bar{N}_{app} , \tag{3.15}$$

$$\overline{Q} = \frac{\overline{\left\langle K^2 \right\rangle} - \overline{\left\langle K \right\rangle}^2 - \overline{\left\langle K \right\rangle}}{\overline{\left\langle K \right\rangle}} = \gamma_2 \bar{\varepsilon}_{app} . \tag{3.16}$$

Expressing the moments $\left\langle k \right\rangle$ and $\left\langle k^2 \right\rangle$ in terms of the brightness and occupation number of each species and evaluating Eq. 3.14 yields the photon count moments $\overline{\left\langle K^r \right\rangle^s}$. The first moment is given by $\left\langle k(t) \right\rangle = \sum_i \varepsilon_i N_i(t)$ and its square is $\left\langle k(t) \right\rangle^2 = \sum_{i,j} \varepsilon_i \varepsilon_j N_i(t) N_j(t)$. The second moment is written as $\left\langle k^2(t) \right\rangle = \gamma_2 \sum_i \varepsilon_i N_i(t) + \left\langle k(t) \right\rangle^2 + \left\langle k(t) \right\rangle$, where we made use of the factorial cumulant $\kappa_{[2]} = \gamma_2 \sum_i \varepsilon_i N_i(t)$.[32] To write the result in a compact format we introduce

$$\eta_i(T) = \frac{1}{T}\int_{t=0}^{T} N_i(t)\, dt , \tag{3.17}$$

$$\eta_{ij}(T) = \frac{1}{T}\int_{t=0}^{T} N_i(t) N_j(t)\, dt . \tag{3.18}$$

The general relationship between the correct and biased brightness of the sample in the presence of photodepletion is

$$\overline{Q} = \gamma_2 \bar{\varepsilon}_{app} = \frac{\gamma_2 \sum_i \varepsilon_i^2 \eta_i + \sum_{i,j} \varepsilon_i \varepsilon_j \left(\eta_{ij} - \eta_i \eta_j \right)}{\sum_i \varepsilon_i \eta_i} . \tag{3.19}$$

For a single monomeric species with brightness ε_1 Eq. 3.19 simplifies to

$$\overline{Q} = \frac{\gamma_2 \varepsilon_1^2 \eta_1 + \varepsilon_1^2 \left(\eta_{11} - \eta_1 \eta_1\right)}{\varepsilon_1 \eta_1} = \gamma_2 \varepsilon_1 \left(1 + \frac{1}{\gamma_2} \frac{\eta_{11} - \eta_1^2}{\eta_1}\right) = \gamma_2 \overline{\varepsilon}_1 . \qquad (3.20)$$

We will now assume an exponential decay model, $p(t) = e^{-kt}$, and determine the normalized brightness of a monomer (n = 1) from Eq. 3.20,

$$\overline{b}_1(r_I) = 1 + \frac{N_0}{\gamma_2}\left(\frac{r_I + 1}{2} - \frac{r_I - 1}{\ln r_I}\right), \qquad (3.21)$$

with r_I defined by Eq. 3.9. The same method is used for determining the brightness of any n-mer of interest. Because the expressions become lengthy as n increases, only the Taylor expansion of the brightness to 2nd order in T is given

$$\overline{b}_{n-\mathrm{mer}}(r_I) = n + \frac{n-1}{2}\ln r_I - \left(3(n-1) + \frac{N_0}{\gamma_2} n\left(1 - 2n + 2n^2\right)\right)\frac{\left(\ln r_I\right)^2}{12}. \qquad (3.22)$$

3.3 Materials and methods

Experimental Setup. A mode-locked Ti:sapphire laser (Tsunami, Spectra-Physics, Mountain View, CA) pumped by an intracavity doubled Nd:YVO_4 laser (Millenia, Spectra Physics) serves as a source for two-photon excitation. The experiments were carried out with a modified Zeiss Axiovert 200 microscope (Thornwood, NY) as previously described.[29] 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, Rockingham, VT). Two photon measurements were taken for ~50 seconds with a 63x Plan Apochromat oil immersion objective (NA = 1.4). The excitation power measured at the objective ranged from 0.8-1.0 mW at 905 nm and 0.65-0.9 mW at 970 nm. These powers were chosen to be sufficiently low to avoid saturation. Photon counts were detected with an avalanche photodiode (APD) (Perkin-Elmer, SPCM-AQ-14) and recorded by a data acquisition card (ISS, Champaign, IL), which stores the complete sequence of photon counts using sampling frequencies ranging from 20 to 200 kHz. The recorded photon counts were stored and analyzed with programs written in IDL 6.0 (Research Systems, Boulder, CO).

Microdroplets. Alexa-488 (Molecular Probes, Eugene, OR) was diluted in water to a concentration of 160 nM. A volume of 100 µL of this solution was combined with 900 µL of silicon oil (Fisher Scientific, Fairlawn, NJ), pippetted for 5 seconds and then vortexed for 20 seconds. The emulsion was allowed to stand for three minutes while the larger droplets settled before removing 10 µL from the top and transferring it onto a two inch square glass slide (Chemglass, Vineland, NJ). A coverglass (12-544F, Fisher Scientific, Fairlawn, NJ) was pressed down on top and affixed at the corners with nail polish to prevent shear motion.

Expression Vectors, Cell Lines, and Cell Measurements. Construction of the tandem dimeric EGFP ($EGFP_2$) was described previously.[6] COS-1 cells were obtained from American Type Culture Collection (ATCC) and maintained in 10% fetal bovine serum (Hyclone Laboratories, Logan, UT) and DMEM (without phenol red) media. MRC-5 and CV-1 cells were obtained from ATCC and similarly maintained. For CV-1 and MRC-5 cells, transfection was carried out using TransFectin reagent (Bio-Rad, Hercules, CA) according to the manufacturer's instructions 24 hours before measurement. For COS-1 cells, G418-resistant cells were selected by culturing cells in media with 250 µg/mL gentamicin sulfate (Mediatech, Hernandon, VA) and measured the second week after transfection. All cells were subcultured into eight-well coverglass chamber slides (Nagle Nunc International, Rochester, NY) with the media replaced by Leibovitz L15 medium (Gibco, Auckland, NZ) immediately before measurement.

Data Analysis. We used moment analysis[31,32] with deadtime corrections[8] to determine brightness and occupation number in this work. We also calculated the brightness of cell data using Q-analysis.[44]

3.4 Results

We measured the fluorescence of EGFP inside the nucleus of living cells with two-photon excitation at 905 nm. The decrease in the fluorescence intensity with time reveals the presence of photobleaching (Figure 3.1A). We also performed *in vitro* measurements of Alexa 488 in microdroplets with sizes similar to that of cells. Measuring in microdroplets ensures a limited fluorescent population so that we can examine the effects of photodepletion on a model system. The decaying fluorescence intensity of Alexa 488 in a droplet is shown in Figure 3.1D. For further analysis we divide the recorded fluorescence signal into data segments of ~1.6 s each. Because the fluorescence intensity is approximately constant within a segment, brightness analysis is performed separately on each segment. The number of molecules N of EGFP in the observation volume (Figure 3.1B) decreases, because bleaching reduces the overall concentration of fluorophores in the sample. The brightness λ, however, remains constant throughout the measurement (Figure 3.1C), because the bleached molecules are not contributing to the fluorescence, while the brightness of the remaining fluorophores is unchanged as illustrated in Figure 3.2A. The number of molecules of Alexa 488 and their brightness is shown in Figures 3.1E-F. Again, the brightness is time-independent, while the number of fluorophores decreases as a consequence of bleaching. Alexa 488 is significantly brighter than EGFP, which leads to a better signal-to-noise ratio, and is useful for comparing data with the theoretical model. Also, because we are able to expose microdroplets to higher laser powers than used in cell measurements, the effects of bleaching are more pronounced in the droplets than in cells.

FFS analysis of the data set as a whole without segmenting leads to an apparent brightness of 50 kilocounts per second. This value is 800% higher than the brightness recovered from the segmented data and illustrates the bias introduced into the analysis by the non-stationary fluorescence signal. The bias is similar in the case of the microdroplet sample, where analysis of the whole data set leads to a ~700% increase in the brightness compared to the value recovered from the segmented data. Thus, segmenting of the data is an efficient method for eliminating the bias introduced by the non-stationary signal.

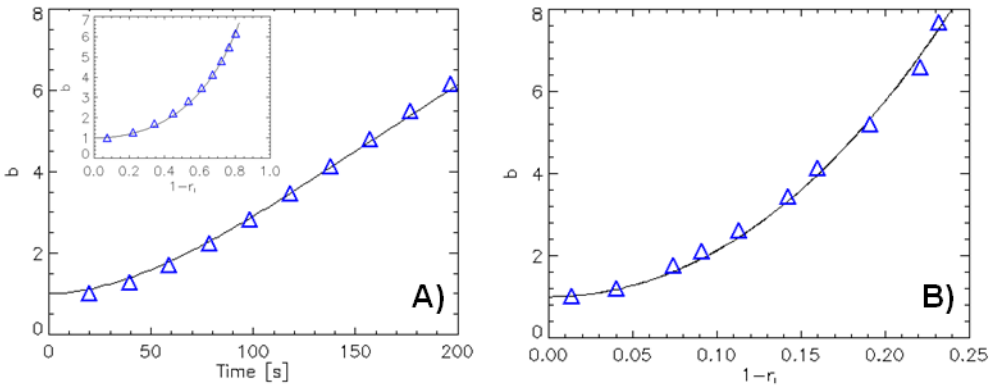

Figure 3.3 **Brightness bias as a function of time and lost intensity fraction.**

The triangles represent the normalized brightness of increasingly longer data segments of Alexa 488 dye in a microdroplet and EGFP in the nucleus of a cell. The brightness values are determined using moment analysis. (A) The brightness bias from the Alexa data of Figure 3.1 is plotted as a function of time. The inset shows the same data plotted as a function of lost intensity fraction since the behavior of the brightness with respect to this parameter is independent of bleaching rate. (B) The EGFP brightness, also from Figure 3.1, is plotted with normalized brightness against the lost intensity fraction and fit with Eq. 3.21.

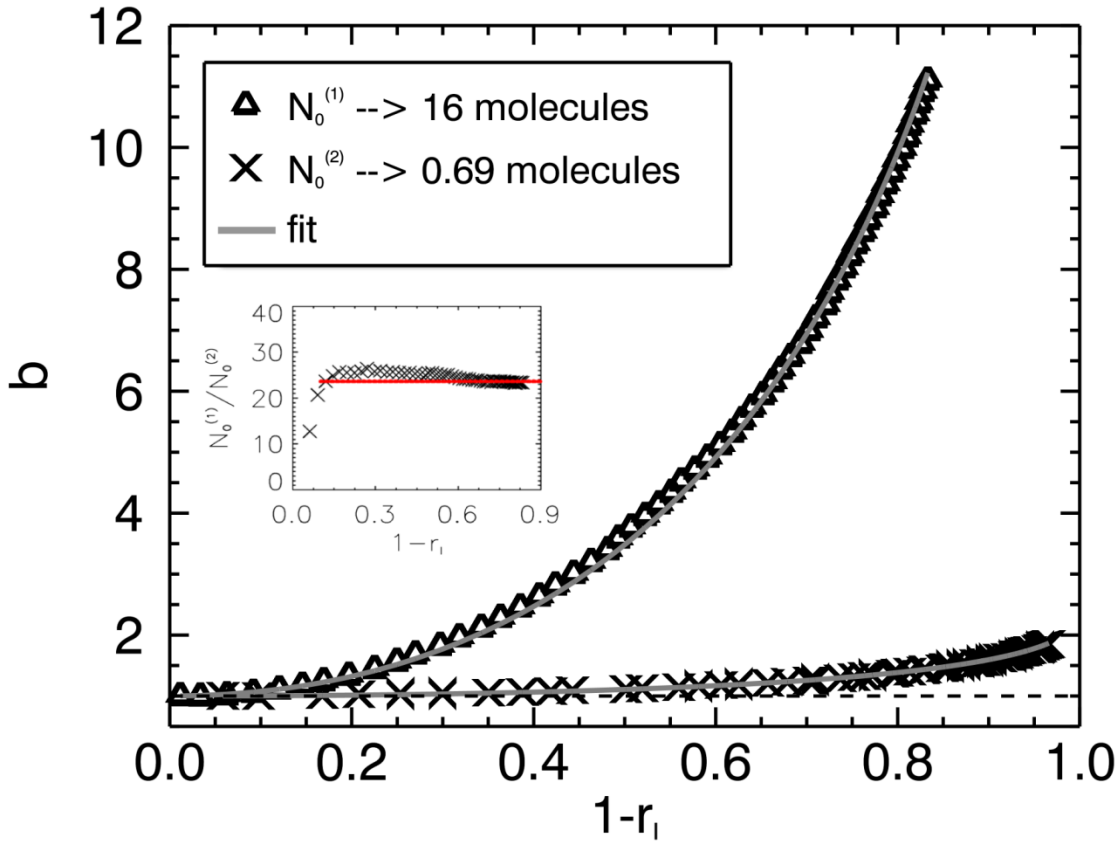

Figure 3.4 **Brightness bias at two different fluorophore concentrations.**

The normalized brightness of two Alexa 488 microdroplets with different concentrations is plotted versus the lost intensity fraction, $1-r_I$. The legend provides the measured average occupation number, prior to bleaching, for each droplet. In the inset, the ratio of the two occupation numbers is calculated from the biased brightnesses according to Eq. 3.23. At low bleaching levels, the calculated ratio is unreliable since the normalized brightness of the low concentration droplet has virtually no bias.

To illustrate the bias on brightness caused by the decreasing fluorescence intensity, we recalculated the brightness of the first segment of the Alexa 488 measurement presented in Figure 3.1, while progressively increasing the duration of the data segment. The bias in the calculated brightness as a function of the length of the data segment is shown in Figure 3.3A. If the data segment includes the whole data set, the brightness bias is ~800%. The increase in brightness with data acquisition time is reproduced by a fit (*solid line*, Figure 3.3A) of the data to Eq. 3.21 with the remaining intensity fraction determined from Eqs. 3.8 and 3.9. The theory predicts that two measurements of the same sample with different bleaching rates will report the same brightness as long as the remaining fluorescence intensity fraction r_I defined by Eq. 3.9 is identical. Even if the fluorescence intensity decay is not exponential, the brightness $b(r_I)$ is according to Eq. 3.21 determined by the remaining intensity fraction r_I. Thus, from now on we graph the brightness as a function of $1\text{-}r_I$ as shown in the inset of Figure 3.3A. We also reanalyzed the cell data of Figure 3.1D as a function of increasing segment time. The resulting brightness bias as a function of $1-r_I$ is shown in Figure 3.3B together with a fit to Eq. 3.21. The agreement between data and fit demonstrates the feasibility of brightness analysis of cell experiments in the presence of photodepletion.

We measured two different droplets containing Alexa 488 at different concentrations. The biased brightness of the droplet with low concentration is graphed together with the brightness of the droplet with high concentration as a function of the lost intensity fraction $1\text{-}r_I$ (Figure 3.4). The figure demonstrates that for equal fractional intensities the brightness bias is much stronger for the sample at high concentration. Both curves are fit to Eq. 3.21 which leads to an initial occupation number $N_{10}=16$ for the high concentration sample, while $N_{20}=0.69$ for the other droplet. This dependence on concentration is predicted by Eq. 3.21 as illustrated by the inset of Figure 3.4, where we plot

$$\frac{b_1(r_I)-1}{b_2(r_I)-1}=\frac{N_{10}}{N_{20}} \tag{3.23}$$

with b_1 and b_2 as the normalized brightness of the two samples.

We now focus on the effects of photodepletion in cell measurements. Data will be analyzed on segments short enough that brightness bias is negligible. For the previous cell data, Figures 3.1 and 3.3, the cell was exposed to two-photon excitation to simultaneously bleach EGFP and monitor its brightness. However, because these experiments require rather long data acquisition times, in this part of the work the cells are measured by exposing the cell repeatedly for short periods to epifluorescence light. After each exposure the instrument performs a short two-photon FFS measurement to record the brightness and the remaining intensity fraction r_I.

First, we wish to establish whether the fluorescent protein EGFP is well-behaved in photodepletion experiments. The data collected from many different cells are summarized in Figure 3.5 for $r_I > 0.3$, which shows that the brightness of the fluorescent protein is stable, as expected from Eq. 3.11 for a monomeric fluorophore. However, the inset to Figure 3.5 shows that the brightness decreases for $r_I < 0.3$ in contrast to the simple bleaching model which predicts a normalized brightness of 1 all the way down to zero remaining intensity. This divergence indicates that the fluorescent protein is not bleaching in a single step, but involves at least one additional state with dim fluorescence. From a practical point of view, this divergence from the model only occurs at extreme levels of photobleaching, and the data demonstrate that EGFP is well described by our simple model up to 70% bleaching of the initial intensity ($r_I = 0.3$).

Now, we examine a dimeric fluorescent protein in a photodepletion experiment by expressing a tandem construct of the fluorescent protein EGFP in CV-1 cells. The normalized brightness of the tandem construct, referred to as $EGFP_2$, decreases as a function of the remaining intensity fraction (Figure 3.6). This is expected, because photobleaching of different chromophores is statistically independent, which leads to a mixture of $EGFP_2$ with two, one, or no fluorescing units as illustrated in Figure 3.2B. The normalized brightness in Figure 3.6 is fit to a straight line with a slope of 0.95 ± 0.05

and an intercept with the y-axis of 1.05 ± 0.07. This result is in agreement with theory, which predicts a linear function with an intercept of one for $r_I = 0$. The theoretical slope of Eq. 3.11 is $n-1$, where n represents the number of fluorophores within the protein complex. Thus, our experiment predicts that the measured protein contains 1.95 fluorophores, which agrees within experimental error with the expected result, $n = 2$, for a dimer. As is the case for the monomeric fluorescent protein, the brightness of the tandem protein diverges from the simple, two-state model for $r_I < 0.3$ (inset of Figure 3.6). EGFP undergoing severe and prolonged photobleaching experiments in cells appear to require a more complex model to fully describe the data, a point we will address briefly in the discussion. However, our present data and experimental approach lack the resolution to unambiguously distinguish between potential three-state models, and furthermore, such extreme photobleaching conditions are unlikely to occur in most fluorescence studies.

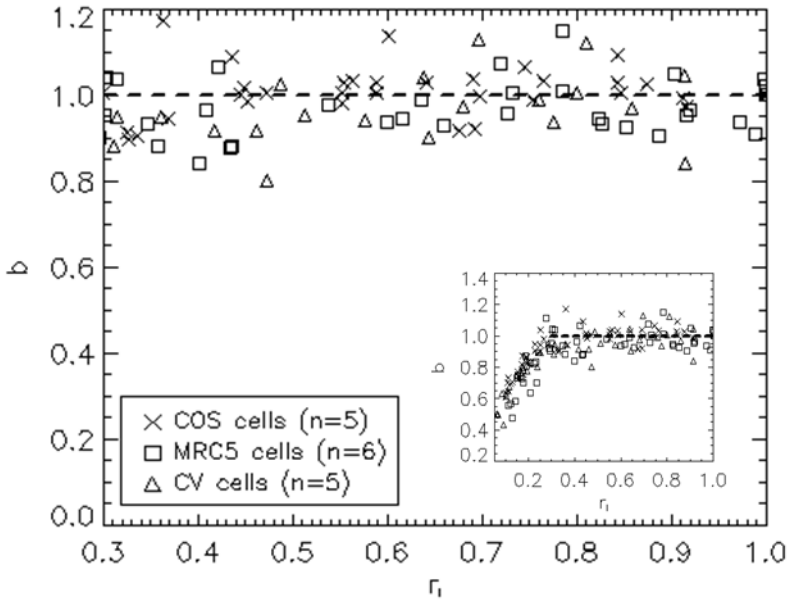

Figure 3.5 **EGFP brightness of cells in the presence of photodepletion.**

EGFP in the nucleus of three different cell lines is bleached using epifluorescence light and monitored via two-photon excitation. The normalized brightness as a function of the remaining intensity fraction r_I is shown for the range where the simple two-state bleaching model (Eq. 3.11) describes the data. The inset graphs the normalized brightness for all measured r_I. The dashed line indicates the regime where the simple two-state bleaching model is applicable and monomeric brightness is stable. Below that regime, the behavior of the brightness is more complicated, but is still universal with respect to the remaining intensity fraction. The dashed line at $b = 1$ marks the range of fractional intensities where the simple analysis model is valid.

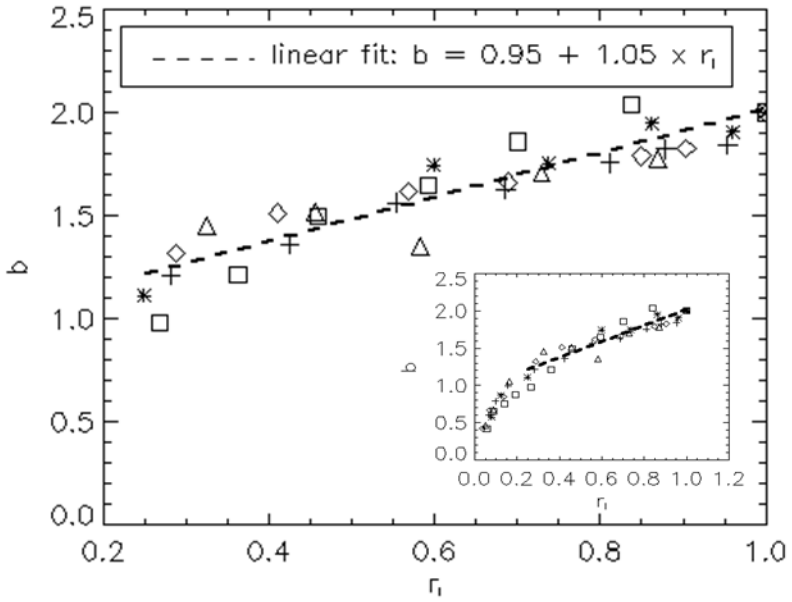

Figure 3.6 **Dimeric EGFP brightness of cells in the presence of photodepletion.**

Dimeric EGFP is bleached with epifluorescence light and monitored by two-photon excitation. The apparent brightness decreases linearly as a function of remaining fractional intensity. The regression line agrees well with our model's prediction of $1+(n-1)\cdot r_I$ (Eq. 3.11). The inset shows that for $r_I < 0.3$ the simple bleaching model fails, and there is evidence of a dim state.

Finally, with the theory developed and validated, we show the strength of the technique. *E. coli* is a broadly used model organism throughout the biological sciences, as the host organism for recombinant DNA, a test system for gene expression, and as a living test tube for still other lab experiments.[50–52] However, due to the extremely small size of *E. coli,* most measurement methods are unavailable and any excitation light applied to an *E. coli* cell will cause photobleaching and therefore deplete a significant fraction of the sample during the measurement. Thus, absolute and quantitative measurements of the concentration and copy number of fluorescent proteins in living *E. coli* cells are feasible for the first time. Here, we show an initial application of the theory and technique to successfully extract correct brightness values.

We investigate the oligomerization state of NTF2 directly in living *E. coli* cells. Although there has been some disagreement in the literature regarding the affinity of the NTF2 homodimer in cellular versus *in vitro* measurements, NTF2 has always behaved as a tight dimer in our hands, as will be explored further in the following chapters. However, most NTF2 solution experiments are performed on proteins purified from *E. coli* cells. Here, we are able to take advantage of our photodepletion corrections to directly measure NTF2 homo-oligomerization in the living test tube of *E. coli*.

Figure 3.7A shows the normalized brightness of NTF2 in *E. coli* for increasingly longer data segments (triangles), plotted as a function of the lost intensity or bleaching fraction. In our experience, 25% bleaching presents enough of a bleaching curve for reliable fitting without requiring the excessive data acquisition times required to approach complete bleaching. The data is then fit (solid line) using the non-stationary signal theory to acquire the corrected brightness of NTF2 for the given *E. coli* cell. For Figure 3.7B, we repeat this procedure for multiple *E. coli* expressing NTF2 (squares) as well as for *E. coli* expressing EGFP (triangles), to act as a control. The results of these experiments are plotted as brightness versus protein concentration. For the sake of simplicity, we will not discuss that in addition to photodepletion correction, geometric corrections need to be considered to achieve absolute concentration and brightness values. The foundation underlying geometric corrections are introduced in the next chapter. We conclude these

results by observing that over the concentration range measured, NTF2 shows good brightness doubling, indicating that NTF2 is dimeric in *E. coli*.

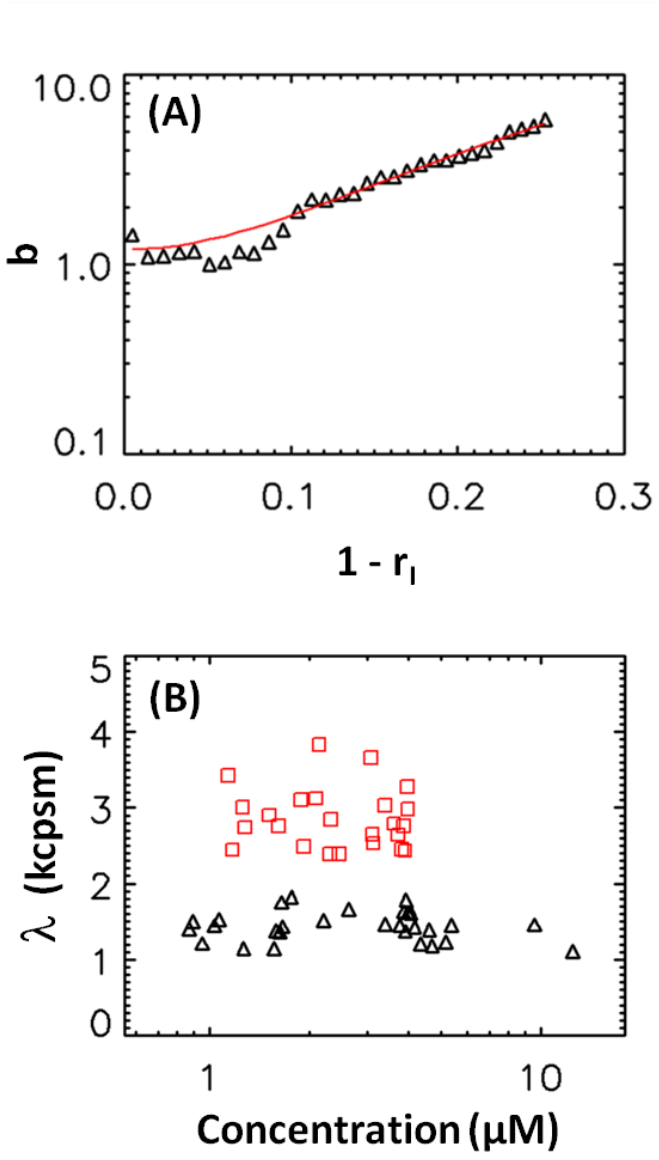

Figure 3.7 **NTF2 Oligomerization in *E. coli*.**

(A) The normalized brightness of EGFP-labeled NTF2 in an *E. coli* cell plotted for increasingly longer data segments versus lost intensity fraction, as opposed to the remaining intensity fraction. These experiments and analysis were performed by Kwang Ho Hur. As before, longer data segments suffer from an increased photodepletion bias. By fitting the data to theory (solid line), we are able to recover the unbiased brightness. (B) Performing this type of experiment for many *E. coli* cells expressing both NTF2 (squares) and EGFP (triangles), we are able to plot brightness of these two proteins in *E. coli* versus their concentration. The doubling of NTF2 brightness over EGFP indicates that NTF2 is dimeric.

3.5 Discussion

We introduced the term photodepletion to refer to experiments where the primary effects of photobleaching, namely reduction in diffusion time and brightness, are negligible. In this case only the cumulative reduction in the concentration of fluorophores affects the interpretation of the data. While the depletion of fluorophores is insignificant when dealing with a large sample volume, it is a relevant issue when measuring cells, droplets, and other closed systems with volumes on the order of picoliters.

Fluorophore depletion leads to a bias in the calculated brightness because the decreasing fluorescence signal is non-stationary. Formulas that account for the bias have been developed and experimentally verified. While we focused here on brightness only, the fluctuation amplitude $g(0)$ is similarly affected by photodepletion, and it is straightforward to arrive at equations that describe the bias on $g(0)$ by repeating the derivation using $g(0) = Q\langle k \rangle$. Theory and data show that the bias introduced by fluorophore depletion increases with concentration. This result is especially important for cell experiments where measurements are often performed at much higher fluorophore concentrations than typically found in aqueous solution.

The easiest way to avoid the bias due to photodepletion is to analyze short segments of the data as illustrated in Figure 3.1. A practical method of determining the proper segment size is to start with a long segment that is successively shortened until the calculated brightness of the segment is constant. Further shortening of the segment is counterproductive as the signal-to-noise ratio of the calculated brightness depends on the length of the data segment. If the experiment demands a higher signal-to-noise ratio, the length of the segment must be increased. In this case any bias introduced by lengthening the segment must be accounted for by the existing theory.

Long measurement times in cells are desirable because they improve the signal-to-noise ratio, which is the crucial factor determining if protein mixtures or interactions are detectable. But photodepletion effects become important when increasing the data acquisition time beyond a critical value, which depends on the photostability of the

protein. While EGFP is relatively photostable, other proteins, such as red fluorescent proteins, are much more photolabile. Thus, photodepletion of red fluorescent proteins is readily observed in two-photon excitation measurements.[47] 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 artifacts introduced by photodepletion.

FFS experiments often require the repeated measurement of the same cell under different conditions, such as after adding ligand or at different excitation wavelengths.[7] If these additional FFS or imaging measurements introduce bleaching, the brightness of the remeasured sample is affected by the degree of photodepletion present, which leads to a bias in the interpretation. To avoid this problem it is possible to measure the brightness as a function of the remaining intensity fraction r_I as shown in Figure 3.6. The slope plus one of the data equals the undistorted brightness even if the sample experienced some bleaching, as long as the original intensity of the sample is known. The slope recovered from Figure 3.6 predicts a protein complex containing two EGFP molecules. The same principle also works for mixtures of different oligomeric states. The apparent brightness *b* of the mixture is, according to Eq. 3.12, equal to one plus the slope of the $b(r_I)$ curve.

In this chapter we consider photodepletion of a single type of fluorophore, such as EGFP. While this assumption is adequate in many cases, situations exist where more than one species bleaches. In the latter case, the theory needs to be expanded to account for the individual bleaching rate of each species. We further assumed a simple two-state bleaching model with a brightness that is ε in the fluorescent state *A* and which drops to zero in the bleached state *D*. The bleaching of EGFP is not described by this simple model (inset of Figure 3.5), which predicts a constant brightness for monomeric EGFP. The experimentally observed drop in brightness for $r_I > 0.3$ is consistent with a three-state bleaching model, where state *A* is converted into an intermediate state *B* before bleaching occurs. The observed drop in Figure 3.5 requires that the brightness of state *B* is lower than that of state *A*. Because the contributions to fluctuations by a single

molecule in state *B* are lower than that of a molecule in state *A*, the presence of state *B* is not observed until a majority of the proteins are converted. For this reason our analysis with the simple bleaching model is successful as long as we limit ourselves to experimental situations where $r_1 > 0.3$. The presence of more than one brightness state of EGFP is supported by additional bleaching experiments. Patterson et. al. measured the properties of EGFP and other fluorescent proteins by irradiating small isolated droplets of protein solution.[53] Another study used immobilized EGFP on a surface to measure the decay characteristics of one- and two-photon photobleaching.[54] In both of these studies, a multi-exponential model was used to describe the photobleaching data suggesting the existence of multiple states. Didier and Bardou developed a Lévy statistics model of photobleaching that suggests the existence of an unstable brightness state with an infinite lifetime.[55] Finally, it is possible to formulate a more complex model that includes intermediate brightness states. Comparing the predictions of such a model with the data presented in the inset of Figure 3.5 would provide information about the presence of intermediate states and their brightness, which is hard to obtain by other methods.

The success of the photodepletion analysis is the ability to quantitatively establish accurate brightness and absolute concentration of fluorescent particles in living *E. coli* and other microorganisms or microdroplets. The ability of FFS to perform brightness titrations is a powerful tool, but lacks meaning if it cannot be correctly related to protein stoichiometry and concentration. Photodepletion theory brings the full strength of FFS and brightness analysis to the world of *E. coli* and other small sample compartments.

4. Z-scan FFS

FFS, as introduced so far, has not included any discussion of the shape of the sample or of the excitation volume geometry. In this chapter, the contents of which have been previously published,[56,57] we will discuss two common situations where geometry – either of the excitation light or the sample – leads to a biased or misinterpreted brightness. The first issue is related to instrumentation and arises from the index of refraction mismatch and the spherical aberrations typically found when performing FFS experiments with an oil-immersion objective.[58,59] The focus of the chapter is the second issue dealing with thin layer sample geometries, like cell cytoplasm. When the focal volume of the excitation light is taller than the sample, the brightness of the sample appears larger than the actual value.

4.1 Introduction

The application of brightness analysis in cells requires caution because the cellular environment is far more complex than that in aqueous solution. Here we focus on the thickness of the cell and its influence on brightness. We show that this brightness artifact appears once the sample thickness approaches the axial size of the observation volume. The thickness of spreading cells, which are widely used in fluorescence microscopy, is sufficiently thin that brightness bias is of concern. In fact, we experimentally observe a thickness dependent bias of brightness in the cytoplasm of COS-1 cells. This bias is problematic, because it obscures the correct interpretation of protein interactions from the FFS data.

We introduce a new data acquisition and analysis protocol to eliminate the thickness-dependence of FFS experiments and to restore the quantitative interpretation of brightness. Our approach relies, in addition to the FFS measurement, on a z-scan of the excitation light across the sample. Z-scan approaches to FFS have previously been used to study the diffusion of proteins in lipid bilayers and cell membranes.[60–62] Here we use the z-scan to gain information about the geometry of the sample which is subsequently incorporated into FFS theory. We further characterize the axial shape of the excitation

light of the two-photon microscope and introduce a point spread function model to quantify z-scan FFS measurements. We experimentally verify the technique on cells expressing EGFP and demonstrate the elimination of the brightness bias in the cytoplasm. In addition, we apply z-scan FFS to investigate the oligomerization of the nuclear transport factor 2 (NTF2) in the cytoplasm of COS-1 cells.

4.2 Demonstration of brightness bias

The enhanced green fluorescent protein (EGFP) expressed by cells is found in both the cytoplasm and in the nucleus. We have previously shown by brightness analysis that EGFP exists as a monomer in the nucleus.[6] Conventional FFS measurements were taken at different positions in the cell, starting in the nucleus and moving outwards to the edge of the cell, as depicted in Figure 4.1A. For each position, the excitation volume was focused at mid-height in the sample. The nucleus provides a thick region which, for many cell lines, completely contains the excitation volume and is free from bias. Beyond the nucleus, sample height falls off quickly as measurements are taken further out into the cytoplasm. We discuss later how the thickness of cells is experimentally determined. Figure 4.1B plots the normalized brightness of EGFP as a function of sample thickness. Normalized brightness, $b = \lambda / \lambda_{monomer}$, is determined by dividing the measured brightness λ by the brightness of a single fluorophore tag in the nucleus; $b = 1$ corresponds to a monomeric protein and $b = 2$ corresponds to a dimeric protein. In the thicker parts of the cell, FFS measurements identify monomeric EGFP ($b = 1$). As the cytoplasmic regions get thinner, the brightness values increase. At the thinnest part of the cytoplasm, conventional FFS data indicate dimeric EGFP ($b \approx 2$). The data suggest a thickness dependent dimerization of the protein. We show in the following that this result is an artifact introduced by ignoring the finite geometry of the sample in conventional FFS analysis.

Brightness involves a spatiotemporal average of the photons emitted by a single fluorophore. Conventional FFS assumes that the fluorophore visits all regions of the

excitation volume with equal probability. This situation is realized by a sample which contains the entire excitation volume, which we refer to as the infinite sample case. The assumption of an infinite sample is problematic when considering cellular FFS experiments in which part of the excitation volume resides outside of the thin sample and excites no fluorophore. Focusing on the mid-section of a thin sample excludes access of the fluorophore to the outer edges of the excitation light, where the intensity is lowest. As a consequence, the fluorophore spends most of the time in the higher intensity, central areas of the excitation volume, and the resulting spatial average skews the brightness upwards. This brightness artifact poses a significant challenge for the quantitative interpretation of protein interactions in the cytoplasm.

4.3 Formulation of z-scan FFS

It is not difficult to modify the theory to account for finite sample geometry, but two practical difficulties are encountered. A priori, the specific geometry of the sample is frequently unknown, particularly in the case of cell experiments. Additionally, the exact position of the excitation volume relative to the sample along the optical- or z-axis is difficult to obtain. We will perform z-scan measurements to address these challenges and develop a quantitative method for measuring brightness in thin geometries.

Scanning the beam uniformly along the z-axis extracts information about the z-geometry of the sample via the resulting intensity trace. Figure 4.2A shows the fluorescence intensity trace for a z-scan through the cytoplasm of a cell expressing EGFP. A single intensity peak is observed, which represents the fluorescent layer of cytoplasm. Because the z-scan is based on a systematic movement of the excitation beam across the sample, modeling of the intensity trace determines the thickness and the z-position of the beam relative to the sample as a function of time. The overlaid cartoon in Figure 4.2A illustrates the position of the excitation light with respect to the sample.

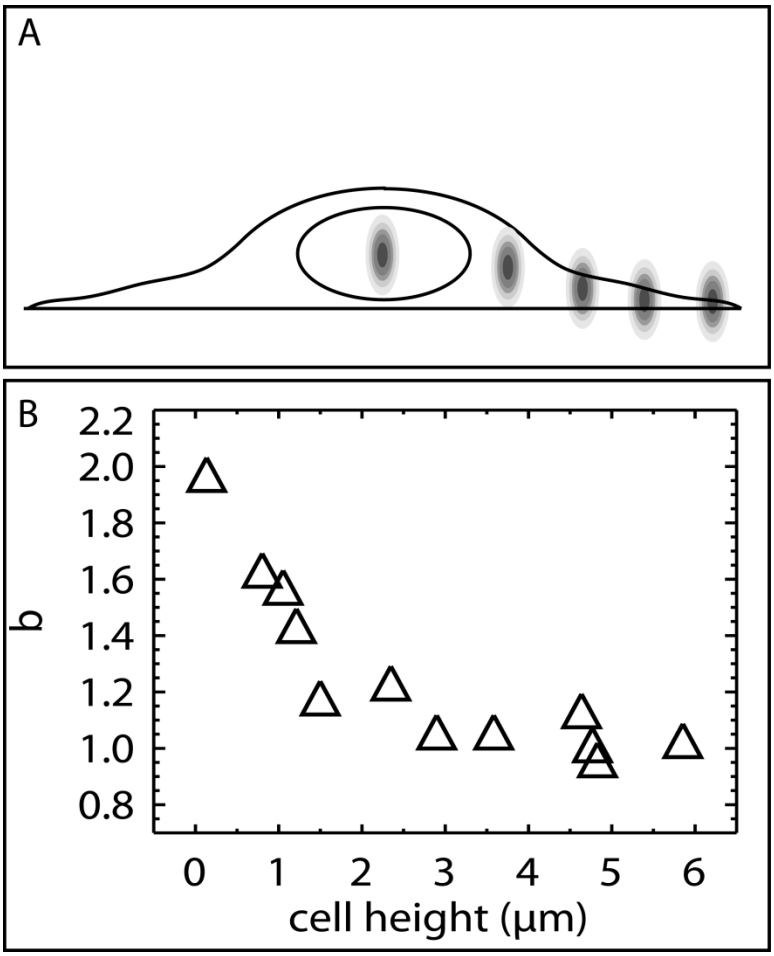

Figure 4.1 **Thin samples bias brightness in conventional FFS data**.

(A) Conventional FFS assumes a homogenous sample environment that contains the entire excitation volume. These conditions are easily met in the center of most cell nuclei. Measurements in thin geometries like the cytoplasm return biased results. (B) A cell containing EGFP is measured at different thickness positions. Normalized brightness, *b*, appears to increase by a factor of 2 as the sample thickness decreases. This is an artifact of conventional FFS measurement in thin geometries.

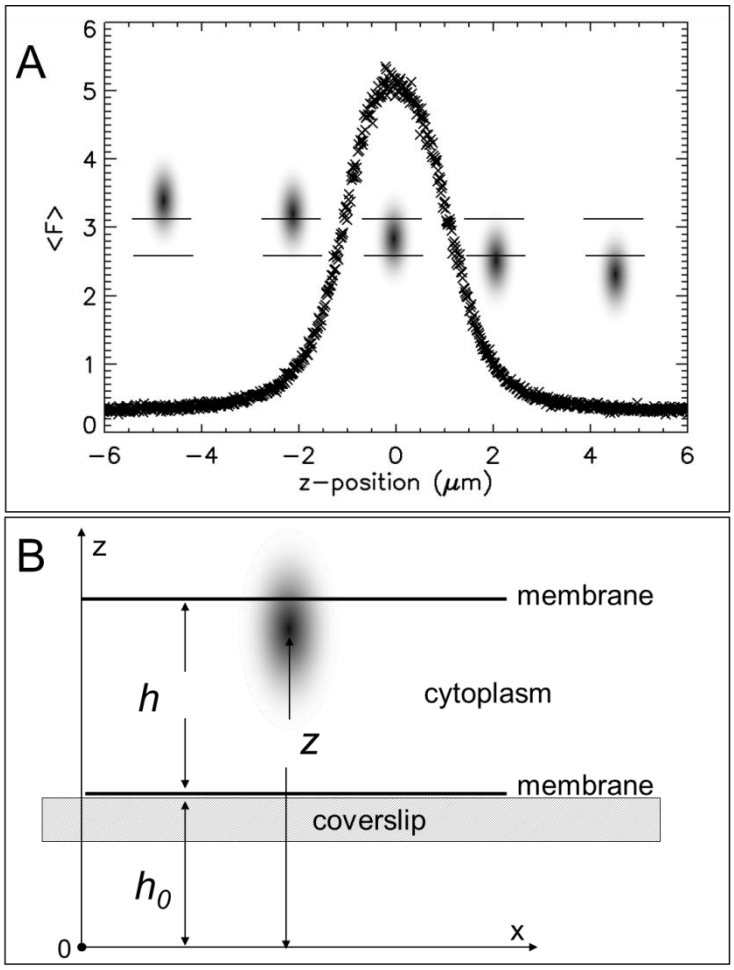

Figure 4.2 **Z-scan approach to finite sample geometries of the living cell**.

(A) A COS cell containing EGFP was z-scanned through the cytoplasm. The average fluorescence intensity is plotted as a function of z-position. The overlaid cartoon indicates the position of the excitation volume with respect to the cytoplasmic slab. A z-scan acquires information for many z-positions and provides an accurate description of the fluorescent signal from a thin sample which a single-point conventional FFS measurement cannot. (B) A cytoplasmic cell section is located a height h_0 above the start of the z-scan and has a thickness h. The position of the excitation light is described by the value of z.

We model each cytoplasmic section enclosed by the observation volume as a fluorescent slab. The diameter of a typical COS cell is ~50 μm in the x-y plane with a thickness that decreases from the center (~5 μm) to the periphery. The intensity of the excitation light is azimuthally symmetric. The axial direction is parallel to z, and the radial direction lies in the x-y plane (Figure 4.2). Compared with the size of the cell along the x-y plane, the radial beam waist (~0.5 μm) of the excitation light is small enough that the change in thickness of the cytoplasm section within the excitation light is negligible. Therefore, we model the cytoplasm section as a slab of constant height. The z-scan begins well below the cell such that sample is first encountered at a height h_0. The cytoplasmic slab has a height h, and the position of the excitation light is described by the z-coordinate of its center.

We now introduce the geometry function $S(z)$ into FFS theory to account for a confined sample environment. For the specific example of a slab (Figure 4.2B) as described above, the function $S(z)$ is given by

$$S(z;h_0,h)=\begin{cases}1, & h_0<z<h_0+h\\ 0, & \text{otherwise}\end{cases} \tag{4.1}$$

The parameter h_0 is important for fitting experimental data, because it determines the difference between the start position of the scan and the start position of the slab. However, including the offset parameter h_0 into the theory development unnecessarily complicates the notation. We therefore set h_0 to zero and hereafter leave it out of the notation. It is straightforward to include the offset in the final equations by a linear transformation of z to $z - h_0$.

We define the volume of the point spread function (PSF) raised to the r-th power by

$$V_{PSF^r}(z)=\int \mathrm{PSF}^r(\rho,\zeta)\cdot S(z+\zeta)\,d\Omega \tag{4.2}$$

The variables ζ and ρ represent the local coordinates measured with respect to the center of the PSF. We explicitly make use of the cylindrical symmetry of the PSF with $d\Omega = 2\pi\rho\, d\rho\, d\zeta$. The position ζ in the local PSF-coordinate system translates to $z+\zeta$

in the experimental setup coordinate system. In addition, we adopt the convention that in the absence of explicit integration limits variables are integrated over all space.

This definition of $V_{PSF^r}(z)$ explicitly includes the geometry function $S(z)$ to reflect that only part of the PSF is accessible to the fluorophore. The conventional FFS theory is recovered by choosing $S(z) = 1$, which corresponds to the infinite sample case, where the complete PSF is accessible. We explicitly refer to the infinite sample case by using the subscript ∞. Thus, $V_{\infty,PSF}$ is the volume of the PSF for the infinite case. The value of $V_{PSF^r}(z)$ for the restricted geometry depends on $S(z)$ and the focus position *z* of the PSF. The fluorescence intensity $F(t,z)$ of a solution of molecules with brightness λ confined in a finite sample is given by[4]

$$F(t,z) = \lambda \int PSF(\rho,\zeta) \cdot c(\rho,\zeta,t) \cdot S(z+\zeta) d\Omega \tag{4.3}$$

with concentration $c(\rho,\zeta,t)$ at position (ρ,ζ) and at time *t*. The geometry function $S(z)$ serves to restrict the sample size.

We now specifically consider the slab-geometry defined by Eq. 4.1. The PSF volume depends for this case on the height *h*, $V_{PSF^r}(z;h)$. For a stationary process the average fluorescence and its variance is determined from Eq. 4.3 as

$$\langle F(z;h) \rangle = \lambda \langle c \rangle V_{PSF}(z;h) \tag{4.4}$$

and

$$\left\langle \Delta F^2(z;h) \right\rangle = \lambda^2 \cdot \langle c \rangle \cdot V_{PSF^2}(z;h) = \gamma_2(z;h) \cdot \lambda^2 \cdot \langle c \rangle \cdot V_{PSF}(z;h) \,. \tag{4.5}$$

While the derivation of Eq. 4.4 is straightforward, Eq. 4.5 can be derived using the cumulants of the fluorescence[32] or by following the method used by Thompson.[39] Eq. 4.5 also introduces a geometry dependent shape-factor

$$\gamma_2(z;h) = \frac{V_{PSF^2}(z;h)}{V_{PSF}(z;h)} \,, \tag{4.6}$$

which is a generalization of the conventional shape factor.[39] All of the above equations are equivalent to the equations of conventional FFS, which are obtained by replacing V_{PSF} by the value for the infinite sample case, V_{∞,PSF^r}.

Brightness and concentration are independent of geometry and thus are good parameters for characterizing the sample of interest. The mean fluorescence intensity $\langle F \rangle$ is an easily measured experimental quantity which is used to determine the size of the slab. Furthermore, we use Mandel's Q parameter[26,49]

$$Q(z;h) = \frac{\left\langle \Delta F^2(z;h) \right\rangle}{\left\langle F(z;h) \right\rangle} = g(0) \cdot \left\langle F \right\rangle = \gamma_2(z;h) \cdot \lambda \,, \tag{4.7}$$

because it is a directly measurable quantity that relates to the brightness through the model-dependent shape factor γ_2. Q is determined by the ratio of the first two moments of the fluorescence intensity, or, alternatively, by the product of the fluctuation amplitude $g(0)$ and the average fluorescence intensity as measured in an FCS experiment.

While we discussed the slab-model in the above section, it is straightforward to extend the model to other geometries by evaluating Eqs. 4.4 to 4.7 for other functions of $S(z)$. Of particular interest is a sample thin enough that it resembles a layer with negligible thickness. We refer to this as a delta-layer sample, which has the corresponding shape-factor

$$S(z) = \delta(z) \,. \tag{4.8}$$

The average fluorescence intensity for a delta-layer is

$$\left\langle F(z) \right\rangle = \lambda \cdot \left\langle \sigma \right\rangle \cdot \mathrm{RIPSF}_1(z) \tag{4.9}$$

where $\left\langle \sigma \right\rangle$ is the area concentration of fluorophores within the delta-layer and $\mathrm{RIPSF}_r(z)$ defines the radially-integrated PSF raised to the r-th power

$$\mathrm{RIPSF}_r(z) = \int_0^\infty \mathrm{PSF}^r(\rho, z) \cdot 2\pi\rho d\rho \,. \tag{4.10}$$

Thus, $\mathrm{RIPSF}_1(z)$ of Eq. 4.9 represents the radially-integrated PSF, which we also refer to as $\mathrm{RIPSF}(z)$ in this paper.

4.4 PSF model

The z-scan fluorescence intensity curve given by Eqs. 4.4 and 4.2 involves the convolution of the PSF and the sample geometry, as it is this overlap volume which

generates the fluorescent signal. To usefully model the convolution, it is first necessary to establish an accurate PSF of the experimental setup from fluorescence intensity curves. However, convolution of the PSF with the sample thickness obscures the features of the experimental PSF. To avoid this complication we choose a delta-layer sample, which according to Eq. 4.9 results in a fluorescence intensity curve that is directly proportional to the radially-integrated PSF, $\mathrm{RIPSF}(z)$. This procedure directly evaluates the quality of a PSF-model in reproducing experimental data.

For this study, z-scans are performed on a cytoplasmic section thin enough that it resembles a delta layer. In order to identify a cytoplasmic delta layer, we take a series of z-scans across the cell moving towards increasingly thinner sections of the cytoplasm. While intensity z-scans in the nucleus are broad, the width of the curve narrows as the cell height decreases. Below a certain thickness, the shape of the intensity z-scan ceases to change. When this condition is met, we have reached an effective delta layer.

FFS theory employs simple model functions in order to approximate the experimental PSF. In the following, the parameter n is used to account for the difference in the PSF for two-photon excitation (TPE) and one-photon excitation (OPE). We define the PSF models where $n = 1$ for OPE and $n = 2$ for TPE. Two PSF-models widely used are the 3-dimensional Gaussian (3DG) PSF

$$PSF_{3DG}(\rho,\zeta) = \exp\left(-2n\left(\frac{\rho^2}{w_0^2}+\frac{\zeta^2}{z_0^2}\right)\right) \tag{4.11}$$

and the Gaussian-Lorentzian (GL) PSF

$$PSF_{GL}(\rho,\zeta) = \left(\frac{w_0}{w(\zeta)}\right)^{2n} \exp\left(-2n\frac{\rho^2}{w^2(\zeta)}\right) \tag{4.12}$$

with

$$w(\zeta) = w_0\left(1+\frac{\zeta^2}{z_0^2}\right)^{1/2} \tag{4.13}$$

The radial and axial beam waists of the PSF are given by w_0 and z_0.While knowledge of the PSF is sufficient for TPE, the description of OPE is further complicated by the need to account for the effect of a pinhole on the collected fluorescence emission.

This effect can be approximated by multiplying the excitation PSF by a collection profile function $\Omega(\rho,\zeta)$.[39,63] Because our experimental work uses TPE we analyze the experiments using a PSF-model with $n = 2$.

We now evaluate the suitability of these two model functions for z-scan FFS. It is straightforward to integrate each PSF over ρ to determine their RIPSF. Figure 4.3 contains the intensity z-scan of an effective delta layer based on a very thin cytoplasmic section of a COS cell expressing EGFP. Figures 4.3A and 4.3B display fits of the intensity curve to the RIPSF using the 3DG and GL model, respectively, plotted with a linear and logarithmic intensity axis to emphasize the tail of the radially-integrated PSF. Simple visual inspection of the fits reveals that neither the 3DG nor GL model lead to a correct description of the experimental RIPSF.

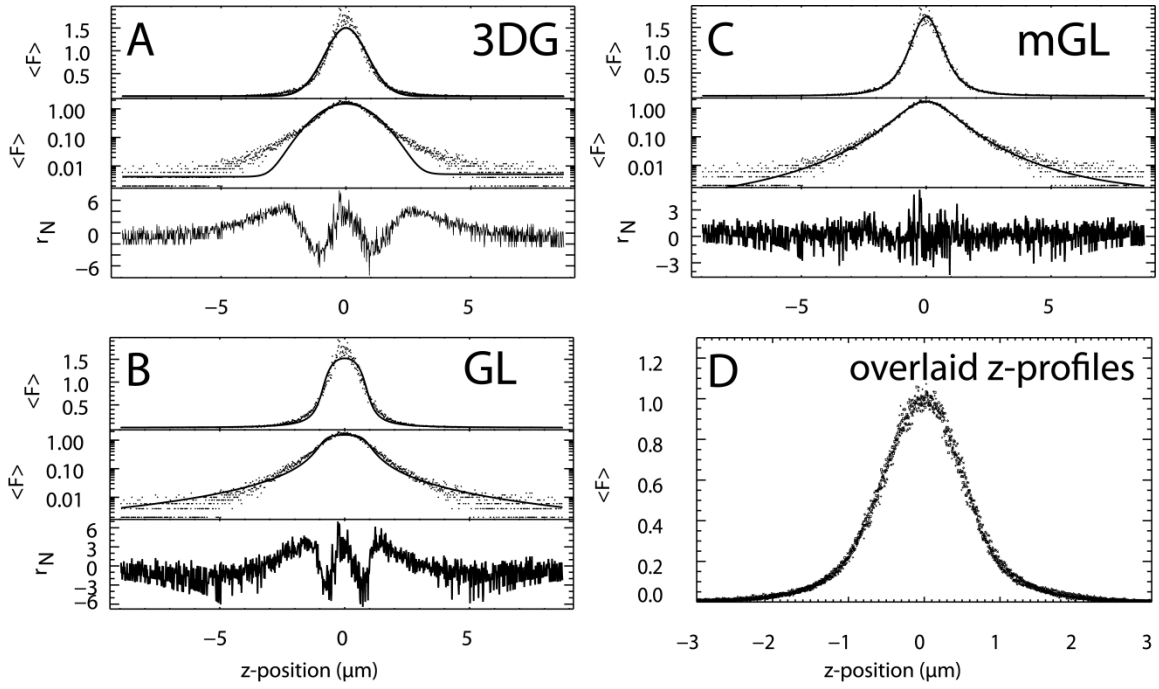

Figure 4.3 **PSF models.**

A z-scan intensity trace is acquired from a thin section of cytoplasm in a cell containing EGFP. The same intensity trace is fit with three different point spread function (PSF) models. The intensity fits are plotted on a linear (upper panel) and semi-log scale (middle panel), along with the normalized residuals (lower panel). (A) The 3D Gaussian model and (B) the Gaussian-Lorentzian model are two PSF models commonly used in conventional FFS, but do not accurately fit z-scan data. (C) We show that our modified Gaussian-Lorentzian (mGL) model permits the more accurate fit necessary to successfully implement the z-scan technique. The mGL fit parameters for this data are z_0=1.1 ± 0.048 μm and y=1.9 ± 0.056. (D) Z-scans were performed across thin sections of five cells and normalized to a peak intensity of 1. The overlaid data demonstrate that the z-scan profile is stable and repeatable.

It is desirable to establish a PSF model that faithfully reproduces the z-scan intensity curves. Such a model is a prerequisite for employing a chi-square goodness-of-fit test to accept or reject the presence of protein interactions from brightness data. A far more sophisticated approach of modeling the PSF involves evaluating the EM-wave propagation through the microscope objective.[63,64] However, such an approach requires extensive modeling and is numerically expensive. Ideally, we would like to establish a heuristic model function which has a simple analytical form and successfully reproduces the experimental data. Figure 4.3 shows that both the 3DG and GL fits have strong correlations in their residuals. These correlations are typical for all cells scanned (not shown). Correspondingly, the reduced chi-squared values for 3DG and GL fits are consistently greater than 4. Closer inspection of the fits reveals that the tail of the 3DG function decays too rapidly, while the tail of the GL function decays too slowly. This observation motivated the introduction of a modified Gaussian-Lorentzian (mGL) PSF

$$PSF_{mGL}(\rho,\zeta) = \left(\frac{w_0}{w(\zeta)}\right)^{(1+y)n} \exp\left(-2n\frac{\rho^2}{w^2(\zeta)}\right) \tag{4.14}$$

The mGL-PSF contains the additional parameter *y*, which varies the exponent of the first factor of the GL-PSF. This heuristic model allows us to tune the steepness of the z-decay of the PSF while maintaining a Gaussian cross-section of the PSF. Notice that for $y = 1$, the modified GL-PSF reduces to the original GL-PSF. The mGL-PSF successfully fits the experimental intensity curve (Figure 4.3C) with largely uncorrelated residuals and a reduced chi-square close to 1. The fit-parameters are $z_0 \approx 0.95$ μm and $y \approx 1.9$. Figure 4.3D shows overlaid z-scans through very thin slices of five different cells. The peak intensities have been normalized to 1 to demonstrate that the RIPSFs all have the same shape and are reproducible. Thus, we will use the mGL-PSF for modeling of all further z-scan FFS experiments.

4.5 Oil- and water-immersion objectives

In a practical consideration, it is important to consider what objectives are suitable for the technique before attempting z-scan experiments. FFS experiments are typically

undertaken using high numerical aperture water- or oil-immersion objectives, but properly designed water objectives are required for z-scan FFS, because oil-immersion objectives exhibit z-dependent changes in the PSF. We demonstrate that FFS parameters, such as intensity, brightness and diffusion, have a strong focal height dependence when measured by an oil-immersion objective making them unsuitable for z-scan FFS. Even setting z-scanning aside, this oil-immersion objective height dependence is not optimal for standard measurements which may involve focus changes as different locations in a sample are observed. Therefore, we also present a heuristic correction for brightness measurements when employing an oil-immersion objective.

High numerical aperture (NA) objectives require a substance to fill the gap between the objective and the coverslip, so that NAs of greater than 1 can be obtained. Immersion objective must therefore be designed carefully with regards to index of refraction. The relevant layers are the immersion fluid (index of refraction n_1), the microscope glass (n_2), and the sample (n_3) and objectives are designed with the criterion that $n_1 = n_3$.[65] The condition is broken when using an oil-objective to measure aqueous solution. For short focus heights (a few micrometers), the light path is not long enough to experience significant divergence of the rays, and so there isn't much effect on the focal volume. As a result, oil-immersion objectives may be used with relative impunity for FFS measurements in the region close to the coverslip. However, the index of refraction mismatch leads to spherical aberrations which become seriously problematic when trying to focus deeper into solution, living cells, or tissue. The effect is a distortion of the PSF and so a change in the recovered FFS parameters. Other studies have demonstrated that the PSF is stretched dramatically in the axial direction and significantly in the lateral direction[66] which has a very noticeable effect on brightness. We recall that brightness λ is defined as:

$$\frac{\langle F \rangle}{\bar{N}} \propto \lambda = \frac{\langle F \rangle G(0)}{\gamma_2}. \tag{4.15}$$

$\langle F \rangle$ is average intensity; $\bar{N}$ is the average number of molecules in the excitation volume and is inversely proportional to $G(0)$.

The enlarged PSF decreases the efficiency of the two-photon excitation and leads to decreasing intensity $\langle F \rangle$ as a function of focal depth. Figure 4.4A shows intensity measurements at increasing focal depths for Texas Red dye in aqueous solution, using a Zeiss 63x Plan Apochromat Oil-immersion objective (NA =1.4). The intensity effect is pronounced, with a 15% decrease within 5 μm and a 70% loss of intensity at a depth of 60 μm. Additionally, the larger PSF covers more area and excites a larger number of molecules. This is reflected in autocorrelation analysis through a decrease in the correlation amplitude $G(0)$ (Figure 4.4B). Similarly, it takes longer for a particle to cross the inflated excitation volume, resulting in increasing diffusion times as shown in Figure 4.4C. Since average brightness is reduced by both the decrease in intensity and the increase in number of molecules, brightness values are extremely sensitive to focal depth. Texas Red brightness values fall by 20% within 5 μm. Focal depth differences of ± 1 micrometer cause brightness deviations that fall within the 10% noise typically associated with brightness experiments in living cells. However, for samples beyond a few micrometers in thickness, the effects of the changing PSF cannot be ignored.

We note an observed relationship between average intensity and the correlation amplitude. Figure 4.4D shows $\lambda/\langle F \rangle^x$ plotted as a function of focal depth, where x is an experimentally determined parameter dependent on objective type, magnification and numerical aperture. For the Zeiss 63x oil objective, we find $x = 2.0 \pm 0.1$. In the case of a 100x oil objective (not shown), $x = 1.4 \pm 0.15$. Performing the above calibration on an objective provides a quick method for correcting the depth-dependent brightness bias in subsequent experiments.

We perform the same experiments as above but now use a Zeiss 63x C-Apochromat water-immersion objective (NA = 1.2). Note that the water objective comes equipped with a correction collar that is tuned to account for the thickness of the coverglass between the objective and the sample. This collar must be set to the proper position for the objective to function as designed.[67] Intensity (Figure 4.5A), correlation amplitude (Figure 4.5B), and diffusion time (Figure 4.5C) are essentially constant over the measured range of 60 μm. The brightness measured with this water-immersion objective is independent of depth as illustrated in the final panel (Figure 4.5D). Thus,

water-immersion objectives that have been corrected for spherical aberrations are advantageous for brightness experiments where the focal depth is changed. The rest of the experiments in this section are performed with such an objective to take advantage of FFS measurements that are independent of focal depth.

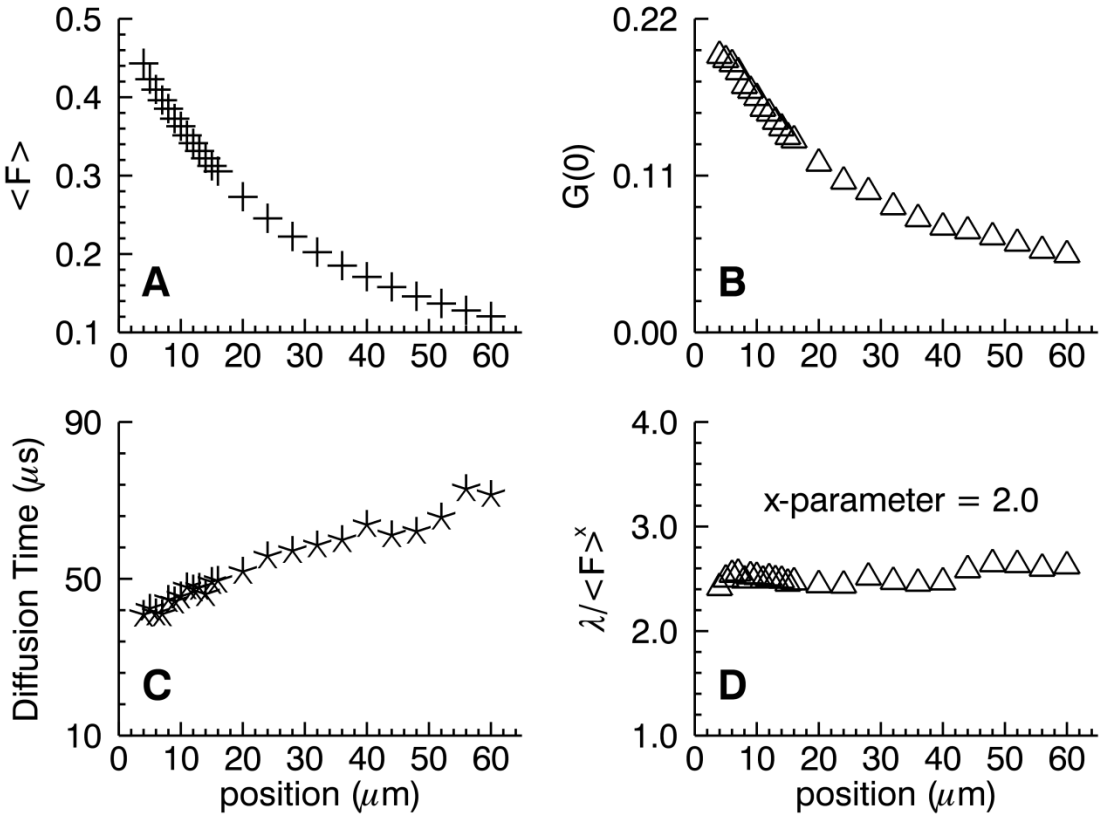

Figure 4.4 **Focal depth dependence of oil-immersion objective**.

A dye solution is measured with two-photon excitation FFS using an oil immersion objective. (A) The intensity falls off strongly as the laser is focused more deeply into the sample. This is the result of an index of refraction mismatch and spherical aberrations that change the shape of the excitation PSF. (B) The autocorrelation amplitude is also plotted as a function of penetration depth into solution. (C) The PSF stretches both in the lateral and axial directions, and fluorescent particles require a longer time to traverse the PSF. (D) Brightness values also decrease sharply as a function of depth, but a calibration experiment can be used to find a relationship between brightness and intensity for a given FFS instrument. For our setup, the brightness falls as the square of the fluorescence intensity.

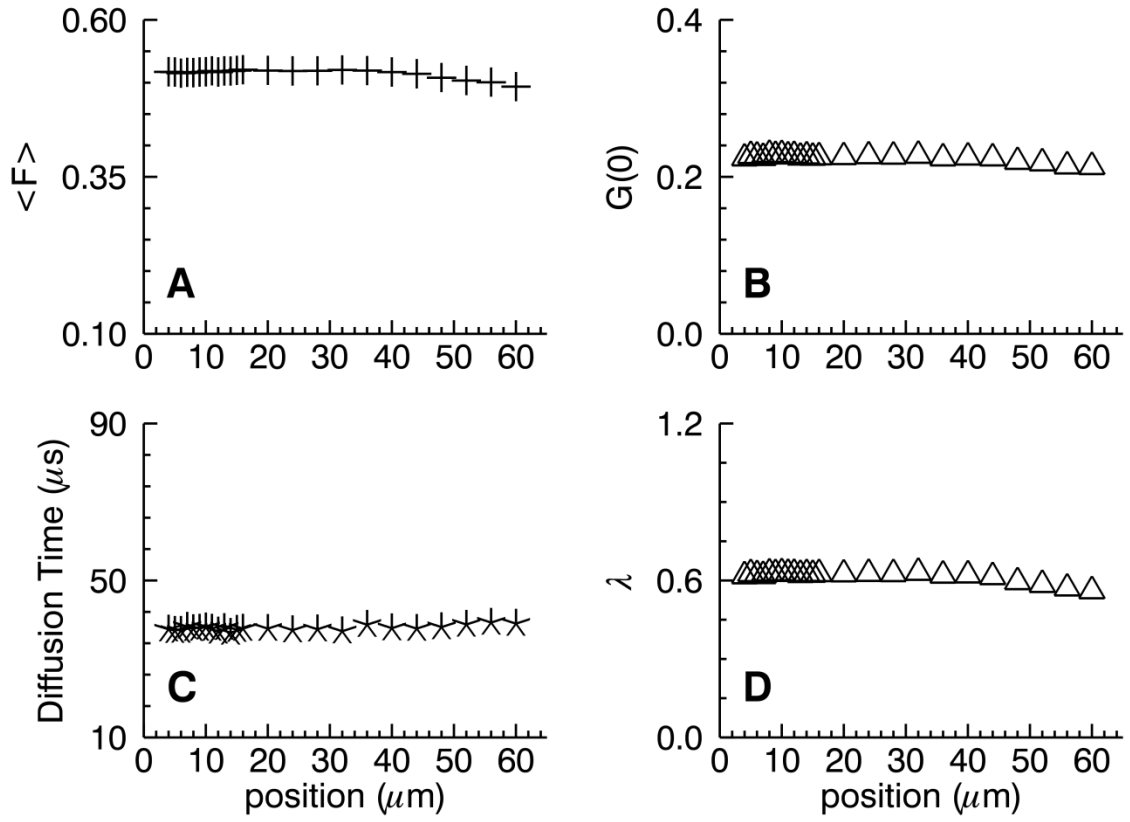

Figure 4.5 **Focal depth dependence of water-immersion.**

The dye solution experiment is repeated using a water-immersion objective. (A) Fluorescence intensity is shown to be independent of focal depth. (B) Autocorrelation amplitude G(0) and (C) diffusion time are also constant over focal depth. (D) Brightness is essentially constant throughout the 60-μm range measured, and is plotted directly as a function of focal depth.

4.6 Materials and methods

Experimental Setup. A mode-locked Ti:sapphire laser (Tsunami, Spectra-Physics, Mountain View, CA) serves as a source for two-photon excitation of a modified Zeiss Axiovert 200 microscope (Thornwood, NY) as previously described.[6] Each FFS measurement lasts 30 to 60 s and uses two-photon excitation of the sample at 905 nm. Excitation light is focused through a Zeiss 63x C-Apochromat water immersion objective (NA = 1.2). Control experiments (data not shown) confirm that 0.3 mW excitation power is sufficiently low to avoid saturation and photobleaching effects. Photon counts were detected with an avalanche photodiode (APD) (Perkin-Elmer, SPCM-AQ-14) and recorded by a data acquisition card (ISS, Champaign, IL), which stores the complete sequence of photon counts using sampling frequencies ranging from 20 to 200 kHz. The photon counts were analyzed with programs written in IDL 6.0 (Research Systems, Boulder, CO).

Z-scan Setup. Intensity z-scans were obtained using a PZ2000 piezo stage (ASI, Eugene, OR) to move the sample in the z-direction. Scan voltages were controlled by an Agilent 33250A arbitrary waveform generator (Agilent Technologies, Santa Clara, CA) running a linear ramp signal with a frequency of 100 mHz and a peak-to-peak amplitude of 1.2 V. This voltage corresponds to an axial travel of 24.1 μm.

Expression Vectors, Cell Lines, and Cell Measurements. A tandem dimeric EGFP ($EGFP_2$) was constructed as previously described.[6] NTF2 was amplified from human NTF2 (Genebank 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 pEGFP-C1 plasmid (Clontech, Mountain View, CA). COS-1 cells were obtained from American Type Culture Collection (ATCC) and maintained in 10% fetal bovine serum (Hyclone Laboratories, Logan, UT) and DMEM media. Transfection was carried out using TransFectin reagent (Bio-Rad, Hercules, CA) according to the manufacturer's instructions 24 hours before measurement. Cells were subcultured into eight-well coverglass chamber slides (Nagle Nunc International, Rochester, NY) with the media

exchanged for Dulbecco's PBS with calcium and magnesium (Biowhittaker, Walkersville, MD) immediately before measurement.

4.7 Z-scan fluorescence intensity curves of cells

We demonstrated that the mGL-PSF describes the z-scan intensity curve for a delta-layer geometry. Theoretically, we expect that knowing the radially-integrated PSF is sufficient to describe any sample geometry $S(z)$, because the volume of PSF^r is related to RIPSF via

$$V_{PSF^r}(z) = \int \mathrm{RIPSF}_\mathrm{r}(\zeta) \cdot S(z+\zeta)\, d\zeta \,. \tag{4.16}$$

Thus, the z-scan fluorescence intensity curve may be written as

$$\langle F(z;h) \rangle = \lambda \langle c \rangle \int \mathrm{RIPSF}(\zeta) \cdot S(z+\zeta)\, d\zeta \,. \tag{4.17}$$

To test our prediction experimentally, multiple intensity z-scans are taken at different positions in a single cell. Each cell position provides a different sample thickness, but brightness and concentration are expected to be identical within a given cell. Figure 4.6 shows eight intensity z-scans taken from a single cell expressing EGFP. The locations of the scans are roughly: {0} at the center of the nucleus, {1} next to the nuclear envelope at the nucleoplasmic side, {2} next to the nuclear envelope at the cytoplasmic side, {3} – {6} progressing radially outward through the cytoplasm, and {7} near the outer edge of the cell. Because EGFP is uniformly distributed in the cytoplasm and nucleus, all eight scans probe a geometry with a single fluorescent layer. The eight intensity z-scan experiments are simultaneously fit to a slab model with a variable thickness for each scan, but globally linked PSF-parameters z_0 and y. The global fit resulted in a reduced chi-square of 1.2 with PSF-parameters $z_0 = 0.986 \pm 0.009$ μm and $y = 2.00 \pm 0.016$. The thicknesses for cell scans {0 – 7} are 6.50, 4.03, 2.46, 2.02, 1.40, 1.18, 0.95, and 0.30 μm. Note that samples thinner than ~0.5 μm become effective delta layers, and the slab model cannot accurately determine their thickness. The same experiment was repeated on other cells and on individual cell slices (data not shown) with similar results. The PSF parameters that best describe all z-scan experiments are $z_0 = 0.95 \pm 0.10$ μm and $y = 1.9 \pm 0.14$.

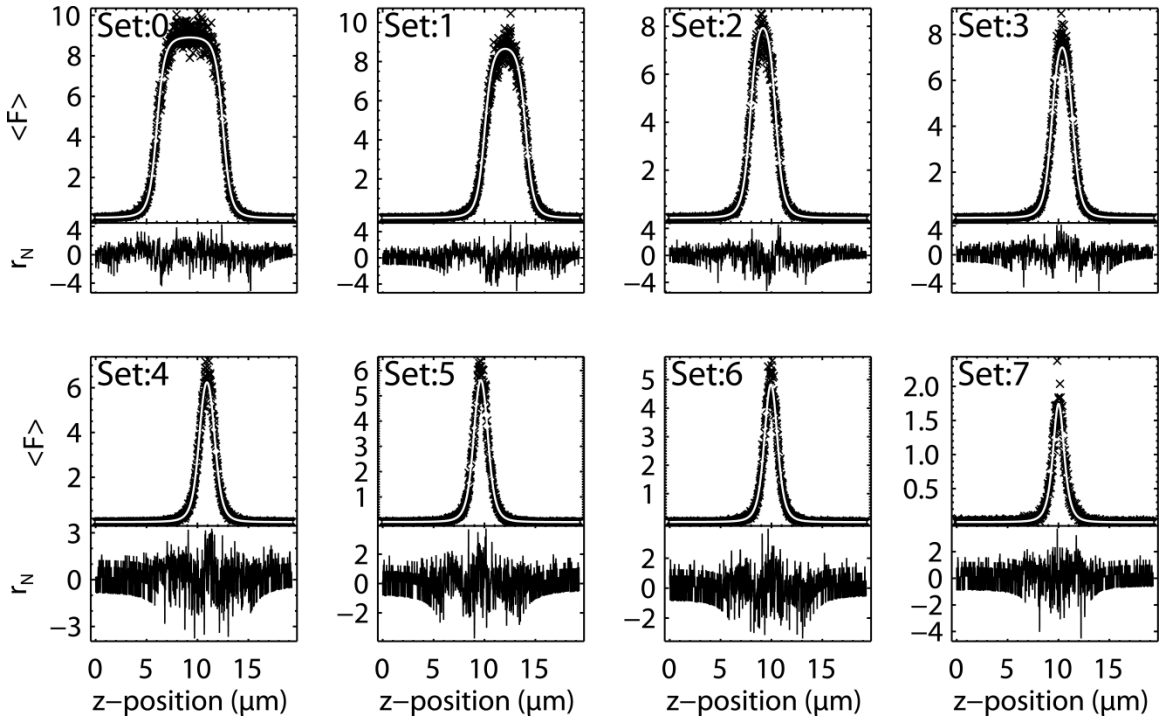

Figure 4.6 **Intensity z-scans fit with mGL PSF model.**

Eight intensity z-scans were taken in the same cell at different thicknesses. The locations are roughly: {0} at the center of the nucleus, {1} just inside the nuclear envelope, {2} just outside the nuclear envelope, {3} – {6} progressing radially outward through the cytoplasm, and {7} near the outer edge of the cell. All sections are fit using identical parameters for the modified GL PSF model, z_0=0.986 ± 0.009 μm and y=2.000 ± 0.016. Thicknesses for cell sections {0-7} are 6.50, 4.03, 2.46, 2.02, 1.40, 1.18, 0.95, and 0.30 μm. The overall reduced chi-squared is 1.19.

4.8 Gamma factor

We previously discussed that conventional FFS analysis leads to a spurious increase of the brightness of EGFP with decreasing sample thickness. This behavior is readily explained by considering Mandel's Q-parameter, which in conventional FFS is related to the brightness by $Q = \gamma_{2,\infty}\lambda$.[26] Inspection of Eq. 4.7 illustrates that the failure of conventional FFS lies in assuming a gamma-factor $\gamma_{2,\infty}$ of an infinite sample, which neglects the actual geometry-dependence of the gamma-factor $\gamma_2(z;h)$. Thus, the brightness increase reflects the ratio of $\gamma_2(z;h)/\gamma_{2,\infty}$, which is 1 for a thick sample, but grows larger as the sample thickness decreases and reaches a limiting value as the thickness approaches a delta layer. A suitable PSF-model needs to reproduce the experimentally observed change of $\gamma_2(z;h)/\gamma_{2,\infty}$ with thickness. Because γ_2 cannot be directly measured, we determine the Q-parameter according to Eq. 4.7 from the first two fluorescence intensity moments. The ratio of the gamma-factors is simply the ratio of the corresponding Q-parameters,

$$\frac{\gamma_2(z;h)}{\gamma_{2,\infty}} = \frac{Q(z;h)}{Q_\infty}. \tag{4.18}$$

We measure this ratio of the Q-parameter by taking FFS measurements in ten cells expressing EGFP. Each cell is measured at various positions while focusing the beam at mid-height on the cellular slab, which corresponds to $z = h/2$, where h is the height of the cellular slab. The experimental Q-parameter $Q(h/2;h)$ is divided by Q_∞, which is obtained from a measurement in the nucleus or in bulk solution. The experimental ratio for Q-parameters is identical to the experimental ratio of gamma-factors which are plotted in Figure 4.7 as a function of thickness *h*.

Next we derive equations (see published[56] supporting material) for the theoretical ratio $\gamma_2(h/2;h)/\gamma_{2,\infty}$ using Eq. 4.6 as a function of thickness *h* for the GL-, the 3DG-, and the mGL-PSF and fit the expressions to the experimental data. The fits are shown together with the experimental gamma ratio in Figure 4.7. Neither the 3DG- nor the GL-model fit the experiment. Only the mGL-PSF model reproduces the experimental data.

The gamma-factor ratio between a delta-layer and an infinite sample, $\gamma_2(h/2;0)/\gamma_{2,\infty}$, provides a sensitive test of the PSF model since its value can be determined from the experimental data and compared to theory. The gamma ratios of the 3DG and GL models are $\sqrt{2}$ and $8/3$, respectively, while the experimental value approaches ~1.8 in the limit as h goes to 0. The mGL-model with $z_0 \approx 0.95$ μm and $y \approx 1.9$, on the other hand, yields a gamma-factor ratio of 1.86, which is in close agreement with the experimental data. These results show that the mGL-PSF is successful in reproducing the experimental gamma-factors for all different heights h of the sample.

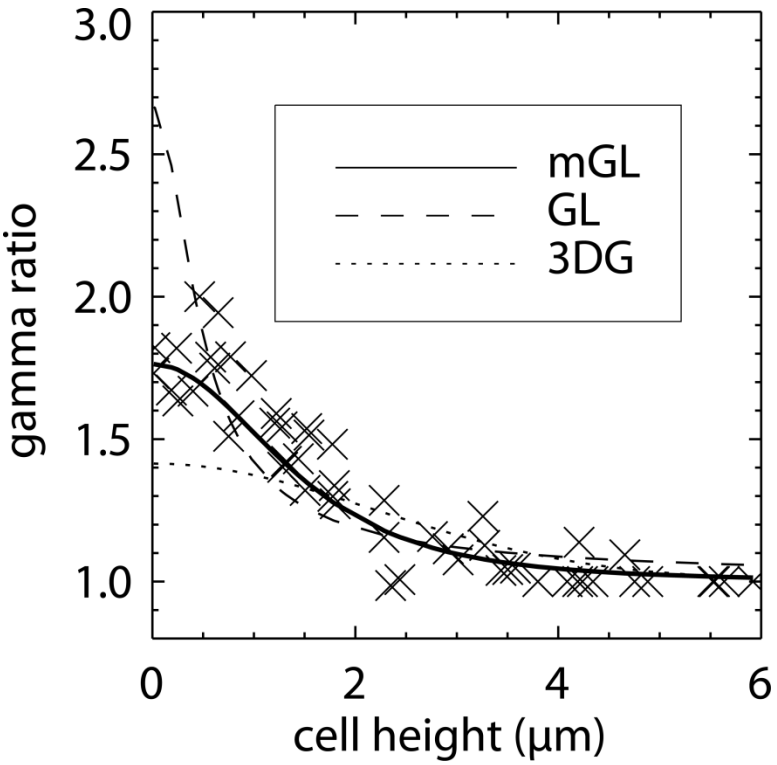

Figure 4.7 **Gamma ratio: experiment and theory.**

Gamma ratio data is plotted for different positions in several cells containing EGFP. The gamma ratio is defined as $\gamma_2(h/2;h)$ divided by $\gamma_{2,\infty}$, where $\gamma_2(h/2;h)$ is a geometry-dependent shape factor and $\gamma_{2,\infty}$ is the infinite sample shape factor assumed by conventional FFS. Thick sample sections, like the nucleus, have an effectively infinite volume, and thus a gamma ratio equal to 1. The opposite limiting case, when the sample becomes a delta layer, is sample independent and thus a useful measure to test the PSF model we have developed. The traditionally chosen 3D Gaussian and Gaussian-Lorentzian models do not fit the experimental data, especially in approach to the delta layer limit. The mGL PSF model matches the behavior of experimental data.

4.9 Brightness by z-scan FFS

The experiments described so far indicate that the mGL-PSF provides an accurate description of the experimental PSF. We therefore expect to furnish unbiased brightness values from thin cell sections by analyzing FFS experiments with the mGL-model. The most straightforward approach to achieve an accurate brightness measurement from cell data requires an intensity z-scan measurement followed by an FFS measurement at the midsection of the cellular slab to determine the Q-parameter. The intensity z-scan provides the thickness h and axial-offset h_0 of the cellular slab, which is used to determine the gamma-factor for the mGL-PSF at the midsection of the cell, $\gamma_{2,mGL}(h/2;h)$. By using the proper gamma-factor, the correct brightness λ is extracted from the Q-parameter, $Q = \lambda \cdot \gamma_{2,mGL}(h/2;h)$. If the gamma-factor $\gamma_{2,\infty,GL}$ for an infinite sample geometry is used instead, a biased brightness λ' is recovered, $Q = \lambda' \cdot \gamma_{2,\infty,GL}$

The above strategy is now tested experimentally. We refer to the intensity z-scan followed by an FFS measurement at the midsection of the cell as z-scan FFS. We report brightness in normalized form, $b = \lambda/\lambda_{monomer}$, where $\lambda_{monomer}$ is the brightness recovered for monomeric EGFP from a sample thick enough to avoid any bias in its brightness. For comparison we first show in Figure 4.8A the biased EGFP data (triangles) originally displayed in Fig 4.1B. The overestimation of the brightness with decreasing thickness of the sample is a consequence of conventional FFS analysis, which erroneously indicates the presence of dimeric EGFP in the cell. We now determine the brightness of the same data using z-scan FFS. The fits of the z-scan intensity curves confirm that EGFP in the cell nucleus and cytoplasm is represented by a single, uniform layer that is accurately modeled by the mGL-PSF applied to a rectangular slab of height h (data not shown). The gamma-factor $\gamma_{2,mGL}(h/2;h)$ was determined from the z-scan intensity fit parameters and used to determine the brightness from the experimentally measured Q-parameter. The recalculated brightness (crosses) is displayed in Figure 4.8A after accounting for the geometry-dependent gamma-factor. As expected, the normalized brightness is 1 at all measured cell locations, meaning that EGFP is a monomer

throughout the cell. This result provides strong evidence that z-scan experiments provide a successful method for the unbiased determination of brightness in the cytoplasm.

We repeated the z-scan FFS measurement on cells expressing the dimeric construct $EGFP_2$. Because $EGFP_2$ covalently links two EGFP molecules together, a normalized brightness of 2 is expected from this artificial dimer. Z-scan FFS measurements are performed at 5 different positions in each cell measured. In addition, we performed z-scan FFS measurements on cells expressing EGFP as a control. The corrected brightness for each protein was determined using the previously described method. Figure 4.8B plots the unbiased EGFP brightness (crosses) and the unbiased $EGFP_2$ brightness (diamonds) as controls. As expected, these normalized brightnesses are all ~1 and ~2, indicating monomers and dimers, respectively.

We now apply z-scan FFS to measure the oligomeric state of EGFP-labeled NTF2 in the cytoplasm of COS-1 cells. NTF2 plays an important role in maintaining the Ran-gradient, which is crucial for nucleocytoplasmic transport.[68] *In vitro* data report that NTF2 exists as a dimer at concentrations exceeding 1 μM,[69] but a direct confirmation of the dimeric state of NTF2 in cells is not yet available. We therefore performed cytoplasmic z-scan FFS experiments on EGFP-NTF2 at concentration exceeding 1 μM and demonstrate that the brightness (asterisks) falls on top of the $EGFP_2$ data, showing that NTF2 is a dimer (Figure 4.8B). Conventional FFS analysis of the data would lead to a large scatter of brightness values with an average that exceeds the value expected for a dimer. This behavior complicates proper identification of the degree of oligomerization, which is avoided by using z-scan FFS.

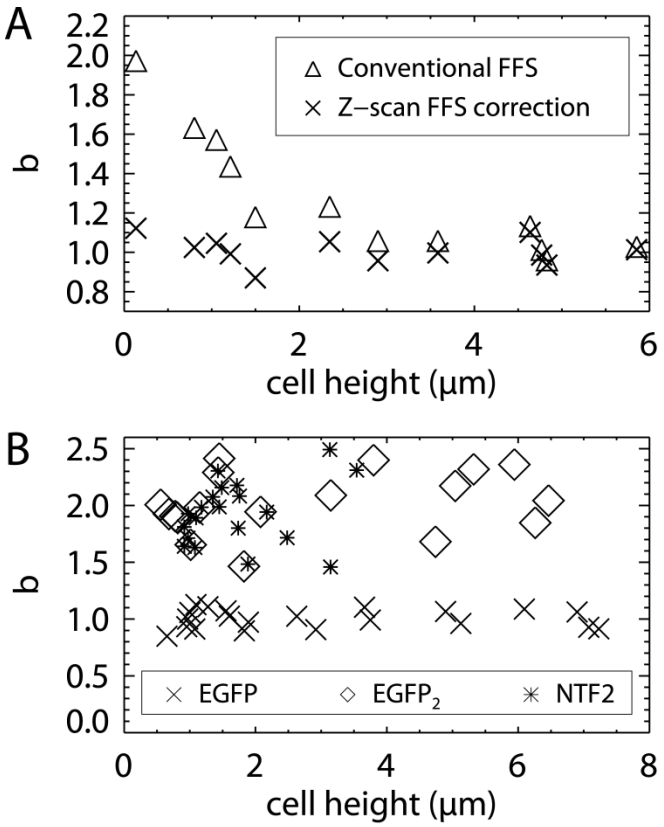

Figure 4.8 **Z-scan corrected brightness analysis.**

(A) EGFP exists as a monomeric protein in cells (normalized brightness, b=1), and distributes uniformly throughout the nucleus and cytoplasm. Data was taken at different locations in the cells, from the center of the nucleus out to the edge of the cell. Conventional FFS ignores sample thickness and yields biased results (triangles). Z-scan FFS corrects for sample geometry and recovers the unbiased brightnesses (crosses). The corrected normalized brightnesses show a value of 1, accurately reporting the existence of monomeric EGFP throughout the cell. (B) Additional experiments were conducted on cells containing the artificial dimer construct $EGFP_2$ (diamonds) and nuclear transport factor 2 (NTF2, asterisks). Z-scan corrected brightnesses are equal to 2 for NTF2 measured in high concentration cells (> 1 μM), which indicates that NTF2 forms a homodimer in the cytoplasm of living cells.

4.10 Discussion

This study demonstrates that once the sample thickness approaches the axial dimension of the optical observation volume, the brightness determined by conventional FFS experiences a sample thickness-dependent bias. The degree of bias depends on the optical setup and the cell type studied. Spreading cells typically have a thickness of a few micrometers at the center, which drops with distance from the center. The axial beam waist of the excitation light depends on the optical set-up, but is typically ~1 to a few μm. Thus, the potential for brightness bias in cell experiments is a concern. Because the cell is usually thickest at the center, the likelihood of brightness artifacts increases for cell measurements taken at off-center positions.

We define the thickness which leads to a 20% bias in brightness as critical thickness. The critical thickness depends on the instrument and needs to be evaluated on an individual basis. Our instrumental setup leads to a critical thickness of 2 μm. We reported thicknesses of COS-1 cells in the cytoplasm of ~2.5 μm to less than 1 μm. Thus, we observe brightness bias in cytoplasmic FFS measurements. Only measurements conducted directly next to the nuclear envelope on the cytoplasmic side of COS-1 cells are generally thick enough to avoid significant brightness bias. However, this location contains the endoplasmic reticulum and Golgi complex. If these organelles need to be avoided experimentally, measurements have to be taken further away from the nuclear envelope, where the thickness falls below the critical value. For example, the phosphatidylinositol type II kinase (P14KII) interacts with endosomal membranes,[70] which prevents the quantitative FFS measurement of cytoplasmic P14KII next to the nuclear envelope. Thus, cytoplasmic P14KII has to be measured further away from the nucleus, where brightness corrections are important. The cell line used in the experiments is another important factor. We have found that CV-1 cells exhibit a shallower height profile than COS-1 cells. Thus, cytoplasmic FFS measurements in CV-1 cells are at the critical thickness even when conducted directly next to the nuclear envelope. It is also important to recognize that while a cell population has a typical height profile, cell-to-cell

variations exist. Thus, the same relative location may be above or below the critical thickness in different cells.

The excitation beam profile is the chief factor in determining the critical thickness. The long axial tail of the GL-PSF makes it especially prone to thickness artifacts, which is an important consideration for 2-photon excitation experiments. For example, a GL-PSF with an axial beam waist of 1 μm leads to a critical thickness of 5 μm. Because the thickness of COS-1 cells at the center is ~5 μm, a GL-PSF would introduce a brightness bias for all cell measurements. Thus, underfilling the back aperture of the objective, which leads to a GL beam profile, should be avoided when performing cell measurements. Since the optics of each instrument differ and the presence of aberrations influences the PSF, it is necessary to characterize the critical thickness of each instrument separately.

The presence of brightness bias is not necessarily easy to recognize. It becomes apparent when graphing the brightness versus thickness as shown in Figure 4.1B, but typical FFS experiments do not record this information. Because FFS studies in cells require the measurement of a large number of cells to acquire sufficient statistics, the variations of thickness of the sample simply increase the scatter of the measured brightness values. Because the bias always results in larger brightness values, the interpretation of the data are systematically skewed towards identifying stronger protein interactions than actually exist. For example, in a system where protein-protein interactions are absent, the bias results in a normalized brightness exceeding 1. This leads to the erroneous conclusion that the proteins do interact. Similarly, quantitative interpretation of binding curves and protein stoichiometries from brightness titration data is also compromised. Brightness corrections are especially important for FFS measurements of proteins at the periphery of the cell, such as would be required in the case of focal adhesion studies. A similar example is FFS in bacteria, such as *E. coli*., which have a diameter smaller than the critical thickness. In addition to FFS measurements at a single location, brightness measurements using imaging approaches[71] are susceptible to the same artifact.

To perform quantitative brightness experiments in thin environments, it is necessary to address the brightness bias. The problem is solved by performing a fluorescence intensity z-scan before the FFS measurement to characterize the thickness and correct the FFS analysis. It only takes ~5 s to execute the intensity z-scan, which is short compared to the FFS data acquisition time (~60 s). Thus, the additional measurement time for carrying out z-scan FFS is negligible. Z-scan FFS only requires a motorized microscope stage and is straightforward to implement. Z-scan FFS experiments should be carried out with a water-immersion instead of an oil-immersion objective, because spherical aberrations need to be minimized.

We applied z-scan FFS to investigate the oligomeric state of NTF2 in the cytoplasm. Biochemical studies and an ultracentrifugation experiment report that NTF2 forms a dimer.[69,72] Our work confirms that NTF2 exists as a dimer in the cytoplasm of living cells. NTF2 is also known to interact with nucleoporin proteins at the nuclear envelope,[68,73] which leads to nuclear rim staining. As a consequence, the fluorescence intensity at the nuclear envelope is enhanced, which prevents us from measuring the brightness of NTF2 in the nucleus. We performed cytoplasmic FFS measurements away from the nuclear envelope to avoid signal contributions from nuclear rim staining. Therefore, all these measurements had to be corrected for the finite thickness of the sample.

NTF2 interacts with the nuclear pore complex and mediates active transport across the pore by an unknown mechanism. We are interested in extending z-scan FFS in the future to measure protein interactions and transport rates directly at the nuclear membrane in order to investigate the transport process. This will require an expansion of z-scan FFS to a multilayer geometry and should open up a number of interesting applications of the technique to the study of internal membrane interfaces. A multi-layer sample can no longer be characterized by a single FFS measurement at the midsection as performed in this paper. A natural extension of this work takes several FFS measurements as a function of the z-position. The geometry function $S(z)$ is easily modified to describe multiple layers, and z-scan FFS can be implemented for complex samples, only being limited by the resolvability of the layers through a fit to the model.

Thus, z-scan FFS should be able to make direct measurements of brightness at internal membrane interfaces like the nuclear envelope; these regions have been previously inaccessible to quantitative brightness analysis.

Z-scans methods have been previously implemented into FFS to study diffusion in membranes like giant unilamellar vesicles and supported phospholipids bilayers.[60,62] These studies were performed in isolated membrane systems, but related z-positioning FCS techniques are now being used to study raft diffusion properties in the plasma membrane of living cells.[61,74] Here, we extend the use of z-scanning to study the brightness in thin sample geometries. While this work focuses mainly on cell measurements, the same problem of brightness bias is encountered in micrometer and submicrometer fluidic devices. Thus, z-scan FFS has potential applications in thin microfluidic channels.

While thin sample geometry limits the access of fluorophores to the excitation light, it also limits the possible movement of the particles. As the sample flattens out, diffusion of proteins is no longer accurately described by a 3D diffusion model. Gennerich and Schild deal extensively with the effect confined geometry has on diffusion, as calculated by FCS.[75] Fortunately, brightness is independent of diffusion time provided that no undersampling occurs.[32] All FFS measurements in this study are sampled sufficiently fast that undersampling is avoided. If necessary, brightness values may be corrected for undersampling effects as previously described.[32]

In summary, this paper describes a sample thickness-dependent artifact of FFS data which leads to an increase in brightness. We report a brightness bias of cytoplasmic FFS measurements in COS-1 cells, which in the worst case leads to an approximate doubling of the brightness value. As brightness analysis is frequently used to detect protein association, a factor of 2 bias is a critical problem for quantitative measurements. The paper describes how the critical thickness is identified and provides advice on performing FFS measurements in the cytoplasm of cells. We introduced z-scan FFS as a simple method to assess the thickness of the sample and provide brightness values free of artifact. Z-scan FFS provides a general method for brightness determination in cases

where the finite thickness of the sample is of concern. We expect that z-scan FFS should prove especially useful for FFS applications in the cytoplasm and other thin geometries.

5. Chromophore Maturation and Cell-Free Expression

Fluorescence measurements are founded upon the fluorophore. Brightness analysis, in particular, requires well-behaved, fluorophore brightness or FFS is no longer a useful tool to quantify protein stoichiometry. The literature continues to clarify the mechanisms by which fluorescent proteins mature into their fluorescent state, but it has been particularly conflicted over the rates of maturation. EGFP has been shadowed with a reputation for very slow maturation. In such a case, a significant fraction of the EGFP-labeled sample might not fluoresce leading to a breakdown of the brightness-stoichiometry relationship. Similarly, other reports allege that a large fraction of EGFP fluorophores simply never mature.[76] Neither of these issues can be investigated in cells, as we have no control over the expression conditions. In this chapter, we seek to employ FFS in a cell-free expression system to address these types of questions.

As indicated above, the main problem with measurements in cells is a lack of experimental control. The researcher cannot simply double the concentration or add a second protein or change the pH. Studies conducted *in vitro* do allow this control, but bring a host of other challenges. For instance, in this artificial environment, additional factors required for proper folding or function may be missing. The purification process to isolate the protein of interest is quite time consuming, and specifically labeling the protein 1:1 with an organic dye is a problem in itself, as its very difficult to effect 100% labeling and simultaneously prevent free-floating dye labels. Here, we investigate FFS in the cell-free expression environment as a bridge between cellular and aqueous protein measurements, and as a tool to better probe protein interactions with the benefits of both genetic labeling and direct control. This chapter has been previously published in Analytical Biochemistry.[77]

5.1 Introduction

Cell-free protein expression is a simple and flexible method for the rapid synthesis of folded proteins. These systems represent an attractive alternative for producing difficult-to-express proteins, such as membrane proteins or proteins that seriously interfere with cell physiology.[78] Recently, cell-free systems have been used to

engineer complex biological processes.[79] The inherent open nature of the cell-free system allows direct access and control of the biological system, which cannot be attained by cell experiments. Thus, the cell-free system is emerging as a versatile platform for systems biology experiments.[80] In addition, cell-free expression has been used lately to study the maturation kinetics of fluorescent proteins.[80,81]

The principle of brightness analysis should be directly applicable to FFS measurements of cell-free systems. Just as in cells, expression of EGFP-tagged proteins provides a convenient method to introduce fluorescent markers. However, there are a number of potential challenges for FFS studies in cell-free systems. The fidelity of cell-free expression and of chromophore maturation needs to be checked, because the presence of non-fluorescent or incompletely expressed protein leads to biased interpretation of FFS experiments.[45] We develop a droplet-based assay for FFS experiments in a cell-free extract. The droplet format not only ensures a small sample volume, but also provides sufficient oxygenation of the sample for efficient chromophore maturation.

The widespread use of fluorescent proteins as markers in biological and medical research rests on the unique ability of the expressed protein to form a chromophore on its own. The formation of the chromophore is referred to as maturation and only requires molecular oxygen as an external reagent.[12] Although the kinetics of chromophore formation has been the subject of numerous studies, there is significant inconsistency in the reported maturation times of individual fluorescent proteins. Here we investigate the temperature dependence of the maturation process, which has received little attention so far. Specifically, we conduct FFS experiments in batch-mode on EGFP, EYFP, and mCherry using an *Escherichia coli*-based cell-free system. We observe a pronounced temperature dependence of the maturation rates of the fluorescent proteins, model the kinetics by transition state theory, and compare our results with other maturation studies based on *de novo* protein synthesis. Because EGFP exhibits a very fast maturation at 37^{o}C, we consider EGFP to be a suitable reporter for many kinetic studies in the cell-free system.

We further use EGFP and a tandem-dimer EGFP ($EGFP_2$) to check the fidelity of expression and maturation of the *E. coli*-based cell-free system and to demonstrate the feasibility of protein titration experiments with the droplet-based assay. Nuclear transport factor 2 (NTF2) is used as a model protein to explore brightness characterization of protein-protein interactions in the cell-free system. NTF2 plays a key role in nucleocytoplasmic transport and maintains the cellular Ran gradient which drives the transport process.[82] Depletion of NTF2 affects the Ran gradient and can lead to cell death.[83] The mechanism by which NTF2 performs its function is not well understood. Because there are conflicting reports regarding the oligomeric state of NTF2,[69,84,85] we examine the stoichiometry and concentration dependence of NTF2 in cell-free solution and compare the result with the literature.

5.2 Materials and methods

Cell-Free Sample Preparation. We use the S30 T7 High-Yield Expression System (Promega, Madison, WI) as our cell-free solution system. We mix DNA into 5-15 μL samples of cell-free solution according to the Promega protocol, and let the synthesis reaction run at room temperature for 2-4 hrs. Reactions were subsequently stopped by the addition of 0.1% RNase A (Sigma-Aldrich, St. Louis, MO) and then spun down at 18,000g for 20 minutes to remove any large particles. Samples were transferred into a ring of vacuum grease (Dow Corning, Midland, MI) to prevent spreading inside an eight-well coverglass chamber slide (Nagle Nunc International, Rochester, NY). The eight-well chamber is sealed with a rubber stopper or an additional coverslip and vacuum grease seal. Glass surfaces are treated with SigmaCote (Sigma-Aldrich) and measurements are taken 10 μm above the coverglass to avoid surface adhesion effects.

Expression Vectors, Purified Protein and Buffers. *Cell-free:* EGFP 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 result was cloned into the pRSET B vector (Invitrogen, Carlsbad, CA). We refer to this plasmid as pB-G. A

tandem dimeric EGFP ($EGFP_2$) was constructed by cloning a second EGFP between the EcoRI and HindIII restriction sites of the pB-G plasmid. NTF2 was amplified from human NTF2 (Genbank accession number: BC002348) and cloned into the EcoRI/HindIII site of pB-G. A monomeric NTF2 was generated by mutating the methionine residue 118 to glutamate.[69] *Solution:* EGFP was expressed in *E. coli* and purified according to protocol using QiaExpress Ni-NTA Fast Start kit (Qiagen, Valencia, CA). Cell-free and aqueous dilution experiments were performed using HBS-EP buffer (Biacore, Uppsala, Sweden) or Dulbecco's PBS with calcium and magnesium (Biowhittaker, Walkersville, MD).

Chromophore Maturation. Protein expression is stopped with 0.1% RNase and quickly transferred to the sample chamber. The observed increase in fluorescence intensity reflects chromophore maturation of previously synthesized proteins and is fit by

$$F(t) = F_0 + \Delta F\left(1 - e^{-t/\tau}\right) \tag{5.1}$$

where F_0 is the average fluorescence intensity at the time the reaction is stopped, ΔF is the change in average fluorescence intensity from the subsequently maturing chromophores, and τ is the characteristic maturation time. The maturation rate coefficient k is the inverse of the maturation time, $k = 1/\tau$.

The temperature dependence of the maturation rate coefficient is modeled using transition state theory. The Eyring equation relates the reaction rate coefficient k to temperature T,

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta H^{\ddagger} - T\,\Delta S^{\ddagger}}{RT}\right) \tag{5.2}$$

where $\Delta H^{\ddagger}$ is enthalpy of activation, $\Delta S^{\ddagger}$ is entropy of activation, k_B is the Boltzmann constant, h is Planck's constant, and R is the gas constant. The reaction rate data are plotted as $ln(k/T)$ vs $1/T$ and fit to a straight line. The slope and intercept of the fitted line determine the activation enthalpy and entropy. Sample temperatures are adjusted using an ASI 400 air stream incubator (Nevtek, Williamsville, VA) or a VWR Polyscience chiller recirculator (Niles, IL) connected to homebuilt coils around the microscope objective and sample holder. The sample temperatures reported in this work

were recorded using a thermocouple placed in a droplet or water reservoir adjacent to the measured sample.

5.3 Cell-free synthesis of EGFP

Because cell-free samples are typically 5-50 µL, the measurement protocols must be adapted for small volumes. Our first approach was to place the cell-free solution between two coverslips. Upon expressing EGFP we observed that the sample was only fluorescing around the edges of the sample, which is in contact with the surrounding air. Because oxygen is required for maturation of the chromophore, our result indicated that good contact of the sample with the surrounding air is crucial. To address this challenge we adopted a configuration where a drop of cell-free solution is placed on a chambered coverglass slide. Before placing the droplet a small circle of vacuum grease is applied on the glass surface to prevent spreading of the solution across the slide. This protocol results in a well-formed droplet with a large surface area in contact with air for efficient uptake of oxygen. Expression of EGFP in the droplet solution resulted in the appearance of fluorescence throughout the sample as confirmed by an axial scan[56] through the sample with a long-working distance objective.

Because the measurement of cell-free expression reactions may take several hours, we evaluated the long-term stability of the droplet sample. A droplet sample containing an aqueous solution of the fluorescent dye Alexa 488 was measured as a control by FFS for several hours (data not shown). We observed a continuous increase in the fluorescence intensity. FFS analysis showed that the brightness of the dye remained constant, but the dye concentration increased. The increase in concentration is consistent with the loss of sample volume due to evaporation. To circumvent this problem we place the droplet into a sealed chamber together with an additional water droplet (Figure 5.2D). The water serves to establish vapor pressure equilibrium after the top opening has been sealed with a rubber stopper, thereby preventing the evaporation of the sample. The sealed chamber and water reservoir are prepared in advance and the cell-free reaction mixture is added later via syringe through the rubber stopper. We tested the sealed chamber using an aqueous droplet sample containing the dye Alexa 488, which was

measured for more than 6 hours. The fluorescence intensity remained constant over the whole measurement period, demonstrating evaporation effects are negligible. Thus, all experiments described here have been performed using the sealed chamber configuration.

We study cell-free expression of EGFP using FFS to monitor the course of protein synthesis in real time. Initial experiments established that the cell-free extract contains aggregates that interfere with later FFS analysis of the data. Fortunately, such complexes are relatively easily removed through centrifugation. Centrifuging the sample at 18,000g for 5-10 minutes pellets the detritus, and the clean cell-free supernatant is easily removed. The centrifugation step can be conducted before or after the protein expression reaction. Immediately after adding the DNA for EGFP to the centrifuged reaction mixture, the sample is placed on the microscope. Sixty second measurements are taken at regular intervals over the next five hours (Figure 5.1A). The intensity increases after an initial lag phase. The reaction eventually begins to run down as the synthesis process exhausts the available energy.[86] Brightness analysis of this data allows us to extract time courses for the average number of molecules in the observation volume and the brightness of EGFP (Figures 5.1B&C). The number of molecules increases over time demonstrating that proteins are being produced as expected. More importantly, the brightness (Figure 5.1C) is constant throughout the reaction which shows that there are no transitional or alternate brightness states. Unfolded and unmatured EGFP proteins are non-fluorescent and so invisible to FFS measurements. Once the protein is completed, it "turns on" at the characteristic brightness of EGFP and remains stable and independent of protein concentration. This is further confirmed by comparing FFS measurements on cell-free expressed EGFP, recombinant EGFP purified from transformed *E. coli*, and EGFP in living mammalian cells (Table 1). Measured under the same excitation power, all three EGFP samples have the same brightness. This robustness of EGFP brightness makes it a useful and reliable tool for quantitatively examining protein interactions. Table 1 also displays the diffusion time or average residence time within the optical volume. This information is acquired through autocorrelation analysis of the data and reflects the differing solution environments of the three methods. Purified EGFP in aqueous buffer solution has the fastest diffusion time. Both EGFP in cells and in cell-free solution have a

slower diffusion time. Fluorescent dye added to both the buffer solution and to basic cell-free solution shows a similar change in diffusion time (data not shown). This indicates that the difference in diffusion time reflects the changes in viscosity of the solutions.

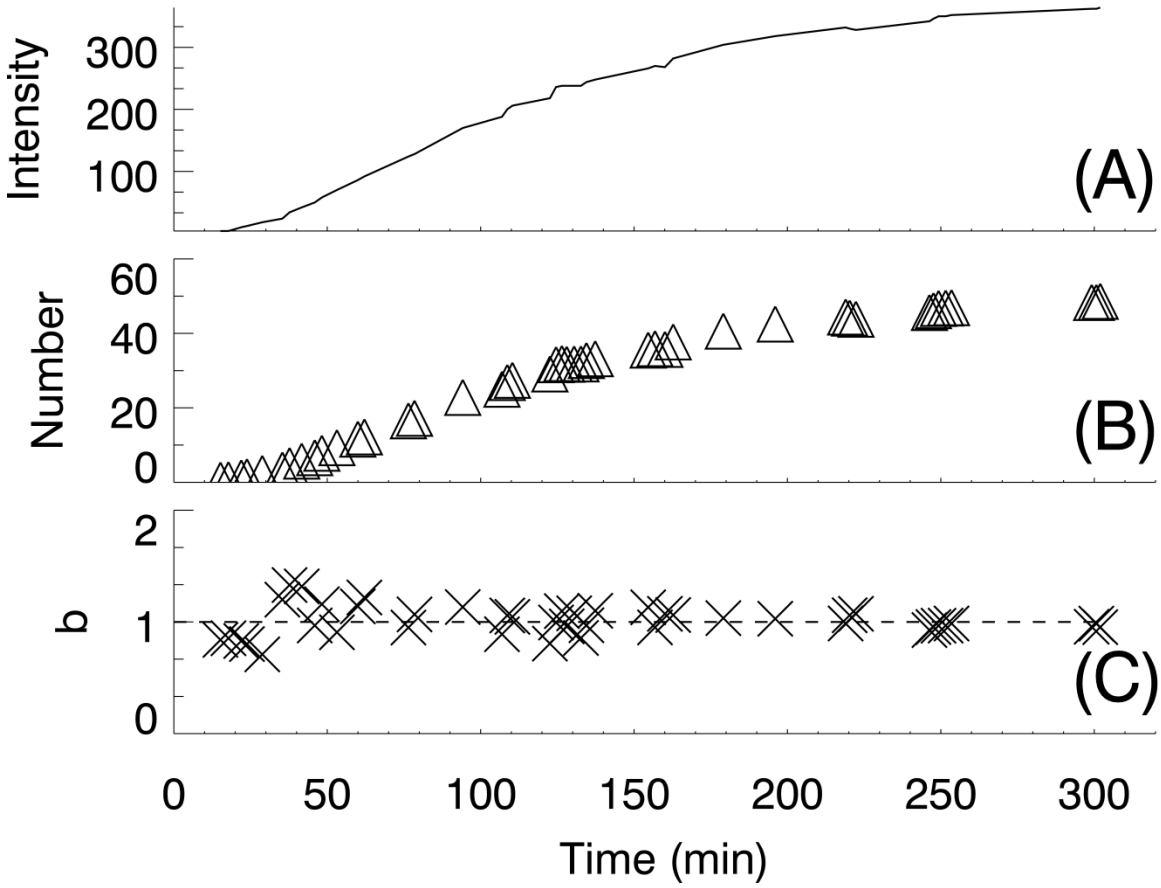

Figure 5.1 **Cell-free protein synthesis**.

(A) EGFP DNA plasmid is added to the cell-free reaction mixture and FFS measurements are taken at intervals. The average fluorescence intensity is plotted as a function of time. (B) Further analysis of the data shows that the number of EGFP molecules tracks with intensity and increases with time as new proteins are synthesized. (C) The normalized brightness, however, remains stable throughout the reaction demonstrating a single, robust EGFP brightness state.

Sample	Brightness (photon counts/s)	Diffusion time (ms)
Cell-free EGFP	3300 ± 110	0.75 ± 0.09
Purified EGFP	3200 ± 200	0.27 ± 0.05
Cellular EGFP	3200 ± 330	1.26 ± 0.38

Table 5.1 **EGFP brightness in different environments**.

EGFP is expressed in cell-free solution, is His-tag purified from *E. coli* into PBS buffer, or is expressed in U2OS cells. FFS measurements on all three samples report the same brightness within error. The changes in the diffusion time reflect the changes in viscosity of each sample.

5.4 Kinetics of chromophore maturation

We investigate the kinetics of chromophore maturation using cell-free expression in a free-standing droplet (~10 µL). After spinning down the cell-free solution, DNA is added and left to synthesize for 10-30 minutes. The reaction is stopped with 0.1% RNase and quickly transferred to the sample chamber. By adding RNase to the droplet the synthesis reaction is rapidly stopped.[80] Any subsequent increases in the fluorescence intensity can be attributed to chromophore maturation of previously synthesized proteins. Since each chromophore's maturation is independent, fluorescence intensity increase can be described by a single-exponential model. Figure 5.2A shows the EGFP maturation intensity as a function of time fit with Eq. 5.1. At room temperature (20˚C), the characteristic maturation time, $\tau_{\mathrm{EGFP}} = 15 \pm 3.5$ min, is established by repeating the experiment several times (n = 6). In addition to EGFP, we also investigated the maturation of EYFP (n = 3) and mCherry (n = 4). We found EYFP at 19˚C to have a significantly longer maturation time, $\tau_{EYFP} = 78 \pm 12$ min, and mCherry at 20˚C to be rather slow at $\tau_{mCh} = 155 \pm 10$ min (Figures 5.2B&C).

Oxygen diffuses within ~100 s from the edge to the center of a 10 µl droplet. As such, we anticipate that oxygen availability is not a limiting factor for maturation reactions with reaction times longer than 100 s. As a control experiment we measure maturation of EGFP in a smaller droplet (1 µl). With a diffusion time to the center of only ~20 s the rate to replenish oxygen is much faster than for the large droplet. The maturation time of EGFP measured in the 1 µl droplet at 19^{o}C (τ_{EGFP} = 15 min) is identical to the time observed in the 10 µl droplet. This result demonstrates that oxygen availability is not a limiting factor for our assay.

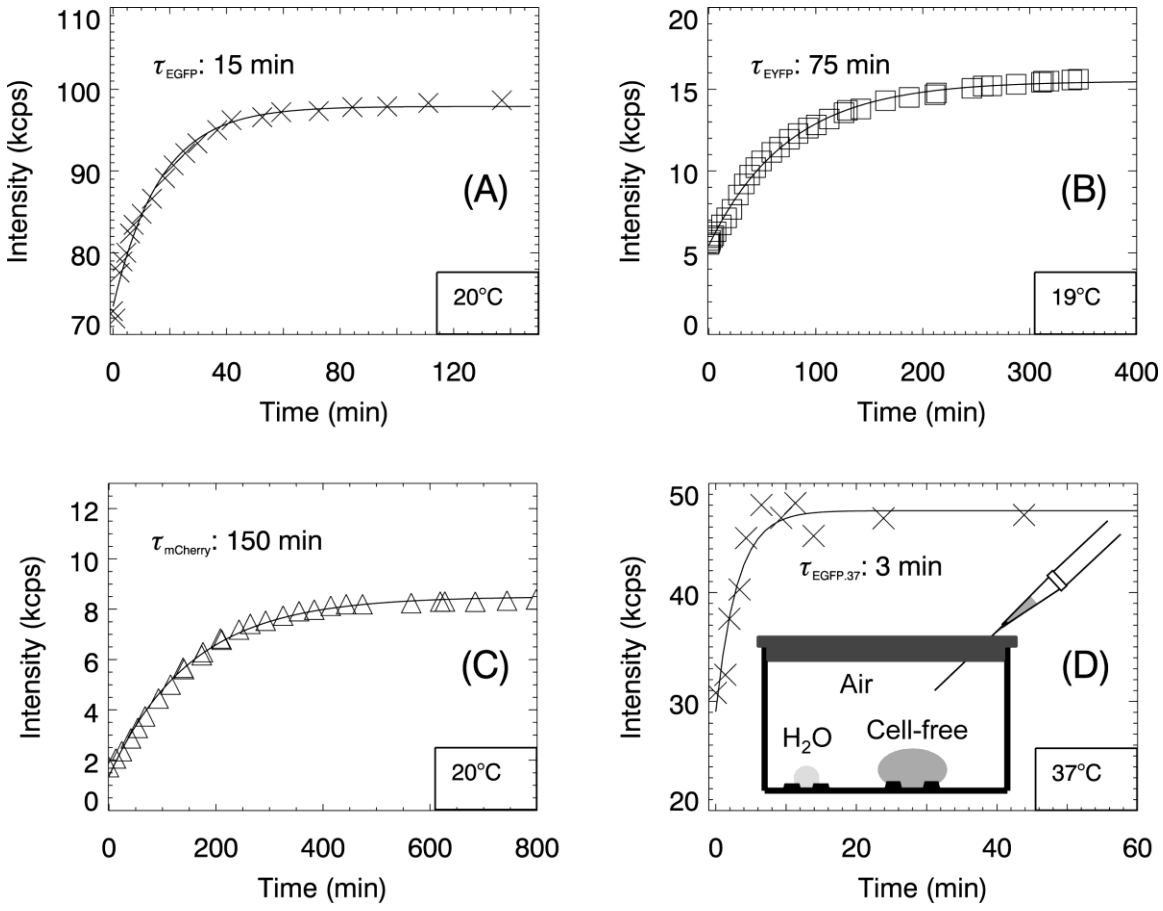

Figure 5.2 **Chromophore maturation time**.

Cell-free EGFP synthesis is initiated at room temperature and then stopped with RNase. Any subsequent intensity increase is the result of a previously synthesized protein completing the maturation process. The curve is fit with a single exponential rate process. (A) Single maturation curve of EGFP with fit. (B) Single maturation curve of EYFP with fit. (C) Single maturation curve of mCherry with fit. The mean maturation time of each protein and its standard deviation are determined from repeated measurements at the same temperature and are quoted in the text. (D) The final panel shows an EGFP sample maturing at 37°C with τ = 3 min. Overlaid is a cartoon of the sample setup. A small sample droplet (~10 μL) is injected into a sealed chamber slide containing a water reservoir to establish vapor pressure equilibrium.

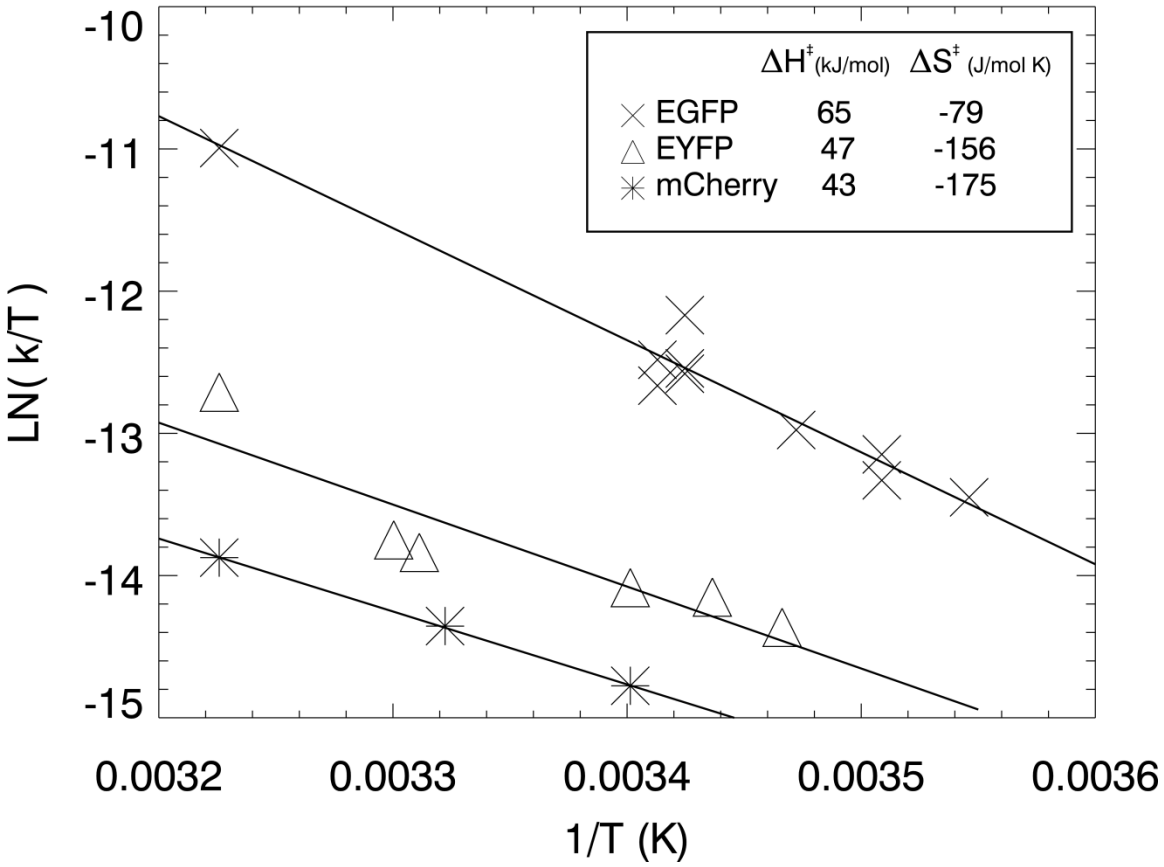

Figure 5.3 **Temperature dependence of maturation rate coefficient *k***.

The fitted maturation rate coefficient k and sample temperature T are graphed in an Eyring plot for EGFP (crosses), EYFP (triangles), and mCherry (asterisks). The fitted curve determines the activation enthalpy $\Delta H^{\ddagger}$ and activation entropy $\Delta S^{\ddagger}$ of the reaction.

Literature values for chromophore maturation times of EGFP cover a broad range of values and involve a variety of experimental approaches.[13,87–90] To further complicate matters individual studies have been performed at different temperatures. However, a quantitative study that evaluates the influence of temperature on maturation is not readily available. We decided to expand our study in order to address the temperature dependence of chromophore maturation. Measurements of the fluorescence intensity of the maturation process were taken at selected temperatures up to 37°C and fit with Eq. 5.1 to recover the characteristic maturation time τ and rate coefficient $k = 1/\tau$. The results of the temperature study are plotted as $ln(k/T)$ versus $1/T$ (Figure 5.3). The Eyring plot shows that maturation rates increase with temperature. Maturation of EGFP at 37° C proceeds sufficiently fast (Figure 5.2D) that we approach the practical limit of the experimental setup. Thus, the temperature-dependence of EGFP maturation was performed by cooling the solution below room temperature. The straight lines through the data points of Figure 5.3 represent a fit of the rate coefficients to transition state theory. The slope and y-axis intercept determine the activation enthalpy $\Delta H^{\ddagger}$ and activation entropy $\Delta S^{\ddagger}$ of the maturation process of each protein. The activation enthalpies for EGFP, EYFP, and mCherry are 65±1.6, 46±4.3 and 43±1.7 kJ/mol, respectively. These values fall in the range expected for solution oxidation reactions. The differences in activation enthalpy between the proteins are relatively small, particularly between EYFP and mCherry. The differences in the activation entropies (–79±5, –156±15, –175±6 J/(mol K) for EGFP, EYFP, and mCherry) follow a similar pattern. Overall, the ratios of the maturation kinetics for different fluorescent proteins are approximately independent of temperature, because of the similarity of the activation enthalpies. Calculating from the fits, the characteristic maturation times at 20˚C are τ_{EGFP} = 15 min, τ_{EYFP} = 82 min, τ_{mCh} = 157 min, and at 37˚C these values become τ_{EGFP} = 3 min, τ_{EYFP} = 27 min, τ_{mCh} = 57 min. Note that the fast maturation rate measured for EGFP at 37°C (Figure 5.2D) agrees with the prediction from transition state theory (upper leftmost cross in Figure 5.3).

5.5 Brightness and titration studies

After a protein is expressed and matured, its properties can be studied by FFS. Here we focus on demonstrating the feasibility of brightness measurements as a function of protein concentration with the droplet setup. For the simple addition of a ligand, enzyme or other chemical, the rubber stopper and syringe method is very effective. However, this approach is inconvenient for performing repeated dilutions of the droplet. Thus, instead of the rubber stopper a coverslip was placed as a lid on top of the chambered coverslide. A thin layer of vacuum grease applied to the contact area of the coverslip seals the sample chamber. A 10 μL droplet of purified EGFP is placed onto the coverslide. Dilutions are performed by unsealing the coverslip, adding additional buffer with a pipette, pipetting up and down, and finally removing an equivalent volume of the mixture and replacing the coverslip. At each dilution step an FFS measurement is taken to determine brightness and concentration (Figure 5.4 inset). The brightness is concentration independent and reflects that EGFP is a monomeric protein. The same result is obtained for EGFP expressed in cell-free solution (Figure 5.4). An important control for brightness experiments is the demonstration of brightness doubling with the tandem-dimer $EGFP_2$, formed by encoding two EGFPs connected with a short linker sequence.[6] After expressing $EGFP_2$ in cell-free solution, a droplet was placed on the coverslide followed by FFS measurements in between dilution steps. The brightness of $EGFP_2$ is concentration independent and twice the brightness of monomeric EGFP, because of the presence of two EGFP molecules. These results demonstrate that the droplet sample provides a quick and reliable method for titration experiments. It further establishes that the quality and fidelity of the cell-free expression system is suitable for brightness studies.

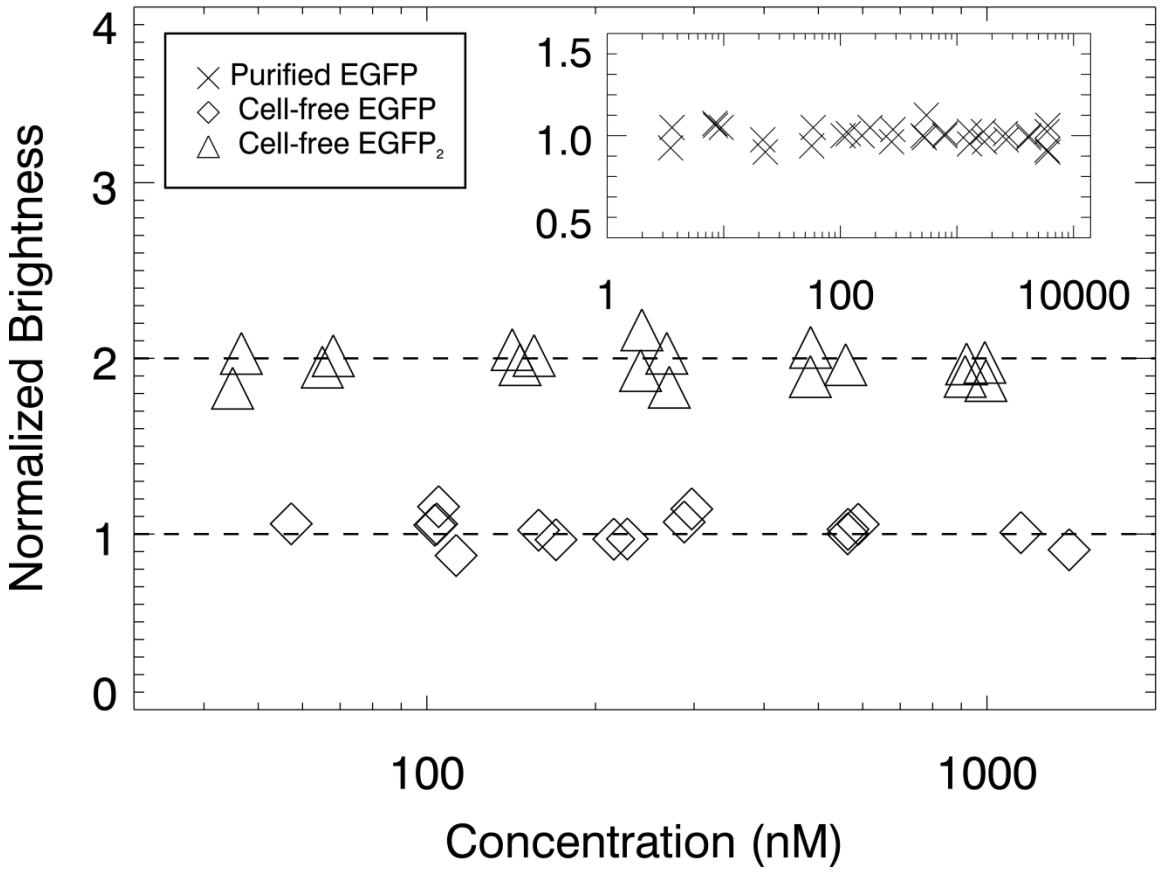

Figure 5.4 **Droplet dilutions**.

Cell-free expressed and *E. coli* purified EGFP are measured in 10 µL droplets. The dilutions are performed using HBS buffer. The inset shows *E. coli* purified EGFP (crosses) with a normalized brightness of ~1 plotted over 4 orders of magnitude of concentration. The FFS measured concentrations are in good agreement with the 1.5x and 2x dilutions that were executed. The main plot is the dilution of cell-free expressed proteins. EGFP (diamonds) and $EGFP_2$ (triangles) are plotted with normalized brightness versus concentration. $EGFP_2$ has $b = 2$ independent of concentration, demonstrating that cell-free brightness is a robust parameter that reports protein stoichiometry.

With the cell-free system characterized and brightness analysis validated, we apply this method to a biological system. The protein NTF2 establishes the cellular Ran gradient, which is crucial for nucleocytoplasmic transport.[91] Ran exists in the form of RanGTP at high concentration in the nucleus and at low concentration in the cytoplasm, where RanGTP is converted to RanGDP. This gradient provides directionality for nucleocytoplasmic transport. The transport factor NTF2 maintains the gradient by returning RanGDP to the nucleus, where it is converted back into RanGTP. NTF2 has been identified as a dimer in previous studies[84,85] with one study reporting a monomer / dimer equilibrium with a dissociation constant of ~1 μM.[69] A recent FFS study of NTF2 from our lab finds NTF2 to be a dimer in cells.[56] However, only cells with high protein concentration were used for the cell study. We now take advantage of the cell-free system to measure the oligomeric state of NTF2 over a wide concentration range using the same FFS approach as the cell study. We perform brightness titration studies on EGFP and $EGFP_2$ expressed in cell-free solution (Figure 5.5) as control experiments. Next, we repeat the dilution experiment on EGFP-NTF2 expressed with the cell-free system. The brightness of EGFP-NTF2 aligns with $EGFP_2$ (Figure 5.5), indicating that NTF2 is dimeric over the entire measured concentration range (~30 – 1400 nM). We also performed a brightness titration of the mutant EGFP-NTF2.M118E expressed in cell-free solution, because it has been previously reported that the point mutation M118E prevents dimerization of NTF2.[69] Our brightness data confirm that the point mutation abolishes dimerization and show that EGFP-NTF2.M118E is a monomer at all measured concentrations (Figure 5.5). Thus, the brightness analysis of NTF2 in the cell-free system agrees with the earlier cell data[56] and indicates that NTF2 continues to be a dimer well below 1 μM.

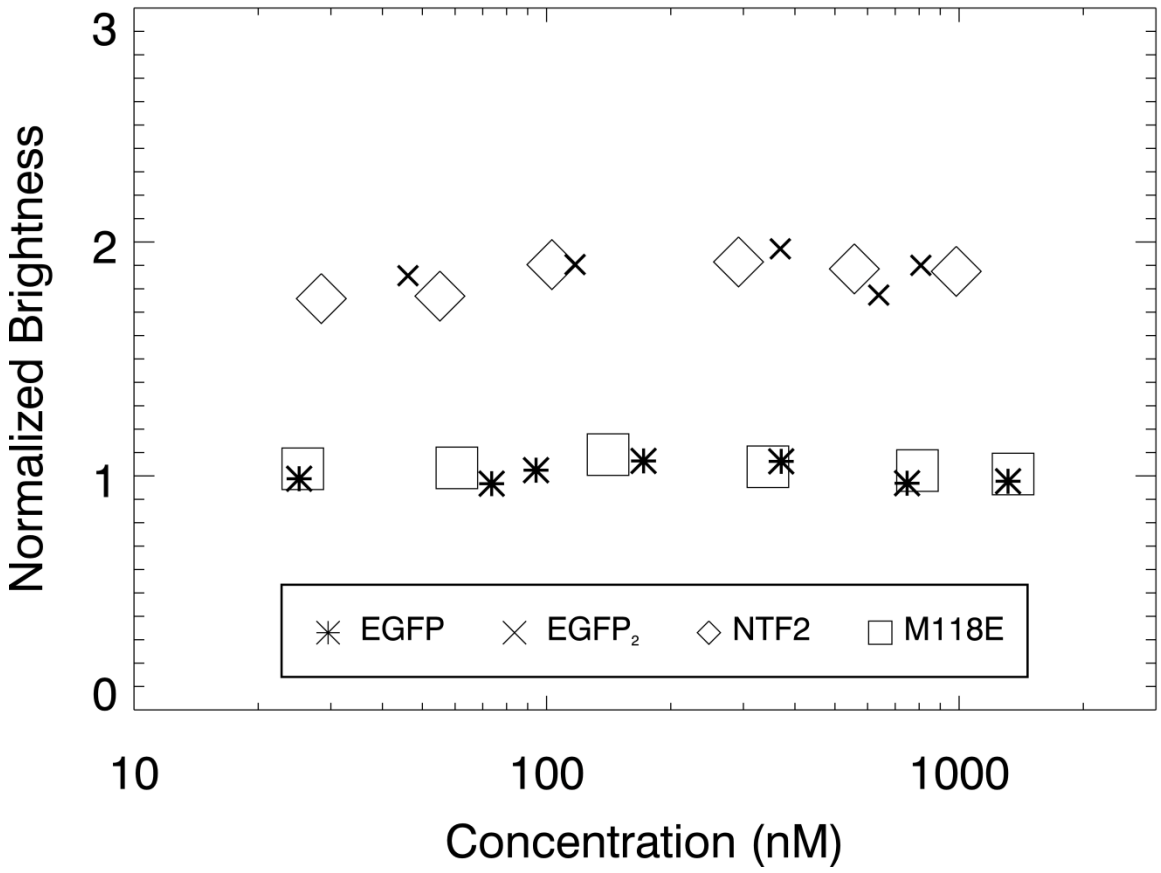

Figure 5.5 **Cell-free NTF2**.

EGFP (asterisks) and $EGFP_2$ (crosses) are expressed in cell-free solution and act as brightness controls plotted as normalized brightness versus concentration. We show that EGFP-NTF2.M118E (squares) is also monomeric in our cell-free system. EGFP-NTF2 data (diamonds) lie on top of $EGFP_2$ revealing that NTF2 is dimeric over the observed concentration range (~30-1400 nM).

5.6 Discussion

There are several factors that are crucial for FFS measurements on cell-free systems. The presence of aggregates in cell-free expression systems is an obstacle to FFS measurements. Some of the aggregates are fluorescent, which leads to spikes in the fluorescence intensity and prevents meaningful brightness analysis. In the case of two-photon excitation we occasionally trapped large fluorescent aggregates in the laser beam. Furthermore, some of the non-fluorescent aggregates are large enough to exclude sufficient volume when passing through the observation volume, so that the fluorescence signal is reduced. While this effect is less noticeable in the raw data, because no spikes are generated, it biases brightness analysis. The presence of large particulates ($\geq$ 1 µm) was directly confirmed by viewing the sample solution through a 63x objective under bright-field conditions. Aggregates seem to be present in most cell-free systems. It is speculated that some of these aggregates build up because cell-free solutions do not have the active processes which clean up unused or discarded reaction material. The bulk of the aggregates can be removed by an additional centrifugation step after reaction solution has been reconstituted and/or after the reaction has finished. Without this step, it is difficult to perform FFS experiments in the cell-free environment.

A unique feature of fluorescent proteins is their ability to form a chromophore by posttranslational modifications. This maturation process only requires molecular oxygen as an external reagent. The amount of oxygen dissolved in air-saturated water under normal atmospheric conditions is ~200 µM at room temperature, which exceeds the concentration of expressed fluorescent protein by more than two orders of magnitude. Despite this apparent excess of oxygen over protein, we observe that a cell-free expression solution sandwiched between two coverslips failed to develop fluorescence except at the edges of the solution that were in contact with air. This observation indicates that other processes exhaust the oxygen reservoir of the solution. In fact, it has been previously reported that oxygen is efficiently depleted within minutes in a cell-free reaction.[92] We choose a droplet as our sample geometry to ensure efficient oxygen transfer at the solution/air interface. The volume of droplets (typically 10-15 µl) permits oxygen to diffuse from the interface to the center of the droplet in ~100 s. Oxygen

diffuses from the periphery to the center of a 1 µL droplet 5 times faster than through a 10 µL droplet. Measurement of EGFP maturation under both conditions returns the same results and demonstrates that oxygen is not a limiting factor in our experiments. In fact, we succeeded in measuring maturation rates as fast as 3 min with the droplet setup (Figure 5.2D). A hydrophobic ring prevents spreading of the sample droplet across the hydrophilic coverglass. This setup is important for the titration experiments, because by preserving the droplet shape it is straightforward to add, mix and remove solution. Measurements on droplets can be performed for many hours (Figures 5.1 and 5.2) without loss of sample. Evaporation is prevented by sealing the measurement chamber as described earlier.

The process of chromophore formation, particularly of GFP, has been carefully studied. The green chromophore pathway (GFP, EGFP, EYFP) has 3 primary stages: cyclization, oxidation and dehydration. The red chromophore pathway (DsRed, mCherry) includes an additional oxidative step for the formation of an acylimine linkage in the polypeptide backbone. Studies suggest that the primary rate-limiting step is oxidation for the green pathway. While the steps for the red pathway are less distinctly time separated, a recent report states that the final dehydration is partially rate-limiting.[93–95] We model EGFP and EYFP maturation with a single exponential rate process, consistent with prior studies. The maturation kinetics of mCherry is also well-modeled by an exponential process. However, the intensity traces of chromophore maturation are affected by experimental imperfections, such as drift in the axial focus and variations in the laser power. These imperfections, although minor, currently limit our ability to investigate whether any small deviation from the single-exponential rate model exists. Nevertheless, the analysis of the maturation kinetics and its temperature dependence demonstrates that a single-exponential kinetic model provides a satisfactory description of the data.

While a coherent mechanism of chromophore formation has emerged, a consistent picture of the overall kinetics is not yet available. The maturation kinetics of fluorescent proteins has been studied by various methods, such as i) *de novo* protein expression, ii) triggered protein folding of solubilized inclusion bodies, and iii) reoxidation after chemical reduction of the chromophore. The results obtained by each method are

inconsistent with each other. For example, the maturation time of EGFP is reported as ~60 min (by folding of solubilized inclusion bodies), as ~130 min (by reoxidation), and as ~8 min (by *de novo* protein expression).[80,90,96] Although it is not clear why the reported rates differ that widely, each method requires unique sample conditions, which may contribute to the diversity of reported values. It seems prudent to compare studies that utilize the same experimental method. Because our study is based on cell-free expression, we focus in this paper on *de novo* protein expression methods.

However, even studies based on the same method report different maturation rates for the same fluorescent protein. Thus, additional experimental factors influence the maturation process. Although it has been noted that the kinetics of maturation of GFP is different at 25°C and 37°C,[97] the influence of temperature on chromophore maturation has not received further attention. We performed, to the best of our knowledge, the first quantitative characterization of maturation kinetics as a function of temperature. Published maturation studies have been conducted at specific temperatures, ranging from room temperature to 37^{o}C. Because the rate of maturation changes ~5-fold over this temperature range, one needs to account for this change when comparing maturation experiments.

Chromophore maturation by *de novo* protein expression is usually performed in one of two ways. Most studies employ anaerobic expression conditions followed by the addition of oxygenated solvent to complete the maturation process. Some studies utilize aerobic expression conditions and terminate protein expression while monitoring the subsequent increase in fluorescence due to maturing protein, as was done in this case. Adding RNase A to the solution stops gene expression almost instantly and allows the determination of maturation times of less than one minute as has been previously demonstrated.[80]

A large number of experiments probing maturation kinetics exist in the literature. To ensure a meaningful comparison we only consider studies based on *de novo* protein expression of the proteins EGFP, EYFP, or mCherry. We make no distinction between *de novo* protein expression by bacterial, cell-free, or eukaryotic cell expression systems. For EYFP a maturation time τ of 56 min measured at 25^{o}C has been reported, while a second

study determined a time constant of 23 min measured at 37°C.[81,98] These published data are in excellent agreement with the values determined from Figure 5.3 with τ = 58 min at 25°C and τ = 27 min at 37°C. The maturation of mCherry has been measured independently by two groups at 37°C, who reported values of 22 min and 57 min.[15,99] Our maturation time for mCherry at 37°C is τ = 57 min, which matches the value reported by Merzlyak et al.[99] Finally, the maturation lifetime of EGFP by cell-free expression has been reported as τ = 14 min at 37°C and as τ = 8 min at 29°C,[80,81] while our own data yield lifetimes of τ = 3 min at 37°C and as τ = 7 min at 29°C. Thus, our result is in close agreement with the value reported by Shin et al.[80] Surprisingly, EGFP is the fastest maturing protein among the ones tested. This observation is in contrast to its reputation of being a relatively slow maturing protein as established by previous denaturation and refolding studies. Because EGFP exhibits very fast maturation kinetics with *de novo* protein synthesis, it is important to ensure that a sufficiently fast rate of solvent oxygenation is guaranteed. Otherwise, the measured kinetics could be slowed down by the limited availability of oxygen.

EGFP's rapid maturation makes it an excellent reporter for studies of protein reactions and interactions conducted in a cell-free expression system, ensuring that the delay between expression and onset of fluorescence is minimal. In addition to traditional fluorescence methods, such as fluorescence lifetime and polarization, it is feasible to conduct fluorescence correlation experiments with cell-free systems. Expanding the repertoire of techniques in cell-free systems to FFS measurements of brightness requires a few precautions. The quality of protein expression of the cell-free system has to be fairly good. Misfolding of the fluorescent protein or failure to mature creates a non-fluorescent component that interferes with the interpretation of brightness experiments. Similarly, the presence of long-lived dark states would bias the interpretation of brightness studies. Observing the brightness of $EGFP_2$ provides a sensitive test of the completeness of protein maturation. For example, if 30% of EGFP molecules did not mature or existed in a long-lived dark state, a subpopulation of $EGFP_2$ would exist with only one of the two EGFPs fluorescing. This, in turn, lowers the brightness from 2 to 1.7. A detailed discussion of the influence of dark states and flickering of fluorophores on

brightness experiments is available.[45] The brightness of EGFP and $EGFP_2$ in mammalian cells is routinely measured as a control to establish that maturation and protein expression are not compromised. The brightness doubling observed with $EGFP_2$ (Figure 5.4) ensures that virtually all EGFP molecules of the cell-free system mature and acquire fluorescence.[45] The result further confirms the absence of long-lived dark states for the EGFP-fluorophore in two-photon excitation. In addition, our brightness experiments on NTF2-EGFP establish that protein interactions can be quantitatively assessed in a cell-free system over a wide concentration range.

In summary, cell-free expression offers a quick and convenient method to perform fluorescence studies of protein-protein interactions with no need for further purification of the product. The commercial *E. coli*-based cell-free system has proven to be suitable for brightness studies. Combining cell-free expression with the droplet setup ensures sufficient oxygenation of the sample and provides an excellent platform for chromophore maturation studies of fluorescent proteins as a function of temperature. Maturation kinetics exhibits a pronounced temperature dependence, which is conveniently characterized by determining the activation enthalpy and entropy. This information should prove useful for future comparison of maturation measurements and hopefully provides a first step towards establishing definitive rate coefficients for individual fluorescent proteins.

6. APOBEC

This chapter is motivated by a biological application. Building on some of the technical advances from the previous chapter, FFS experiments are conducted in aqueous solution to characterize the oligomerization behavior of an important enzyme and its interactions with DNA and RNA. We demonstrate that RNA induces strong oligomerization of A3G, depending on the RNA sequence, but that DNA, while also showing sequence dependence, does not cause significant oligomerization. These studies were performed in collaboration with the Reuben Harris lab at the University of Minnesota, from which Ming Li, Michael Carpenter and Judd Hultquist were the main contributors, as well as with Jinhui Li and Yan Chen from our own lab.

6.1 Introduction

APOBEC3G (apolipoprotein B mRNA-editing catalytic-like 3G) is the first human enzyme protein discovered that is capable of restricting HIV-1. This makes it an exciting target for research, although APOBEC3G (A3G) successfully restricts HIV-1 only in the absence of the HIV-1 virion infectivity factor (Vif), which HIV-1 has evolved to block this immune defense. Nevertheless, in Vif-deficient HIV-1, A3G is a key inhibitory factor and has been the focus of a great deal of study.[100,101] It has since been discovered that there are a family of APOBEC3 proteins, A3A, A3B, A3C, A3D/E, A3F, A3G, and A3H, several of which have innate immune activity against retroviruses, notably HIV-1 and HTLV.[101,102] Because of the importance of the APOBEC3 enzymes and the detrimental effects of Vif on their function, it is very important that we understand their immune defense mechanisms at the molecular level. This chapter focuses on A3G and also includes measurements of A3A, which does not restrict HIV-1. In its primary function against HIV-1, A3G is a DNA cytidine deaminase that changes C to U, thus causing degradation or debilitative mutation in the retrovirus. A3G is known to package into viral particles,[103,104] although the mechanism is not understood. It is also known to form high molecular mass (HMM) complexes, although they are reported to be enzymatically inactive.[105]

RNA is believed to play an important role in both the viral packaging and the formation of these HMM complexes.[106,107] Indeed, it has been shown that treatment of the HMM A3G complexes with RNase can restore the enzymatically active low molecular mass (LMM) form.[108] From our lab's cellular studies, described in the following section, we believe that HMM A3G and interaction with RNA play a role in the viral packaging of A3G. Furthermore, it has been shown that A3 proteins which do not package into the virion do not restrict HIV-1.[109] We therefore apply FFS to quantitatively study the oligomerization of A3G, its interactions with RNA and DNA, and to develop a clearer understanding of the functional mechanism of this important enzyme.

6.2 A3G in cells

Two members of the Mueller lab, Jinhui Li and Yan Chen, employed FFS to study the oligomerization of APOBEC3 proteins in living cells. Analysis of the data revealed that the APOBEC3 family members essentially fall into two categories of behavior. One set, including A3G, shows strong oligomerization as a function of protein concentration, saturating at approximately an average stoichiometry of a 20-mer homocomplex. The other set, including A3A, show almost no oligomerization and predominantly exists in a monomeric state. We single out A3G and A3A because they have been successfully purified for *in vitro* experiments and have also, consequently, received greater study. Figure 6.1 shows the average oligomerization of A3G and A3A as a function of concentration to illustrate the two behavioral categories. Collaborative efforts with the lab of Reuben Harris, Department of Biochemistry, Molecular Biology, and Biophysics at the University of Minnesota, revealed that the APOBEC3 proteins which show oligomerization in cells are packaged into the assembling HIV-1 virus.[109] The monomeric APOBEC3 members are not packaged and fail to restrict the virus. Since packaging is a prerequisite for HIV-1 restriction, these observations leads us to hypothesize that the oligomerization has a necessary and functional role in the packaging of APOBEC3 proteins into the budding virus and thus ultimately in the restriction of HIV-1.

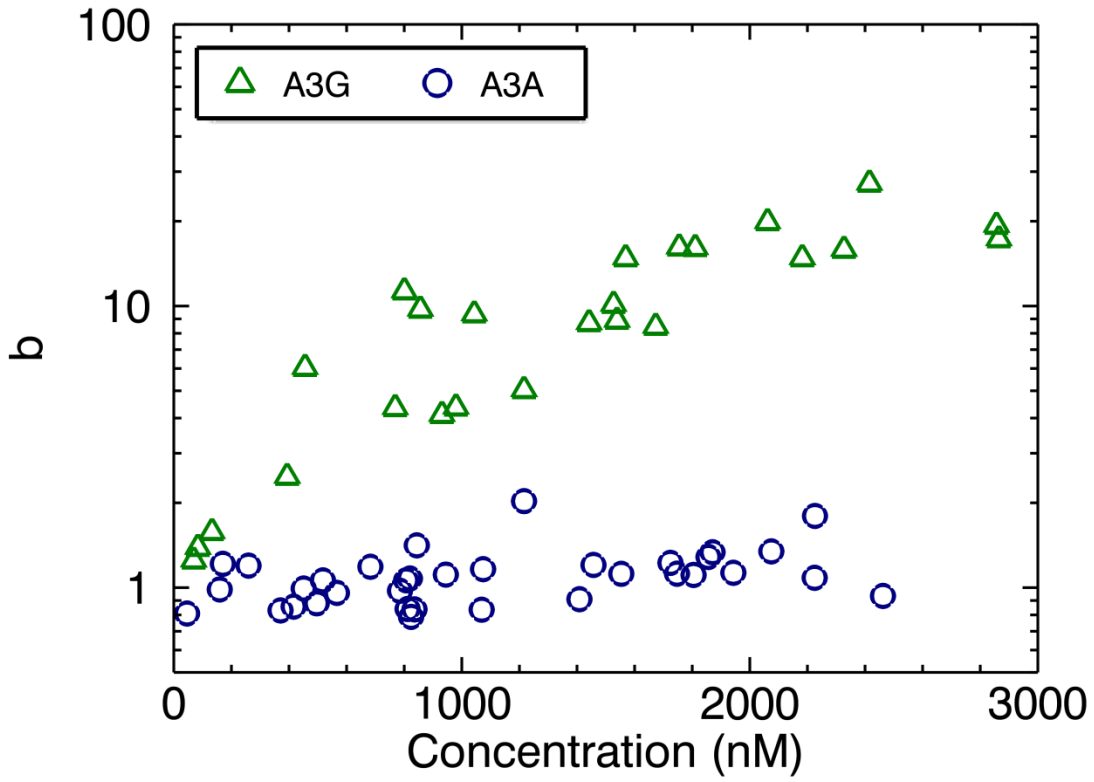

Figure 6.1 **A3G and A3A in cells.**

The plot displays normalized brightness measurements of EGFP-labeled A3G and A3A in U2OS cells in experiments performed by Jinhui Li. The inherent variation in protein expression levels creates a wide concentration range, and each point represents a different cell. A3G shows concentration dependent oligomerization in cells. A3A, contrastingly, remains essentially monomeric.

Because of the previously mentioned results that show RNA is involved in A3G oligomerization and packaging, we wish to make a closer study of A3G-RNA interactions. However, the cell is full of RNA and various other factors which may be interacting with A3G. Therefore, as previous chapters have discussed, we do not have the necessary control in the cell environment to conduct the desired experiments. We have no influence on the RNA population in cells, nor are those RNAs labeled. We therefore apply the droplet protocols from chapter 5 to study purified A3G and its interactions with RNA and DNA in an aqueous solution environment.

6.3 Materials

Purified proteins: A3G-H248A/H250A-Myc-His was transiently transfected into HEK 293T cells and purified by Ming Li of the Harris lab as previously described.[110] EGFP-labeled A3G, and labeled and unlabeled A3A samples were also provided by the Harris lab. Solution experiments were performed using the following buffers. *Protein buffer:* 50 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10% glycerol, 0.5% Triton-X 100. *Oligo buffer:* 50 mM Tris-Cl, pH 7.4, 10 mM EDTA. *NEB3 buffer:* 50 mM Tris-Cl, pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol (New England Biolabs, Ipswitch, MA). Experiments are conducted in a 1:1 mixture of protein and oligo buffer or in purely NEB3 buffer. No differences were observed between these two buffer systems. *RNA/DNA oligonucleotides:* All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA).

Single-stranded RNA oligonucleotides -

Test+ RNA (27b): GUGAGCAAGG GCGAGGAGCG GGAGGGA

is a short sequence from EGFP with the addition of two GGGA repeats.

Test- RNA (25b): GACAAGGUCU UUACUGGGUG UAUUU

is derived from the bar domain of APPL1 membrane adaptor protein.

CA RNA (25b): AACCAACCCAAACAACCCACACACC

GA RNA (25b): AAGGAAGGGAAAGAAGGGAGAGAGG

GU RNA (25b): UUGGUUGGGUUUGUUGGGUGUGUGG

CU RNA (25b): UUCCUUCCCUUUCUUCCCUCUCUCC

Single-stranded DNA oligonucleotides -

TCCCA ssDNA (21b): AGATCCCATAATAGATAATGT

is the A3G deaminase target oligo used in Shlyakhtenko et al.[110]

GT ssDNA (25b): TTGGTTGGGTTTGTTGGGTGTGTGG

CT ssDNA (25b): TTCCTTCCCTTTCTTCCCTCTCTCC

Fluorescently-labeled oligonucleotides –

5'-labeled Texas Red. TCCCA ssDNA

5'-labeled Texas Red.Test+ RNA

5'-labeled Cy5.GA RNA

5'-labeled Cy5.CA RNA

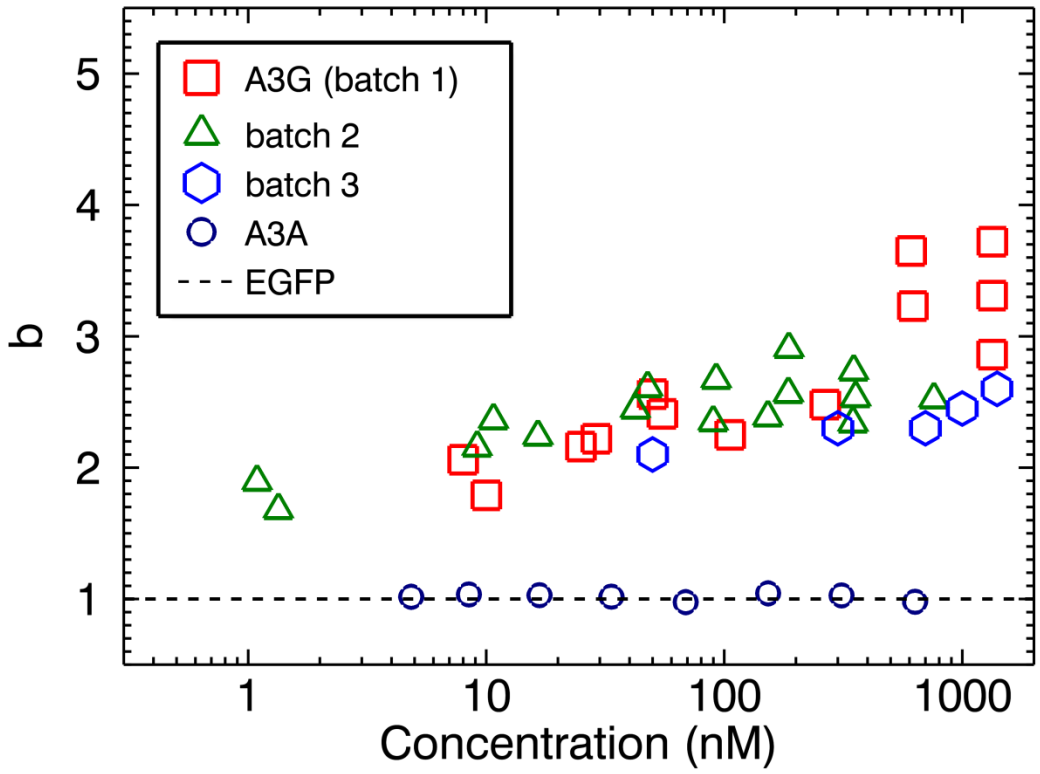

Figure 6.2 **A3G in solution.**

In solution, A3G appears to be a tight dimer. The average normalized brightness value measured is approximately 2.3, which we interpret as a predominantly dimeric population with a small percentage of higher order complexes mixed in. There is a weak concentration dependence as well as some variation from one batch of purified A3G to another, but average brightness values slightly greater than 2 are very repeatable. A3A measurements fall directly on top of the EGFP values, indicating that A3A is monomeric and independent of concentration. Thus, in aqueous solution as in cells, differences are observed in the behavior of A3G and A3A.

6.4 A3G in solution

In order to study A3G-RNA interactions in solution, we must first characterize the behavior of A3G alone in solution. Initially, we perform dilution studies of purified A3G in aqueous buffer. Figure 6.2 shows that from nanomolar to micromolar concentrations, the brightness stays close to 2 with only a weak concentration dependence. A3G in solution is therefore primarily dimeric. A close inspection of the data reveals that the normalized brightness values are slightly above 2, which we interpret as the presence of a small population of higher order complexes. For instance, an 8% population of tetramers would return the observed apparent brightness, as would lower percentages of higher order complexes. Individual purification batches may generate slightly different percentages of the larger complexes in the stock solution, explaining the brightness variations at the highest concentrations. A3G in solution can also be sensitive to temperature (see Appendix B). Nevertheless, below ~1 μM, the brightness values are independent of preparation and are very repeatable. We note that Salter et al. also found A3G to be predominantly dimeric in solution with a small percentage of monomers and tetramers, and that their mix of A3G oligomers would also yield an apparent brightness slightly above 2.[111] Figure 6.2 also shows a dilution study of A3A in solution. Throughout this chapter, we will use A3A as a control and to emphasize that A3G and A3A exhibit different behaviors. As in the cell measurements, A3A is revealed to be a monomer, independent of concentration, provided the solution is a reducing environment (Appendix A).

6.5 A3G interactions with RNA

As previously mentioned in the introduction, A3G has been shown to interact with RNA, although little has been reported about specific RNA targets.[112] A few studies have suggested that certain broad RNA sequences have a preferred association with A3G, such as 7SL RNA and HIV-1 RNA, which are also packed into the HIV-1 virion.[113–115] We were intrigued by a paper which reported a more general RNA-sequence dependence to interactions with A3G. Bogerd and Cullen found that the interaction of A3G and the nucleocapsid (NC) domain of HIV-1 is dependent on single-stranded RNA and, more

specifically, on RNA containing guanine bases.[116] However, their method could only probe the effect of RNA on A3G-NC complexation. Inspired by this result, we have constructed our study to explore the direct effect of RNA on A3G. We wish to determine if RNA sequence plays a functional role in A3G oligomerization and therefore, as hypothesized, with viral packaging and HIV-1 restriction. Thus, we conduct solution experiments of purified A3G with short RNA oligonucleotides. Studies have shown that A3G has difficulty binding oligonucleotides shorter than 10 bases,[108,116] so we select sequences of greater than 20 bases. The experiments are conducted by mixing different RNA sequences with A3G in a series of 4 μL droplets. A3G is always added to a final concentration of 200 nM, and RNA to final concentrations of 200 nM or 2 μM, for a 1:1 and 1:10 ratio of A3G to RNA. The first two sequences displayed in Figure 6.3 are simply test cases, Test+ RNA and Test– RNA. The sequence of Test+ RNA is mostly drawn from EGFP and is selected to have a high percentage of guanine bases. Test- RNA is part of the APPL1 gene sequence and is selected to have a more uniform distribution of the four RNA bases: adenine (A), guanine (G), cytosine (C), and uracil (U). Figure 6.3 reveals that, for this experiment, A3G alone has a normalized brightness of 2.1, but in the presence of Test+ RNA, the oligomerization noticeably increases. Interestingly, 10 times excess of Test+ RNA over A3G leads to an even stronger multimerization of A3G than the 1:1 mixture. Test– RNA, on the other hand, effectively has no influence on the oligomerization of A3G.

Because of this observed sequence dependence, we perform experiments involving more precise control in the composition of the RNA oligonucleotides. We construct a series of 25 base RNA strands containing combinations of only two nucleobases: CA, CU, GU, and GA RNA. Each sequence has an identical pattern of repeats, only varying the bases involved, as described in the Materials section. These limited-base sequences are similar to the sequences used in the Bogerd and Cullen paper. Similarly to Bogerd and Cullen's report that "G-rich" RNA is required for NC domain binding, we observe that "G-rich" RNA induces A3G oligomerization. The two sequences without any guanines fail to cause oligomerization, but closer inspection of the data reveals divergent behavior for the two different sequences. While CU RNA exhibits

no effect on A3G oligomerization, CA RNA appears to cause dissociation of the A3G dimer. This is a surprising and exciting result, for it demonstrates that not only does the composition of the RNA affect A3G's response, but that different sequences can generate entirely different A3G behaviors.

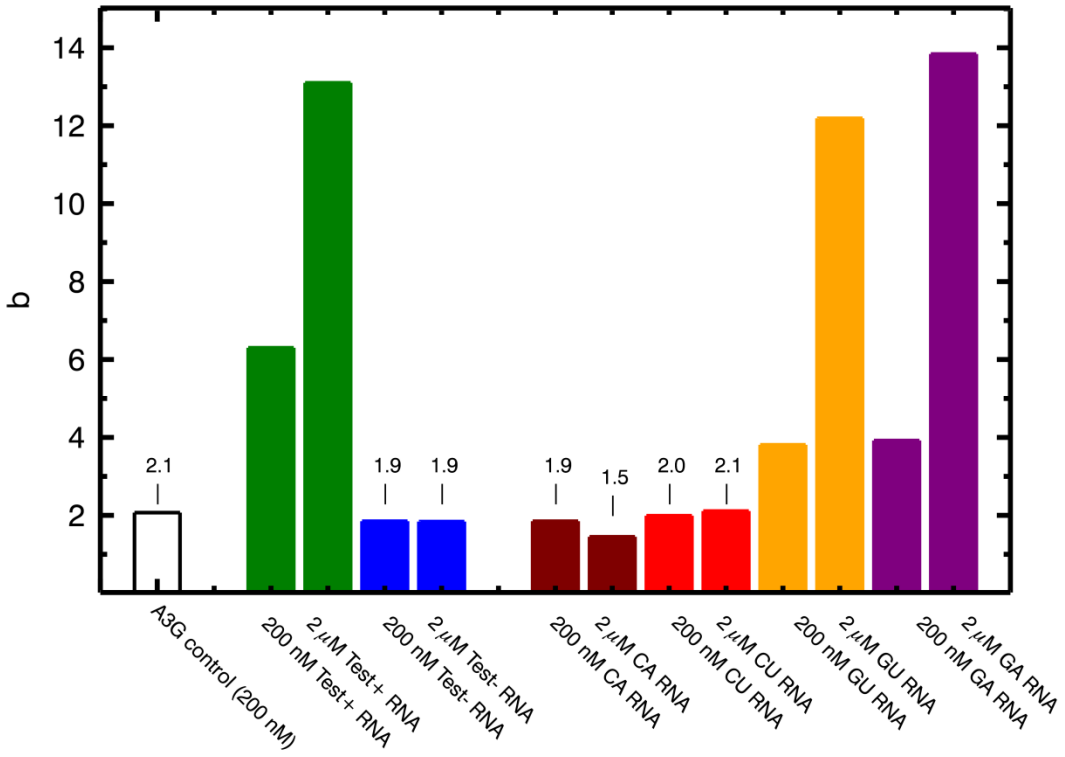

Figure 6.3 **Sequence dependent RNA interactions with A3G.**

In this plot, we show the response of EGFP-labeled A3G to the introduction of different RNA oligonucleotides. These experiments are conducted in solution droplets at room temperature. Each RNA sequence (25-27 bases long) is mixed with A3G at a 1:1 and a 1:10 ratio. The normalized brightness value, or height of the column, indicates the average number of A3Gs that complex together through interaction with RNA. Here, the control measurement of A3G returns a value of 2.1, again confirming that A3G alone forms a dimer. The addition of Test+ RNA, derived from an EGFP sequence, causes significant oligomerization and appears correlated with the amount of RNA. On the other hand, Test– RNA, derived from an APPL1 protein sequence, show effectively no change in oligomerization. Test+ RNA has a predominance of guanine bases, while Test– RNA is much more evenly distributed amongst the four: guanine (G), cyostine (C), adenine (A), and uracil (U). The remaining four RNA sequences are constructed of only two bases each, and again the two “G-rich” sequences show oligomerization behavior, while the other sequences do not. In fact, the CA sequence appears to have a dissociative effect.

To investigate this response more carefully and to explore the role of RNA concentration, we single out Test+ RNA and CA RNA for droplet titration experiments as described in chapter 2. EGFP-labeled A3G and A3A are held fixed at a 200 nM concentration in a series of droplets with varying concentrations of RNA. Figure 6.4 shows that at low concentrations of Test+ RNA, the A3G oligomerization (diamonds) is slightly above 2, just as it is in the absence of RNA. However, starting at about a 1:1 ratio of A3G to Test+ RNA, complexes form that grow with increasing RNA concentration to an average size of about 20 copies of A3G per complex. We note that this complex size falls in a similar range to that found in the A3G cell experiment (Figure 6.1). Figure 6.4 also shows that A3A (circles) remains monomeric in the presence of Test+ RNA. A3A behaves as expected, based on the cellular measurements, but the details of the A3G response do not match the hypothesized model. This will be addressed in the discussion section.

In startling contrast, Figure 6.5 shows that CA RNA does not induce oligomerization of A3G (squares), but rather has precisely the opposite effect. Increasing concentrations of CA RNA appear to dissociate the A3G dimer, which we believe to be very tight (see Appendix C). At low RNA concentrations, A3G has a brightness slightly above 2 as expected, but the average oligomerization of the A3G sample shifts towards the monomer as CA RNA achieves 100 times excess concentrations. At very high concentrations of RNA (> 30 μM), this trend breaks down, possibly due to Mg^{2+} dependent RNA homo-interactions.[117] We note that Chelico et al. report that the DNA substrate sequence ‘AAACCCAAA’ shows the highest deamination activity in their sequence-dependent study,[108] a point that may be relevant to the above experiment and will be discussed in a later section. Finally, we again include A3A (circles) as a control which does not show any effect from CA RNA.

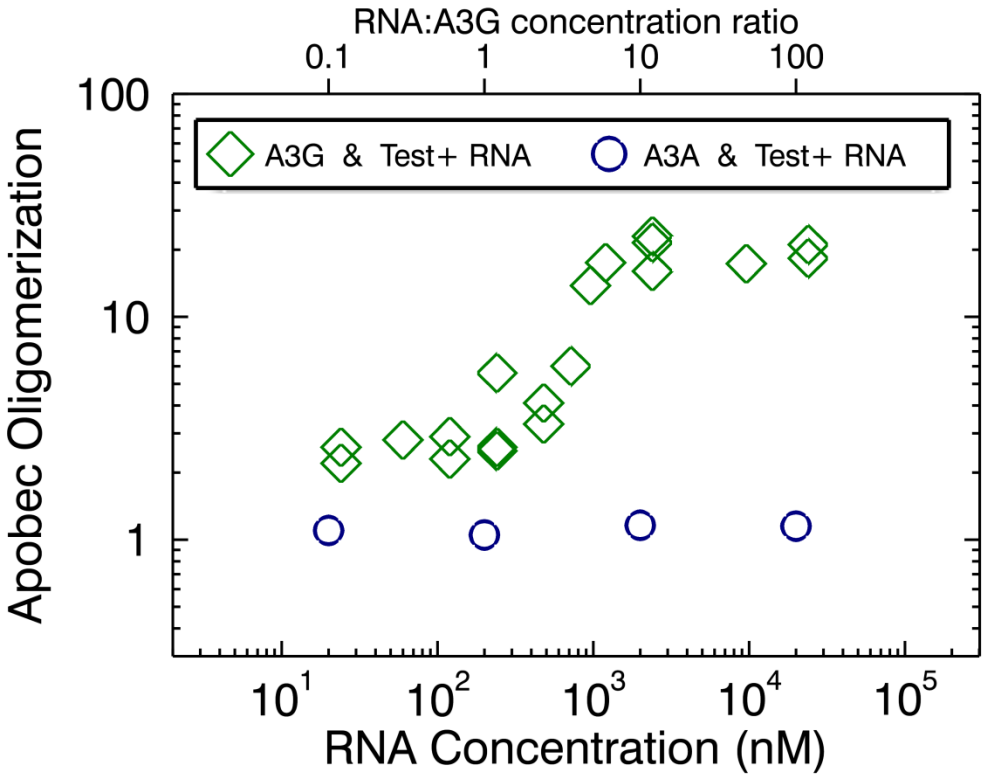

Figure 6.4 **EGFP-labeled APOBEC and Test+ RNA droplet titration.**

Purified, EGFP-labeled A3G and EGFP-labeled A3A are fixed at 200 nM in a droplet titration with Test+ RNA. The bottom axis shows the final concentration of RNA in each droplet while the top axis shows the ratio of RNA to APOBEC. As before, A3G shows strong oligomerization that appears to initiate at about a 1:1 ratio of A3G to RNA and to saturate at 1:10. We also show A3A data which exhibits no response to the presence of the RNA, and remains monomeric.

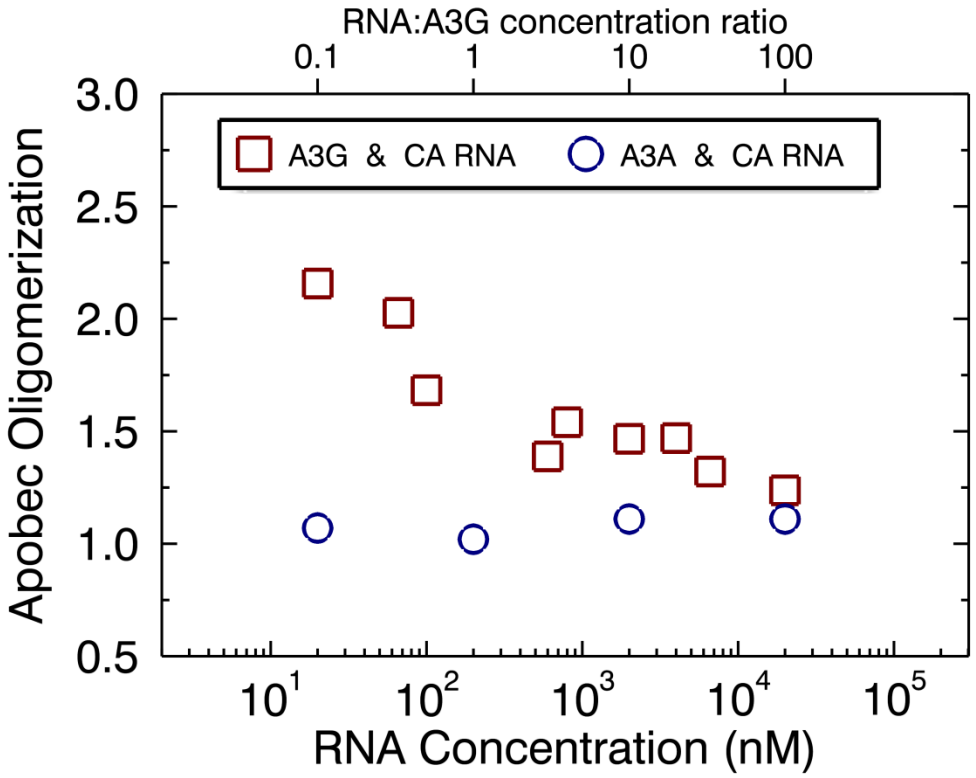

Figure 6.5 **EGFP-labeled APOBEC and CA RNA droplet titration.**

Purified, EGFP-labeled A3G and EGFP-labeled A3A are fixed at 200 nM in a droplet titration with CA RNA. Here, we see that CA RNA does induce dissociation of the A3G dimer. At the lowest RNA concentration measured, A3G has a normalized brightness of 2.2 similar to A3G alone. Increasing the CA RNA concentration leads to clear dissociation of the A3G dimer and begins to approach a monomeric value. Again, A3A shows effectively no response to the presence of RNA.

6.6 A3G interactions with ssDNA

Although our study has dealt thus far with the interaction of A3G and RNA, the role of A3G in restricting HIV-1 is based on its interaction and enzymatic deamination of single-stranded DNA (ssDNA). For deamination of ssDNA, A3G's preferred target (5'-…CC<u>C</u>) is the third C in a sequence coming from the 5' end.[104,118,119] Shlyakhtenko et al. have reported a small increase in oligomerization as a result of interaction with ssDNA and, as mentioned above, Chelico et al. found that the sequence of bases flanking the CCC target have significant effects on deamination efficiency.[108,110] We now use FFS and brightness analysis to study the oligomerization response of A3G to the ssDNA versions of the previously used RNA sequences (CA, CT, GT, GA). We see in Figure 6.6 that while there is a sequence dependent response, it does not precisely match that of RNA. Only the GT ssDNA sequence appears to induce oligomerization, and the effect is rather weak compared to RNA. The GA ssDNA sequence does not seem to have any effect at all. We note, however, that GU ssDNA, containing deoxyuracil, causes oligomerization to a comparable level with "G-rich" RNA. Figure 6.7 shows the earlier A3G oligomerization data for CU/GU RNA plotted alongside CU/GU ssDNA and CT/GT ssDNA. It becomes apparent that ssDNA is not influencing A3G oligomerization, except in the guanine-deoxyuracil sequence. This may be evidence that it is not merely guanine, but a combination of bases that triggers oligomerization. As expected, neither CT nor CU ssDNA have any oligomerization effect on A3G. Interestingly, CA ssDNA induces dissociation of A3G (Figure 6.6), which agrees with the observed effect of CA RNA.

Overall, it seems that ssDNA does not induce A3G oligomerization, and indeed that certain sequences dissociate the dimeric A3G subunit. On the other hand, we have strong evidence that "G-rich" RNA sequences induce large oligomeric A3G complexes. These results appear to fall in line with the two disparate A3G oligomeric roles discussed in the introduction. We have hypothesized that A3G oligomers are necessary for packaging into the virion which, in turn, is required for restriction. Yet A3G enzymatic deamination, the primary restrictive function, can only be efficiently completed by LMM A3G complexes. Taken together with our results, these data suggest a model. "G-rich"

RNA assists A3G to oligomerize and package into the virion. As the retrovirus undergoes reverse transcription and degrades RNA via ribonuclease H,[120] the oligonucleotide population shifts from RNA to DNA. DNA does not cause A3G oligomerization and may cause A3G dissociation. Thus, the A3G oligomers will reduce to LMM complexes and can complete their enzymatic deamination to restrict HIV-1.

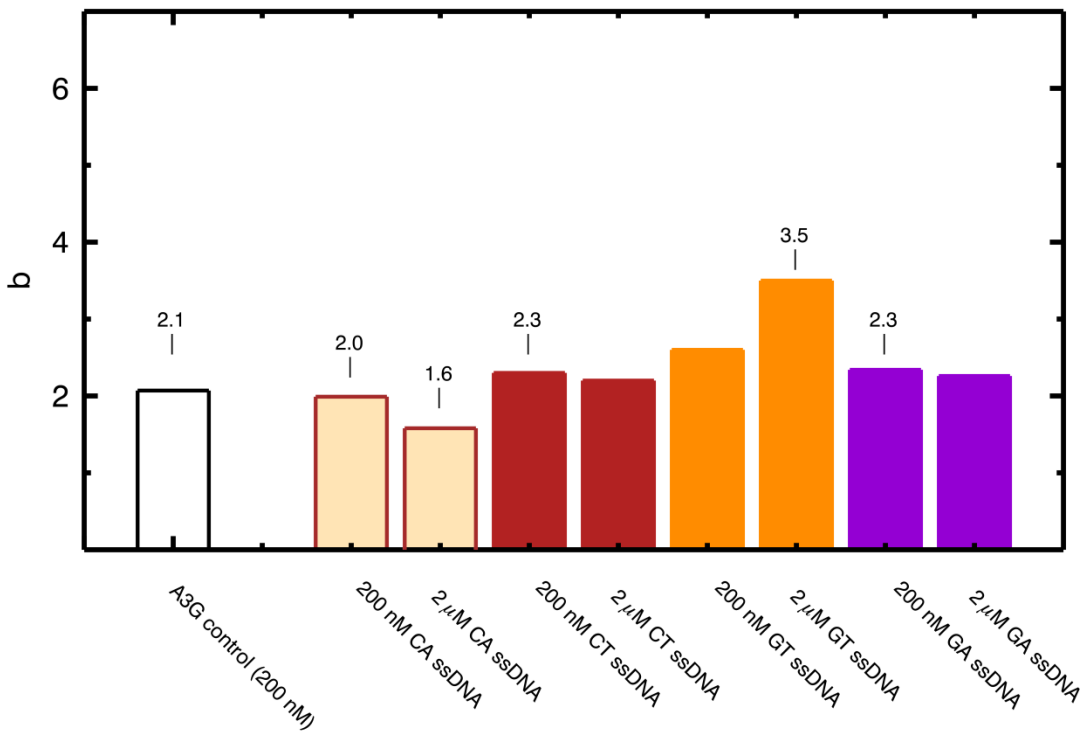

Figure 6.6 **Sequence dependent ssDNA interactions with A3G.**

In this plot, we show the response of EGFP-labeled A3G to different ssDNA constructs. The four sequences are the DNA analogs of the RNA sequences described previously, with thymine (T) replacing uracil. We see that CA ssDNA also has a dissociative effect on the A3G dimer. Only GT ssDNA increases oligomerization, but it is a much weaker effect than GU RNA.

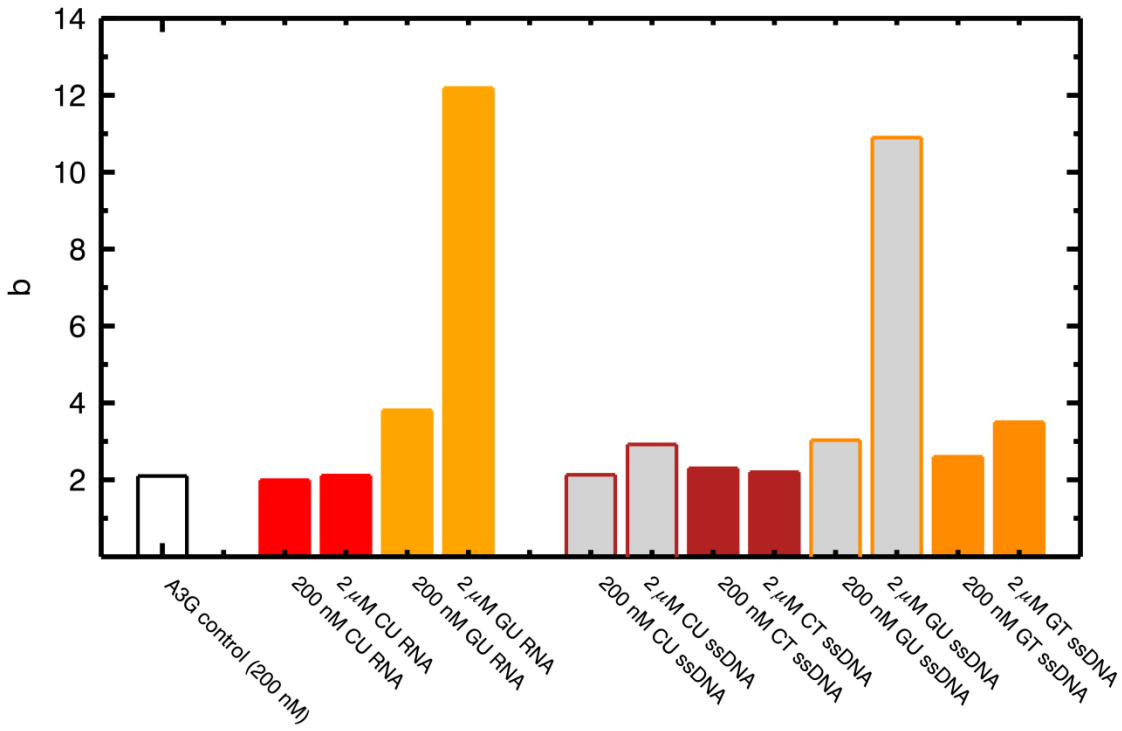

Figure 6.7 **GU ssDNA versus GT ssDNA A3G oligomerization.**

The A3G oligomerization in the presence of CU/GU RNA (from Figure 6.5) is plotted alongside A3G in the presence of CU/GU ssDNA, containing deoxyuracil bases, as well as the natural ssDNA analogs CT/GT (from Figure 6.6). Overall ssDNA does not have much effect on oligomerization except when it contains deoxyuracil. This may indicate that "G-rich" sequences require particular base combinations, or perhaps that the deoxyuracil causes ssDNA to bind in the RNA pocket and thus generates the corresponding oligomerization effect.

6.7 FFS diffusion studies

The previous sections focus on the main goal of this study, answering the question of A3G oligomerization as dependent on RNA and ssDNA. The next two sections contain preliminary experiments exploring the direct interaction of A3G with RNA and DNA, which will be necessary for constructing a more realistic interaction model of these biomolecules. Thus far in our APOBEC studies, we have always had fluorescently-labeled APOBEC and unlabeled oligonucleotides. This means all of our data only provide information about APOBEC copy numbers, and furthermore, those data never directly measure the interaction with nucleotides, but only assume interaction with RNA/DNA because of a change in the A3G stoichiometry. Therefore, we change gears and work with unlabeled protein and fluorescently-labeled RNA and ssDNA. In these experiments, brightness values will provide information about RNA/DNA copy numbers, and we can take advantage of diffusion information to study interactions. The 25-base oligonucleotides we use in this experiment are small (~ 9 kDa) and therefore diffuse quickly. EGFP-labeled A3G (~ 73 kDa) and EGFP-labeled A3A (~ 50 kDa) move much more slowly through solution, so that when the fluorescent oligonucleotide binds to the larger protein, the slowing in diffusion rate is detectable. Figure 6.8 shows autocorrelation curves for EGFP-labeled A3G and Texas Red-labeled RNA, along with their respective fits. This method for studying interactions is much less sensitive than brightness because diffusion time goes approximately as the cube root of molecular weight, $\tau_D \sim (MW)^{1/3}$, a relationship that assumes hydrodynamically spherical proteins. The interaction between two similar weight proteins leads to only a 25% change in the diffusion coefficient, which is close to the intrinsic uncertainty in diffusion coefficient measurements. In the case of interaction between two objects with sufficiently different molecular weights, however, it is a very useful method. For an oligonucleotide binding dimeric EGFP-labeled A3G, the molecular weight relationship predicts nearly a 3-fold increase in the time required for the complex to diffuse through the observation volume compared to that of the free oligonucleotide.

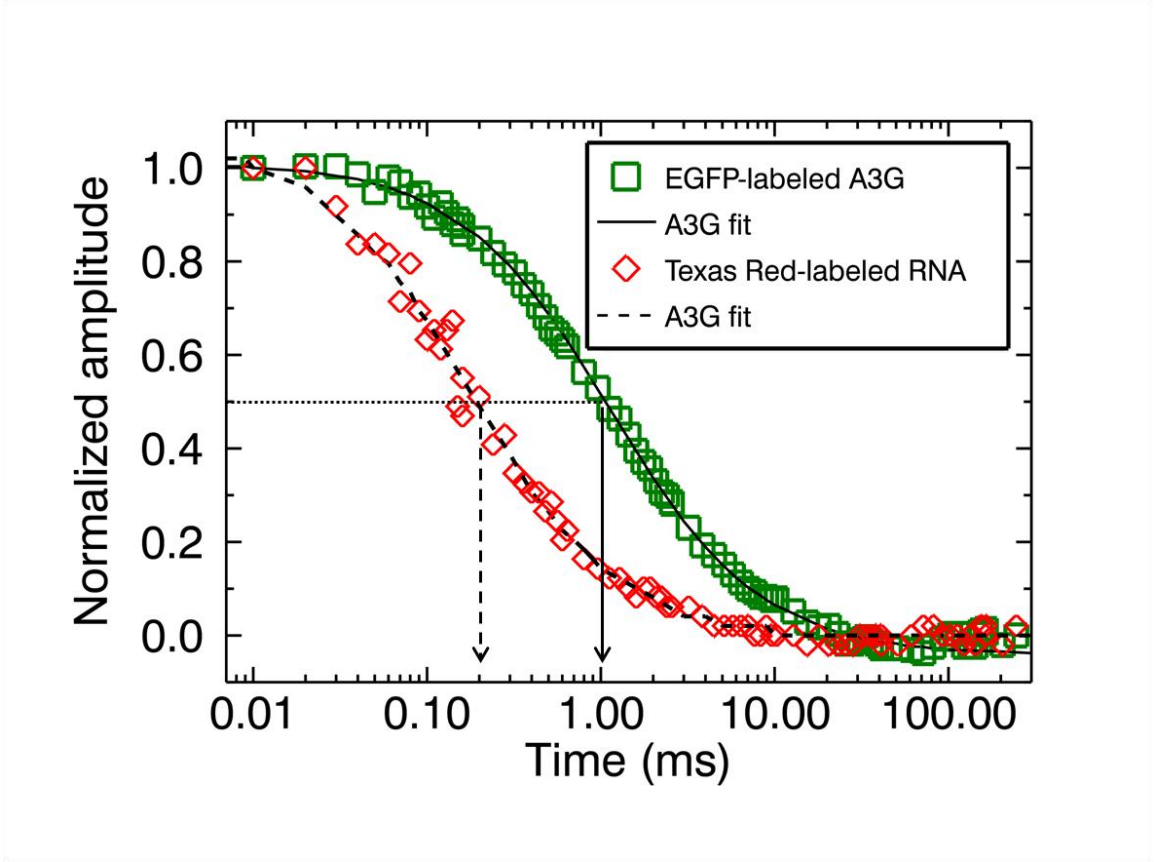

Figure 6.8 **Autocorrelation curves for A3G and RNA.**

We display the normalized autocorrelation curves for EGFP-labeled A3G (squares) and Texas Red-labeled Test+ RNA (diamonds). Note the strong difference in the characteristic diffusion time for the RNA oligonucleotide and the A3G dimer. This change makes it possible to use diffusion to study interaction.

We begin with a mainly qualitative experiment to confirm the interaction of A3G with RNA and ssDNA, and to assess the corresponding interactions with A3A, for which our previous experiments have not provided any direct information. Figure 6.9 displays the diffusion time controls for ssDNA (dotted outline) and RNA (solid outline). The TCCCA ssDNA is labeled with a Texas Red organic dye and is derived from the deamination assay used in Shlyakhtenko et al.[110] The Test+ RNA is the same sequence used in the previous sections but now including a Texas Red label on the 5' end. Both labeled oligonucleotides have a diffusion time $\tau_D \sim 0.13\, ms$ in their unbound state. If the ssDNA or RNA binds APOBEC, we expect to see an increase in the diffusion time since the larger complex takes longer to pass through the observation volume. We emphasize that these preliminary experiments are somewhat qualitative. Because both the bound and unbound fluorescently-labeled oligonucleotides will contribute to the signal, the measured diffusion time represents an average, and a single measurement cannot resolve the two subpopulations. Nevertheless, an increase in diffusion time is a firm indication of interaction between A3G and oligonucleotide since the effect of the unbound population is to reduce the average diffusion time back toward the control value. Samples contain 5-fold excess of proteins over oligonucleotides to shift the oligonucleotide population towards the bound state.

The next set of columns in Figure 6.9 show the interaction of A3G with TCCCA ssDNA and Test+ RNA. Having measured a diffusion time of $\sim 1\, ms$ for an EGFP-labeled dimer (Figure 6.8), we estimate $\sim 0.8\, ms$ for the unlabeled A3G dimer. Therefore, the diffusion time of $0.4\, ms$ for the 5:1 A3G/ssDNA sample clearly demonstrates the presence of interactions, but since the value is below $0.8\, ms$, not all ssDNA is bound. The interaction of A3G and Test+ RNA, on the other hand, reveals that the RNA is almost completely bound, having a diffusion time of $0.75\, ms$, which suggests that A3G has a stronger affinity for RNA than for ssDNA. However, these experiments have to be confirmed with RNA and DNA that share the same sequence.

A3A is known to interact with ssDNA because it is a very efficient deamination enzyme.[121] However, we were unable to detect interaction until we increased the

A3A/ssDNA ratio to 100:1 in order to shift the equilibrium towards the bound state. Under such a condition, the ssDNA is always surrounded by binding partners and thus spends most of its time in the bound state, even if the average duration of the binding is very short. A3A's transient binding is likely an important aspect of its enzymatic efficiency, in that it is always free to bind a new substrate. Finally, we are able to establish that A3A does indeed bind to Test+ RNA, despite the fact that there is no change in oligomerization. This supplies further evidence that A3G and A3A substantially differ in their behavior. The diffusion time of $\sim 0.4\, ms$ indicates that RNA binds to A3A with less affinity than to A3G. We note that A3A also shows a stronger affinity for RNA than ssDNA, but again, this needs to be confirmed with RNA and DNA of identical sequence.

The labeled oligonucleotide method can also be used to gain additional information about the interaction of A3G and Test+ RNA. We know Test+ RNA induces oligomerization and that oligomerization initiates at about a 1:1 ratio of protein to RNA. The questions become whether RNA oligonucleotides are, in fact, included in the large A3G complexes and why the 1:1 ratio is required for complex formation. A more quantitative diffusion study with labeled RNA can address these points using diffusion to identify RNA interaction with the complexes, and oligonucleotide brightness to investigate RNA copy number.

In Figure 6.10 we display a droplet titration in which A3G is unlabeled and thus invisible to FFS. The data is plotted as diffusion time versus the ratio of RNA to A3G; the top axis shows absolute RNA concentration which is varied while A3G concentration is held fixed. A control measurement was made in a sample droplet containing only the labeled RNA construct, and the resulting diffusion time is represented by the gray line in the figure. Clearly, even at the lowest RNA/A3G ratio measured, the observed diffusion time is significantly larger than the control. Therefore, when there is excess A3G protein the Test+ RNA is binding, but apparently not causing oligomerization. Normally, we would expect the diffusion time to reduce or stay approximately constant as the ratio of labeled RNA increases, since we are increasing the level of unbound RNA. Recalling the ~20% uncertainty, the data show a roughly constant diffusion time until we reach the 1:1

ratio. We do, however, expect an increase in diffusion time around the 1:1 ratio because A3G starts to oligomerize. The measured diffusion time of $\sim 2.6\ ms$ clearly indicates the inclusion of RNA in large complexes. Additionally, the diffusion value is consistent with the size of oligomer complexes observed in the EGFP-labeled A3G experiments, although we mustn't overinterpret this numerical value because it is averaged over a mixture of A3G-RNA complexes and free RNA oligonucleotides. The experiment should be repeated over a broader range of mixtures in which we expect the higher RNA ratios to show reducing average diffusion times as the free oligonucleotides dominate the signal. Nevertheless, the data do show that RNA is bound into the large A3G complexes and agree with the labeled A3G experiments (Figure 6.4) that this oligomerization occurs at ~1:1 ratio. The simplest explanation for this transition point is that more than one RNA per A3G are required for complex formation, a hypothesis we will return to in the next section of this chapter.

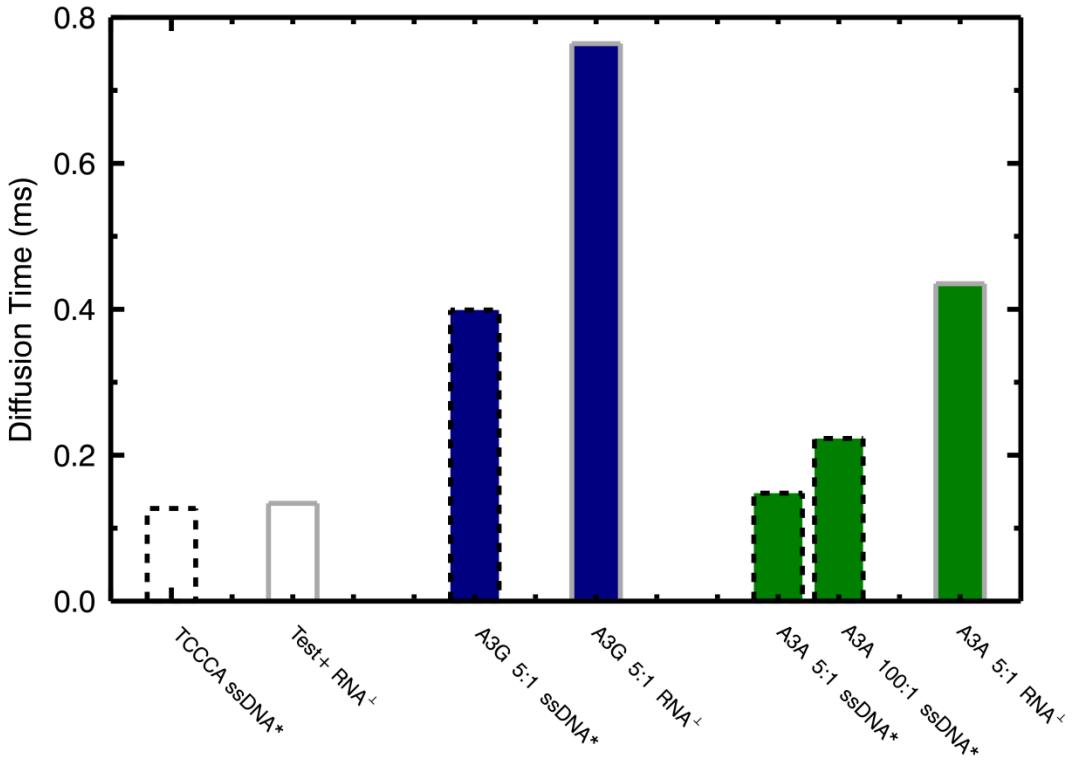

Figure 6.9 **Diffusion studies: APOBEC-oligonucleotide interaction**

The first two columns are the control measurements for the fluorescently-labeled ssDNA (dotted outline) and RNA (solid outline) samples. TCCCA ssDNA and Test+ RNA are used throughout this experiment. Both have similar diffusion times, $\tau_D \sim 0.13\, ms$. These preliminary experiments are conducted with 5 times excess APOBEC over oligonucleotide to reduce the probability of unbound fluorescent oligonucleotides. A3G shows clear interaction with the deamination target, TCCCA ssDNA, by evidence of the slower diffusion time. By comparison, the significantly slower motion of the A3G-Test+ RNA sample indicates a stronger affinity and therefore a higher fraction of bound oligonucleotide. The interaction of A3A with ssDNA displays an even weaker affinity, and requires a higher concentration of protein to "catch" the oligonucleotides in a bound state. This very transient binding is likely an important aspect of A3A's enzymatic efficiency. Finally, we observe that A3A does indeed bind Test+ RNA, though we know from previous experiments that this binding does not change its oligomerization state.

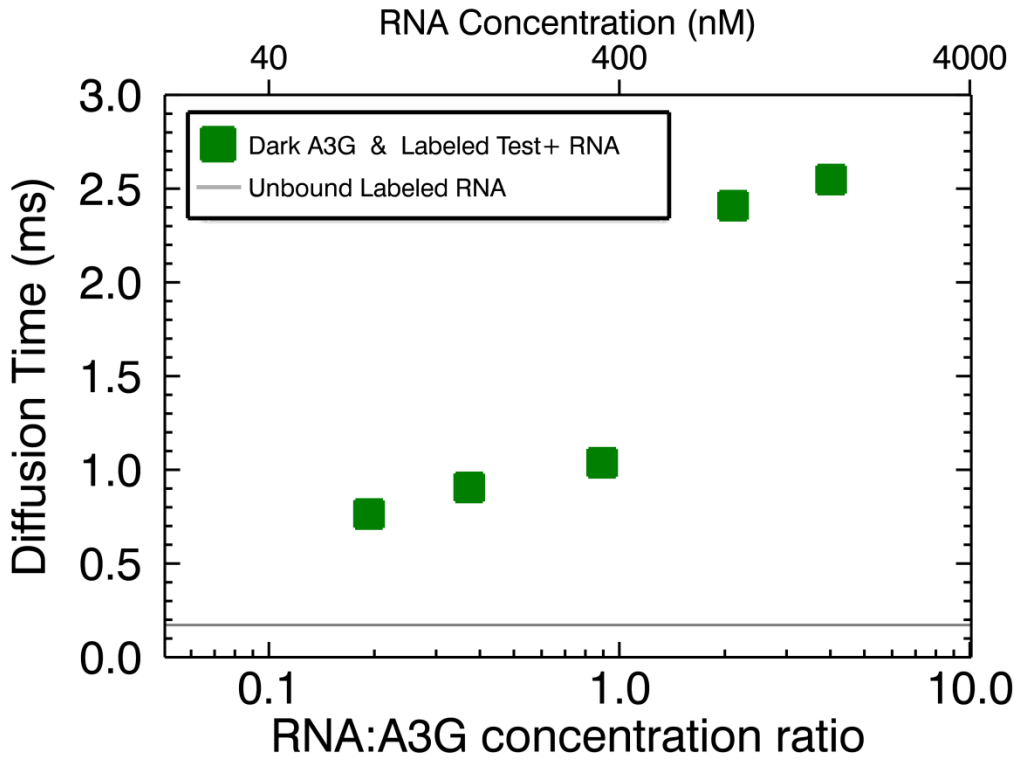

Figure 6.10 **Diffusion study binding of A3G and Test+ RNA.**

Unlabeled or "dark" A3G is mixed in droplets with varying concentrations of fluorescently labeled Test+ RNA. The data are plotted as diffusion time versus the RNA to A3G concentration ratio, while the top axis shows the absolute concentration of RNA. Since even the smallest RNA/A3G ratio measured shows a diffusion time much slower than that of "free" labeled RNA control, it is clear that A3G and Test+ RNA are bound prior to the oligomerization response. Again, at roughly a 1:1 ratio, we observe a dramatic increase in diffusion time as the large, slow complexes are formed. However, this experiment has been performed only once and the data are rather noisy. Thus, it will be important to repeat the experiment.

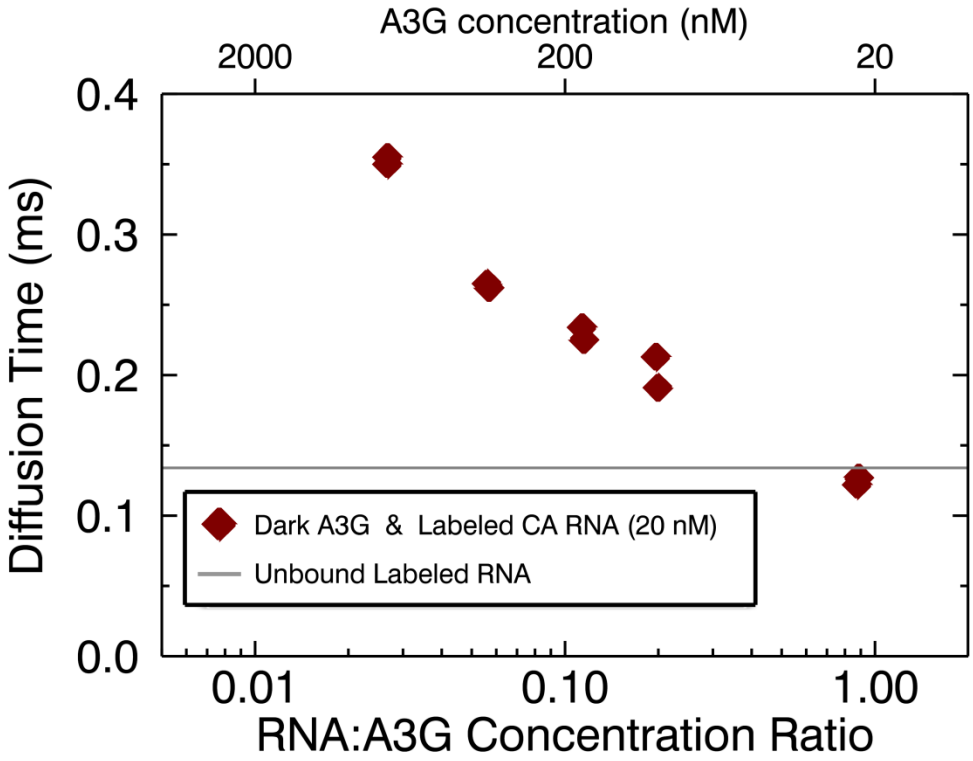

Figure 6.11 **Diffusion study binding of A3G and CA RNA.**

The labeled CA RNA concentration (20 nM) was held fixed across the droplets, and the dark A3G concentration was varied. The data are plotted as a function of RNA/A3G ratio to correspond with the previous figure. We see that on the left side of the plot, in the case of excess protein, most of the CA RNA is bound because of the slow diffusion time of the complex. As we decrease the concentration of unlabeled A3G, there are fewer binding targets and more free RNA, thus the measured diffusion times eventually fall to the control value of free RNA. Note that this data is convolved with the A3G dissociation reaction induced by CA RNA.

We also performed a droplet titration diffusion study on the interaction between unlabeled A3G and labeled CA RNA (Figure 6.11). The CA RNA sequence is the same as before, now with a Cy5 dye attached to the 5' end. Here, we hold the concentration of the labeled CA RNA fixed and vary the A3G concentration, which differs from the experimental designs from the first half of the chapter. This approach prevents variation in the fluorescence signal strength from droplet to droplet. The diffusion times are plotted as a function of RNA/A3G concentration ratio to correspond with the previous figure, and the top axis shows the absolute A3G concentration. In this case, we observe a simpler and more typical diffusion binding model. At an excess protein or low RNA/A3G ratio, most of the labeled CA RNA is bound, and we measure a slow diffusion time. As the concentration of dark A3G decreases and there are fewer protein binding targets, the signal is dominated by free, unlabeled CA RNA, and the measured diffusion times drop to the control value. We've previously demonstrated that CA RNA causes dissociation of A3G, a process which also reduces the diffusion time. However, the difference between diffusions times of the dimer and monomer is only ~25%, thus this effect on the above experiment is small. The lowest RNA/A3G ratio measured is about 30:1 at which ratio the A3G is still fully dimeric according to Figure 6.5. Therefore the measured diffusion time at this ratio $\sim 0.35\, ms$, compared to the fully bound estimated diffusion time of $\sim 0.8\, ms$, indicates that CA RNA is bound to A3G with weaker affinity than either Test+ RNA or TCCCA ssDNA.

6.8 Oligonucleotide brightness

The previous section focused only on diffusion times. However, each of those experiments also recorded brightness values for the oligonucleotides during interaction with APOBEC. These brightness results are summarized Table 6.1. Because A3A has weak affinity, we show the ssDNA brightness for the 100-fold excess A3A case. We estimate from the diffusion time ratio $(0.2/0.8)$ that about 25% of the ssDNA is bound. If that 25% ssDNA had a possible two binding sites per A3A and the remaining 75% ssDNA stayed unbound, the average measured brightness would have a value of $b \sim 1.4$.

The flat brightness $b = 1.0$ is a strong indication that A3A only has one ssDNA binding site. This appears to hold true for Test+ RNA and CA RNA as well, although additional measurements should be taken to determine CA RNA brightness in the case of excess A3A. These data are consistent with the facts that A3A is monomeric and has only a single catalytic domain.[122,123] A3G has two domains, CD1 which is necessary for RNA binding, and CD2 which catalyzes ssDNA.[115,124] Additionally, A3G exists as a dimer in our solution studies, yet this dimer only appears to bind one ssDNA strand. One might speculate that in the formation of the dimer, one of the CD2 domains is blocked. Salter et al. have performed small-angle x-ray scattering (SAXS) experiments on A3G and suggest a "tail-to-tail" model that puts the two CD2 domains in close proximity.[125] Finally, there is the interaction of A3G with RNA. Both CA and Test+ RNA, in cases of excess protein, show a brightness greater than 1. As the inherent uncertainty in brightness measurements is $\sim 10\%$, brightness values of 1.3 and 1.2 indicate that some A3G complexes are binding more than 1 RNA. We confirm this observation by measuring the CA RNA brightness in a 1:1 CA/A3G mixture. The brightness $b = 2.2$ matches precisely with the brightness of A3G alone, which seems to indicate that each A3G dimer has bound two molecules of CA RNA. The 1:1 A3G/Test+ RNA experiment indicates the same binding stoichiometry, but this datapoint (*, Table 6.1) is more difficult to interpret. The 1:1 ratio is also the regime in which Test+ RNA causes strong oligomerization of A3G which would tend to bias the average brightness upwards.

To identify RNA copy number, we have labeled the RNA and left the A3G dark. Therefore, in higher Test+ RNA/A3G ratio experiments, we can no longer identify the A3G copy number, and hence the number of potential RNA binding sites. Furthermore, triggering A3G oligomerization requires excess RNA which means a large amount of unbound RNA. All the non-interacting or unbound RNA will have a brightness of 1, thus reducing the measured copy number of RNAs in complexes. Nevertheless, in several experiments containing excess Test+ RNA, and therefore presumably A3G oligomers, we have observed RNA copy numbers of 2-4 (data not shown). Further studies are required, but this preliminary evidence strongly suggests that the A3G dimer binds more than 1 RNA and that large A3G complexes contain multiple copies of "G-rich" RNA.

	Normalized Oligonucleotide Brightness				
	TCCCA ssDNA	CA RNA		Test+ RNA	
A3G (400 nM)	$b = 1.1$ (1 µM) A3G 100:1 ssDNA	$b = 1.3$ (30:1)	$b = 2.2$ (1:1)	b=1.2 (10:1)	$b = 2.2$* (1:1)
A3A (400 nM)	$b = 1.0$ (1 µM) A3A 100:1 ssDNA	$b = 1.0$ A3A 1:2 RNA		$b = 1.0$ A3A 4:1 RNA	

Table 6.1 **RNA/DNA copy number in APOBEC complexes.**

This table reports the copy numbers for RNA and ssDNA as they interact with A3G and A3A. The experiments are conducted at approximately 400 nM protein, except in the case of TCCCA ssDNA which required a higher protein concentration to achieve the large APOBEC/ssDNA ratio and still have detectable ssDNA signal. A3A shows evidence of only a single binding site. Note that additional A3G-RNA experiments should be conducted at higher A3A/RNA ratios to confirm this observation. The A3G dimer appears to bind only 1 ssDNA, but data indicate that it binds more than 1 RNA. The 1:1 A3G/CA RNA displays 2 RNAs per A3G dimer. The 1:1 A3G/Test+ RNA shows a similar result, but the interpretation is potentially complicated by the onset of A3G oligomerization in this regime.

6.9 Discussion

Using both A3G brightness experiments and RNA diffusion studies, we have shown that "G-rich" RNA induces the formation of complexes in A3G. The next step is to model this A3G-RNA interaction. Our original hypothesis was that multiple A3Gs would bind an RNA strand like "beads on a string." Under this model, however, Figure 6.4 should have shown A3G oligomerization when A3Gs outnumbered RNAs rather than the reverse. We also performed an experiment with a shorter RNA strand—only the first half of the Test+ RNA (not shown)—but the final oligomer size did not differ from that with the full length Test+ RNA. Therefore, the "beads on string" model is conclusively eliminated.

The other potential models involve multiple RNAs. One such model is that A3G-RNA chains are being formed, perhaps with the A3G dimer acting as a connector between two RNA segments. In this "co-polymer chain" model, the upper limit of the A3G complexes, ~20 copies, would represent the average of a distribution of A3G chains limited in length by the increased probability of "link" failure in the larger complexes. Most of the A3G experiments in this study have been conducted at 200 nM, but we have varied concentration from 40 to 1200 nM with little, if any, effect on the final average oligomer size in the presence of Test+ RNA. However, other models involving multiple RNAs are feasible as well. We have established that "G-rich" RNA helps A3G to form complexes, but we do not have experimental data to identify by what mechanism or via what connections the RNA strands and A3G assemble.

Regardless of the molecular details of the A3G-RNA model, we still cannot satisfactorily explain the data from Figure 6.4. All the evidence indicates that the RNA strands are providing binding targets for A3G which somehow 'glue' the complex together. Furthermore, excess RNA over A3G appears to be required to form the complexes. However, at very high ratios of excess RNA, we would expect to see a reduction in A3G complexes. Very high numbers of RNA strands should effectively 'screen' the A3G dimers from one another, providing a wealth of targets for protein-RNA interactions, but drastically decreasing the probability that individual RNAs can encounter and facilitate protein-protein interactions between 2 or more A3Gs. However,

at the highest RNA/A3G ratio in our experimental data (>100), the oligomer size remains at about 20 copies of A3G. As a result, we feel it necessary to propose cooperativity between protein-protein and protein-RNA interactions that make the formation of such large complexes in the presence of excess RNA energetically favorable. Additional experiments will be required to fully characterize the A3G-RNA interaction model.

Alternative methods are required to further elucidate a binding model for A3G-RNA complexes. Dual-color FFS studies offer an attractive route that will provide additional information about the composition of the complex. If both A3G and the RNA have labels, it is easier to identify the interacting species, although Förster resonance energy transfer (FRET) between the two fluorescent labels may complicate such measurements. Very preliminary dual-color data indicate approximately two RNA copies per A3G monomer in the large A3G-RNA complexes (Appendix E). However, these data have not been corrected for FRET or quenching effects and should be taken with a grain of salt. The A3G-RNA binding mechanism is clearly non-trivial, and developing an understanding of the model could provide valuable information about the biological function of this interaction.

Another important aspect to understanding A3G oligomerization is to develop an interaction model of the A3G subunit. Our studies, along with other reports from the literature, have observed A3G to form dimers in solution.[107,111,125] In fact, prior to discovering the effect of CA ssDNA and RNA, we found it to be such a tight dimer that is was difficult to take apart. We performed mixing experiments between labeled and unlabeled A3G that did not yield the expected results (see Appendix C). This also complicates matters because the labeled and unlabeled, or bright and dark, A3G mixing experiment is one of our best tools to eliminate the possibility of label interactions. We have performed studies in which Test+ RNA is added to a 1:1 mixture of bright and dark A3G, and the resulting oligomers yielded roughly half the brightness of a 100% labeled sample (Appendix C), which demonstrates that both bright and dark A3G form complexes with RNA. We know from other work that the fluorescent protein label does not interfere with A3G deamination,[126] nor do label interactions appear to cause the aggregation observed at higher temperatures (Appendix B). However, we are still left

with some questions about the A3G dimer, as we have discovered that when purified A3G is placed in *E. coli* extract or HeLa cytosol, the average normalized brightness reduces from $b \sim 2.3$ to $b \sim 1.3$ (Appendix D). It is not clear what element or elements of the lysate environments lead to this result. Our present knowledge on the nature of the monomeric and dimeric protein state remains woefully incomplete, and the set of current experimental data is difficult to reconcile with a physical model of binding interactions. Additional work is clearly needed to tease out the underlying mechanism governing A3G oligomerization.

The most significant result of this study is the sequence dependence of A3G interactions with RNA and DNA. "G-rich" RNA (Test+ RNA, GA RNA, GU RNA) cause A3G to form higher order oligomers, while ssDNA does not—except for GU ssDNA, containing deoxyuracil. CA RNA and CA ssDNA cause dissociation in the A3G dimeric subunit. We add to these results the previously mentioned literature reports that HMM A3G complexes are enzymatically inactive. Finally, sequence dependent deamination experiments of Chelico et al.[108] show almost a 20-fold increase in deamination activity for 5'-AAACCCAAA as compared to 5'-TTTCCCTTT, and several other flanking combination sequences containing adenines also showed high deamination activity. Our own CA ssDNA oligonucleotide does not have this identical sequence, but does contain two separate target sequences, 'AACCCAAA' and 'AACCCA', which fairly closely fit the description. Therefore, the CA ssDNA sequence mimics the preferred deamination target of A3G and also causes dissociation of dimeric A3G. We observed from the diffusion studies (Figures 6.10 & 6.11) that CA RNA had a lower binding affinity than Test+ RNA. This probably means that CA sequences have a higher off rate, or shorter binding duration, leaving the A3G protein enzyme free to catalyze more targets, which would agree with the results of Chelico et al. In addition, CA-induced dissociation creates two A3G molecules in the place of one dimer, increasing the effective concentration of independent A3G enzymes, which would also lead to an increase in deamination activity.

We have shown that CA RNA can cause dissociation of the A3G dimer. However, our hypothesis is that the large oligomeric complexes are the form packaged

into the virion. Therefore, the important question becomes whether this dissociation effect also applies to HMM A3G complexes. We perform an experiment in which we mix labeled A3G with unlabeled Test+ RNA at a 1:4 ratio (200 nM A3G, 800 nM RNA). Figure 6.12 shows that, for this experiment, the final A3G oligomerization is $b \sim 27$. Subsequently in individual droplets, we add CA RNA and CA ssDNA (800 nM, 1.6 μM) for 1:1 and 1:2 ratios of Test+ RNA to CA oligonucleotide. The A3G oligomerization is reduced in all 4 cases (Figure 6.12), but it was particularly exciting to see the very low value of $b \sim 4$ in the final CA ssDNA condition. Note that all four samples are very inhomogeneous, containing some intensity spikes which are presumably large A3G complexes that did not dissociate, thus the data sets presented in Figure 6.12 were intentionally selected to avoid those containing large spikes that skew the average brightness upwards. Nevertheless there is a clear dissociative response for HMM A3G complexes in the presence of their deamination target ssDNA.

With these results, we strengthen our earlier hypothesis. "G-rich" RNA in the cytoplasm causes A3G to oligomerize which, in a process that is not yet understood, allows it to package into the virion. Once inside the virion, reverse transcription causes a shift from an RNA-containing environment to a DNA-containing environment which does not promote complex formation. Additionally, it appears that "CA-rich" ssDNA can actively out-compete the effects of "G-rich" RNA. These processes, presumably, result in the dissociation of the HMM A3G complexes back into the LMM form which are enzymatically active, mutate the viral DNA, and are thus able to restrict HIV-1.

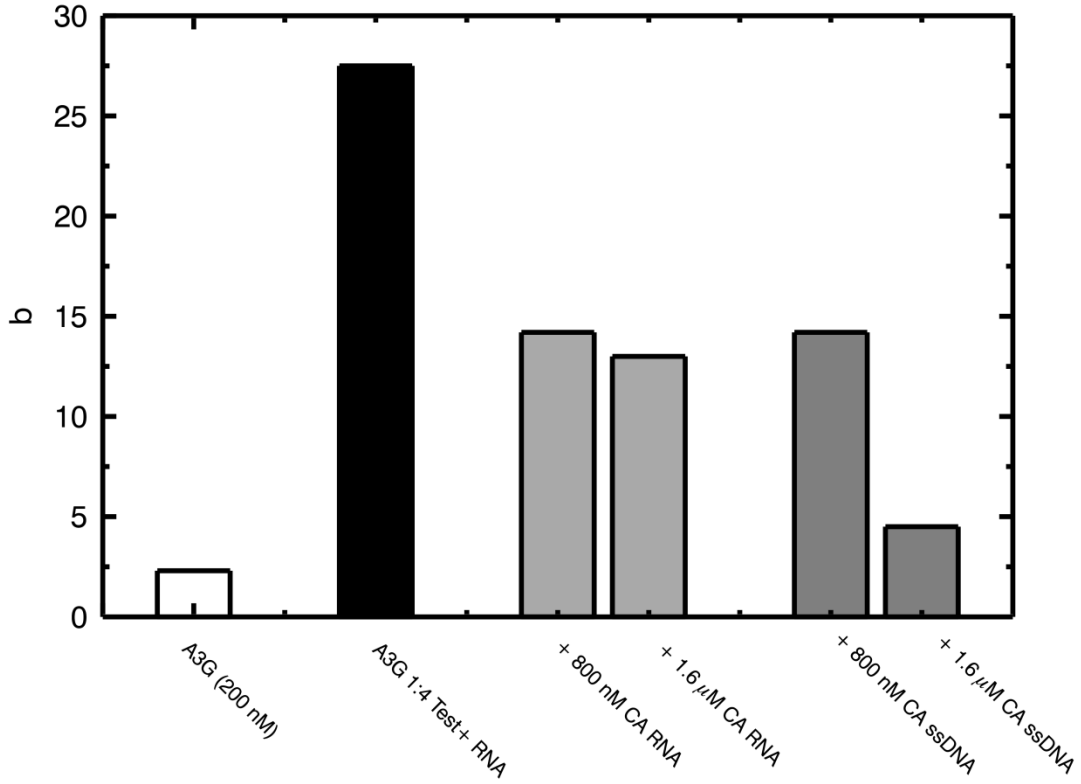

Figure 6.12 **CA ssDNA dissociates A3G oligomers.**

These data shows that the dissociation effect of the CA sequences can assist in taking apart the HMM A3G complexes formed with Test+ RNA. The first two columns show the A3G control (200 nM) and A3G-RNA complexes formed with Test+ RNA (800 nM). Subsequently, CA sequences were added to droplets of the A3G-RNA sample in four different conditions: CA RNA at equal and twice the amount of Test+ RNA, and CA ssDNA at equal and twice Test+ RNA. All four conditions show A3G-RNA complex dissociation, but 1:2 Test+ RNA/CA ssDNA shows that a small excess amount of the deamination target can effectively take apart the complex and reduce it to a, presumably, enzymatically active form. The above data represent a single experiment which should be repeated, along with similar experiments to test the effect of other ssDNA sequences on HMM A3G complexes.

7. Summary

FFS is a powerful technique for studying biomolecular interactions inside living cells. Because the technique is quantitative, non-invasive and measures in real time, it has a very broad range of applications and broader scope yet to be realized. We demonstrated that quantitative FFS can be performed in new territory, such as small sample reservoirs and thin sample geometries, and further established that the same technique can be applied in a consistent manner across multiple sample environments (cellular, cell-free and aqueous solution).

We extended the reach of quantitative FFS to small sample volumes where photodepletion is present. Our analysis showed that photodepletion clearly biases FFS brightness measurements. Of particular interest is the extension of quantitative FFS to *E. coli* cells which are the most popularly studied prokaryotic model organism. As an example, systems biology questions often depend on knowledge of protein copy numbers per cell and the variation in those numbers across a cell population.[127,128] FFS in *E. coli* presents a useful tool for providing exactly this type of information and should be broadly applicable for quantitative studies in the fields of biotechnology and microbiology.

We demonstrated that thin geometry biases FFS brightness measurements by as much as a factor of 2, a fact which ought to be noted by anyone using FFS in *E. coli*, microfluidic devices, or in thin cell sections. The same bias appears not only in stationary FFS experiments, but also in related fluorescence imaging techniques applied to thin samples. It is often more difficult to detect in these circumstances. We developed modifications to FFS theory, and coupled the theory with a z-scan approach that allows us to identify thin samples by their intensity profile, fit the data, and recover an unbiased brightness. An exciting new direction for this technique is its application to cell membranes. Fluorescent membrane studies are very difficult to conduct because the membranes are so much thinner than the excitation light. Total internal reflection fluorescence (TIRF) microscopy uses an evanescent field to limit excitation depth to ~100 nm, but this technique can only be applied to the membrane in contact with the coverglass. Z-scan FFS provides access to internal membranes for the first time. Other

members of the lab are already pursuing z-scan FFS to differentiate between cytoplasmic and membrane fraction of proteins, as well as to study protein interaction directly at the nuclear membrane. These applications require further development, but z-scan FFS can bring quantitative measurement of protein interaction to the wide-ranging and important field of membrane proteins, and do so directly in living cells.

Although the ability to take quantitative measurements in living cells is the chief strength of FFS, when trying to carefully characterize a protein, control over the environment to systematically change conditions is an important component. Since our ability to change cellular conditions is extremely limited, we combined FFS and cell-free expression to create an environment that offers the same convenience of genetic labeling as a cell, while providing direct access and control of the sample. We exploited these new capabilities by conducting a maturation study of three common fluorophores (EGFP, EYFP, and mCherry) as a function of temperature. The result revealed that much of the scattered literature values may be in closer agreement than is apparent at first glance, and that EGFP is a fast maturing protein and a good candidate for FFS and brightness experiments. We have also confirmed that temperature and oxygenation are important factors in maturation, and that fluorescence researchers must consider these conditions since very slowly maturing fluorophores present a problem. mCherry is the only red fluorescent protein we've found that is suitable for FFS brightness experiments. Having established its slow maturation, the development of a fast red protein is important. FFS and cell-free expression provide a quick and simple test that would be useful in screening new red fluorescent protein candidates.

Additionally, FFS and cell-free expression can be used to address the role of the medium in protein-interaction assays. An enormous number of standard protein assays are currently conducted in vitro or in aqueous buffers containing only a few salts. However, cell cytoplasm is a crowded environment of biomolecules, and cell-free solution maintains some of this characteristic. The literature has demonstrated that molecular crowding can affect binding affinities. The cell-free environment is partially composed of extracted cytoplasm and, for the purposes of molecular crowding experiments, can be enhanced with the addition of biocompatible macromolecules like

polyethylene glycol (PEG), so that FFS and cell-free solution can study crowding phenomena. The combination of FFS and cell-free expression also lends itself to the exploration of gene regulation networks. An important and still developing approach is the attempt to build these networks from the ground up. FFS in cell-free expression can play an important role in identifying components of such a minimal model, as proteins and transcription factors assemble into the machine complexes that make the networks function.

We also applied FFS to study APOBEC3G oligomerization and its interactions with DNA and RNA. Surprisingly, we found that A3G oligomerization in the presence of DNA and RNA differs and, additionally, depends on the sequence of the oligonucleotide. This has not been previously demonstrated, and could not have been identified in cellular studies. The data presented in the chapter 6 and the appendices demonstrate that A3G exhibits complex behavior. In the absence of DNA and RNA, A3G demonstrated oligomerization behaviors in response to increased temperature and mixing experiments that we have been unable to explain and require further attention. However, the differing oligomerization responses, thus far identified, of A3G to DNA and RNA might be a very important part of the mechanism by which A3G packages into the virus and executes its enzymatic function against HIV-1. A3G and selected members of the APOBEC protein family are promising drug targets for the fight against HIV-1, and FFS provides a powerful tool in the characterization and quantification of APOBEC proteins.

In this work, we have developed FFS techniques for small volumes and thin samples. We've used FFS to perform brightness measurements in *E. coli*, mammalian cytoplasm, cell-free expression systems and aqueous buffers, thus improving the potency and reach of FFS for the study of protein interactions. Proteins are the building blocks of the biological world, and FFS can identify how many there are in the system and how they are put together, building up our understanding to the larger protein machines, the assembly lines, and perhaps to the network connecting them all, that makes the whole system run—and produces life.

References

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

2. Berland, K. M., P. T. So, and E. Gratton. Two-photon fluorescence correlation spectroscopy: method and application to the intracellular environment. *Biophys. J.* 68:694–701, (1995).

3. Magde, D., E. Elson, and W. W. Webb. Thermodynamic fluctuations in a reacting system -- measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* 29:705–708, (1972).

4. Elson, E. L., and D. Madge. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers* 13:1–27, (1974).

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

6. Chen, Y., L.-N. Wei, and J. D. Mueller. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 100:15492 –15497, (2003).

7. Chen, Y., and J. D. Mueller. Determining the stoichiometry of protein heterocomplexes in living cells with fluorescence fluctuation spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 104:3147–3152, (2007).

8. Chen, Y., M. Tekmen, L. Hillesheim, J. Skinner, B. Wu, and J. D. Mueller. Dual-color photon-counting histogram. *Biophys. J.* 88:2177–2192, (2005).

9. Shimomura, O., F. H. Johnson, and Y. Saiga. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* 59:223–239, (1962).

10. Shimomura, O. Structure of the chromophore of *Aequorea* green fluorescent protein. *FEBS Lett.* 104:220–222, (1979).

11. Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science* 263:802–805, (1994).

12. Heim, R., D. C. Prasher, and R. Y. Tsien. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* 91:12501 –12504, (1994).

13. Heim, R., A. B. Cubitt, and R. Y. Tsien. Improved green fluorescence. *Nature* 373:663–664, (1995).

14. Heim, R., and R. Y. Tsien. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* 6:178–182, (1996).

15. Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. . Giepmans, A. E. Palmer, and R. Y. Tsien. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* 22:1567–1572, (2004).

16. Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier, and others. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111:229, (1992).

17. Bhattacharjee, Y. How bad luck & bad networking cost Douglas Prasher a Nobel prize. , (2011).at <http://discovermagazine.com/2011/apr/30-how-bad-luck-networking-cost-prasher-nobel/>

18. Chen, Y., J. D. Mueller, Q. Ruan, and E. Gratton. Molecular brightness characterization of EGFP in vivo by fluorescence fluctuation spectroscopy. *Biophys. J.* 82:133–144, (2002).

19. Zhou, H. X., G. Rivas, and A. P. Minton. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Ann. Rev. Biophys.* 37:375, (2008).

20. Muramatsu, N., and A. P. Minton. Tracer diffusion of globular proteins in concentrated protein solutions. *Proc. Natl. Acad. Sci. U.S.A.* 85:2984–2988, (1988).

21. Miyoshi, D., S. Matsumura, S.-I. Nakano, and N. Sugimoto. Duplex dissociation of telomere DNAs induced by molecular crowding. *J. Am. Chem. Soc.* 126:165–169, (2004).

22. Kim, Y. C., R. B. Best, and J. Mittal. Macromolecular crowding effects on protein-protein binding affinity and specificity. *J. Chem. Phys.* 133:205101, (2010).

23. Martin, J., and F. U. Hartl. The effect of macromolecular crowding on chaperonin-mediated protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 94:1107–1112, (1997).

24. Van Orden, A., K. Fogarty, and J. Jung. Fluorescence fluctuation spectroscopy: a coming of age story. *Appl. Spectrosc.* 58:122A–137A, (2004).

25. Rigler, R., U. Mets, J. Widengren, and P. Kask. Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion. *Eur. Biophys. J.* 22:169–175, (1993).

26. Chen, Y., J. D. Mueller, P. T. So, and E. Gratton. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys. J.* 77:553–567, (1999).

27. Saffarian, S., Y. Li, E. L. Elson, and L. J. Pike. Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. *Biophys. J.* 93:1021–1031, (2007).

28. Petersen, N. O., P. L. Hoddelius, P. W. Wiseman, O. Seger, and K. E. Magnusson. Quantitation of membrane receptor distributions by image correlation spectroscopy: concept and application. *Biophys. J.* 65:1135–1146, (1993).

29. Chen, Y., J. D. Mueller, S. Y. Tetin, J. D. Tyner, and E. Gratton. Probing ligand protein binding equilibria with fluorescence fluctuation spectroscopy. *Biophys. J.* 79:1074–1084, (2000).

30. Palmer, A. G., and N. L. Thompson. High-order fluorescence fluctuation analysis of model protein clusters. *Proc. Natl. Acad. Sci. U.S.A.* 86:6148–6152, (1989).

31. Qian, H., and E. L. Elson. On the analysis of high order moments of fluorescence fluctuations. *Biophys. J.* 57:375–380, (1990).

32. Mueller, J. D. Cumulant analysis in fluorescence fluctuation spectroscopy. *Biophys. J.* 86:3981–3992, (2004).

33. Wu, B., and J. D. Mueller. Time-integrated fluorescence cumulant analysis in fluorescence fluctuation spectroscopy. *Biophys. J.* 89:2721–2735, (2005).

34. Denk, W., J. H. Strickler, and W. W. Webb. Two-photon laser scanning fluorescence microscopy. *Science* 248:73–76, (1990).

35. So, P. T. C., C. Y. Dong, B. R. Masters, and K. M. Berland. Two-Photon Excitation Fluorescence Microscopy. *Annu. Rev. Biomed. Eng.* 2:399–429, (2000).

36. Johnson, J., Y. Chen, and J. D. Mueller. Characterization of brightness and stoichiometry of bright particles by flow-fluorescence fluctuation spectroscopy. *Biophys. J.* 99:3084–3092, (2010).

37. Chen, Y., L.-N. Wei, and J. D. Mueller. Unraveling protein-protein interactions in living cells with fluorescence fluctuation brightness analysis. *Biophys. J.* 88:4366–4377, (2005).

38. Duckworth, B. P., Y. Chen, J. W. Wollack, Y. Sham, J. D. Mueller, T. A. Taton, and M. D. Distefano. A universal method for the preparation of covalent protein-DNA conjugates for use in creating protein nanostructures. *Angew. Chem. Int. Ed. Engl.* 46:8819–8822, (2007).

39. Thompson, N. L. Fluorescence correlation spectroscopy. In: Topics in Fluorescence Spectroscopy, edited by J. R. Lakowicz. New York: Plenum, pp. 337–378, (1991).

40. Klingler, J., and T. Friedrich. Site-specific interaction of thrombin and inhibitors observed by fluorescence correlation spectroscopy. *Biophys. J.* 73:2195–2200, (1997).

41. Mueller, J. D., Y. Chen, and E. Gratton. Resolving heterogeneity on the single molecular level with the photon-counting histogram. *Biophys. J.* 78:474–486, (2000).

42. Finn, M. A., G. W. Greenless, T. W. Hodapp, and D. A. Lewis. Real-time elimination of dead-time and afterpulsing in counting systems. *Rev. Sci. Instrum.* 59:2457–2459, (1988).

43. Hillesheim, L. N., and J. D. Mueller. The photon counting histogram in fluorescence fluctuation spectroscopy with non-ideal photodetectors. *Biophys. J.* 85:1948–1958, (2003).

44. Sanchez-Andres, A., Y. Chen, and J. D. Mueller. Molecular brightness determined from a generalized form of Mandel's Q-parameter. *Biophys. J.* 89:3531–3547, (2005).

45. Chen, Y., J. Johnson, P. Macdonald, B. Wu, and J. D. Mueller. Observing protein interactions and their stoichiometry in living cells by brightness analysis of fluorescence fluctuation experiments. *Meth. Enzymol* 472:345–363, (2010).

46. Wu, B., Y. Chen, and J. D. Mueller. Heterospecies partition analysis reveals binding curve and stoichiometry of protein interactions in living cells. *Proc. Natl. Acad. Sci. U.S.A.* 107:4117–4122, (2010).

47. Hillesheim, L. N., and J. D. Mueller. The dual-color photon counting histogram with non-ideal photodetectors. *Biophys. J.* 89:3491–3507, (2005).

48. Saffarian, S., and E. L. Elson. Statistical analysis of fluorescence correlation spectroscopy: the standard deviation and bias. *Biophys. J* 84:2030–2042, (2003).

49. Mandel, L. Sub-Poissonian photon statistics in resonance fluorescence. *Opt. Lett.* 4:205–207, (1979).

50. Lee, S. Y. Systems Biology and Biotechnology of *Escherichia coli*. Springer, 2009, 488 pp.

51. Bhomkar, P., W. Materi, and D. S. Wishart. The bacterial nanorecorder: engineering *E. coli* to function as a chemical recording device. *PLoS ONE* 6:e27559, (2011).

52. Kashiwagi, A., I. Urabe, K. Kaneko, and T. Yomo. Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. *PLoS ONE* 1:e49, (2006).

53. Patterson, G. H., S. M. Knobel, W. D. Sharif, S. R. Kain, and D. W. Piston. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* 73:2782–2790, (1997).

54. Drummond, D. R., N. Carter, and R. A. Cross. 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.* 206:161–169, (2002).

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

56. Macdonald, P. J., Y. Chen, X. Wang, Y. Chen, and J. D. Mueller. Brightness analysis by z-scan fluorescence fluctuation spectroscopy for the study of protein interactions within living cells. *Biophys. J.* 99:979–988, (2010).

57. Macdonald, P. J., J. Johnson, E. Smith, Y. Chen, J. D. Mueller. Brightness analysis. In: Fluorescence Fluctuation Spectroscopy (FFS), Part A: 518, edited by S. Y. Tetin. Academic Press, pp 71-98, (2013).

58. Frisken Gibson, S., and F. Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J. Opt. Soc. Am. A* 9:154–166, (1992).

59. Müller, C. B., T. Eckert, A. Loman, J. Enderlein, and W. Richtering. Dual-focus fluorescence correlation spectroscopy: a robust tool for studying molecular crowding. *Soft Matter* 5:1358–1366, (2009).

60. Benda, A., M. Beneš, V. Mareček, A. Lhotský, W. T. Hermens, and M. Hof. How to determine diffusion coefficients in planar phospholipid systems by confocal fluorescence correlation spectroscopy. *Langmuir* 19:4120–4126, (2003).

61. Humpolíčková, J., E. Gielen, A. Benda, V. Fagulova, J. Vercammen, M. vandeVen, M. Hof, M. Ameloot, and Y. Engelborghs. Probing diffusion laws within cellular membranes by z-scan fluorescence correlation spectroscopy. *Biophys. J.* 91:L23–L25, (2006).

62. Korlach, J., P. Schwille, W. W. Webb, and G. W. Feigenson. Characterization of lipid bilayer phases by confocal microscopy and fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 96:8461–8466, (1999).

63. Hess, S. T., and W. W. Webb. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophys. J.* 83:2300–2317, (2002).

64. Gregor, I., and J. Enderlein. Focusing astigmatic Gaussian beams through optical systems with a high numerical aperture. *Opt. Lett.* 30:2527–2529, (2005).

65. Nasse, M. J., and J. C. Woehl. Realistic modeling of the illumination point spread function in confocal scanning optical microscopy. *J. Opt. Soc. Am. A.* 27:295, (2010).

66. De Grauw, C. J., J. M. Vroom, H. T. M. van der Voort, and H. C. Gerritsen. Imaging properties in two-photon excitation microscopy and effects of refractive-index mismatch in thick specimens. *Appl. Opt.* 38:5995–6003, (1999).

67. Enderlein, J., I. Gregor, D. Patra, and J. Fitter. Art and artefacts of fluorescence correlation spectroscopy. *Curr. Pharm. Biotechnol.* 5:155–161, (2004).

68. Bayliss, R., K. Ribbeck, D. Akin, H. M. Kent, C. M. Feldherr, D. Görlich, and M. Stewart. Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. *J. Mol. Biol.* 293:579–593, (1999).

69. Chaillan-Huntington, C., P. J. G. Butler, J. A. Huntington, D. Akin, C. Feldherr, and M. Stewart. NTF2 monomer-dimer equilibrium. *J. Mol. Biol.* 314:465–477, (2001).

70. Balla, A., G. Tuymetova, M. Barshishat, M. Geiszt, and T. Balla. Characterization of type II phosphatidylinositol 4-kinase isoforms reveals association of the enzymes with endosomal vesicular compartments. *J. Biol. Chem.* 277:20041–20050, (2002).

71. Digman, M. A., R. Dalal, A. F. Horwitz, and E. Gratton. Mapping the number of molecules and brightness in the laser scanning microscope. *Biophys. J.* 94:2320–2332, (2008).

72. Corbett, A. H., and P. A. Silver. The NTF2 gene encodes an essential, highly conserved protein that functions in nuclear transport in Vivo. *J. Biol. Chem.* 271:18477–18484, (1996).

73. Lane, C. M., I. Cushman, and M. S. Moore. Selective disruption of nuclear import by a functional mutant nuclear transport carrier. *J. Cell Biol.* 151:321–332, (2000).

74. Bacia, K., D. Scherfeld, N. Kahya, and P. Schwille. Fluorescence correlation spectroscopy relates rafts in model and native membranes. *Biophys. J.* 87:1034–1043, (2004).

75. Gennerich, A., and D. Schild. Fluorescence correlation spectroscopy in small cytosolic compartments depends critically on the diffusion model used. *Biophys. J.* 79:3294–3306, (2000).

76. Jung, G., and A. Zumbusch. Improving autofluorescent proteins: Comparative studies of the effective brightness of Green Fluorescent Protein (GFP) mutants. *Microsc. Res. Tech.* 69:175–185, (2006).

77. Macdonald, P. J., Y. Chen, and J. D. Mueller. Chromophore maturation and fluorescence fluctuation spectroscopy of fluorescent proteins in a cell-free expression system. *Anal. Biochem.* 421:291–298, (2012).

78. He, M. Cell-free protein synthesis: applications in proteomics and biotechnology. *New Biotechnol.* 25:126–132, (2008).

79. Simpson, M. L. Cell-free synthetic biology: a bottom-up approach to discovery by design. *Mol. Syst. Biol.* 2:69, (2006).

80. Shin, J., and V. Noireaux. Study of messenger RNA inactivation and protein degradation in an *Escherichia coli* cell-free expression system. *J. Biol. Eng.* 4:9, (2010).

81. Iizuka, R., M. Yamagishi-Shirasaki, and T. Funatsu. Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem.* 414:173–178, (2011).

82. Ribbeck, K., G. Lipowsky, H. M. Kent, M. Stewart, and D. Görlich. NTF2 mediates nuclear import of Ran. *EMBO J.* 17:6587–6598, (1998).

83. Casanova, M., P. Portalès, C. Blaineau, L. Crobu, P. Bastien, and M. Pagès. Inhibition of active nuclear transport is an intrinsic trigger of programmed cell death in trypanosomatids. *Cell Death Differ.* 15:1910–1920, (2008).

84. Paschal, B., and L. Gerace. Identification of NTF2, a cytosolic factor for nuclear import that interacts with nuclear pore complex protein p62. *J. Cell Biol.* 129:925–937, (1995).

85. Bullock, T. L., D. W. Clarkson, H. M. Kent, and M. Stewart. The 1.6 Å resolution crystal structure of nuclear transport factor 2 (NTF2). *J. Mol. Biol.* 260:422–431, (1996).

86. Kim, D., and C. Choi. A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. *Biotechnol. Prog.* 12:645–649, (1996).

87. Katranidis, A., D. Atta, R. Schlesinger, K. H. Nierhaus, T. Choli-Papadopoulou, I. Gregor, M. Gerrits, G. Büldt, and J. Fitter. Fast biosynthesis of GFP molecules: a single-molecule fluorescence study. *Angew. Chem.* 121:1790–1793, (2009).

88. Evdokimov, A. G., M. E. Pokross, N. S. Egorov, A. G. Zaraisky, I. V. Yampolsky, E. M. Merzlyak, A. N. Shkoporov, I. Sander, K. A. Lukyanov, and D. M. Chudakov.

Structural basis for the fast maturation of arthropoda green fluorescent protein. *EMBO Rep.* 7:1006–1012, (2006).

89. Megerle, J. A., G. Fritz, U. Gerland, K. Jung, and J. O. Rädler. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.* 95:2103–2115, (2008).

90. Sniegowski, J. A., J. W. Lappe, H. N. Patel, H. A. Huffman, and R. M. Wachter. Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J. Biol. Chem.* 280:26248 –26255, (2005).

91. Wente, S. R., and M. P. Rout. The nuclear pore complex and nuclear transport. *Cold Spring Harb. Perspect. Biol.* 2:, (2010).

92. Jewett, M. C., K. A. Calhoun, A. Voloshin, J. J. Wuu, and J. R. Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* 4:220–220, (2008).

93. Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* 38:2865, (2009).

94. Wachter, R. M. Chromogenic cross-link formation in green fluorescent protein. *Acc. Chem. Res.* 40:120–127, (2007).

95. Wachter, R. M., J. L. Watkins, and H. Kim. Mechanistic diversity of red fluorescence acquisition by GFP-like proteins. *Biochemistry* 49:7417–7427, (2010).

96. Reid, B. G., and G. C. Flynn. Chromophore Formation in Green Fluorescent Protein. *Biochemistry* 36:6786–6791, (1997).

97. Siemering, K. R., R. Golbik, R. Sever, and J. Haseloff. Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr. Biol.* 6:1653–1663, (1996).

98. Gordon, A., A. Colman-Lerner, T. E. Chin, K. R. Benjamin, R. C. Yu, and R. Brent. Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Methods* 4:175–181, (2007).

99. Merzlyak, E. M., J. Goedhart, D. Shcherbo, M. E. Bulina, A. S. Shcheglov, A. F. Fradkov, A. Gaintzeva, K. A. Lukyanov, S. Lukyanov, T. W. J. Gadella, and D. M. Chudakov. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* 4:555–557, (2007).

100. Harris, R. S., and M. T. Liddament. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* 4:868–877, (2004).

101. Cullen, B. R. Role and Mechanism of Action of the APOBEC3 Family of Antiretroviral Resistance Factors. *J. Virol.* 80:1067–1076, (2006).

102. Ooms, M., A. Krikoni, A. K. Kress, V. Simon, and C. Münk. APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. *J. Virol.* 86:6097–6108, (2012).

103. Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418:646–650, (2002).

104. Harris, R. S., K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, and M. H. Malim. DNA deamination mediates innate immunity to retroviral infection. *Cell* 113:803–809, (2003).

105. Chiu, Y.-L., V. B. Soros, J. F. Kreisberg, K. Stopak, W. Yonemoto, and W. C. Greene. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. *Nature* 435:108–114, (2005).

106. Khan, M. A., R. Goila-Gaur, S. Opi, E. Miyagi, H. Takeuchi, S. Kao, and K. Strebel. Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. *Retrovirology* 4:48, (2007).

107. Huthoff, H., F. Autore, S. Gallois-Montbrun, F. Fraternali, and M. H. Malim. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog.* 5:e1000330, (2009).

108. Chelico, L., P. Pham, P. Calabrese, and M. F. Goodman. APOBEC3G DNA deaminase acts processively 3’ --> 5’ on single-stranded DNA. *Nat. Struct. Mol. Biol.* 13:392–399, (2006).

109. Hultquist, J. F., J. A. Lengyel, E. W. Refsland, R. S. LaRue, L. Lackey, W. L. Brown, and R. S. Harris. Human and Rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H Demonstrate a Conserved Capacity To Restrict Vif-Deficient HIV-1. *J. Virol.* 85:11220–11234, (2011).

110. Shlyakhtenko, L. S., A. Y. Lushnikov, M. Li, L. Lackey, R. S. Harris, and Y. L. Lyubchenko. Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J. Biol. Chem.* 286:3387–3395, (2011).

111. Salter, J. D., J. Krucinska, J. Raina, H. C. Smith, and J. E. Wedekind. A hydrodynamic analysis of APOBEC3G reveals a monomer-dimer-tetramer self-association that has implications for anti-HIV function. *Biochemistry* 48:10685–10687, (2009).

112. Iwatani, Y., H. Takeuchi, K. Strebel, and J. G. Levin. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J. Virol.* 80:5992–6002, (2006).

113. Wang, T., C. Tian, W. Zhang, K. Luo, P. T. N. Sarkis, L. Yu, B. Liu, Y. Yu, and X.-F. Yu. 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. *J. Virol.* 81:13112–13124, (2007).

114. Bach, D., S. Peddi, B. Mangeat, A. Lakkaraju, K. Strub, and D. Trono. Characterization of APOBEC3G binding to 7SL RNA. *Retrovirology* 5:54, (2008).

115. Friew, Y. N., V. Boyko, W.-S. Hu, and V. K. Pathak. Intracellular interactions between APOBEC3G, RNA, and HIV-1 Gag: APOBEC3G multimerization is dependent on its association with RNA. *Retrovirology* 6:56, (2009).

116. Bogerd, H. P., and B. R. Cullen. Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. *RNA* 14:1228–1236, (2008).

117. Beuning, P. J., M. R. Tessmer, C. G. Baumann, D. A. Kallick, and K. Musier-Forsyth. Sequence-dependent conformational differences of small RNAs revealed by native gel electrophoresis. *Anal. Biochem.* 273:284–290, (1999).

118. Yu, Q., R. König, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, and N. R. Landau. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* 11:435–442, (2004).

119. Beale, R. C. L., S. K. Petersen-Mahrt, I. N. Watt, R. S. Harris, C. Rada, and M. S. Neuberger. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J. Mol. Biol.* 337:585–596, (2004).

120. Klarmann, G. J., M. E. Hawkins, and S. F. J. Le Grice. Uncovering the complexities of retroviral ribonuclease H reveals its potential as a therapeutic target. *AIDS Rev.* 4:183–194, (2002).

121. Carpenter, M. A., M. Li, A. Rathore, L. Lackey, E. K. Law, A. M. Land, B. Leonard, S. M. D. Shandilya, M.-F. Bohn, C. A. Schiffer, W. L. Brown, and R. S. Harris. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem.* 287:34801–34808, (2012).

122. Jarmuz, A., A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott, and N. Navaratnam. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* 79:285–296, (2002).

123. Bulliard, Y., I. Narvaiza, A. Bertero, S. Peddi, U. F. Röhrig, M. Ortiz, V. Zoete, N. Castro-Díaz, P. Turelli, A. Telenti, O. Michielin, M. D. Weitzman, and D. Trono. Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.* 85:1765–1776, (2011).

124. Chiu, Y.-L., and W. C. Greene. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* 26:317–353, (2008).

125. Wedekind, J. E., R. Gillilan, A. Janda, J. Krucinska, J. D. Salter, R. P. Bennett, J. Raina, and H. C. Smith. Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *J. Biol. Chem.* 281:38122–38126, (2006).

126. Nowotny, B., T. Schneider, G. Pradel, T. Schirmeister, A. Rethwilm, and M. Kirschner. Inducible APOBEC3G-Vif double stable cell line as a high-throughput screening platform to identify antiviral compounds. *Antimicrob. Agents Chemother.* 54:78–87, (2010).

127. Singh, A., B. S. Razooky, R. D. Dar, and L. S. Weinberger. Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. *Mol. Syst. Biol.* 8:, (2012).

128. Maier, T., A. Schmidt, M. Güell, S. Kühner, A.-C. Gavin, R. Aebersold, and L. Serrano. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* 7:511, (2011).

Appendices

Appendix A. APOBEC and reducing environments

In cells, A3A is monomeric, and we expected similar results in solution. However, our initial measurements showed A3A appearing with a normalized brightness of nearly 3, which would indicate a trimer. After consultation with Michael Carpenter, we suspected that the exposed cysteine residues of A3A might lead to disulfide bonds between A3A molecules. Providing a reducing environment in solution would prevent these bonds as it does inside the cell. We used both mercaptoethanol β and dithiothreitol (DTT) to show that this was, in fact the case. Below we present the data for the DTT experiment.

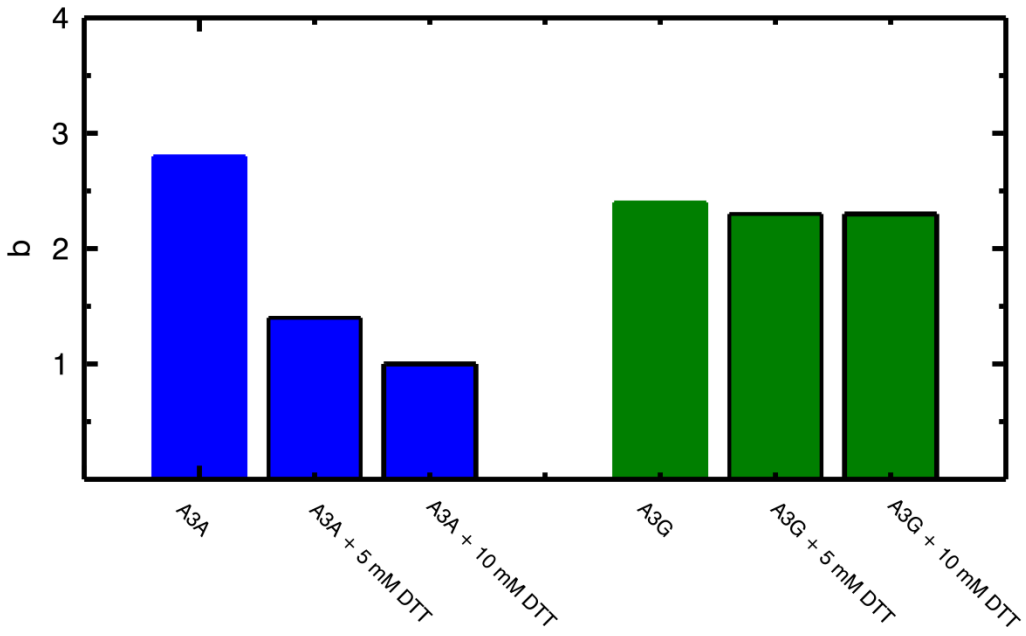

Figure A.1 **A3A and A3G in a reducing environment.**

It is clear that in a reducing environment A3A is monomeric as expected, but A3G remains unaffected. Thus, all of the A3A experiments in chapter 6 are conducted in a solution containing DTT, as are most of the A3G experiments (NEB3 buffer).

Appendix B. Temperature-induced A3G oligomerization

We observe a temperature-induced aggregation phenomenon of EGFP-labeled A3G in solution (50 mM Tris-HCl). No significant alteration in the phenomenon were observed when using Tris-HCl buffer with additional salts, or when using a 10 mM HEPES buffer. At room temperature, we consistently recovered an average brightness $b = 2.3$ as previously stated. However, if we incubate the A3G-EGFP sample at 37°C, and then cool and measure it at 20°C, we see an increase in oligomerization (Figure B.1).

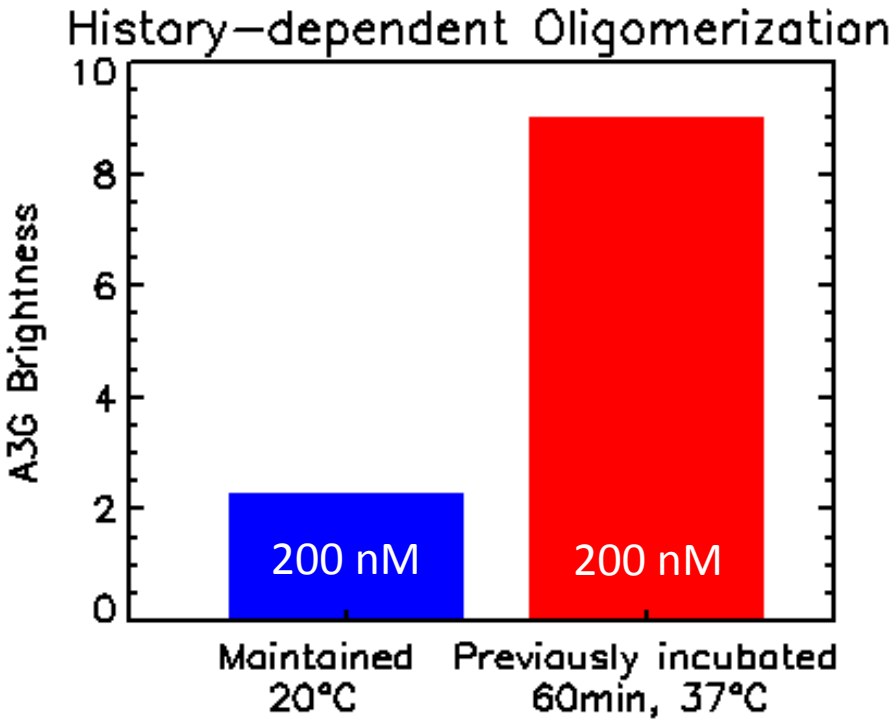

Figure B.1 **Temperature-induced A3G aggregation.**

A higher starting concentration of A3G-EGFP and/or a longer time spent incubating results in larger aggregates. The trend of this phenomenon has been very consistent although successive trials return scattered brightness results. Note again that the aggregate does not dissociate when the sample is cooled back down to room temperature (or below), indicating that the aggregation is irreversible.

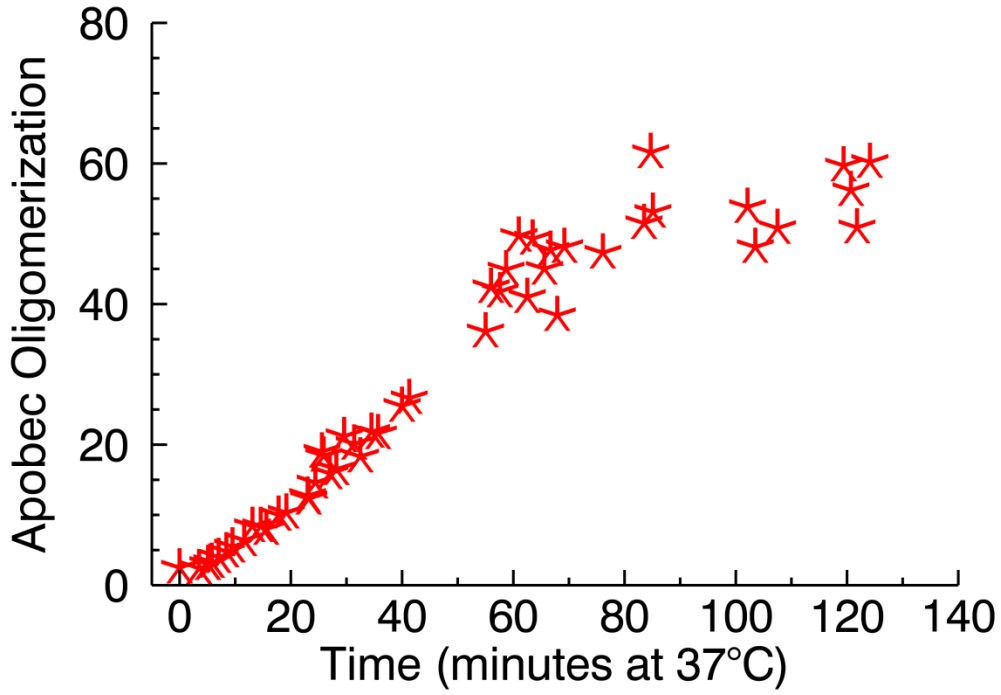

Figure B.2 **37°C A3G runaway aggregation.**

Figure B.2 shows a sample of room temperature A3G placed into a 37°C environment at time 0. The sample rapidly begins to aggregate with no obvious endpoint, except perhaps a lack of additional A3G molecules. This response suggests that A3G experiments performed in 37°C solution may be difficult to conduct or interpret. We also perform an experiment in which the temperature is slowly increased while the sample is being measured. Figure B.3 shows that the aggregation reaction does not appear to initiate until the temperature rises above ~35°C. The temperature of the sample was determined by inserting a thermocouple wire into a buffer droplet located in the well neighboring the actual sample.

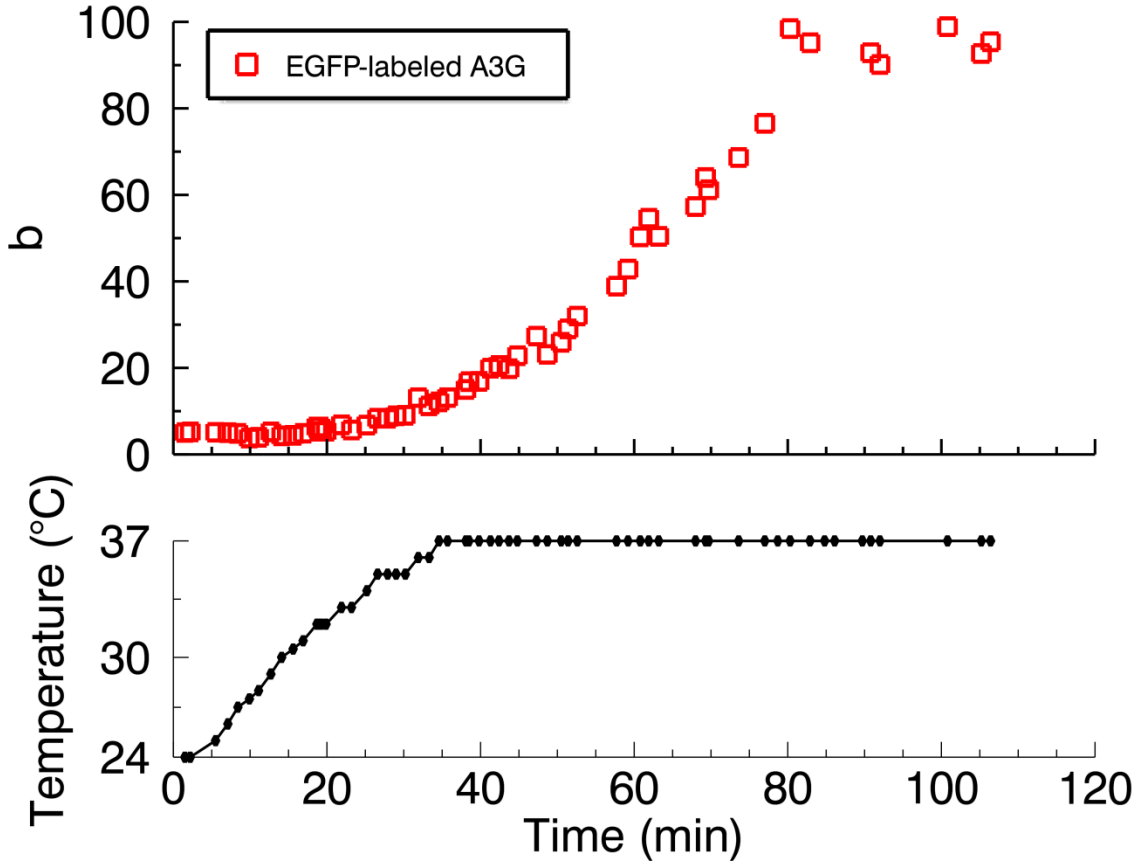

Figure B.3 **Temperature threshold of A3G oligomerization.**

To explore whether this temperature-induced aggregation might be the result of the EGFP label, we would like to test the temperature effect on an A3G sample without the EGFP label. However, without the label, there is nothing for our fluorescence technique to measure. Our strategy was to create a sample that is mostly unlabeled but has a small percentage of EGFP-labeled A3Gs. The bulk of the interactions should take place between unlabeled A3G proteins, but the smattering of labeled A3G proteins still permit us to extract information using FFS. To this end, we created a mixture of labeled and unlabeled A3G at a 1:9 ratio, and conducted an experiment using two samples:

600 nM A3G – 100% labeled with GFP

600 nM A3G – 10% labeled with GFP

At room temperature, the 100% labeled sample has a brightness of 2.4 and the diffusion time is 1 ms ± 20%. After a 60 minute incubation at 37°C, the sample now has a brightness of 33, and a diffusion time of 3.2 ms (Table B.1). The normalized brightness of 33 indicates large aggregates, as does the increased diffusion time. A quick estimate

(assuming spherical particles) relates the change in the diffusion time to increased mass (Stokes-Einstein relation). Because of the 20% error in measured diffusion times, the observed increase from $1\,ms$ to $3.2\,ms$ could indicate anything from a 12x to 68x increase in mass, the low end of which is in approximate agreement with the brightness data. It is important to recall, however, that this is a crude estimate based on the assumption that the A3G complexes are spherical in shape.

Sample	**Non-incubated Measured at 20°C**		**Incubated for 60 min Measured at 20°C**	
	Normalized Brightness	Diffusion Time (ms)	Normalized Brightness	Diffusion Time (ms)
100% labeled A3G (600 nM)	2.4	1	33	3.2
10% labeled A3G (600 nM)	1.4	0.7	3.1	2.5

Table B.1 **Labeled and mostly unlabeled A3G shows temperature aggregation.**

When we examine the 10% labeled sample at room temperature, we find a brightness of 1.4 and a diffusion time of $0.7\,ms$. According to theoretical calculations, a 10% fluorescently-labeled dimer should have a brightness of 1.1. Appendix D addresses the fact that the A3G dimer does not exchange subunits as expected. Nevertheless, the reduction from $b = 2.4$ to $b = 1.4$ confirms that the A3G sample is only partially labeled. We now consider the diffusion parameter. A3G has a molecular weight of about $46\,kD$, and GFP has a molecular weight of $27\,kD$, so EGFP-labeled A3G has $MW = 73\,kD$. From molecular weight arguments, we expect a diffusion time of the fractionally-labeled sample to be reduced because many of the A3G moleucles are smaller and faster. A theoretical estimate of the reduced diffusion time, based on the assumption of hydronamic spheres, yields $\sim 0.85\,ms$ for the 10% labeled A3G sample. This is approximately in agreement with the measured result $0.7\,ms$.

We will demonstrate in Appendix C that while mixing labeled and unlabeled A3G dimers should result in a predictable distribution of labeled, unlabeled, and half-labeled A3G dimers, we do not observe the expected behavior. However, this issue will not obscure the results of the following experiment, in which we are forming large aggregates. The temperature-induced aggregate will contain a statistical mix of labeled,

unlabeled and mixed A3G dimers. Because the number of A3G molecules in the aggregate is large, the fraction of labeled and unlabeled A3G molecules within the complex will reflect the population fractions of labeled and unlabeled A3G molecules of the sample at large. For example, large oligomers generated from a 50/50 mix of labeled, unlabeled and mixed A3G dimers will contain an average of 50% labeled and 50% unlabeled A3G molecules, regardless of the specific distribution of the A3G dimer populations.

After incubating the sample at 37°C for 60 minutes, we observe a strong increase in the diffusion time of the sparsely labeled sample (Table B.1). The increase indicates the formation of large complexes and, within uncertainty, agrees with the increase observed for the 100% labeled sample. More importantly, when we compare the final diffusion times of the aggregated 10% and 100% labeled samples, we see that the sparsely labeled aggregates move about 20% faster than the fully labeled aggregates. A diffusion time of $\sim 2.5\ ms$ is consistent with same-sized aggregates as the fully labeled sample (~3.2 ms), once we factor in that most of the A3Gs are not carrying the extra mass of an EGFP label. Most convincingly, however, is the comparison of brightness values from the 10% labeled and 100% labeled A3G samples. The 100% labeled A3G form temperature-induced aggregates with an average brightness $b = 33$. The incubated 10% A3G sample has a brightness $b = 3.1$ which is approximately 10% of 33 and indicates that the mostly unlabeled A3G is also producing temperature-induced aggregates of approximately the same size as the fully labeled sample.

After incubation, the 10% labeled sample showed an increase in diffusion time and in brightness, both indicative of an increase in aggregation. In short, this experiment provides two independent measures that suggest (mostly) unlabeled A3G is still producing large aggregates when subject to 37°C incubation.

Appendix C. Bright and Dark A3G studies

In this appendix we present results that we have thus far been unable to satisfactorily explain regarding the nature of the A3G dimer in solution. This research is an ongoing project, and further experiments are being designed to carefully investigate the issues involving the stoichiometry of A3G. The following experiments should serve as a foundation to guide the design of future experiments.

In FFS measurements of the oligomeric state of A3G in solution (50 mM Tris-HCl), we have consistently recovered an average brightness slightly above 2. We interpret this as population composed of dimeric A3G with a small percentage of higher order oligomers mixed in. Since the field has reported a range of oligomers for A3G in solution, from monomeric to tetrameric, we wish to confirm our observations. Because we are unable to take apart the dimer through dilution, we perform a quick check to assure ourselves that the brightness $b = 2.3$ really does indicate a predominantly dimeric population. In chapter 3, we discussed that oligomeric brightness species demonsrate a predictable decrease in brightness as a function of photobleaching. Therefore, we used epifluorescence light to bleach an EGFP-labeled A3G sample in stages, using 2-photon FFS to take monitoring measurements between each photobleaching step. The normalized brightness values of A3G as a function of intensity relative to the unbleached intensity are plotted in Figure C.1 below.

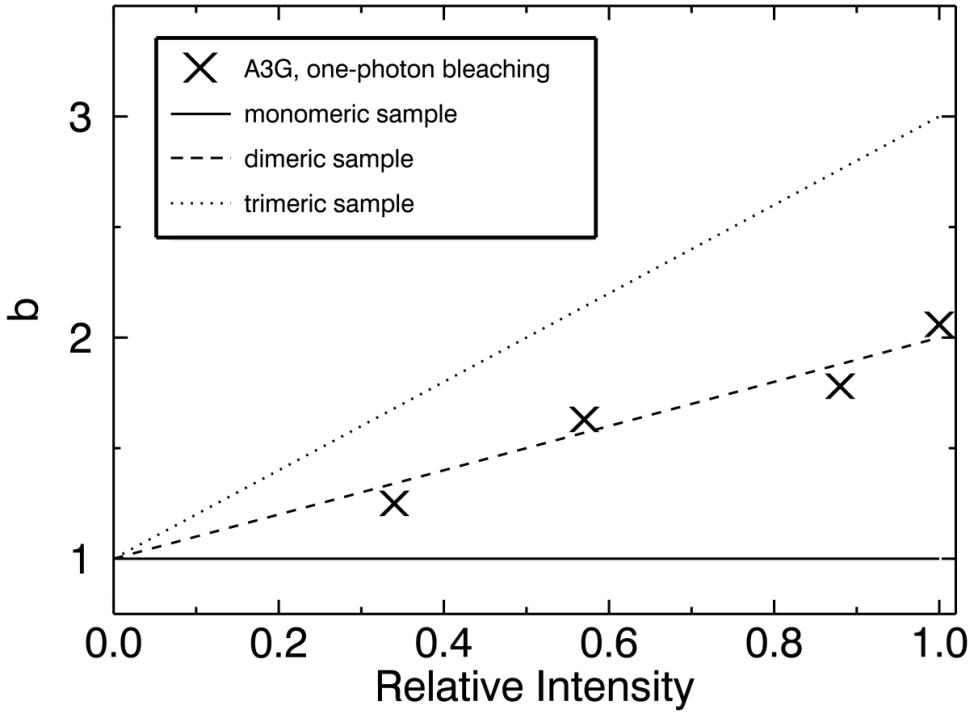

Figure C.1 **A3G sample photobleaches as a dimer.**

The plot also includes theoretical bleaching curves for pure monomer, dimer, and trimer samples. The A3G data falls very nicely along the dimeric curve, confirming that A3G is mostly dimeric as our initial experiments indicated.

To further characterize this dimeric population, we conduct mixing experiments with "bright", or EGFP-labeled, A3G and "dark", or unlabeled, A3G. By observing the exchange of monomer units between the dimers through brightness, we can gain information about the affinity of the dimers. Recall that apparent brightness is given by

$$\lambda^{app} = \frac{\sum_i f_i \lambda_i^2}{\sum_i f_i \lambda_i}$$

where λ_i is the brightness of each species and f_i is the fraction of the total population that species represents. For a 50/50 mix of bright and dark A3G (Figure C.2), we expect a mix of dimers with two labeled A3G proteins (BB), 1 labeled and 1 unlabeled (BD), and two unlabeled proteins (DD). From binomial probability, the population should be approximately

$$0.25\,(BB) + 0.50\,(BD) + 0.25\,(DD)$$

which gives an apparent brightness of

$$\lambda^{app} = \frac{0.25\,(2)^2 + 0.50\,(1)^2 + 0.25\,(0)^2}{0.25\,(2) + 0.50\,(1) + 0.25\,(0)} = 1.5$$

for a 1:1 mixture.

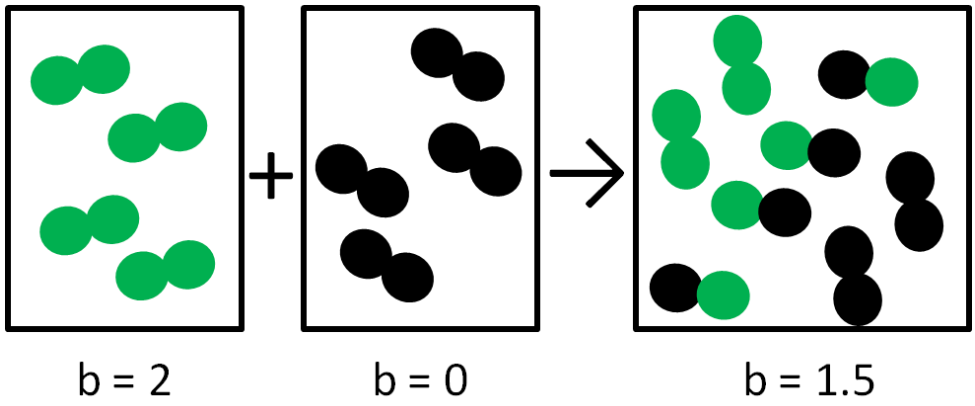

Figure C.2 **Bright & Dark dimeric protein mixing.**

Similarly, for a mix of 1 bright A3G to 10 dark A3G we get an apparent brightness

$$\lambda^{app} = \frac{0.01\,(2)^2 + 0.18\,(1)^2 + \ldots + (completely\ dark\ species)(0)^2}{0.01\,(2) + 0.18\,(1) + \ldots + ds(0)} = 1.1$$

reflecting the 1:10 bright to dark sample.

However, our measurements for a 1:1 mixture of bright and dark A3G recover an apparent brightness of ~1.9 which, within uncertainty, would indicate the absence of mixing. Further experiments with a 1:10 mixture resulted in a brightness of ~1.5 which indicates the presence of mixing. Both these values are much higher than expected for thermodynamic mixing, and the values do not change over a period of 6 hours. Table C.1 shows one example of a mixing experiment which demonstrates that the measured brightness does not change appreciably with time. (Note: the increase in brightness observed at the 20 hour timepoint is likely due to significant sample evaporation which results in increased protein concentration as well as substantial increases in the salt and glycerol concentrations of the solution buffer.) More careful experiments to explore mixing kinetics are still to be undertaken, but the current results are baffling and seem irreconcilable with thermodynamic equilibrium models.

Time (hr)	Labeled A3G	Labeled: Unlabeled 1:1	1:2	1:4	1:9	1:30
1	2.21	2.04	1.97	1.88	1.69	1.61
5	2.20	1.94	1.80	1.70	1.67	1.49
6	2.10	1.90	1.80	1.72	1.70	1.54
7	2.20	1.84	1.80	1.64	1.75	1.51
10	2.20	1.90	1.80	1.70	1.80	1.50
20	2.40	2.03	2.01	1.95	2.10	1.65

Table C.1 **Bright & Dark A3G mixing study.**

Below, Figure C.3 shows a compilation of brightness results from several bright and dark mixing experiments. Viewed in this figure, there seems to be a clear trend in brightness reduction as a function of dark/bright A3G ratio. However, we have no direct evidence that the reduction in brightness is a result of monomer protein exchange as opposed to some other mechanism causing dissociation or quenching. Furthermore, especially in combination with the kinetic data, we can offer no thermodynamically-sound explanation for the observed results.

The most confusing and most recent experiment involved the comparison of brightness between the standard EGFP-labeled A3G and a copurified labeled/unlabeled A3G sample. Ming Li cotransfected unlabeled and EGFP-labeled A3G into mammalian cells and then purified the result. Having run the copurified sample on a gel, he reports that it is approximately 28% labeled. Bafflingly, we observe a brightness $b = 2.3$ for this sample just as we do for the 100% sample. At present, there is no explanation for this result.

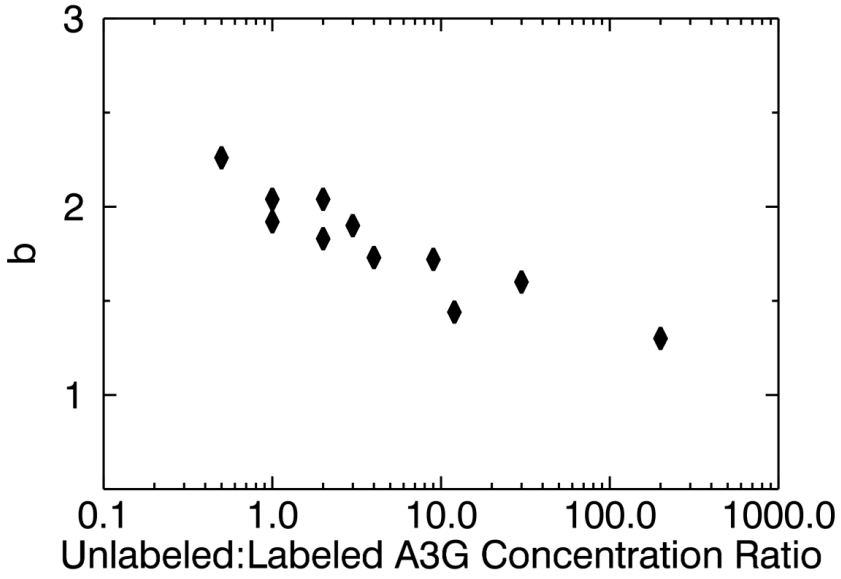

Figure C.3 **Compiled Bright & Dark A3G mixing data.**

Finally, we present a related experiment that does behave as expected. In appendix B, we performed temperature-induced aggregation studies with fractionally labeled samples. In that case, the large A3G aggregates were composed of multiple A3G dimers. We explained that the distribution of labeled, unlabeled, and mixed dimers is irrelevant for predicting the brightness of the aggregate. The average labeled protein fraction of these complexes is simply equal to the fraction of labeled A3G molecules in the sample. Similarly, the formation of large A3G complexes with "G-rich" RNA should also be independent of the dimer mixing issue. The following experiment (Figure C.4) shows 100% labeled A3G interacting with excess Test+ RNA. The same experimental conditions were then applied to a 50% labeled and 28% labeled A3G sample. The latter of these is the copurified sample from the previous paragraph, and the results demonstrate that this sample does contain unlabeled A3G. As the figure shows, the measured brightnesses are reduced by approximately the correct fractions, indicating that RNA complexes are composed of both labeled and unlabeled A3G dimers.

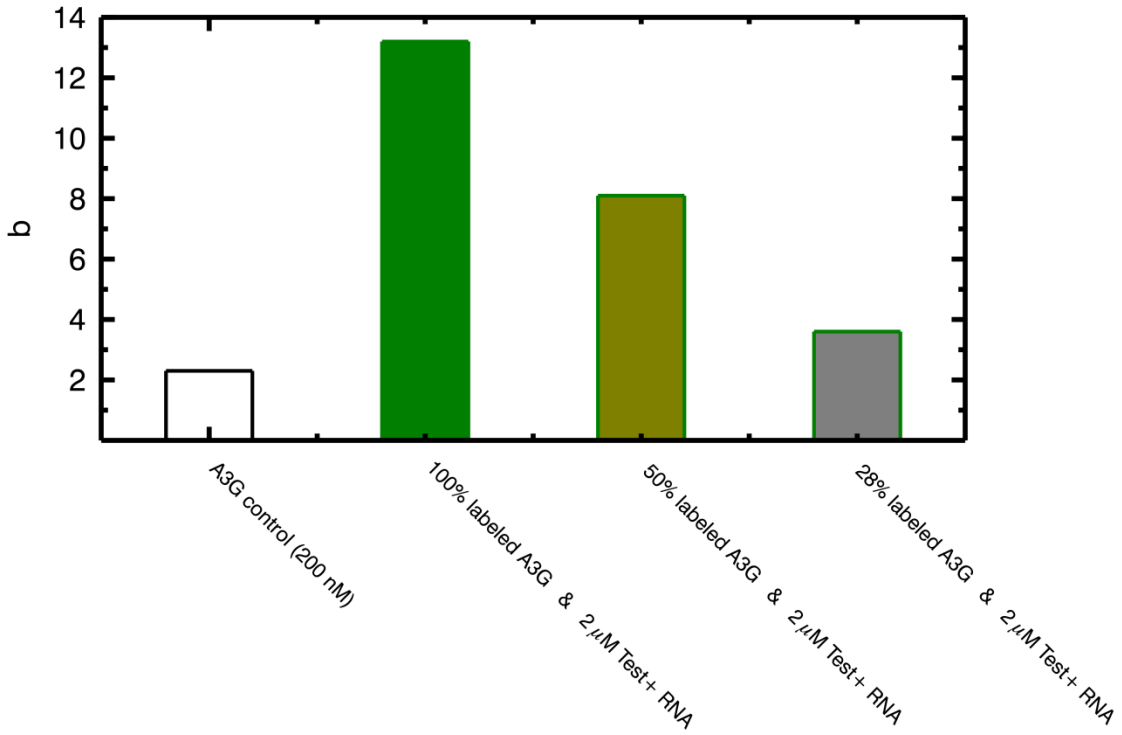

Figure C.4 **Bright & Dark A3G mixtures complex with Test+ RNA.**

Appendix D. A3G dimer dissociation in cytosol analogs

We have been unable to take apart A3G dimers using dilution, mixing with unlabeled A3G, and even adding relatively high concentrations (0.3%) of sodium dodecyl sulfate (SDS), a surfactant we have used to disassemble dimers in the past. However, looking more closely at our low concentration cellular A3G measurements seemed to reveal A3G with an average oligomerization of less than 2.

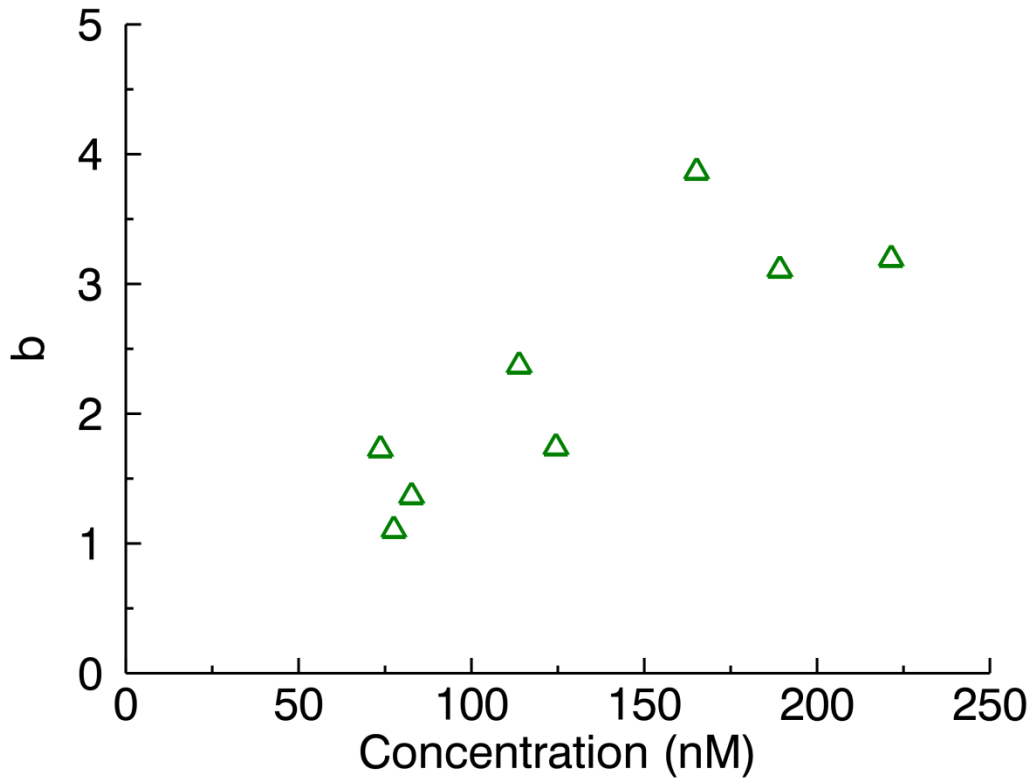

Figure D.1 **Low concentration cellular A3G.**

We must be careful about interpreting the above data. Because this experiment focuses on low concentrations of A3G, the fluorescence intensity signal strength is weak. As a result, the cell's autofluorescent background contributes a large fraction of the measured signal, which has a tendency to reduce the apparent brightness. Nevertheless, based on our estimates of the background bias effect, the data seem to indicate brightness

values less than 2. However, the data shown does not give irrefutable evidence of A3G monomers.

We therefore wondered if some quality or element of the cytoplasmic environment might lead to the dissociation of the A3G dimer. To that end we placed purified A3G—the same material we have been measuring in aqueous solution—into cell free *E. coli* lysate and HeLa cell cytosol extract. Neither of these solutions is identical to cell cytoplasm but, being derived from the cytosol, contain many cytoplasmic proteins. In the following experiment, we dilute aqueous solution A3G into the cytoplasmic extracts by volume ratios, keeping the final A3G concentration fixed at 200 nM.

EGFP controls				
aqueous solution		75% *E. coli* extract		75% *HeLa cytosol*
1.00		1.01		1.02

A3G oligomerization				
aqueous solution	25% *E. coli* extract	75% *E. coli* extract	25% *HeLa cytosol*	75% *HeLa cytosol*
2.10	1.50	1.35	1.80	1.45

Table D.1 **A3G dimer dissociation in cell extract.**

The upper panel of Table D.1 shows EGFP brightness controls in the two lysate environments. Both lysates have slightly greater autofluorescent background compared to aqueous buffer, but for 200 nM A3G this background has a negligible effect on brightness. In the bottom panel, A3G shows clear dissociation in both *E. coli* and mammalian cell extracts, although we are currently unaware of the cause. The following contain descriptions of the extract preparations acquired through personal communication:

> The HeLa cell cytosol extract (S-100) was prepared from HeLa S3 cells. There is no addition of micrococcal nuclease in the preparation protocol. A HeLa cell suspension of high concentration was lysed by osmotic pressure in the presence of a high salt concentration, followed by dialysis against HEPES buffer in the presence of 20% of glycerol. The extract contains splicing factors and has a limiting amount of SR proteins.
>
> The *E. coli* lysate is a cytoplasmic extract from *E. coli*, containing all of the inherent proteins. It is resuspended in 5-10 mM Tris buffer with 14 mM magnesium, 60 mM potassium, and 1mM DTT, pH 8.2.

Since the *E. coli* extract was provided to us by the lab of Vincent Noireaux, University of Minnesota, rather than being purchased, we have a better knowledge of its contents. The *E. coli* extract contains 1 mM spermidine, a charged (3+) polyamine, which we considered might be interfering with a charge interaction or might sequester small RNAs necessary for A3G dimerization. However, we observed only minimal dissociation at 20 mM of spermidine, a concentration well above that in the extract. We have not yet tested other candidates, but this dissociation of A3G in lysates does provide us with an experimental avenue to explore the behavior of the A3G dimer and its monomer exchange in the labeled/unlabeled studies.

Appendix E. Dual-color FFS of A3G-RNA

In the discussion section of chapter 6, we addressed the model of A3G-RNA interaction and complex formation. One of the difficulties in the experiments presented in that chapter is that we only labeled either the A3G or the RNA, and thus could only infer the behavior of the unlabeled reactant. The solution is to conduct more advanced studies using dual-color FFS. The underlying principles are identical, but now both reactants are labeled with differently colored fluorescent labels, and the collected fluorescence light is split into two channels using a dichroic mirror and sent to two separate APD detectors.[33,47]

We have conducted one preliminary experiment with EGFP-labeled A3G and Texas Red-labeled Test+ RNA. Similarly to the initial experiments in chapter 6 (Figure 6.4), we keep the A3G concentration fixed and mix a series of droplets containing different amount of Test+ RNA. We record the data from each detector, perform brightness analysis, and then run the resulting data through an additional filter that separates out the signal for heterospecies complexes. Because both the red and the green channels are synced in time, we can identify and separate the signal contributions that contain both a red and green component. This is called heterospecies partition analysis and allows us to analyze protein complexes separately from the unbound protein background.[46]

Thus the data presented in Figure E.1 show brightness values only for A3G (squares) and RNA (triangles) that are part of an A3G-RNA complex. As before, we observe that A3G starts at slightly above a dimer when the RNA concentration is low, and then forms oligomers at approximately a 1:1 ratio of RNA/A3G. Now, however, we can also see the copy number of RNA. The lower panel of the figure shows the ratio of red to green brightness from the main plot, or in other words, the ratio of RNA to A3G. At low RNA concentrations, there is less than 1 RNA per A3G, as expected from the relative concentrations in the mixture. In regime of large A3G-RNA complexes, we see that the RNA/A3G ratio falls between 2 and 3. This indicates that for every A3G monomer in the complex there are 2 or more Test+ RNA oligonucleotides.

When performing dual-color FFS studies, one must contend with two primary issues that affect quantitative analysis. Most label pairs, including EGFP and Texas Red, do not have complete spectral separation, which means that in our experiment, some of the light from EGFP (~ 3%) ends up in the red detection channel. We have purposely constructed our system (610 nm dichroic mirror) with a red bandpass filter (HQ670, 630-710 nm, Chroma) to minimize the EGFP spillover into the red channel, and with a green bandpass filter (FF02-520, 506-534 nm, Semrock) to prevent any Texas Red contribution

to the green channel. Secondly, the EGFP and Texas Red fluorescent labels will undergo FRET if they are in close proximity (<10 nm). In FRET, energy is transferred from the donor (EGFP) to the acceptor (Texas Red) which results in an increase in Texas Red brightness and a decrease in EGFP brightness. Finally, organic dyes, like Texas Red, are vulnerable to quenching effects when the dye environment changes or when two dyes are brought into close proximity. Figure E.1 contains corrections for none of the above issues, so quantitative results must be taken with a grain of salt. However, since all of these effects will reduce the measured RNA copy number, the data are still suggestive that more than 1 RNA per A3G is to be found in A3G-RNA complexes.

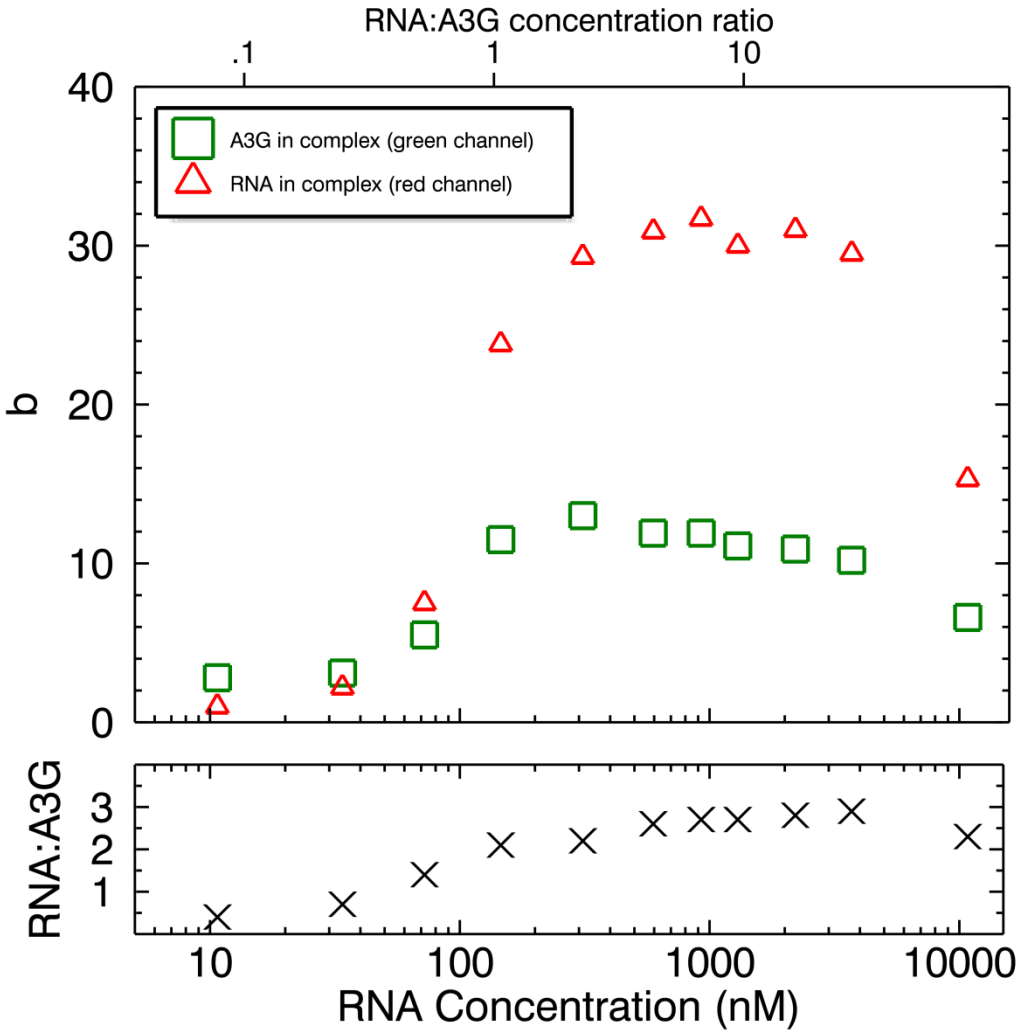

Figure E.1 **Average A3G and RNA copy numbers in A3G-RNA complex.**