Single-molecule fluorescence spectroscopy techniques such as Fluorescence Correlation Spectroscopy (FCS) and single-molecule Förster Resonance Energy Transfer (smFRET) not only possess an unprecedented high sensitivity but also have high temporal and spatial resolution. Therefore, they have an immense potential both in investigation of fundamental biological principles and in clinical applications.

FCS analyses are based on both theoretical approximations of the beam geometry and assumptions of the underlying molecular processes. To address the accuracy of analysis, firstly the experimental conditions that should be fulfilled in order to obtain reliable physical parameters are discussed and the input parameters are carefully controlled accordingly to demonstrate the performance of FCS measurements on our home-built confocal multiparameter photon-counting microscope in several \textit{in vitro} and \textit{in vivo} applications.

Secondly, we performed a comprehensive FCS analysis of rhodamine family of dyes to evaluate the validity of assigning the correlation relaxation times to the time constant of conformational dynamics of biomolecules. While it is the common approach in literature our
data suggests that conformational dynamics mainly appear in the correlation curve via modulation of the dark states of the fluorophores.

The size and shape of the folded, unfolded and chemically-denatured states of the N-terminal Src-homology-3 of downstream of receptor kinases (DrkN SH3) were investigated by FCS and smFRET burst experiments. Based on the data, we conclude that a considerable subpopulation of the denatured protein is in a closed loop state which is most likely formed by cooperative hydrogen bonds, salt bridges and nonpolar contacts.

As a clinical application, we developed and characterized an ultrasensitive capillary electrophoresis method on our multiparameter confocal microscope. This allowed us to perform Direct Quantitative Analysis of Multiple microRNAs (DQAMmiR) with about 500 times better sensitivity than a commercial instrument. Quite remarkably, we were able to analyze samples of cell lysate down to the contents of a single cell.
“In this parterre, none plucked the rose without the thorn.“

Hafiz, ghazal 64

To the reader:

“Said our Pir: “On the Creator’s pen, passed no error:”

On his pure sight, error-covering, hurrah!”

Hafiz, ghazal 105
Acknowledgments

I would like to sincerely thank all the people who supported, helped and assisted me in the past six years. Because of you I had a wonderful time during my study at the University of Toronto, harvesting not only academic achievements, but also friendship and happiness.

I would like to start by presenting my special appreciations to my supervisor: Dr. Claudiu Gradinaru. Dear Claudiu I thank you for your continuous support and help in my academic research. Without your insightful discussions on the aspects of my projects, I would not have accomplished what I have today. I appreciate your efforts and putting up with me during all these years.

I do not know how to thanks Baoxu Liu. He was the first graduate student in the Gradinaru Lab and every day more than before I came to believe I am very lucky to have him as a friend. He helped me not only during my early years but also later on. I would like to add Dr. Sergei Musikhin to the list as well, for tutoring me in working with laser and optics.

I would like to thank all the members of the Gradinaru group for creating a nice working atmosphere in the lab and for helping me with my projects. Particularly, I would like to thank Abdullah Bahram, for assisting me in various research projects. Abdullah, all your hard work, skill and time are highly appreciated. I would also like to acknowledge other members of the group, past and present: Masood Samim, Zhenfu Zhang, Yuchong Li, Subin Yoo, Jordan Rubelo, Maxim Prigozhin, Denys Marushchak, Susanne Scharf, Daniel Jacobs, Daniel Badali and Anthony Pugliese. It has been my great pleasure meeting you all.

I would like to express my distinctive gratitude and appreciations to our research collaborators: Prof. Julie Forman-Kay, Sergey Krylov, Regis Pomès, Walid Houry, and Patrick Gunning and students Angela Yu, Hong Lin, Sarah Rauscher, Grace Li, Bryan Dodgson, David Wegman and Joel Drewry: without your constructive suggestions and valuable materials and samples provided, the research progress presented here would not have been possible. In addition I would like to thank the members of my supervisory committee, Prof. Virgis Barzda and Daniel James for their insightful helps and discussions throughout the years.
To conclude I also would like to thank my family specially my beloved wife Nehleh Rahmati for her continuous support over the past six years. Your love and patience have contributed significantly to my accomplishments.
## Contents

List of Tables xi
List of Figures xii
List of Appendices xv
List of Abbreviations xvi

1 Single-Molecule Spectroscopy: Theory and Experiment 1
   1.1 Introduction .............................................................................................................. 2
   1.2 Theory .................................................................................................................... 3
      1.2.1 Fluorescence Correlation Spectroscopy ............................................................ 3
      1.2.2 Förster Resonance Energy Transfer ................................................................. 7
      1.2.3 FRET-FCS ....................................................................................................... 11
      1.2.4 Time Resolved Fluorescence Anisotropy ....................................................... 12
   1.3 Experimental ............................................................................................................ 14
      1.3.1 Confocal Laser Scanning Microscope ............................................................... 14
      1.3.2 Photon correlation scheme ............................................................................. 17
   1.4 Sample preparation .................................................................................................. 21
      1.4.1 Protein expression and purification ................................................................. 21
      1.4.2 Protein Labeling .............................................................................................. 23
         1.4.2.1 Amine Group Reactive Cross-Linking Chemistry .................................. 24
         1.4.2.2 Cysteine Reactive Cross-Linking Chemistry ....................................... 25
   1.5 Bibliography .......................................................................................................... 26

2 On the Performance of Bioanalytical Fluorescence Correlation Measurements 30
   2.1 Introduction ............................................................................................................ 31
   2.2 Experimental ......................................................................................................... 31
      2.2.1 Materials ....................................................................................................... 31
      2.2.2 Instrumentation ............................................................................................. 32
3.6 Discussion: the assignment of dark states ................................................................. 78
  3.6.1 Assignment of dark state 1 ................................................................................. 79
  3.6.2 Assignment of dark state 2 ................................................................................. 80
  3.6.3 Assignment of dark state 3 ................................................................................. 81
  3.6.4 Assignment of dark state 4 ................................................................................. 83

3.7 Summary ...................................................................................................................... 85

3.8 Bibliography ............................................................................................................ 86

4 Size, shape and folding of the SH3 Domain of Drk Drosophila Adaptor Protein .... 90
  4.1 Introduction .............................................................................................................. 91

  4.2 Experimental ........................................................................................................... 93
    4.2.1 Materials ........................................................................................................... 93
    4.2.2 Protocols .......................................................................................................... 94
      4.2.2.1 Purification of Urea .................................................................................... 94
      4.2.2.2 Protein expression and purification ......................................................... 94
      4.2.2.3 Protein labeling ....................................................................................... 95
      4.2.2.4 Sample preparation ............................................................................... 95
    4.2.3 Instrumentation ............................................................................................... 96
    4.2.4 Procedures ..................................................................................................... 97
      4.2.4.1 Fluorescence correlation spectroscopy ................................................... 97
      4.2.4.2 smFRET-burst ....................................................................................... 97

  4.3 DrkN SH3 .................................................................................................................. 99
    4.3.1 Motivation: simulations vs. experiments ....................................................... 100

  4.4 FCS: Size and shape of $F_{\text{exch}}, U_{\text{exch}}$ and $U_{\text{den}}$ states .................... 103

  4.5 smFRET: the heterogeneity of disordered conformations ..................................... 107
    4.5.1 Direct observation of folded and unfolded states ........................................ 108
      4.5.1.1 The nature of low FRET efficiency peak ............................................. 109
    4.5.2 The charge and pH effect on the Unfolded state ......................................... 111

  4.6 The effect of a variety of compounds on SH3 conformations ................................. 112
    4.6.1 Folded state stabilizers ............................................................................... 113
    4.6.2 Denaturants: resolving a compact scaffold ............................................... 114
    4.6.3 The high-FRET scaffold is not due to aggregates ...................................... 118
    4.6.4 Conformational clusters ............................................................................. 119
4.7 Looping and knotting .................................................................120
  4.7.1 Fluctuating structures in high denaturant concentration ..........121
    4.7.1.1 Evidence of loop formation from the size scaling and intrinsic viscosity .................................................................123
4.8 Loop formation: Cooperative salt bridges and H-bonds ...............125
4.9 Summary and outlook ...............................................................130
4.10 Bibliography ............................................................................132

5 Direct Analysis of Single-Cell miRNA by Capillary Electrophoresis with Confocal MultiParamater Fluorescence Detection 140
  5.1 Introduction .............................................................................141
  5.2 CE review and our setup design .............................................141
    5.2.1 Sensitivity definitions .......................................................142
    5.2.2 CE-LIF methods ..............................................................142
    5.2.3 CE-CMPF ........................................................................143
  5.3 ECI detection volume and sensitivity ........................................146
  5.4 Signal-to-noise optimization ..................................................149
    5.4.1 Data binning time ............................................................149
    5.4.2 Excitation intensity ..........................................................150
    5.4.3 Polarization and lifetime filtering ......................................151
  5.5 Detection sensitivity ...............................................................152
  5.6 Analysis of miRNA biomarkers ..............................................153
  5.7 Summary ...............................................................................157
  5.8 Experimental ..........................................................................157
  5.9 Bibliography ............................................................................162

Appendix A ..................................................................................165
Appendix B ..................................................................................169
Statement of Contributions ..........................................................171
List of Publications ......................................................................173
List of Tables

1-1 List of parts used for all the experiments described in this thesis ................................. 15

2-1 Fluorescence saturation parameters for Rh110 and EGFP ........................................... 36
2-2 Spectral and temporal filtering of the background signal in FCS ................................. 40
2-3 Fitting results of the EGFP polarized-FCS data ............................................................ 45
2-4 FCS fitting parameters of EGFP diffusion in the nucleus of S2 cells .......................... 52

3-1 FCS fitting results for free dyes in glycerol and in water ............................................. 73
3-2 The effect of oxygen and Gloxy on the dark states of TMR attached to ClpP ............... 74
3-3 FCS analysis of wt-SH3 and the T22G mutant in TRIS and 6 M GdmCl ...................... 75
3-4 FCS fitting parameters for 40-bp ssDNA and dsDNA labeled with TMR .................... 78
3-5 DS population (shaded) and de-population rates .......................................................... 78

4-1 Hydrodynamic and gyration radii of SH3 .................................................................... 106
4-2 FRET parameters of the SH3 at several pH values and in MilliQ water .................... 112
4-3 FRET parameters of the SH3 in 6 M GdmCl .............................................................. 115
# List of Figures

1-1 FCS data acquired for 2 nM EGFP in PBS .................................................................6
1-2 FCS data acquired at different depths into a solution of 2 nM Rh6G .................7
1.3 The dependence of the FRET efficiency on the donor-acceptor distance ..........8
1-4 Fluorescence and absorption spectra of Bodipy-FL and Alexa 647 ......................9
1-5 smFRET histogram of the end-labelled DrkN SH3 domain in TRIS buffer ........11
1.6 General optical layout of our confocal microscope ..............................................14
1-7 The user interface of the software correlator that I developed ..............................19
1-8 Digital correlation calculation .............................................................................21
1-9 Labeling chemistries .........................................................................................23

2-1 The average number of photons per molecule vs. excitation intensity ..........35
2-2 FCS-derived parameters vs. excitation intensity ...............................................37
2-3 FCS fitting results vs. background levels ............................................................39
2-4 (A) Time gating. (B) FCS curves for various filtering modes ...........................40
2-5 The dependence of $N_{	ext{eff}}$ and $\tau_d$ parameters on the sample concentration ....41
2-6 Fluorescence correlation measurement of 2 nM EGFP in PBS .........................45
2-7 TRFA data of EGFP in PBS buffer without and with 50% w/w glycerol ..........46
2-8 Normalized FCS curves measuring the Stat3-peptide binding .........................48
2-9 Titration of protein-ligand binding by FCS and FA ........................................49
2-10 Confocal image of the distribution of EGFP in Drosophila S2 cell .................50
2-11 FCS data measured at a location inside the S2 cell nucleus ............................51

3-1 Simplified 3-level scheme of the electronic states of a fluorophore ..............64
3-2 FCS analysis of Rh6G at different excitation intensities ..................................66
3-3 FCS data for SH3-TMR in 6M GdmCl ..............................................................66
3-4 The dyes investigated in this chapter ...............................................................68
3-5 Lactonic, zwitterionic and cationic forms of TMR ............................................70
3-6 Dark state life times versus viscosity for TMR and Rh6G, as estimated by FCS ....73
3-7 FCS analysis of ClpP-TMR in oxygenated and de-oxygenated buffers ..........74
3-8 The FCS curves of WT SH3 and SH3 T22G in TRIS and in 6 M GmdCl .........76

xii
3-9 Correlation curves of the ssDNA and dsDNA, fitted to a 3-dark-state model ............ 77
3.10 The proposed relaxation scheme for DS2 ................................................................. 81
3.11 The proposed relaxation scheme for DS3 ................................................................. 82
3.12 The proposed relaxation scheme for DS4 ................................................................. 84
3-13 Schematics of the assigned dark states of rhodamine dyes ..................................... 85

4-1 The solution structure of the DrkN SH3 domain (PDB ID: 2A36) ...................... 99
4-2 The R_{EE} distribution of the folded ensemble of WT SH3 calculated from a canonical MD simulation starting from the NMR structure ............................................. 100
4-3 The end-to-end distance distribution of WT SH3 at 300 and 802 K ..................... 101
4-4 Normalized Rg of the disordered/denatured state by the Rg of the folded state for number of proteins ...................................................................................................... 102
4-5 Intensity correlation data of WT SH3 and SH3 T22G labeled with A555 ............. 104
4-6 Correlation curves of TMR-labeled SH3 in Tris buffer and in 6M GdmCl ............ 105
4-7 smFRET histogram of SH3 labeled with and A647 in TRIS buffer ....................... 108
4-8 FRET-FCS analysis of SH3 protein ....................................................................... 110
4-9 smFRET histogram of DrkN SH3 at different pH values ..................................... 112
4-10 smFRET histogram of SH3 in the presence of protein stabilizers ....................... 113
4-11 smFRET histograms of SH3 at various concentrations of GdmCl and urea .......... 114
4-12 smFRET histogram of SH3 in 6M GdmCl ............................................................ 115
4-13 The high-FRET efficiency scaffold of SH3 in GdmCl, urea and TRIS buffer ........ 116
4-14 smFRET histograms of SH3 in the presence of formamide and /or GdmCl ......... 117
4-15 The histogram of photon burst brightness for SH3 in 6M GdmCl ....................... 118
4-16 smFRET histogram of SH3 in 6M GdmCl ............................................................ 120
4-17 The K parameter as a function of molecular weight ........................................... 125
4-18 Contact map of half turn loops for DrkN SH3 ....................................................... 127
4-19 Contact map of full-turn loops .......................................................................... 128
4-20 Three loop formation schemes .......................................................................... 130

5-1 Schematic and Photographs of an embedded capillary interface ....................... 144
5-2 Backscattering image of a dry capillary ................................................................. 145
5-3 Backscattering from a dry capillary and fluorescence from a capillary filled with 100 nM Rhodamine 110 ................................................................. 146
5-4 Images of immobilized beads in a capillary ......................................................... 147
5-5 FCS curve of 10 nM Rhodamine dyes measured on the ECI platform .......................148
5-6 Normalized S/N as a function of data binning time .............................................150
5-7 S/N versus the laser excitation power measured after the objective lens ..............150
5-8 (a) Lifetime histograms of tetraborate buffer and 100 pM fluorescein measured in the ECI configuration. (b) S/N for 100 pM fluorescein versus time-gating window in different polarization channels ........................................................................151
5-9 (a) Electropherograms of fluorescein of different concentrations. (b) The S/N ratio of the electropherograms ........................................................................................153
5-10 CE separation of three miRNA biomarkers by DQAMmiR .................................154
5-11 CE analyses of MCF-7 cell lysate ........................................................................155
5-12 CE Analysis of a single cell content using the CE-CMPF-ECI setup ......................156

B.1 Single-molecule burst analysis of 100 pM fluorescein solution in the capillary ......170
List of Appendices

A  Photon detection efficiency 165
B  Single-molecule bursts in the capillary 169
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>Two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>A555</td>
<td>Alexa 555</td>
</tr>
<tr>
<td>A647</td>
<td>Alexa 647</td>
</tr>
<tr>
<td>BFL</td>
<td>Bodipy fluorescein</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>CE</td>
<td>Capillary Electrophoresis</td>
</tr>
<tr>
<td>cLOD</td>
<td>concentration limit of detection</td>
</tr>
<tr>
<td>ClpP</td>
<td>Caseinolytic protease</td>
</tr>
<tr>
<td>CMPF</td>
<td>Confocal multiparameter fluorescence</td>
</tr>
<tr>
<td>cP</td>
<td>centipoise</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per molecule</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>CW</td>
<td>Continuous-wave</td>
</tr>
<tr>
<td>DrkN SH3</td>
<td>N-terminal SRC Homology 3 domain of downstream of receptor kinase</td>
</tr>
<tr>
<td>DS</td>
<td>Dark states</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>DQAMmiR</td>
<td>Direct Quantitative Analysis of Multiple miRNAs</td>
</tr>
<tr>
<td>ECI</td>
<td>Embedded capillary interface</td>
</tr>
<tr>
<td>EGFP</td>
<td>Enhanced green fluorescence protein</td>
</tr>
<tr>
<td>F/B</td>
<td>Fluorescence to background ratio</td>
</tr>
<tr>
<td>F*-peptide</td>
<td>carboxyfluorescein-labelled phosphopeptide</td>
</tr>
<tr>
<td>FA</td>
<td>Fluorescence anisotropy</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>FCS</td>
<td>Fluorescence correlation spectroscopy</td>
</tr>
<tr>
<td>FP</td>
<td>Fluorescent protein</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescence protein</td>
</tr>
<tr>
<td>GdmCl</td>
<td>Guanidinium hydrochloride</td>
</tr>
<tr>
<td>H-bond</td>
<td>Hydrogen bond</td>
</tr>
<tr>
<td>ID</td>
<td>Inner diameter</td>
</tr>
<tr>
<td>IDP</td>
<td>Intrinsically disordered protein</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropylthio-β-galactoside</td>
</tr>
<tr>
<td>ISC</td>
<td>Intersystem crossing</td>
</tr>
<tr>
<td>LIF</td>
<td>Laser-Induced Fluorescence</td>
</tr>
<tr>
<td>LOD</td>
<td>Absolute limit of detection</td>
</tr>
<tr>
<td>MCE</td>
<td>Molecular Counting Efficiency</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamic</td>
</tr>
<tr>
<td>MG</td>
<td>Molten globule</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>mLOD</td>
<td>Mass limit of detection</td>
</tr>
<tr>
<td>MPD</td>
<td>MicroPhotonDevices</td>
</tr>
<tr>
<td>MQ water</td>
<td>MilliQ water</td>
</tr>
<tr>
<td>NA</td>
<td>Numerical aperture</td>
</tr>
<tr>
<td>NHS</td>
<td>N-Hydroxysulfosuccinimide</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>OD</td>
<td>outer diameters</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>o.u.</td>
<td>optical unit</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PCE</td>
<td>Photon collection efficiency</td>
</tr>
<tr>
<td>PET</td>
<td>Photoinduced electron transfer</td>
</tr>
<tr>
<td>PFG-NMR</td>
<td>Pulsed-field-gradient NMR</td>
</tr>
<tr>
<td>PS</td>
<td>Polystyrene polymer</td>
</tr>
<tr>
<td>$R_{EE}$</td>
<td>end-to-end distance</td>
</tr>
<tr>
<td>$R_g$</td>
<td>Gyration radius</td>
</tr>
<tr>
<td>$R_h$</td>
<td>Hydrodynamic radius</td>
</tr>
<tr>
<td>Rh101</td>
<td>Rhodamine101</td>
</tr>
<tr>
<td>Rh110</td>
<td>Rhodamine110</td>
</tr>
<tr>
<td>Rh6G</td>
<td>Rhodamine 6G</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small angle X-ray scattering</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SH2</td>
<td>SRC Homology 2</td>
</tr>
<tr>
<td>SH3</td>
<td>SRC Homology 3</td>
</tr>
<tr>
<td>SHG</td>
<td>Second-harmonic generation</td>
</tr>
<tr>
<td>SMF</td>
<td>Single-molecule fluorescence</td>
</tr>
<tr>
<td>smFRET</td>
<td>Single-molecule Förster resonance energy transfer</td>
</tr>
<tr>
<td>SPAD</td>
<td>Single-photon avalanche detectors</td>
</tr>
<tr>
<td>SSB</td>
<td>Single-stranded DNA binding protein</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>Stat3</td>
<td>Signal transducer and activator of transcription 3</td>
</tr>
<tr>
<td>STD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>TCSPC</td>
<td>Time-correlated single-photon counting</td>
</tr>
<tr>
<td>TDC</td>
<td>Time-to-digital converter</td>
</tr>
<tr>
<td>TICT</td>
<td>Twisted-intramolecular-charge-transfer state</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>TMR</td>
<td>Tetramethylrhodamine</td>
</tr>
<tr>
<td>TRFA</td>
<td>Time-resolved fluorescence anisotropy</td>
</tr>
<tr>
<td>TTTR</td>
<td>Time-tagged time-resolved</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
Chapter 1

Single-Molecule Spectroscopy: Theory and Experiment
1.1 Introduction

Discovering correlations between the conformational changes and the function of proteins has become one of the most significant areas of research in biology. Single-molecule techniques, unlike ensemble averaging measurements, have a great potential to determine the mechanisms and the pathways of elementary processes that are at the basis of biological activities [1]. Enzymatic activity [2], chemical reactions [3], conformational changes [4], random diffusion and rotational motions [5] and excited-state photophysics [6] are some of the possible sources of fluctuations of the fluorescence signals recorded from biological systems. Fluctuations appear in the intensity, the singlet lifetime, the emission frequency and the polarization anisotropy of the detected fluorescence signal.

Fluorescence correlation spectroscopy (FCS) and Förster resonance energy transfer (FRET) and other complementary techniques such as FRET-FCS and time-resolved fluorescence anisotropy (TRFA) have been developed studying fluorescence fluctuations from biological samples. The experimental layout is very similar for all techniques and the main difference is in the type of analysis performed on the fluorescence trace. FCS is able to resolve dynamics if they are faster than diffusion time [7]. On the other hand, fluorescence burst analysis techniques are suited for the slower dynamics and the analysis of FRET efficiency within the photon burst can provide information about conformational heterogeneity and flexibility, and help characterize the structural populations of the protein ensemble [8].

In particular, in-solution single-molecule fluorescence (SMF) techniques are of significant biological interest among the single-molecule methods. The broad range of time scales accessible (nanosecond-second), the possibility of various complementary analyses on the same data set and the ease of sample preparation make them an ideal choice for bioanalytical and biophysical applications, including protein folding and protein interaction studies.

In this chapter I describe in some detail the SMF techniques used in this thesis, the experimental setups in the Gradinaru lab which I helped build and on which I performed the single-molecule measurements, the digital (photon) correlator software which I developed for FCS studies, and the sample preparation protocols for SMF experiments.
1.2 Theory

1.2.1 Fluorescence Correlation Spectroscopy

Fluorescence fluctuation analysis was developed in the early 1970’s to measure chemical kinetic constants and diffusion coefficients in the equilibrium states [9]. Later improvements of the technique using laser excitation sources and ultrasensitive single-photon avalanche detectors (SPADs) in a confocal microscope have made it a standard spectroscopy tool for diluted samples [10]. FCS proved to be an elegant and powerful technique for a broad range of biophysical applications including: screening for protein binding inhibitors [11], non-invasive measurements of mRNA levels in single cells [12], intracellular viscosity mapping [13], in-vivo enzyme kinetics [14, 15], early detection of apoptosis [16], nanoscale thermal sensing [17] and quantum dot photophysics [18]. FCS has also been applied to the measurement of bioreporter response in live cells or cell lysates, though it was not ideal for high fluorophore concentrations or backgrounds, such as those often encountered in the cell [19].

FCS is based on the analysis of intensity fluctuations of fluorescent molecules diffusing through a (sub)femtoliter detection volume [7, 20]. Due to this small volume, the measurements are optimal for concentrations of 1-10 nM, which ensure that the average number of molecules emitting at any given moment is limited to one or a few. The signal fluctuations arise from rotational and translational diffusion of individual fluorophores through the detection volume and from their reversible transitions to less bright states or dark states, due to singlet-state quenching and triplet-state blinking. FCS maps the timescales of the fluctuations and the local concentration to the decay and the amplitude of the correlation curves, respectively.

The fluorescence intensity is commonly split into two or more channels and the normalized auto- and cross-correlation functions are defined as [21]:

$$G_{i,j}(\tau) = \frac{\langle \delta F_i(t) \delta F_j(t+\tau) \rangle}{\langle F_i(t) \rangle \langle F_j(t) \rangle} = \frac{1}{T_d} \left[ \int F_i(t) \left( F_j(t+\tau) - \langle F_j \rangle \right) dt \right] = \frac{1}{T_d} \left[ \int F_i(t) F_j(t+\tau) dt \right] \frac{1}{\left[ \frac{1}{T_d} \int F_i(t) dt \right]} \frac{1}{\left[ \frac{1}{T_d} \int F_j(t) dt \right]} - 1,$$  (1-1)
where \(i\) and \(j\) are detection channels, \(T_A\) is the acquisition time and brackets denote time averaging. The detection volume is determined by the product of the excitation laser beam profile and the collection efficiency function of the confocal pinhole. When the projection of pinhole diameter on the sample plane is in the order of the excitation beam waist, the detection volume can be approximated by a 3D Gaussian profile [22, 23]. Under these assumptions, the intensity correlation function for free Brownian diffusion of a single species is [7]:

\[
G(\tau) = G_{\text{diff}} \prod_i G_{\text{DS},i} = \frac{1}{N_{\text{eff}}} \left(1 + \frac{\tau}{\tau_d}\right)^{-1} \left(1 + \frac{\tau}{s^2 \tau_d}\right)^{-0.5} \prod_i \left(1 + \frac{f_{\text{DS},i}}{1 - f_{\text{DS},i}} \exp\left(-\frac{\tau}{t_{\text{DS},i}}\right)\right),
\]

(1-2)

where in a background free measurement \(N_{\text{eff}}\) is the number of molecules in the effective detection volume, \(V_{\text{eff}} = \pi^{3/2} z_0 w_0^2\), and represents the inverse of the diffusion-related amplitude, \(G_0 = N_{\text{eff}}^{-1}\). However, when the fluorescence intensity is weak and comparable with the non-correlating background \(B\), i.e., at picomolar sample concentrations or for some measurements in cells, the actual number of molecules in \(V_{\text{eff}}\) is determined using [24]:

\[
N = N_{\text{eff}} \frac{\langle F + B \rangle^2}{\langle F \rangle^2}.
\]

(1-3)

In equation (1-2) \(s\) is the structure parameter namely the ratio between the axial and the lateral radii of the detection ellipsoid, \(s = z_0/\omega_0\), \(\tau_d\) is the diffusion time, which is related to the diffusion coefficient \((\omega_0^2 = 4D \cdot \tau_d)\), \(t_{\text{DS},i}\) is the dark state relaxation time and \(f_{\text{DS},i}\) is the fraction of molecules of the \(i^{th}\) dark state. The diffusion coefficient \(D\) depends on the solvent's temperature \((T)\) and viscosity \((\nu)\) and the solute's hydrodynamic radius \((R_h)\) [7]:

\[
D = \frac{k_B T}{6\pi \nu R_h}.
\]

(1-4)

The geometric parameters of the detection volume, \(s\) and \(w_0\) are estimated using a standard dye solution with a known diffusion coefficient. Typical values for our setup were \(s = 6\) and \(w_0 = 300\) nm \((V_{\text{eff}} \approx 0.9\) fL), which were treated as fixed parameters for fitting all other correlations curves. For samples consisting of a mixture of non-interactive diffusive species, the total correlation function is given by the brightness-weighted sum of the individual correlation functions [20]:
\[ G(\tau) = \sum_i N_i^2 \eta_i^2 G_i(\tau) / \left( \sum_i N_i \eta_i \right)^2, \]  

where \( N_i \) is the number of molecules of species \( i \) in the detection volume and \( \eta_i \) is their molecular brightness. Messeth et al. [25] simulated correlation data with statistical noise and showed that one-color FCS can resolve two diffusive species if their diffusion coefficients are at least 60% different. For smaller differences, the fitting quality for one-component and two-component models were the same and therefore indistinguishable. Their experiments showed that there are even more stringent requirements to be fulfilled in real measurements. For globular proteins to be resolved in a mixture, the ratio between their molecular weights needs to be 4 or larger.

In experiments on thin sample layers, the focal spot is treated as a 2D detection area and the translational diffusion part of the correlation function is simply given by [26]:

\[ G_{\text{diff}} = \frac{1}{N_{\text{eff}}} \left( 1 + \frac{\tau}{\tau_d} \right)^{-1}. \]  

This model is useful in cases such as protein diffusion in membrane [26] or in measurements in tight capillaries (see section 5-3) and micro fluidic channels.

In literature reports, the correlation analysis is typically performed on the \( \mu \text{s}-\text{ms} \) time scale, so that the correlation function is satisfactorily described by translational diffusion and a single triplet blinking component (equation (1-2). However, as long as the time scales are well separated, other sources of fluctuations in the fluorescence intensity can be treated independently and therefore included as multiplication factors in the correlation function. Rotational diffusion (to the second order approximation) and anti-bunching terms dominate (sub)nanosecond time scale [21, 27]:

\[ G_{ab} G_{rot} = \left( 1 - \frac{\tau}{\tau_f} \right) \left( 1 + R_1 \exp \left( -\frac{\tau}{\tau_{rot}} \right) + R_2 \exp \left( -\frac{10}{3} \frac{\tau}{\tau_{rot}} \right) \right). \]  

In this equation \( \tau_f \) and \( \tau_{rot} \) are the lifetime of the first singlet excited state and the rotational correlation time, respectively, and \( R_1 \) and \( R_2 \) are numerical parameters that depend on the polarization setup of excitation and detection beams and the average fluorescence anisotropy of the fluorophore. In the case of a single diffusive species, a parameter \( A \) can be defined in such a
way that \( G(0) = 0 \), irrespective of how many molecules are in the detection volume [27, 28]. To the first order approximation, \( \tau_{rot} = 1/6D_{rot} \), where \( D_{rot} \) is the rotational diffusion coefficient given by the Einstein–Smoluchowski relation:

\[
D_{rot} = \frac{k_B T}{8\pi \nu R_h^3}.
\]  

(1-8)

Nanometer-sized fluorophores show nanosecond range \( \tau_{rot} \) [29, 30], and therefore rotational dynamics is usually neglected when analyzing fluctuations on the microsecond time scale and longer. Figure 1-1 shows the correlation curve of a 2 nM Enhanced Green Fluorescence Protein (EGFP) in 50% glycerol. The red line is the fit using anti bunching, rotational diffusion, three dark-state components and translational diffusion (equation 1-2 and 1-6). The detail of this experiment is given in section 2.4.1.

![Figure 1-1](image)

**Figure 1-1.** FCS data acquired for 2 nM EGFP in PBS (pH 7.4, 50% w/w glycerol) – blue line. Data fitting using the complete model described by eqs. 1-2 and 1-6 – red line. The horizontal bar shows the time scale of each term in the correlation function.

The 3D Gaussian description of the detection volume which is characterized by two structure parameters, \( s \) and \( w \), is the underlying assumption when deriving equation 1-2. Deviations of focal volume from the 3D Gaussian profile are, to a large extent, limited by shaping the excitation laser beam and by selecting the appropriate size of the confocal pinhole (~1 optical unit) [22]. The 50-μm pinhole used in our FCS experiments fulfills this requirement.

Aberrational distortion of the detection volume due to non-optimal coverslip thickness or refraction index mismatch can be minimized by optimizing the focal depth in solution. It is
especially important in our setup because we are using a 1.4 NA/100x microscope oil objective (PlanApochromat, Zeiss) which is optimally-corrected for 170 μm glass thickness, while the thickness of our coverslips (12-542C, Fisher Scientific, Canada) is around 155 μm. In addition, the oil immersion objectives are not designed for examining aqueous solutions. We performed a series of FCS experiments on a solution of 2 nM Rhodamine 6G (Rh6G) (R4127, Sigma Aldrich, Oakville, Canada) at different depth (z) of the focal point in the sample, above the coverslip surface (figure 1-2). In the immediate vicinity of the coverslip surface, G(τ) contains a significant slow phase due to the adsorption of the dye to glass. The amplitude of the correlation curve increases as the focal point moves further into the solution, reaching a maximum around z = 4 μm. The detection volume has its smallest size between z = 4 - 8 μm, before the spherical aberrations start to dominate. Therefore, we performed all our FCS experiments within this range, usually at z = 5 μm.

![Figure 1-2](image)

**Figure 1-2.** FCS data acquired at different depths into a solution of 2 nM Rh6G.

### 1.2.2 Förster Resonance Energy Transfer

FRET is a nonradiative energy transfer process from an excited fluorophore, named donor, to another fluorophore in its ground state, named acceptor, through the dipole-dipole interaction [31]. Consequently the donor molecule will relax to the ground state without fluorescence emission and the acceptor will be promoted to a higher electronic state. The process was described by Theodor Förster in a very famous paper in 1948 [32], in which he demonstrated that the efficiency of energy transfer, $E$, between a pair of molecules has a steep dependence on their separating distance, $R$, and is expressed by [4, 33]:

$$E = \frac{1}{1 + \frac{R^6}{R_0^6}}$$
1.2 Theory

\[ E = \frac{1}{1 + \left( \frac{R}{R_o} \right)^6} \quad (1-9) \]

where \( R_o \) is the Förster radius of the donor-acceptor pair and depends on the spectral properties of the two molecules [34]:

\[
R_0^6 = \frac{9000 \ln 10 \, \kappa^2 \, \Phi D J}{128\pi^5 n^4 N_A} \quad (1-10)
\]

Here \( n \) is the refractive index of the solution, \( N_A \) is the Avogadro number and \( \kappa^2 \) is the dipole-dipole orientation factor given by \( \kappa^2 = (cos\theta_T - 3cos\theta_D cos\theta_A)^2 \). \( \theta_T \) is the angle between the interacting dipoles, and \( \theta_D \) and \( \theta_A \) are the angles that donor and acceptor dipoles make with the line connecting them. A rapid reorientation of dyes on the scale of fluorescence lifetime is often assumed, which leads to \( \kappa^2 = 2/3 \) [34]. \( J = \int_0^\infty F_D(\lambda) \epsilon_A(\lambda) \lambda^4 \, d\lambda \) is the overlap integral of donor emission (\( F_D \)) and acceptor absorption spectra (\( \epsilon_A \)).

**Figure 1.3.** The dependence of the FRET efficiency on the donor-acceptor distance. The highlighted area is the distance range which can be estimated reliably based on experimental data using this FRET pair. The red symbols represent the end-to-end distances estimated from smFRET data for several conformational states of the SH3 protein studied in chapter 4.

FRET efficiency because of its strong dependence on \( R \) is used as an intramolecular ruler in protein research. Figure 1-3 shows the transfer efficiency as a function of inter-dye distance for the FRET pair, Bodipy Fluorescein (BFL) and Alexa 647 (A647), used in chapter 4 of this thesis. The highlighted area depicts the range of donor-acceptor distances for which \( E \) falls between 5% and 95%. Outside this range, \( R \) cannot be determined accurately from the measured FRET
efficiency. As an example, the estimated end-to-end distances of the folded, natively unfolded and fully denatured states of a SH3 protein that is labeled with BFL and A647 at its two ends are depicted by red symbols in the figure (for details, see chapter 4).

![Fluorescence and absorption spectra of Bodipy-FL and Alexa 647. The blue- and red-filled areas represent the spectral detection windows for the donor and acceptor channels, respectively.](image)

**Figure 1-4** Fluorescence and absorption spectra of Bodipy-FL and Alexa 647. The blue- and red-filled areas represent the spectral detection windows for the donor and acceptor channels, respectively.

In FRET experiments, the collected fluorescence from the sample is split by a dichroic mirror into blue and red, or donor and acceptor channels. Bandpass filters further narrow down the detection spectral range for each channel. The figure 1-4 shows the fluorescence and absorption spectra, as well as the spectral detection windows of the FRET pair, Bodipy Fluorescein (BFL) and Alexa 647 (A647). The fluorescence quantum yields of the two dyes conjugated to the protein were $\Phi_D \approx 0.55$ and $\Phi_A \approx 0.37$, respectively. Using the measured donor yield in eq. 1-10 results in a small Förster radius, $R_0 \approx 44.4 \text{ Å}$, which is primarily due to the small spectral overlap of this dye pair. The quantum yields and emission spectra of both dyes varied upon addition of denaturants, so that $R_0$ needed to be estimated under various solvent conditions to account for these spectral changes.

Single-molecule FRET (smFRET) measurements can be performed on both surface-immobilized and on freely-diffusing molecules. In the first case, an intensity-time trajectory is acquired for seconds (even minutes) from a single molecule and used to identify FRET efficiency levels representing protein conformational states, as well as map the transition probabilities among them. Biotin-streptavidin interactions are used to tether the single molecules on the glass surface [33]. This requires conjugation of biotin to the biomolecule which may affect its thermodynamics. Alternatively, the encapsulation of biomolecules within surface-immobilized
lipid vesicles makes it possible to reduce both the conjugation perturbation and the surface effects, while providing a controlled environment for biophysical studies [33].

On the other hand, smFRET measurements on freely-diffusing molecules, which are presented in this thesis, do not require multi-step, complex sample preparation protocols and do not suffer from the artifacts and restrictions that may affect the immobilized scheme. However, this is achieved at the expense of losing the underlying millisecond-second dynamics information [33]. To obtain single-molecule resolution, the molecular concentration will be diluted to such a low concentration (~50 pM) that only about 1-5% of the time during a measurement, a molecule passes through the focal detection volume and produces a short (ca. 100 μs) burst of fluorescence photons. Intensity bursts will be recognized using a search scheme such as the sliding window burst search algorithm [1]. For each burst, the FRET efficiency can be calculated from the number of detected photons in the donor \(n_D\) and acceptor \(n_A\) channels by [35]:

\[
E = \frac{n_A}{n_A + \gamma n_D}
\]

where \(\gamma\) is the ratio of the detection efficiencies \((\xi)\) and the quantum yields \((\Phi)\), \(\gamma = \xi_A \Phi_A / \xi_D \Phi_D\). Practically, the background in each channel should be subtracted from \(n_D\) and \(n_A\) to obtain the actual FRET efficiency. In addition, simple corrections should be applied to account for the spectral cross-talk when required [36]. The histogram of the FRET efficiencies calculated from individual bursts contains information about the underlying conformational states of the host biomolecule. In addition, eq. 1-9 can be used to calculate the donor-acceptor distance of each FRET sub-population. Figure 1-5 shows the smFRET burst histogram obtained for an unstable SH3 domain under normal conditions. Two states can be easily distinguished, folded and natively unfolded, which correspond to the 97% and 52% FRET efficiency peaks. The details of this experiment are presented in chapter 4 of this thesis.

In eq. 1-11, \(\gamma\) has to be determined from control experiments performed in advance. Briefly, the detection efficiencies \(\xi_A / \xi_D\) have both spectral and spatial contributions. The spectral contribution is evaluated for our FRET pair to be ~0.55 by taking into account the fluorescence spectra, the spectra of the filter sets and the quantum efficiency of detectors in both channels. The spatial contribution is mainly due to the different spatial filtering of the donor and acceptor photons by the confocal pinhole, different detection path alignments and different size of the photosensitive chip of the two SPAD detectors.
1.2 Theory

Figure 1-5. smFRET histogram of the end-labelled DrkN SH3 domain in TRIS buffer (see section 4-5-1).

The bandpass filters were removed and the ratio of fluorescence intensities in acceptor and donor channels was measured from a 1 μM solution of Alexa 555 dye (~1.19). In addition, the dye fluorescence spectrum in conjunction with the dichroic filter spectrum and the spectral quantum efficiency of the SPAD detectors were used to calculate the expected ratio of intensities in both channels neglecting the spatial filtering effect (~0.76). The measured value (1.19) was divided by the calculated value (0.76) to obtain the spatial contribution of $\frac{\varsigma_A}{\varsigma_D}$ (~1.57), which results in $\frac{\varsigma_A}{\varsigma_D} \approx 0.55 \times 1.57 = 0.86$. Finally, $\gamma \approx 0.58$ was determined using this value and the quantum yields of the two dyes measured in TRIS buffer (pH 7.5).

1.2.3 FRET-FCS

It is also possible to perform correlation analysis on FRET-pair-labeled biomolecules by cross-correlating the donor and acceptor signals. A FRET-FCS model has been developed to describe two-state conformational dynamics [37]. The formalism is quite general and can be applied to any case that the FRET efficiency fluctuates between two different values $E_1$ and $E_2$. Under some simplifying assumptions, the correlation functions, $G$, can be expressed as a product of the familiar translational diffusion term, $G_{\text{diff}}$, and a kinetic term, $X$, that represents the kinetics of interconversion between the two FRET states [37].

$$G_{xy} = G_{\text{diff},xy}(\tau)X_{xy}(\tau)$$  \hspace{1cm} (1-12)

The indexes $x$ and $y$ are D or A, representing the donor and the acceptor channels. $X_{xy}(\tau)$ is given by [37]:

$$X_{xy}(\tau) = \frac{1}{\gamma} \frac{\varsigma_A}{\varsigma_D} \left[ 1 - e^{-\gamma \tau} \right]$$
1.2 Theory

\[ X_{DD}(\tau) = 1 + \frac{k_{12}k_{21}(E_1 - E_2)^2}{k_{21}(1 - E_1) + k_{12}(1 - E_1)} e^{-(k_{12} + k_{21})\tau} \]

\[ X_{AA}(\tau) = 1 + \frac{k_{12}k_{21}(E_1 - E_2)^2}{k_{21}E_1 + k_{12}E_2} e^{-(k_{12} + k_{21})\tau} \]

\[ X_{DA}(\tau) = 1 - \frac{k_{12}k_{21}(E_1 - E_2)^2}{k_{21}(1 - E_1) + k_{12}(1 - E_1)(k_{21}E_1 + k_{12}E_2)} e^{-(k_{12} + k_{21})\tau} \]

(1-13)

\[ k_{12} \text{ and } k_{21} \text{ are the conversion rates between the FRET states. The power of the technique is when diffusion parameters of the two states can be considered the same hence the diffusion components cancel out in the correlation ratios and ideally correlation analysis extends beyond the diffusion time. In this case the three correlation ratios are:} \]

\[ \frac{G_{AA}}{G_{DD}} = \frac{X_{AA}(\tau)}{X_{DD}(\tau)} , \quad \frac{G_{AA}}{G_{DA}} = \frac{X_{AA}(\tau)}{X_{DA}(\tau)} \text{ and } \frac{G_{DD}}{G_{DA}} = \frac{X_{DD}(\tau)}{X_{DA}(\tau)} \]

(1-14)

They can be fitted globally to estimate the FRET efficiencies $E_1$ and $E_2$ and the conversion rates, $k_{12}$ and $k_{21}$.

1.2.4 Time-Resolved Fluorescence Anisotropy

The probability of exciting a molecule by linearly polarized light is proportional to $\cos^2 \theta$, where $\theta$ is the angle between the absorption transition dipole moment and the polarization vector of the incident light. When the two vectors are collinear the absorption is maximum. On the contrary, when they are perpendicular the absorption is very poor. Due to this photo-selection rule, the excited molecules are initially oriented in a direction close to that of the electric field. If by some magic mechanism we could freeze the molecules immediately after the excitation event, their fluorescence also will be preferentially polarized along the same direction. In reality, the Brownian rotational diffusion randomly shuffles the orientation of the molecules, which results in a decrease of polarization purity of the emitted fluorescence. Typically, organic dyes have a rotational diffusion (correlation) time much shorter than the fluorescence lifetime, hence their fluorescence are almost non-polarized [38]. However, they may become spatially restricted upon attachment to larger molecules like proteins. As a consequence, the degree of fluorescence polarization is determined by the rotational mobility of the host molecule instead of that of the
fluorophore itself. The fluorescence polarization is commonly used in ligand-protein binding applications, including drug screening applications [39, 40].

The fluorescence polarization anisotropy is defined as [31]:

\[ r = \frac{I_\parallel - g I_\perp}{I_\parallel + 2g I_\perp} \] \hspace{1cm} (1-15)

where \( I_\parallel \) and \( I_\perp \) are detected fluorescence signals in two orthogonally polarized channels and \( g \) is a correction factor for different sensitivity of the two channels. For measurements in an isotropic solution, the limiting cases of very fast and very slow molecular tumbling molecules yield anisotropy values of 0 and 0.4, respectively. In reality, the measured anisotropy depends on the rotational diffusion time of the molecule and it falls between these two limiting cases. This dependence is described by Perrin formula [31]:

\[ \frac{r_0}{r} = 1 + \frac{\tau_f}{\tau_{rot}} \] \hspace{1cm} (1-16)

where \( r_0 \) is the initial anisotropy (typically 0.4), and \( \tau_f \) and \( \tau_{rot} \) are the fluorescence lifetime and the rotational correlation time, respectively.

When the Brownian rotation is the only fluorescence depolarization process, eq. 1-16 can be used to estimate the rotational correlation time from the steady-state anisotropy. However, this can only be strictly applied to spherical molecules and it is not applicable to other molecular shapes and to more complex cases. For instance, when a fluorophore is attached to a protein, the anisotropy decay reflects fluorophore tumbling, protein rotation and other protein dynamics such as segment rotation. In such circumstances, the time-resolved modality provides the complete information of dynamic anisotropy [31, 41]. In a pulsed excitation scheme, the fluorescence intensities in eq. 1-15 (and therefore the anisotropy) are a function of time and decay exponentially after the excitation pulse. By fitting the measured time-resolved anisotropy \( r(t) \) using multiple exponential decays it is possible to determine the time constants of the underlying processes.
1.3 Experimental

1.3.1 Confocal Laser Scanning Microscope

All measurements in this thesis were performed on our custom-built multiparameter confocal microscope. A simplified schematic optical layout of the setup is shown in the figure 1.6. The parts used in this scheme for each experiment are given in table 1.1. The excitation light can be selected from three continuous-wave (CW) lasers with narrow lines at 473, 532 and 635 nm, respectively, and a broadly tunable Ti-sapphire laser (TsunamiHP, Spectra Physics) that outputs 100-fs pulses at a repetition rate of 80 MHz.

![General optical layout of the home-built confocal multiparameter fluorescence (CMPF) microscope in Gradinaru lab. Pinholes, SPADs, dichroics (Di), long-pass (LP) and band-pass (BP) filters, beam splitters (Sp), and other additional optics (O) used for the experiments in this thesis are given in table 1.1. A broadband polarizer (O1) and a broadband half-wave plate (O2) were installed for rotational diffusion measurement by FCS (see section 2.4.1).](image)

Figure 1.6. General optical layout of the home-built confocal multiparameter fluorescence (CMPF) microscope in Gradinaru lab. Pinholes, SPADs, dichroics (Di), long-pass (LP) and band-pass (BP) filters, beam splitters (Sp), and other additional optics (O) used for the experiments in this thesis are given in table 1.1. A broadband polarizer (O1) and a broadband half-wave plate (O2) were installed for rotational diffusion measurement by FCS (see section 2.4.1).
Table 1.1. List of parts used in figure 1.6 for all the experiments described in this thesis. The blue and green laser lines are ~480 nm and ~530 nm, respectively, either from CW or from pulsed laser sources.

<table>
<thead>
<tr>
<th>Parts</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FCS</td>
</tr>
<tr>
<td>Laser color</td>
<td>CW/Femto blue</td>
</tr>
<tr>
<td>Di</td>
<td>FF495-Di02 Semrock</td>
</tr>
<tr>
<td>BP1</td>
<td>HQ530/50 Chroma</td>
</tr>
<tr>
<td>BP2</td>
<td></td>
</tr>
<tr>
<td>BP3</td>
<td></td>
</tr>
<tr>
<td>Sp1</td>
<td></td>
</tr>
<tr>
<td>Sp2</td>
<td></td>
</tr>
<tr>
<td>Sp3</td>
<td>Polarizer cube</td>
</tr>
<tr>
<td>Pinhole</td>
<td>50 μm</td>
</tr>
<tr>
<td>SPAD 1</td>
<td></td>
</tr>
<tr>
<td>SPAD 2</td>
<td></td>
</tr>
<tr>
<td>SPAD 3</td>
<td>MPD</td>
</tr>
<tr>
<td>SPAD 4</td>
<td>MPD</td>
</tr>
</tbody>
</table>

The Tsunami laser is continuously tunable across a broad spectrum, from 700 nm to 1100 nm, but three wavelengths (810, 960 and 1054) are of special interest in single-molecule applications due to the spectra of fluorophores commonly used in these studies [42-44]. In order
to obtain visible pulses with wavelengths centered at ~405, 480 nm and 527 nm, three nonlinear β-BBO crystals in parallel optical paths are utilized for second-harmonic generation (SHG). Flipping mirrors are used to switch between different optical paths with high repeatability. The laser power is adjusted by a rotating half-wave plate followed by a polarization cube in the principal laser beam. All laser lines are combined using dichroic filters and flipping mirrors.

A 1.4NA/100x Plan-Apochromat microscope oil objective (420790-9900, Carl Zeiss, Canada) focuses the laser excitation beam onto the sample, which is mounted on top of a thin glass coverslip. The coverslip is mounted on a three-axis piezo scanner (T225, MadCity Labs, Maddison, WI, USA), which controls the position and depth of the focus into the liquid sample. The emitted fluorescence is collected through the same objective and the scattering and out-of-focus contributions are removed by a series of dichroic, long-pass optical filters and a confocal pinhole (50 or 75 μm diameter) [45].

The fluorescence is further split into two or more component beams as required, using dichoric mirrors and/or polarization cubes. Each resulting beam is spectrally filtered using band-pass filters and then focused onto separate SPAD detectors. SPADs are either fast-response MicroPhotonDevices (MPD) modules (PD5CTC, Optoelectronic Components, Kirkland, QC, Canada) or red-sensitive modules (COUNT-100C, Laser Components, Hudson, NH, USA), which have ~48% and ~73% photon detection efficiency peaks at 542 and 660 nm, respectively.

The only deviation from this scheme was in the measurement of rotational diffusion by FCS (see section 2.4.1), for which a broadband polarizer was inserted after the confocal pinhole to select the parallel polarization of fluorescence. After the polarizer a broadband half-wave plate was inserted in order to rotate the polarization by an angle of ca. 45° so that a polarizer cube can split it almost evenly between two detection channels.

The SPADs are connected to one of the two independent input ports of a PicoHarp300 single-photon counting module (PicoQuant, Germany), either directly or via a router module (PHR800, PicoQuant). Each PicoHarp300 input channel has an independent time-to-digital converter (TDC) with a ~80 ns dead time and picosecond timing precision while the four input ports of the router module share the same TDC. Therefore, in order to perform correlation analysis for time scales faster than the TDC dead time, the two cross-correlating detector outputs should be connected directly to PicoHarp300 to avoid sharing the same TDC (see section 2.4.1).
1.3 Experimental

Data acquisition is performed in either Time-tagged (TTR) or Time-tagged time-resolved (TTTR) photon counting modes of PicoHarp300. In the TTR mode both input ports are connected to the SPAD outputs. In the TTTR mode, which is used in combination with pulsed excitation, the first input port is reserved for the reference signal from a high-speed photodiode (PHD-400-N, Becker & Hickl) in the excitation path to accomplish time-correlated single-photon counting (TCSPC) detection. In this mode, in addition to the absolute photon arrival times, the delay between the excitation pulses and the detected photons (start-stop time) in each channel are recorded with a resolution of 4 ps, therefore it is best suited for applications such as time-resolved fluorescence anisotropy and fluorescence lifetime measurements.

A LabView code developed in Gradinaru Lab by Baoxu Liu was used to control photon data acquisition on the CMPF microscope and to construct and visualize intensity-time trajectories and start-stop (lifetime) histograms. I wrote a LabView code to accomplish digital (photon-by-photon) correlation analysis using logarithmic time binning and 24 bins per temporal decade (see below). A LabView code for identifying and analyzing single-molecule intensity bursts was written by Baoxu Liu.

1.3.2 Photon correlation scheme

In order to perform fluorescence correlation analysis, the correlation curves need to be calculated from the fluorescence intensity-time trajectories. This can be done with both hardware and software based correlators [46]. Hardware correlators are commercially available and easy to use, hence they are very popular. However, they suffer from several drawbacks including their inability in rejecting unwanted signals such as scattering of the excitation light and giant fluorescence bursts from impurities in the sample. While (rare) giant bursts in the intensity data could be removed in software, they can seriously affect the correlation curve produced by hardware correlators and lead to discarding the entire acquisition. In addition, whereas fluorescence photons emitted by typical fluorophores are spread over several ns after the excitation event, (Rayleigh and Raman) scattering photons arrive very soon after the excitation pulse, within ca. 100 ps, and therefore they can be removed by software-gating in time-resolved experiments (see section 2.3.2). The combination of this capability with software correlation allows us to obtain reliable FCS data in dilute (pM) samples and in live cells, for which either the scattering or the fluorescence background may be dominant.
Ideally the measured standard deviation squared (SD$^2$) of the correlation curve at each time bin should be used as weighting factors for fitting the data to a theoretical model (eqs. 1-2, 1-5, 1-6 and 1-7). SD can be obtained experimentally either from a sequence of identical measurements or by dividing a single, long acquisition into shorter time slices (typically 10 s) and calculating a FCS curve for each slice. To avoid the requirement for multiple measurements in the hardware correlating scheme, several models have been developed to estimate the noise of the correlation curves. However, all these models deviate from the experimentally-measured SD at long correlation times [46]. Correct data weighting is essential for the accuracy of parameter evaluation by fitting FCS curves [46].

I developed a LabView code that builds the correlation curves from the time-resolved raw data, which is a list of photons and their absolute detection time. This scheme also allows the calculation of SD for weighting in the curve fitting. Typically, the correlation analysis is performed over a time scale spanning at least five orders of magnitude (1 μs – 100 ms). Therefore, the linear binning of the time axis is not efficient for practical purposes. Hardware correlators are designed based on the proposed architecture by Schatzel, where the time axis is binned using a mixed linearly and logarithmically increasing sampling times [46]. This way, for example, for the correlation analysis in the range of 0.2 μs - 50 ms the number of sampling times is reduced from 2.5×10$^5$ in the linear sampling design to only 128 time bins in the Schatzel design. Similarly, in our software correlator design, we used a logarithmic time binning and 24 bins per temporal decade, equivalent to linear binning of the logarithmic time axis. Unlike the hardware correlators, the number of time sampling bins per decade is not fixed and can be adjusted as required. This could be important in resolving close correlation decay components such as overlapping dark states and diffusive species.

Figure 1-7 shows the graphic user interface of our correlator software code. It consists of three panels: the fluorescence correlation curve, SD$^2$ and the fluorescence intensity count rate. In the figure, they are shown for a 1000 s data set acquired as part of a sequence from a 1 nM Rhodamine 6G dye in 50% glycerol (see section 2.5.1). The program is incorporated in the main single-molecule data analysis LabView program developed in the Gradinaru lab. The correlation curve and its SD$^2$ will be saved in one file in .txt format, which later are used for parameter fitting with an appropriate theoretical model in Origin. The input parameters in my program are the correlation channels, the number of time slices for the SD$^2$ calculation, the number of bins per
1.3 Experimental decade, the time range for calculating the correlation curve and the edge correction factor (see below).

Figure 1-7 The user interface of the software correlator that I developed. The interface has three panels: the correlation curve (up), SD^2 (middle) and the intensity count rate (bottom). The data is 1000 second acquisition in a series of acquisitions on a sample of 1nM Rhodamine 6G in 50% w/w glycerol.
To obtain the digital correlation function for a photon dataset with a binning time \( \Delta \), we substitute the integral by a sum in eq. 1.1, \( F(t) \) by \( \frac{N_{\text{ph}}(n)}{\Delta} \) and \( dt \) by \( \Delta \):

\[
G_{i,j}(m) = \frac{1}{T_A \Delta} \sum_n N_{\text{ph},i}(n) N_{\text{ph},j}(n + m) - 1 = \frac{1}{T_A \Delta} \sum_n N_{\text{ph},i}(n) N_{\text{ph},j}(n + m)

\]

\[
= \frac{1}{T_A \Delta} \sum_n \left( \frac{1}{T_A} \sum_n N_{\text{ph},i}(n) \right) \left( \frac{1}{T_A} \sum_n N_{\text{ph},j}(n) \right) - 1.
\]

(1-17)

where \( n \) and \( m \) are the time bin indexes, \( N_{\text{ph}}(n) \) is the number of photons in the time bin with index \( n \), and \( CR_i \) and \( CR_j \) are average count rates in channels \( i \) and \( j \). Building the correlation curve is therefore equivalent to calculating the sum in the numerator of equation 1-17. Figure 1-8 depicts how this calculation is performed for a virtual photon time sequence for \( m = 2\Delta \). This calculation can be accomplished in two ways. In the first method, which is just exactly what the equation states, the time trajectory is shifted by \( m = 2\Delta \), then multiplied by the original time trajectory and the obtained values at each bin are summed up. In an equivalent approach, as depicted in the figure, we count for each photon in channel \( i \) the photons arriving exactly \( 2\Delta \) later in channel \( j \), and then sum all numbers obtained for all photons in the dataset. Thus, the calculation of the correlation function is reduced to the histograming of all the photon-photon time differences in the raw data, and equation 1-17 can be written as:

\[
G_{i,j}(\tau) = \frac{1}{T_A \Delta} \frac{\text{Hist}_{i,j}(\tau)}{CR_i CR_j} - 1,
\]

(1-18)

where \( \text{Hist}_{i,j}(\tau) \) is the histogram of time differences \( \tau \) between photons in the detection channels \( i \) and \( j \). Unlike equation 1-17, this equation allows us to avoid choosing a binning value of the intensity-time trajectories before deriving the correlation function. In fact, we use the absolute arrival time of photons to calculate the time differences and any time binning of the correlation time axis can be chosen at the end to plot the correlation histogram. In another words, the calculation of equation 1-17 requires a linear binning of the absolute time axis and therefore of the correlation time axis, whereas any binning format can be applied using equation 1-18. For this reason, we use \( \tau \) instead of \( m \) for the correlation time in the equation 1-18. As mentioned above, a logarithmic time binning with 24 bins per temporal decade was used in all studies presented in this thesis.
1.4 Sample preparation

1.4.1 Protein expression and purification

**N-terminal SH3 domain of the Drosophila adaptor protein Drk (DrkN SH3):** A plasmid coding for the isolated N-terminal SH3 domain of downstream of receptor kinase (DrkN SH3), residues 1 through 59 [47], was transfected into *Escherichia coli* HMS 174 cells under the control of...
of the T7 promoter. The expression of the protein was induced for 4 h at OD$_{600}$ = 0.6 by addition of 250 mg/liter IPTG to bacterial growths at 37 °C in M9 minimal medium, supplemented with 0.3% $^{13}$C labeled glucose, 0.1% $^{15}$NH$_4$Cl, 100 mg/L ampicillin, 10 mg/L thiamine, 10 mg/L biotin and 1 mM of each MgSO$_4$ and CaCl$_2$. Cells were lysed by sonication in 50 mM Tris, 2 mM EDTA, 5 mM benzamidine HCl and 7 mM $\beta$-mercaptoethanol. The DrkN SH3 domain was purified on a DE 52 ion-exchange column with a linear gradient of NaCl (0-1 M) followed by Superdex 75 gel filtration column in 0.15 M NaCl, 50 mM Tris, 2 mM EDTA, 5 mM benzamidine HCl and 7 mM $\beta$-mercaptoethanol, followed by a MonoQ ion-exchange column with a linear gradient of NaCl (0-0.3 M). The T22G [48] and the C2C61G62 mutants were expressed and purified in a similar manner to the wild-type protein. The identity and purity of all proteins were confirmed by mass spectrometry. All the protocols were carried out in the Forman-Kay laboratory (Sick Kids, UofT).

_E. coli_ caseinolytic protease (ClpP):_ ClpP E150C was purified as previously described [49]. _E. coli_ ClpP (E150C) mutant was generated from the wild-type ClpP using the QuickChange kit (Stratagene) following manufacturer’s protocol. The construct was then subcloned into pET9a vector and transformed into SG1146(DE3) ΔclpP cells to express untagged ClpP (E150C). A single transformant colony was grown overnight in 10 mL of Luria-Broth (LB) media containing 100 μg/mL ampicillin at 37°C. The overnight culture was inoculated into 1 L LB with 100 μg/mL ampicillin and grown at 37°C. When OD$_{600}$ reached 0.6, isopropyl-1-thio-$\beta$-D-galactopyranoside (IPTG) was added at a final concentration of 1 mM to induce protein expression for 3 hours at 37°C. Cells were then collected by centrifugation and lysed by French press. Cell debris was removed by centrifugation. The supernatant containing ClpP(E150C) was partially purified by ammonium sulphate precipitation at 30-60% saturation. Precipitated ClpP(E150C) was resolubilized and dialyzed in buffer A (50 mM TrisHCl, pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol). The protein was further purified using an anion-exchange column (Q Sepharose). The high salt buffer (buffer B) contained 50 mM TrisHCl, pH 7.5, 1 M KCl, 1 mM DTT, and 10% glycerol. ClpP (E150C) eluted at 250-300 mM KCl. Subsequently, the protein was further purified by size exclusion chromatography using Superdex 200 column equilibrated with buffer A. The protein was then stored in buffer A, in the absence of 1 mM DTT. The yield was typically 60 mg of purified ClpP (E150C) from 1 L culture. All the protocols were carried out in the Houry laboratory (Biochemistry, UofT).
1.4.2 Protein Labeling

Fluorescent proteins (FPs) are rare examples of biomolecules that display strong fluorescence in the visible spectrum [50], suitable for single molecule studies. It is possible to use proteins, such as G(reen)FP as intracellular biomarkers, simply by attaching the GFP gene to the gene of interest through genetic engineering and taking the advantage of protein expression by the cell machinery. These molecular probes are proved to be extremely useful in various applications such as \textit{in-vivo} oxygen imaging [51], quantitative monitoring of gene expression [52], intracellular pH Measurements [53] and visualizing subcellular localization of proteins [54]. In chapter 2 we used FCS to study the local concentration and diffusion properties of the enhanced GFP (EGFP) within the nucleus of live Drosophila cells.

However, fluorescent proteins are exceptions and for many applications the common way to study biological molecules by fluorescence is to tag them with organic fluorophores such as Xanthene, Rhodamine, Cyanine or Alexa dyes [1, 55]. The most common labeling targets are the free amine groups of residues like lysine and the sulfhydryl group of cysteine [55] (figure 1-9).

![Figure 1-9 Amine-reactive N-hydroxysuccinimide ester (NHS ester) and thiol-maleimide cross-linking chemistries. This figure was kindly provided by Baoxu Liu.](image-url)


### 1.4.2.1 Amine Group Reactive Cross-Linking Chemistry

The most common labeling targets in proteins are the primary amines found in Lysine, Arginine and at the N-terminus of the polypeptide chain. N-Hydroxy Succinimide (NHS) activated probes at pH 7-9 are used to specifically label these sites. Due to the abundance of Lys residues on the surface of proteins, this labeling strategy is very efficient but is not site-specific [55]. The labelling is carried out in phosphate, HEPES or borate buffers, as the common TRIS buffer contains primary amine and therefore competes with the dye for the NHS reaction. High-performance dyes for SMF studies are available commercially in the NHS ester activated form.

SH3 DrkN [56-58] is the protein we study in chapter 4. The WT protein does not contain any Cys residues, therefore I used NHS chemistry to label it (non-specifically) as well as the T22G mutant with Tetramethylrhodamine (TMR)-NHS and Alexa555-NHS dyes. The NHS-amine coupling reaction was performed in PBS buffer adjusted to pH = 8. The reaction is started by addition of NHS-ester-activated fluorophores to a 50 μL solution of 200 μM proteins to obtain a final molar ratio of dye:protein = 1:2. The sample was gently shaken for 3 hours in the dark at room temperature. The remaining free dyes were removed by size exclusion chromatography [59] using Sephadex G-25 gels (G2580, Sigma Aldrich, Oakville, Canada) in a BioLogic LP system (731-8300, Bio-Rad, Mississauga, Canada). A typical labeling efficiency of ~5% was estimated from the absorption spectra of the product and was also confirmed by mass spectrometry.

The low dye:protein ratio (1:2) in this protocol was used on purpose to prevent multiple labeling of the protein. This was done to minimize the degree of perturbation of the host protein and to avoid the self-quenching of the fluorophores, which can affect the analysis of correlation curves (chapter 3). Although the low labeling efficiency ensures the irrelevance of doubly labeled proteins, the nonspecific nature makes it inadequate for FRET studies where precise control on the labeling site is required. Therefore, the amine labeling protocol was only used for FCS studies and for fluorescence lifetime analysis.
1.4.2.2 Cysteine Reactive Cross-Linking Chemistry

Due to the very low occurrence frequency (1-2%) of Cys in protein sequences [60], it has become a popular target for site-specific labeling [1]. Maleimide is a thiol-reactive group which is highly specific and efficient in labeling of exposed cysteines at neutral pH conditions [61].

The doubly-mutated SH3-C2C61 protein was labeled with a FRET pair to perform smFRET experiments in chapter 4. Two thiol-reactive dyes, Bodipy Fluorescein (BFL)-maleimide and Alexa 647 (A647)-maleimide, were used for the site specific labeling of the protein at the sites 2 and 61, close to the two termini. BFL is a neutral dye and it has poor solubility in water. Therefore, 10 μL DMSO was added to the protein solution (50 μL, 100 μM protein, TRIS buffer) to increase the BFL solubility and its labeling efficiency. TCEP was added at 20:1 molar ratio excess compared to the protein in order to reduce the possible disulfide bonds [61]. The labeling reaction was initiated upon addition of BFL at 5:1 molar ratio with the protein. In order to avoid disulfide bond formation, oxygen was removed and the sample was flushed with argon in a vacuum desiccator for 5 min. The vial was capped tightly and the reaction was continued by gentle vortexing for 12 hours at room temperature. A647 was then added in 20x molar excess and the solution was sealed under argon gas. The labelling reaction was allowed to continue for 3 days in the dark at room temperature. The excess dye was removed by size-exclusion chromatography. Labeling efficiencies of ~4% and ~70% for BFL and A647, respectively, were estimated from the absorption spectra of the labeled protein using the protocol provided by the manufacturer.

The cysteine-mutated caseinolytic protease (ClpP E150C) from *E. coli* is used in chapter 3 to extend the correlation analysis time scale of the dark states of rhodamine dyes by nearly one order of magnitude, to about 0.5 ms. The 300 kDa protein consisting of two heptameric rings was mutated, expressed and purified in the Houry lab (Biochemistry, UofT). For correlation analysis, a low labeling efficiency <20% was required in order to avoid multiple labeling at the 14 identical subunits. To this end, 10x excess TMR-maleimide dye was used to label ClpP, which was provided in rather low concentration (~2 μM the stock solution). The reaction proceeded by gentle shaking for 3 hours at room temperature and overnight in 4 °C. The remaining excess dye was removed by size-exclusion chromatography [59] using Sephadex G-50 gels (G5080, Sigma Aldrich, Oakville, Canada).
1.5 Bibliography


Chapter 2

On the Performance of Bioanalytical Fluorescence Correlation Measurements

Abstract

The accuracy of fluorescence correlation spectroscopy (FCS) analysis was systematically investigated under different laser excitation and sample concentration conditions for two representative fluorophores, Rhodamine110 and Enhanced Green Fluorescent Protein (EGFP). Reliable concentrations and diffusion constants were obtained by fitting the experimental FCS curves, only when the excitation intensity did not exceed 20% of the saturation level for each fluorophore. Furthermore, accurate results were obtained for sample concentrations varying from pM to μM range, as well as for conditions of very high background signals.

The versatility of our home-built correlation setup and data analysis capabilities were tested both in vitro and in vivo. In vitro, singlet and triplet lifetimes, rotational and translational diffusion rates of EGFP were measured using only 10 million photons acquired in a single 30-minute experiment. In vivo, the diffusion properties of EGFP in the nucleus of Drosophila cells were measured under conditions of high concentration and molecular crowding. As a bioanalytical application, the binding affinity of a novel peptide-based drug to the cancer-regulating STAT3 protein was measured by FCS and the results were corroborated with fluorescence polarization analysis derived from the same photon dataset.

2.1 Introduction

In recent years, novel instrumentation [1] and data analysis models [2, 3] for FCS have been proposed, while the statistical accuracy of correlation measurements and the impact of possible experimental artifacts have been treated in several studies [4-7]. These are of great importance because the data analysis is based on theoretical approximations of the optical paths and simplifications of the underlying molecular processes. For example, focal volume artifacts caused by non-Gaussian beams, thick coverslips and refractive index mismatch have been measured in experiments and simulated theoretically [6, 7].

In this chapter, I am first focusing on which experimental conditions should be fulfilled in order to obtain reliable physical parameters via FCS. We will see how characteristics of the detection system and phenomena, such as fluorophore saturation and photobleaching, as well as the background level restrict the range of acceptable input parameters, such as excitation intensity and sample concentration, on the FCS experiment design and data analysis.

Secondly, I will demonstrate the performance of FCS measurements on our home-built multiparameter photon-counting microscope in several applications. For example, I will show how a 30-minute experiment can yield a myriad of physico-chemical characteristics of fluorophores in solution, such as singlet and triplet lifetimes and populations, rotational and translational diffusion rates. Finally, I will describe biological applications aimed at measuring the binding affinity of a small-molecule drug to a protein that regulates cancer, and the anomalous nuclear diffusion of EGFP in live Drosophila cells under conditions of high local concentration, high background and molecular crowding.

2.2 Experimental

2.2.1 Materials

Samples were diluted in PBS buffer (pH 7.4) and included Rhodamine110 (Rh110) (20310, Biotium Inc., Hayward, CA) and enhanced green fluorescence protein (EGFP) (BioVision, Mountain View, CA). Signal transducer and activator of transcription 3 (S54-54G, SignalChem, Burlington, Canada) and 5-carboxyfluoresceine(F*)-pYLPQTV-NH₂ (QCP100302-17R, CanPeptide) were diluted in a buffer containing 10 mM HEPES (pH 7.5) and 50 mM NaCl. The
binding experiments were performed 30 minutes after adding various amounts of Stat3 protein to a fixed amount of F*-peptide. The Stat3 concentration was varied from 50 nM to 1 µM in the final mixture, while the concentration of F*-peptide was maintained constant at 12 nM. Drosophila Schneider S2 cells (D.Mel-2, Invitrogen) were imaged in a serum-free medium (CCM3, Fisher Scientific) at 25°C. The plasmid pMG6, containing EGFP gene with the actin 5C promoter, was transiently transfected into S2 cells using the Cellfectin reagent (10362-010, Invitrogen). After twenty-four hours transfection, cells were placed on plasma-cleaned microscope coverslips for confocal imaging and FCS spectroscopy.

Each experiment was performed on a ~50 µL sample solution dropped on plasma-cleaned coverslips. When needed, polyethylene glycol (mPEG-SVA-5000, Laysan Bio. Inc.) was used to create a hydrophilic coating, which ensured low adsorption of the EGFP and Stat3 proteins to the glass surface [8]. A closed chamber was created by inserting an adhesive spacer (S24735, Invitrogen) between two coverslips, in order to prevent sample evaporation when long data acquisition times were required.

### 2.2.2 Instrumentation

Time-resolved fluorescence measurements were performed on a multiparameter confocal microscope that was described in detail in Chapter 1, as well as in earlier publications from the Gradinaru lab [9, 10]. The femtosecond laser (Tsunami HP, Spectra Physics) was tuned to 960 nm to obtain 480-nm pulses after frequency doubling and data acquisition was performed in the T3 mode of the PicoHarp photon counting module (PicoQuant). To measure rotational relaxation of EGFP, the measurements were done in the PicoHarp T2 mode using CW laser excitation at 473 nm (Cobolt Blue). In addition, a broadband polarizer was used to select the parallel polarization of fluorescence and then a broadband half-wave plate was inserted to rotate this polarization by 45° and thus divide the measured signal quasi-equally between two detection channels.

Prior to each experiment, the top surface of the glass coverslip was found by maximizing the backscattering signal while moving the stage along the axial (z-) direction. FCS measurements were conducted after the focal point was moved up to a depth of ca. 5 µm into the liquid sample. Custom LabView codes were developed to control photon data acquisition, construct and visualize intensity time-trajectories and start-stop (lifetime) histograms. Digital (photon-by-
photon) correlation curves were built in real-time by a custom-written LabView code using logarithmic time binning and 24 bins per temporal decade (see Chapter 1).

2.2.3 Procedures

FCS is based on the study of temporal fluctuations of the fluorescence signal ($\delta F$) recorded when fluorescently-labeled molecules diffuse through a small detection volume [3, 11]. As described in detail in Chapter 1 of this thesis, typically, the signal fluctuations arise from rotational and translational diffusion of individual fluorophores through the detection volume and from photophysics such as singlet-state quenching and triplet state blinking. The fluorescence intensity is commonly split into two or more channels and the normalized auto- and cross-correlation functions are defined as:

$$G_{i,j}(\tau) = \frac{\langle \delta F_i(t) \delta F_j(t+\tau) \rangle}{\langle F_i(t) \rangle \langle F_j(t) \rangle}$$

(2-1)

where $i$ and $j$ are detection channels and brackets denote time averaging. Under some simplifying assumptions, the intensity correlation function for free Brownian diffusion of a single molecular species with triplet blinking is given by [11]:

$$G(\tau) = \frac{1}{N_{eff}} \left( 1 + \frac{\tau}{\tau_d} \right)^{-1} \left( 1 + \frac{\tau}{s^2 \tau_d} \right)^{-0.5} \left( 1 + \frac{f_{DS}}{1 - f_{DS}} \exp \left( -\frac{\tau}{\tau_{DS}} \right) \right)$$

(2-2)

In equation (2-2), $N_{eff}$ is the average number of molecules in the effective detection volume, $V_{eff} = \pi^{3/2} z_0 w_0^2$, and it is given by the inverse of the diffusion-related amplitude, $G_0 = N_{eff}^{-1}$.

In addition, $s$ is the structure parameter (the aspect ratio of the axial and the lateral radii of the detection volume, $s = z_0/\omega_0$), $\tau_d$ is the diffusion time (related to the diffusion coefficient $D$ by $\omega_0^2 = 4D \cdot \tau_d$), $\tau_{DS}$ is the triplet lifetime and $f_{DS}$ is the average fraction of molecules in the triplet state.

For samples consisting of a mixture of non-interactive diffusive species, the total correlation function is given by the brightness-weighted sum of the individual correlation functions [3]:

$$G(\tau) = \sum_i N_i^2 \eta_i^2 G_i(\tau) / \left( \sum_i N_i \eta_i \right)^2$$

(2-3)
where $N_i$ is the number of molecules of species $i$ in the detection volume and $\eta_i$ is their specific molecular brightness.

### 2.3 Reliability of FCS analysis

In FCS experiments, important physical parameters of the sample are estimated by fitting the correlation curve to an appropriate model. Accurate results can be obtained only if the experimental conditions satisfy some key assumptions of the theoretical model behind eq. (2), such as the 3D Gaussian profile of the detection volume [3]. In Chapter 1 I discussed that for the combination of the oil objective, the coverslip thickness and confocal pinhole diameter used in our setup, a depth of 5-7 μm in solution is optimal for signal collection and for a stable FCS curve lacking focal spot distortion artifacts. However, there exist also restrictions imposed by the photophysical properties of the fluorophore and the properties of the photon detection hardware. For a reliable FCS analysis not only a flawless beam geometry is essential, but also experimental design parameters such as laser intensity, sample concentration and background level should be set carefully and controlled accordingly.

#### 2.3.1 Excitation power conditions

In literature, the laser excitation intensity used for FCS studies varies widely, although it is often set to relatively high values in order to acquire more fluorescence photons in the same time frame and presumably lead to higher quality correlation curves. However, the photobleaching and the saturation of absorption/emission of the fluorophore can seriously influence the results and they must be carefully considered when estimating the optimal excitation conditions for FCS experiments.

##### 2.3.1.1 Saturation

To investigate the effect of the excitation power on the correlation analysis, we start by considering the dependence of the rate of single-molecule fluorescence on the excitation intensity. Assuming a single-emitting species at infinite dilution, the molecular brightness (counts per molecule per second) is given by $\eta = \langle F \rangle / N_{\text{eff}}$. For a sample excited by a pulsed laser with a repetition time $T_0$ that is much longer than the relaxation time of the first singlet excited state, $\tau_s$, $\eta$ depends on the excitation intensity $I$ according to the formula [12]:
For a Gaussian beam of (time-averaged) power $P$ and beam waist $w_0$, the time- and space-averaged excitation intensity at the focus is $\bar{I} = P / \pi w_0^2$. The average intensity in this equation is further divided by 2 to account for the non-uniform distributions of the excitation beam and of the detection efficiency (Appendix A). $\varepsilon_{\text{avg}}$ is the spatially-averaged photon collection and detection efficiency, $\Phi_F$ is the quantum yield of fluorescence and $\Phi_{\text{ISC}}$ is the quantum yield of intersystem crossing (ISC). $\tau_t$ is the triplet lifetime, $\kappa$ is a numerical parameter related to stimulated emission and $I_s$ is the saturation intensity.

\[
\eta(I) = \frac{\varepsilon_{\text{avg}} \cdot \Phi_F \cdot \frac{1 - \exp\left(-\bar{I}/2I_s\right)}{\kappa + \frac{\tau_t \cdot \Phi_{\text{ISC}}}{I_0} \cdot \left[1 - \exp\left(-\bar{I}/2I_s\right)\right]}}{T_0} \tag{2-4}
\]

**Figure 2-1.** The average number of photons per molecule ($n$) vs. excitation intensity ($I$). Data points for EGFP (green circles) and Rh110 (blue solid squares) calculated with $\tau_d$ and $N_{\text{eff}}$ obtained by FCS in the low-intensity limit; red curves are best fits using equation (2-4); open blue squares are data points for Rh110 estimated with $\tau_d$ and $N_{\text{eff}}$ obtained by FCS at each excitation intensity. The dashed vertical line indicates the saturation intensity for Rh110 (see text for details).

Figure 2-1 shows the average number of photons $n$ detected per molecule for freely diffusing Rh110 (10 nM solution) and EGFP (14 nM solution) illuminated at increasingly higher intensities. This quantity is given in counts per molecule (cpm) and is calculated as: $n = \eta \cdot \tau_d = \langle F \rangle \cdot \tau_d / N_{\text{eff}}$. FCS experiments at low excitation power ($I \sim 1 \text{kW} \cdot \text{cm}^{-2}$) are used to determine $N_{\text{eff}}$ and $\tau_d$ for each fluorophore with high accuracy. These values were then used to calculate $n$ based on the average fluorescence count rate $\langle F \rangle$ recorded at different excitation...
2.3 Reliability of FCS analysis

powers. On the other hand, \( n \) and \( \eta \) are related by a multiplication factor, \( \tau_d \) (41 \( \mu \)s for Rh110, 227 \( \mu \)s for EGFP), which implies that \( n(I) \) can be fitted to the equation (2-4), as shown in figure 2-1. Laser pulses at the sample are longer (~500 fs) than vibrational relaxation times of typical fluorophores (~100 fs) [13], so that \( \kappa \approx 1 \) [12]. The fitting results obtained using \( \Phi_{F(Rh110)} = 0.91 \) [14] and \( \Phi_{F(EGFP)} = 0.6 \) [15] are listed in table 2-1.

\[ \text{Table 2-1. Fluorescence saturation parameters for Rh110 and EGFP.}^a \]

<table>
<thead>
<tr>
<th>Sample</th>
<th>( I_s ) (kW cm(^{-2}))</th>
<th>( \Phi_{ISC} \cdot \tau_t ) (ns)</th>
<th>( \varepsilon_{avg} ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhodamine 110</td>
<td>20.9 ± 2.3</td>
<td>4.8 ± 2.1</td>
<td>0.73 ± 0.09</td>
</tr>
<tr>
<td>EGFP</td>
<td>23.8 ± 12.2</td>
<td>46.3 ± 29.9</td>
<td>0.93 ± 0.48</td>
</tr>
</tbody>
</table>

\(^a\) Obtained by fitting the fluorescence saturation data to Eq. (2-4), scaled by the diffusion time estimated by FCS in the low-excitation intensity limit.

The saturation intensities measured for the two fluorophores are: \( I_s(Rh110) = 20.9 \text{ kW}\cdot\text{cm}^{-2} \) and \( I_s(EGFP) = 23.8 \text{ kW}\cdot\text{cm}^{-2} \). The EGFP value is in agreement to that obtained previously using CW excitation, 19 kW\cdot\text{cm}^{-2} [16]. The triplet state of Rh110 has a lifetime \( \tau_t = 10 \pm 2 \mu \text{s} \), as measured by FCS and \( \Phi_{ISC} \cdot \tau_t = 4.8 \pm 2.1 \text{ ns} \) (table 2-1), which leads to a very low intersystem crossing yield \( \Phi_{ISC} = (4.8 \pm 2.3) \times 10^{-4} \), similar to \( \Phi_{ISC} = (5.5 \pm 2.0) \times 10^{-4} \) reported by Kunavin et. al.[17]. Similarly, \( \tau_t = 52.5 \pm 10.5 \mu \text{s} \) and \( \Phi_{ISC} = (8.8 \pm 5.9) \times 10^{-4} \) were obtained from the EGFP data.

2.3.1.2 Photobleaching

The number of fluorescence cycles that occurs for a fluorophore before photobleaching depends on both its photophysics and exposure time. Rhodamine dyes, known for their photostability, are reported to have a probability of photobleaching in the order of \( 10^{-6} \) for intensities below 1 kW\cdot\text{cm}^{-2} [18]. This probability increases considerably for higher intensities. Typical organic dyes emit >100,000 photons before photobleaching occurs [19].

In order to assess the extent of photobleaching we estimated the average number of total fluorescence photons a rhodamine 110 dye emits while it is in the detection volume. The average detection efficiency estimated using the rhodamine dye is \( \varepsilon_{avg} = 0.73 \) (table 2-1). Assuming a Gaussian detection profile, this leads to a peak detection efficiency \( \varepsilon_{max} = 2.3\% \) when the emitter lies at the center of the confocal volume (Appendix A). In the high intensity limit, on average,
around 16 photons are detected from a single fluorophore (figure 2-1). Taking into account the estimated detection efficiency, each molecule emits ca. 2200 photons while passing through the confocal volume, well below the photobleaching limit of rhodamine dyes. Note that measurements involving less stable fluorophores that reside a relatively long time in the detection volume, due for instance to higher viscosity in cells, need extra care (see 2.4.3).

2.3.1.3 Optimum excitation intensity for FCS

To determine the optimal excitation conditions for FCS in the context of optical saturation, experiments were performed at various laser powers on a 10 nM Rh110 solution. The correlation curves were fitted with the function described by equation (2-2) with only one dark state. The results, i.e. the relative concentration $N_{eff}$ and the diffusion time $\tau_d$ are plotted against the excitation intensity (figure 2-2).

![Figure 2-2. FCS-derived parameters vs. excitation intensity: $N_{eff}$ – red squares, $\tau_d$ – blue circles. $N_{eff}$ values are normalized to the value obtained at $I = 1.16$ kW/cm$^2$. Error bars are based on the variance calculated from multiple measurements. Vertical line indicates the saturation intensity for Rh110.](image)

The saturation of fluorescence manifests itself in FCS as a shift to shorter decay times (smaller $\tau_d$) and a reduced correlation amplitude (larger $N_{eff}$). This can be caused by a deformation of the detection volume, a relative increase of background photons and/or by changes in the properties of the dark states of the fluorophore. Triplet state kinetics and translational diffusion are competing terms in the fitting model, therefore an increase of the uncertainty of the triplet parameters directly affects the estimation of $\tau_d$ and $N_{eff}$. The diffusion time was less sensitive, with a maximum deviation of $\sim$ 15% at the highest excitation intensity.
used in our study. However, significant deviations of the FCS-estimated $N_{\text{eff}}$ from the theoretical (low-intensity) value were observed at and above the saturation intensity. For example, at $I = I_s(\text{Rh110}) = 20.9 \text{ kW cm}^{-2}$, the correlation analysis overestimated $N_{\text{eff}}$ by about 50%.

At each excitation intensity, we used the FCS fitting results and the experimental fluorescence count rate $<F>$ to estimate the average number of photons detected per molecule, $n$ (open blue squares, figure 2-1). Compared to the data points calculated using the correct $\tau_d$ and $N_{\text{eff}}$ values derived under low-power conditions (filled blue squares, figure 2-1), deviations above 10% were observed at intensities above $4.2 \text{ kW cm}^{-2}$. This suggests that in order to avoid saturation effects in FCS, the experiments should be performed at excitation intensities $I \leq 0.2I_s$. For Rh110 the upper limit corresponds to an average laser power at the sample of 6.2 $\mu$W and to an average of $n \approx 2$ photons detected per molecule. In general, as a prerequisite for accurate FCS results one should set the excitation intensity at a value not exceeding 20% the saturation level. This means that for most commonly used fluorophores, the laser power should not exceed 10 $\mu$W at the sample. This limitation is independent of the sample concentration, but it is more critical at low concentrations, for which the signal is intrinsically low.

### 2.3.2 Background filtering

The influence of various background contributions on FCS results was investigated with the excitation intensity set below 20% of the saturation value. In a live cell, the background consists of autofluorescence from endogenous fluorophores [20] and Raman scattering of water. In order to simulate these conditions, either ambient light or scattered light from the sample was allowed to reach the detectors. Figure 2-3 shows FCS fitting results obtained from a 1 nM aqueous solution of Rh110 illuminated at an average power of 3 $\mu$W.

Correlation curves were built from all photons detected under various levels of background ambient light and fitted w/wo the correction factor given by eq. 1-3. Upon correction, the error of estimating $N_{\text{eff}}$ is within 10% for fluorescence signals as low as 50% above the background level. As expected, background photons are uncorrelated and $\tau_d$ is rather independent of the background level. In our setup, a fluorescence signal of about 2200 counts-per-second (cps) is typically detected from a 1 nM sample of Rh110 upon 480-nm excitation at 3 $\mu$W ($2 \text{ kW cm}^{-2}$). Under ideal conditions, the only contribution to the background is the detector dark count rate, $B_{\text{min}} \approx 50$.
2.3 Reliability of FCS analysis

cps. Based on the results shown in figure 2-3, FCS measurements on our setup are expected to be accurate at sample concentrations as low as 40 pM (see also below).

![Figure 2-3](image.png)

**Figure 2-3.** FCS fitting results ($N_{\text{eff}}$ and $\tau_d$ (inset)) vs. background levels. $N_{\text{eff}}$ and $\tau_d$ are normalized to the values obtained at the highest fluorescence-to-background ratio (F/B). Red squares – no correction; green circles – background corrections. Error bars are based on the variance of multiple measurements.

In certain experiments, due to overlapping spectra, scattering components cannot be removed without a significant reduction of the fluorescence signal. Pulsed excitation and time-correlated single-photon counting detection uniquely enables the efficient removal of Raman/Rayleigh scattering photons using software gating on the picosecond time scale [21]. The approach relies on the fact that the scattering photons arrive very early after the excitation laser pulse, so that most of them pile up in the start-stop time histogram within the width of the instrument response function of the APD detector.

To investigate the benefits of time-gating, FCS measurements were performed on a 5 nM Rh110 sample w/wo spectral filtering. Spectral filtering was done using two blocking filters (LP-488 Semrock, and BP-530/50 Chroma) in the emission path. With the filters removed, the photons having a start-stop time < 0.8 ns were either discarded (time-gating) or not (raw data) when calculating the correlation curve (figure 2-4). A summary of FCS fitting results for the three cases considered is given in table 2-2.
2.3 Reliability of FCS analysis

Table 2-2. Spectral and temporal filtering of the background signal in FCS.

<table>
<thead>
<tr>
<th></th>
<th>F/B</th>
<th>(N_{\text{eff}})</th>
<th>(\tau_d) (μs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw data</td>
<td>0.34</td>
<td>19.5 ± 2.0</td>
<td>98.2 ± 12.6</td>
</tr>
<tr>
<td>spectral filtering</td>
<td>1.2×10^4</td>
<td>1.9 ± 0.1</td>
<td>68.4 ± 1.3</td>
</tr>
<tr>
<td>time gating</td>
<td>9.7</td>
<td>2.5 ± 0.1</td>
<td>74.4 ± 2.8</td>
</tr>
</tbody>
</table>

*Data measured on a solution of 5 nM Rh110. The background signal was measured from a water sample under identical experimental conditions.*

The results clearly indicate that the offline photon selection based on their start-stop time leads to correct fluorescence correlation curves and robust fitting output even for data acquired at dominant background levels. Such conditions \((F/B = 0.34)\) lead to large overestimations of \(N_{\text{eff}}\) (ca. 10 times!) and to a lesser extent of \(\tau_d\) (ca. 45%). Using time-gating, the errors are reduced to within 30% for \(N_{\text{eff}}\) and 10% for \(\tau_d\). Pyridinic and flavin coenzymes are the main endogenous fluorophores in the cell and they have typically short fluorescence lifetimes (< 0.5 ns) relative to common dyes [20]. The capability of selecting the desired fluorescence signals in a complex environment is crucial for accurate data analysis.
environment by time-gating or more advanced lifetime filtering is of great advantage for quantitative FCS studies in living cells.

### 2.3.3 Sample concentration

The size of most eukaryotic cells is on the order of 10 to 100 microns. Many enzyme proteins are expressed at levels of thousands of copies or more per cell, equivalent to average concentrations in the order of 100 nM or higher. For instance, most of the cell metabolites are found in micromolar concentrations [22]. On the other hand, it is well known that there is an optimal sample concentration for correlation analysis, corresponding to an average of 1-10 molecules in the detection volume. This means that, for standard confocal FCS using high numerical aperture objectives, the optimum concentration lies in the range of 1-10 nM. To extend the use of the technique at higher concentrations, new experimental designs were proposed based on the reduction of the detection volume [23]. Here we focus on understanding and mitigating the sources of poor FCS signal quality obtained under non-optimal sample concentrations, in order to allow us to perform FCS measurements even in the μM range.

**Figure 2-5.** The dependence of $N_{\text{eff}}$ and $\tau_d$ on the sample concentration. The results are normalized relative to the results obtained for 0.5 nM Rh110 in PBS, 5 nM Rh110 in PBS glycerol and 0.7 nM EGFP in PBS, respectively. Error bars are based on the variance calculated from multiple measurements.

FCS experiments were conducted at different fluorophore concentrations and solvent viscosities. The dispersion of the fitting results ($N_{\text{eff}}$ and $\tau_d$) around the expected values is shown in figure 2-5. In each case, the laser excitation intensity was set below the saturation threshold. To avoid instrumental distortions caused by the nonlinearity of the time-to-amplitude convertor and the dead time of the
APD detectors [4], the measured fluorescence count rates were maintained under 250 kcps. To that end, the excitation intensity was gradually decreased in the high concentration range (> 75 nM). In the low concentration regime (< 0.5 nM) the count rate was relatively low (≤ 1 kcps) and photons were acquired for longer times, up to one hour per sample. In a typical experiment, about $10^7$ photons were acquired for calculating the correlation curve, and for each concentration at least two separate experiments were conducted.

Our data show that local concentrations and diffusion rates can be reliably measured by FCS on our custom-built setup for sample concentrations spanning five orders of magnitude. In the sub-nanomolar regime, $N_{\text{eff}}$ was estimated using the corrective background factor (equation 1.3). In agreement with conclusions from the previous section, concentrations as low as 37 pM were estimated within 30% error, whereas $\tau_d$ was estimated within ~5% error. This suggests that increased detection efficiency and lower dark count rate for the photon detectors coupled with offline background filtering can deliver accurate FCS results from picomolar samples with a practically short acquisition time.

At the other end of the spectrum, the FCS fitting parameters of dyes in aqueous solutions exhibit large uncertainty at concentrations of 100 nM or larger. The uncertainty arises from the large noise amplitude in the correlation curve. In the high concentration regime ($N_{\text{eff}} >> 1$), the $S/N$ ratio of the correlation curve, $G(\tau)/\sigma(G(\tau))$, has the form [4, 24]:

$$S/N = \eta \left( t_b U / 2 \right)^{1/2}$$

(2-5)

where $\eta$ is the molecular brightness, $U$ is the total data acquisition time and $t_b$ is the correlation binning time. Accordingly, for two different fluorophores measured under identical conditions (concentration, excitation power, etc) one should expect the best quality of the correlation data for the brightest one. However, the underlying assumption in (eq. 2-5) is that $t_b$ is constant, whereas we use a variable (logarithmic) binning scheme. In addition, the $S/N$ ratio defined by (eq. 2-5) is calculated at $\tau = 0$ and was found to decrease for longer delays [24], therefore it should be regarded as a high-limit value when considering the overall quality of the FCS curve.

To elucidate the physical factors that determine the sample concentration limits on our setup, we performed experiments on Rh110 dissolved in a 70% (v/v) glycerol solution. Changing viscosity provides a direct control on the average number of photons detected per molecule (cpm), $n = \tau_d \times \eta$. We can assume that the molecular brightness $\eta$ remains unchanged in glycerol
and in water, however $\tau_d = 1.3$ ms in 70% v/v glycerol, approximately 30 times larger than the value in water ($\tau_d = 41$ µs). Thus, $n$ appears to increase linearly with the viscosity of the solvent, while $\eta$ stays constant, at least in the first order approximation. Then it seems surprising that at higher concentrations, the FCS parameters are estimated with higher accuracy in glycerol than in water (figure 2-5). According to (eq. 2-5), the quality of the FCS curve should be very similar, because the average molecular brightness is very similar between the two solvents.

In addition, based on the data presented in figure 2-1, it can be estimated that $\eta_{EGFP} \approx 0.64 \cdot \eta_{Rh110}$, which would imply lower $S/N$ and larger error bars in the FCS fit for EGFP than for Rh110 in aqueous solvent. In contrast, the fitting results at high concentrations are superior for EGFP compared to Rh110 (figure 2-5). This supports the idea that, in this regime, the number of photons detected per molecule $n$ is more relevant than the molecular brightness $\eta$ for the quality of the correlation curve. Accurate parameter estimation at (sub)micromolar concentrations is important for FCS measurements performed in the crowded and viscous medium inside living cells (see section 2.4.3).

FCS experiments at pM concentrations, which are important for high throughput genomic and proteomic assays, are mainly limited by the dark signal of the detectors. At µM concentrations, which are relevant for live cell studies, the major restriction is imposed by the photon counting hardware, which limits the total count rate and consequently the number of photons detected from each molecule $n$. However, in more viscous solvents $n$ increases and thus more concentrated samples can be measured by FCS. On our setup, the maximum concentration for reliable FCS analysis increase from around 100 nM in water to 2-3 µM in 70% (v/v) glycerol, at a viscosity resembling that of the cell interior [25-27]. Note that according to equation 2-5 both concentration limits can be somewhat extended by longer acquisition times. For the purpose of this study, the duration of each FCS experiment did not exceed one hour.

2.4 FCS performance in various applications

The strength of FCS analysis in time-resolving bio-chemical processes is minimally affected by the photophysical changes of the fluorescence labels. Therefore, unlike other fluorescence techniques, processes modulating the signal such as nonspecific binding, static and collisional quenching, etc., can be filtered out in FCS analysis. The aim of this section is to show how the
controls developed and discussed above are used to perform more informative and reliable analysis of FCS data measured on "real biological systems" such as proteins and live cells.

2.4.1 Fast time scales: rotational diffusion and anti-bunching

The rotational correlation time ($\tau_{rot}$) is proportional to the cube of the hydrodynamic radius of the molecule (eq. 1-8), and so it is more sensitive than the translational diffusion time ($\tau_d$) to changes caused by structural dynamics and chemical reactions undergone by macromolecules. Time Resolved Fluorescence Anisotropy (TRFA) is the technique of choice to measure the rotational dynamics via fluorescence assays. However, TRFA analysis is less reliable in the presence of quenching or if rotational relaxation is much slower than the excited-state lifetime. In contrast, the rotational dynamics appears quite distinctly in the correlation curve. Our setup enables a direct and instant comparison of rotational diffusion measured by the two methods.

FCS experiments were performed on a 2 nM solution of EGFP in a viscous PBS buffer (pH 7.4, 50% w/w glycerol) using the T2 mode on PicoHarp and CW excitation at 473 nm. The parameters $R_{1,2}$ in eq. 1-7 depend on the experimental scheme and the degree of polarization of the dye [28]. The highest contrast between rotational and translational components is achievable when both channels detect photons of polarization parallel to the polarization of the excitation (the xxx mode). For this configuration $R_2 / R_1$ was calculated to be $\approx 0.09$ [28, 29] and this value was fixed in the fitting of the correlation curve. Figure 2-6 shows the correlation curve calculated from data acquired over a 1-hour interval, consisting of about $10^7$ photons. The red solid line is a fit of the data using the following model function:

$$G = G_{DS(3)} G_{DS(2)} G_{DS(1)} G_{diff} G_{rot} G_{ab}$$  \hspace{1cm} (2-6)

Two additional exponential terms ($G_{DS(2)}$ and $G_{DS(3)}$) were needed for improving the fitting in the range above 100 μs and below 1 μs. In the fit, the ratio $R_2 / R_1$ and the geometrical factors $\omega_0$ and $s$ were the only fixed parameters, while the other parameters in eq. (2-6) were fitted to the values given in table 2-3. Quite remarkably given the noise level in the curve, the anti-bunching component was fitted to a lifetime of $\tau_f=2.71 \pm 0.50$ ns, in excellent agreement with $\tau_f=2.62 \pm 0.02$ ns measured separately by time-correlated single-photon counting.
Figure 2-6. Fluorescence correlation measurement of 2 nM EGFP in PBS (pH 7.4, 50% w/w glycerol) using CW excitation at 473 nm. The red line is a fit by the complete model (eq. 1-2) including anti-bunching and rotational relaxation.

Table 2-3. Fitting results of the EGFP polarized-FCS data (figure 2-6).

<table>
<thead>
<tr>
<th>Neff</th>
<th>τ_d (µs)</th>
<th>R₁</th>
<th>τ_{rot} (ns)</th>
<th>τ_f (ns)</th>
<th>τ_{DS} (µs)</th>
<th>f_{DS}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.38±0.02</td>
<td>1530±78</td>
<td>0.95±0.10</td>
<td>149±20</td>
<td>2.71±0.50</td>
<td>22.5±7.3</td>
</tr>
<tr>
<td></td>
<td>382±74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.27±0.03</td>
</tr>
</tbody>
</table>

The fitting yields three dark states for EGFP in the glycerol buffer, with lifetimes of 382 µs, 22.5 µs and 1 µs. The presence of two dark states for EGFP with lifetimes in the range of 10-1000 µs has been observed in several studies [30-32]. A dark state with a lifetime in order of 100 µs or more was found to be related to protonation of the protein and therefore pH-sensitive [31]. This state was not observed in the aqueous buffer probably because the lifetime is similar or longer compared to the translational diffusion time. At low excitation intensities, resembling our experimental conditions (I~1 kW/cm²), Widengren et. al. observed an intensity-independent exponential term with a lifetime of about 1 µs [33].

The translational diffusion time of EGFP in the glycerol buffer, τ₅₉ = 1.53 ms, is 9 times longer than in the aqueous buffer, τ₅₉ = 170.1 µs (using CW excitation). Similarly, the rotational correlation time of EGFP in glycerol τ_{rot} = 149 ns, is 8.7 times larger than 17.2 ns, the value
measured for EGFP in aqueous buffer by TRFA (see below). These observations are consistent with the linear dependence of both translational and rotational times on the dynamic viscosity $\nu$, according to equations 1.4 and 1.8.

TRFA measurements on EGFP were performed in an aqueous solution and in 50% w/w glycerol (figure 2-7). Due to instrumental limitations, the anisotropy decay was only acquired and fitted for a 12 ns range, which is much shorter than the time scale of rotational diffusion, particularly in glycerol. The curves were satisfactorily fitted using mono-exponential decay functions. The width of the instrument response function (~100 fs) was much smaller than the rotational diffusion time so that the rise in the TRFA curves was neglected in this analysis.

![TRFA data of EGFP in PBS buffer without (squares) and with (triangles) 50% w/w glycerol. The solid lines are the mono-exponential fits. The weighted residuals of the fit for the glycerol data are shown in the lower panel.](image)

**Figure 2-7.** TRFA data of EGFP in PBS buffer without (squares) and with (triangles) 50% w/w glycerol. The solid lines are the mono-exponential fits. The weighted residuals of the fit for the glycerol data are shown in the lower panel.

The EGFP anisotropy decay times in water and glycerol were fitted to $\tau_w = 17.2 \pm 2.9$ ns and $\tau_g = 84.4 \pm 3.8$ ns, respectively. The decay time measured in the aqueous buffer is in agreement with literature values, e.g., $\tau_w \approx 16$ ns [34]. However, the 84 ns anisotropy decay time is somewhat smaller than the rotational correlation time (149 ns, by FCS). This deviation can be attributed to the relatively short fitting window in TRFA and deviations from linear polarization of the excitation beam produced by the high numerical aperture microscope objective [35].
2.4.2 Bioanalytical studies: protein-ligand binding

Fluorescence anisotropy (FA) is a standard fluorescence technique for the measurement of binding affinities. However, FA may not produce reliable results when the binding is accompanied by fluorescence quenching, as observed in a protein inhibition assay using organometallic complexes [36, 37]. In contrast, the FCS analysis tracks the change of diffusion upon binding and is not affected by quenching [36]. We will also show that, unlike the FA assay, FCS analysis can determine the active fraction of binding partners. As a proof of principle, we studied a ligand-protein binding assay under non-quenching condition with both FA and FCS. It is expected that both techniques result in similar dissociation constants. This way, in addition to the direct comparison of the techniques we show the ability of one-color FCS to resolve different diffusion components by titration of ligand-protein binding.

Signal transducer and activator of transcription 3 (Stat3) is a 90 kDa transcription factor whose constitutive activation plays a key role in the onset of cancers such as lymphoma and leukemia [38]. Stat3 in its homodimer active form shuttles from the cytoplasm to the nucleus and causes over-transcription of anti-apoptotic genes. Stat3 homodimers are formed by phosphorylation of the tyrosine residues in SH2 domains. This oncogenic process can be suppressed using SH2-domain binders to prevent the formation of homodimers. The ligand is a 5-carboxyfluorescein-labelled phosphopeptide (F*-peptide), which is known to bind strongly to the SH2 domain of Stat3 [38]. The kinetic equation of the binding reaction is:

\[
[\text{pep}]+[\text{Stat3}] \xrightleftharpoons[k_+]{k_-} [\text{pep-Stat3}]
\]

(2-7)

The fraction of peptide bound to protein, \( f_b \), can be written as:

\[
f_b = \frac{[\text{pep-Stat3}]}{[\text{pep}]_t} = \frac{f[\text{Stat3}]}{K_d+[\text{Stat3}]}
\]

(2-8)

where the subscript \( t \) refers to total concentration and \( K_d \) is the dissociation constant, equal to the ratio between the backward and the forward rates. The peptide can be inactivated by the hydrolysis of the phosphate group on the tyrosine residue, thus only a fraction \( f \) is active for protein binding. In order to measure \( K_d \) by titration of [Stat3]t, at each concentration point both [pep-Stat3] and [Stat3] need to be measured. F*-peptide concentration was maintained constant at 12 nM. This value was chosen to keep [pep-Stat3] and [pep] in the nM range, which is optimal.
for FCS experiments in aqueous solutions. Furthermore, choosing \([\text{pep}]_t \ll K_d\) ensures that free \([\text{Stat}3]\) is always much greater than \([\text{pep-Stat}3]\) and hence \([\text{Stat}3] \approx [\text{Stat}3]_t\).

The free and the Stat3-bound F*-peptide are the only fluorescent species in the solution. The Stat3-bound peptide corresponds to a slow-diffusion component whose concentration increases as Stat3 is added to the solution. This leads to a shift of the measured FCS curves to longer diffusion times at higher protein concentration (figure 2-8). For statistical purposes, at each titration point several sequential acquisitions were performed, each consisting of around 10 million photons.

![Figure 2-8](image)

**Figure 2-8.** Normalized FCS curves measuring the binding between a fluoresceinated peptide and a large protein (Stat3). The peptide concentration was maintained constant at 12 nM. The Stat3 concentration for each curve is given in the legend (in nM).

The FCS data for ligand-protein binding can be analyzed with a two-component diffusion model. Assuming that chemical relaxation is much slower than diffusion, the two diffusion times are those of the free and the bound state of the dye-labeled ligand [39]. Then the two species can be treated as nearly independent and the FCS curve can be fitted using equation (2-3). The concentration of Stat3-peptide found by correlation analysis is plotted against the Stat3 concentration (figure 2-9). These data points are then fitted to equation 2-8 to provide the dissociation constant \(K_d = 285 \pm 84\) nM and the fraction of active peptide \(f = 0.86 \pm 0.07\).

FA analysis was performed simultaneously on the same photon dataset used for FCS. Anisotropy is an additive quantity and therefore can be written in terms of free and bound peptide fractions, \(f_f\) and \(f_b\):
2.4 FCS performance in various applications

\[ r = f/r_{\text{min}} + f_b/r_{\text{max}} = \frac{K_d}{K_d + [\text{Stat3}]} r_{\text{min}} + \frac{[\text{Stat3}]}{K_d + [\text{Stat3}]} r_{\text{max}} \]  

(2-9)

where \( r_{\text{min}} \) is the anisotropy of the dye-peptide construct and \( r_{\text{max}} \) that of the dye-peptide-protein complex.

\[ K_d = 209 \pm 35 \text{ nM} \]

Figure 2-9. Titration of protein-ligand binding by FCS (solid squares), fitted to a curve (red) described by eq. 2-8. The same titration by FA (hollow squares) built from the same photons used for FCS and fitted to a curve (green) described by eq. 2-9.

Figure 2-9 shows the total anisotropy measured at each titration point. The dissociation constant \( K_d \) was found by fitting this data to equation 2-9, with \( r_{\text{min}} \) fixed to the value for the pure peptide solution. The model in equation 2-9 assumes that all the peptide is active for binding. However, if only a fraction \( f \) is active, then it can be shown that \( r_{\text{max}} \) should be replaced by: \( r_{\text{max}} - (1 - f)(r_{\text{max}} - r_{\text{min}}) \), whereas \( K_d \) is unaffected. Therefore, to measure \( f \) from anisotropy data, \( r_{\text{max}} \) should be determined in a separate measurement.

Note that both assays, FCS and anisotropy, were performed simultaneously and yielded similar values for \( K_d \) of peptide-Stat3 interaction, which are comparable to previously published results [38, 40]. Even though a fluorescence polarization assay may be an easier and more straightforward method for quantifying ligand-protein binding [41], it does not provide critical information about the absolute concentrations of reactants and the partially active ligands, which FCS analysis can provide instead.
2.4.3 Mobility of EGFP in the cell nucleus

In order to perform FCS measurements in vivo, EGFP was expressed in Drosophila S2 cells. An S2 cell has a typical size of 10-20 μm and a volume that is ca. $10^3$ larger than our detection volume. This ensures negligible confinement effects and helps mitigate, although not eliminate (see below), the impact of probe photobleaching in FCS measurements. Minimum autofluorescence in non-transfected cells was detected at the nucleus (~4 kcps at 0.27 kW/cm$^2$ excitation). This corresponds to the fluorescence of ~20 nM EGFP and it is about one order of magnitude dimmer than the signal recorded in the cytoplasm or the Golgi complex [42]. The cells were spontaneously attached to the surface upon 15-minute incubation of the cell solution on a plasma-cleaned glass coverslip. Wide-field imaging confirmed that cells were firmly tethered to the surface and also that there was significant intracellular mobility associated with different cellular compartments.

![Confocal image of the distribution of EGFP in Drosophila S2 cell. The large region in red showing high fluorescence is the nucleus.](image)

**Figure 2-10.** Confocal image of the distribution of EGFP in Drosophila S2 cell. The large region in red showing high fluorescence is the nucleus. Point spread function deconvolution of the raw image was done using ImageJ software.

Figure 2-10 shows a confocal fluorescence image of a typical cell after a 24-hour transfection cycle. Surprisingly, a large fraction of EGFP is localized inside the cell nucleus. Seibel et. al. has reported similar behavior of EGFP transfected in other types of cells [43]. According to their explanation, monomeric EGFP diffuses freely through the pores in the nuclear membrane, but EGFP multimers are formed inside the nucleus which become trapped and accumulate. We performed FCS of EGFP in S2 cells to demonstrate the capability of in vivo
measurements of high protein concentrations and to study the aggregation and the diffusion behavior of EGFP inside the nucleus.

Photobleaching is an important factor that must be considered for *in vivo* FCS measurements. The presence of an immobile fraction of fluorescence components and/or non-photostable species in the detection volume manifests itself as a fast initial drop in the fluorescence count rate, followed by a relatively flat plateau. In our nuclear FCS measurements in S2 cells we did not face this problem; however the fluorescence intensity decreased by ~15% in a 60-second acquisition interval, at an excitation intensity of 0.27 kW/cm$^2$. FCS measurements were performed under various intensity conditions, this value being chosen as a good trade-off between minimal photobleaching and maximum number of photons detected per diffusive EGFP molecule.

![Figure 2-11. EGFP FCS data measured at a location inside the S2 cell nucleus: (blue) 1-second correlation time; (red) 0.1-second correlation time; (black) fitting using the obstructed diffusion model. The inset shows the standard deviation (STD) for the two cases.](image)

The signature of moderate photobleaching in FCS curves is a clear positive offset of $G(\tau)$ and a sharp drop at the large time scales [44]. Figure 2-11 shows two different FCS curves constructed from the same nuclear fluorescence data, spanning 100 ms and 1 second, respectively. In both cases, a 50-second photon acquisition set is divided up in ten 5-second sets, which were then used to construct the overall FCS curve and estimate its standard deviation (STD, inset Figure 2-11). The late decay associated with sample photobleaching appears in the 1-second FCS curve at delays longer than 100 ms. Photobleaching causes a different $G_\infty$ and
therefore larger than expected STD values for $\tau > 100$ $\mu$s, which, used as weighting factor in fitting, can yield large errors. Therefore, in our analysis we have restricted the delay time to 100 ms (red curves) to limit the influence of photobleaching effects in the correlation analysis. Also, it turns out that this time scale is sufficient to resolve the diffusion of EGFP in the nucleus.

Table 2-4. FCS fitting parameters of EGFP diffusion in the nucleus of S2 cells.

<table>
<thead>
<tr>
<th></th>
<th>Conc. (nM)</th>
<th>$\tau_d$ (ms)</th>
<th>$\alpha$</th>
<th>Relative brightness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non global</td>
<td>1016±109</td>
<td>5.44 ± 1.09</td>
<td>0.83 ± 0.11</td>
<td>1.03</td>
</tr>
<tr>
<td>Global fitting</td>
<td>993</td>
<td>5.02 ± 0.81</td>
<td>0.80 ± 0.07</td>
<td>0.98</td>
</tr>
</tbody>
</table>

A model including anomalous diffusion in which $\tau/\tau_d$ is substituted by $(\tau/\tau_d)^\alpha$ in $G_{\text{diff}}(\tau)$, $\alpha$ is an exponent between 0 and 1, is normally invoked for describing nucleoplasmic mobility of EGFP in other cells [26, 27, 42]. The dark state parameters and the time exponent $\alpha$ are competing parameters in the correlation curve fitting, leading to less robust results. To ensure that only one dark state component is needed for the cell data, in vivo FCS curves were fitted for $\tau > 30$ $\mu$s (see section 2.4.1). For FCS curve fitting, the EGFP dark state parameters were fixed to values found by Windengren et al. [33], $\tau_{DS} = 900$ $\mu$s and $f_{DS} = 0.13$, under similar excitation conditions. We further verified how fitting parameters vary for a range of $\tau_{DS} = 800$-1000 $\mu$s and $f_{DS} = 0.10$-0.16. The values were robust against the variation in $\tau_{DS}$ but they changed considerably with $f_{DS}$ (see STD values in table 2-4).

A sequence of four consecutive nuclear FCS measurements was analyzed using the obstructed diffusion model and the results are listed in table 2-4. Each acquisition consists of around 10 million photons. For the non-global fitting, the four data sets were fitted individually and their average count rate was used as a scaling factor to obtain the initial EGFP concentration and its uncertainty. For the global fitting we used linear constraints on the local concentrations scaled by the relative count rate. In order to verify the robustness of the fit and the choice of fixed triplet parameters, a similar global analysis but with shared free triplet parameters was performed. We obtained $\tau_{DS} = 950$ $\mu$s and $f_{DS} = 0.05$, which are reasonably close to the fixed values.
2.5 Summary

The average molecular brightness $\eta$ of the nuclear diffusive components divided by that measured for EGFP monomers in vitro yields a relative brightness very close to 1, suggesting that the percentage of multimeric EGFP in nucleus is negligible and therefore it cannot explain the accumulation of EGFP inside the nucleus. Then, assuming to have only EGFP monomers in the nucleus, the diffusion time of $5.02 \pm 0.74$ ms corresponds to a nucleoplasmic viscosity $\nu_n = 23.9 \pm 3.6$ cP and a diffusion coefficient $D_n = 2.27 \pm 0.38 \mu$m$^2$/s. The diffusion coefficient is much smaller than the nuclear diffusion coefficient of EGFP measured in mammalian cells like HeLa (23.0 ± 1.0 $\mu$m$^2$/s) [25], HEK293 (14.9 ± 0.6 $\mu$m$^2$/s) [26] and AT-1 (8.7 ± 1.0 $\mu$m$^2$/s) [27]. Consistently, the measured anomalous factor $\alpha \sim 0.8$ is smaller than the reported value of 0.91 for HEK293 cells [26] and 0.87 for AT-1 cells [27], implying a higher crowding effect or nonspecific binding interactions of EGFP in the nucleus of Drosophila cells.

Using a relatively small (10 million) photon dataset we have demonstrated the capability to perform quantitative FCS measurements in live cells, at $\mu$M concentrations and in the presence of photobleaching. EGFP concentration is measured with an uncertainty of ~10%, while the uncertainty for the diffusion time is higher, around 20%. This is reasonable, considering the complexity of the fitting model, the high level of noise and change of the local environment due to the nuclear movement. Our FCS analysis indicates that the monomeric form of EGFP is abundant in the nucleus. The aggregation-induced mechanism previously suggested to explain the localization of EGFP in the nucleus does not apply to S2 cells and the nature of this phenomenon needs further investigation.

2.5 Summary

Photon correlation measurements performed on a custom-built multiparameter fluorescence setup can, even with a relatively limited number of photons, accurately determine molecular mobility and excited-state photophysics on timescales spanning from nanoseconds to milliseconds. Excitation saturation levels were determined for an organic dye, Rh110, and a fluorescent protein, EGFP, and it was shown that the correlation curve fitting yields accurate local concentrations and diffusion constants only if the excitation intensity is maintained at least a factor of 5 below the saturation level.
At least in the high concentration limit, our data suggest that the FCS fitting accuracy is primarily determined by the total number of photons $n$ detected from a single molecule passing though the observation volume. That is, the higher the product between the emission rate and the diffusion time, the higher the fitting accuracy, as shown by experiments in which the viscosity was used to slow down the diffusion time. For measurements not exceeding 1 hour, our custom-built detection, digital correlation and data analysis provided consistent results for samples concentrations from 50 pM to 2 µM, and for conditions in which the background level matches or even dominates the fluorescence signal. This range can be further extended by longer measurement times or by more complex offline photon filtering algorithms.

We have demonstrated the capability of FCS in live cells at high concentrations by performing measurements inside the nucleus of Drosophila S2 cells. The estimated nuclear viscosity (~24 cP) was considerably higher than in mammalian cell nucleus. In addition, the fluorescence fluctuation analysis indicates that EGFP inside the S2 cell nucleus exists mainly as monomers.

FCS capabilities presented here are suitable for characterizing trace amounts of biosensors in analytical studies and for studying the mobility and the dynamics of proteins in live cells. The multiparameter photon-by-photon detection scheme is a valuable quantitative tool for description of ligand-protein binding interactions by performing fluorescence correlation and polarization analysis simultaneously, on the same photon dataset. Exclusively, FCS measurements provided not only affinity information, but also information about the fraction of (in)active reactants. The simultaneous FCS and anisotropy analysis approach has a great potential to explore the mechanism of action of small-molecule inhibitor drugs (i.e. organometal complexes) and help optimize their design to target transcription factors, such as Stat3, that play a significant role in the onset of cancers.

2.6 Bibliography


Chapter 3
Rhodamine Dark States as Probes for Conformational Dynamics of Biomolecules

Abstract

The dark states of several rhodamine dyes were investigated by Fluorescence Correlation Spectroscopy (FCS). We resolved 2 dark states (DS) for free rhodamines in aqueous solutions, of which the longer-lived state depends on the viscosity of the solution. The correlation curves of protein and DNA samples labeled with tetramethyl-rhodamine display between 3 and 4 kinetic components in the (sub)microsecond time range that can be ascribed to dark states decay. We investigated the nature of these decay components by performing FCS experiments under various experimental conditions and for different samples. Based on the changes recorded in population and lifetime for each dark state with irradiance, viscosity, oxygen and the detection geometry, we assigned them, in order of increasing decay times, to a triplet state, a twisted-intramolecular-charge-transfer state (TICT), a lactonic state, and a photoionized state, respectively. Consequently, our data suggests that none of the observed dark states can be directly assigned to the structural dynamics of the biomolecules to which they are attached. However, we found evidence that the intrinsic conformational dynamics of the biomolecule appears in the correlation curves as a modulation of the aforementioned dark-state photophysics of the dye. This proves the extreme importance of accurate control measurements and appropriate modeling of the dye photophysics in biological FCS experiments, as most studies in literature incorrectly interpret (some) fluorophore dark state relaxation times to directly correspond to folding dynamics in proteins or nucleic acids.
3.1 Introduction

Any source of fluctuations of fluorescence intensity can be studied, in principle, by FCS. Access to a wide range of time scales, from nanoseconds to milliseconds, makes FCS a very attractive technique for studying the photophysics of fluorescent probes, as well as the conformational fluctuations of biological molecules. However, the fluorescence correlation analysis has been primarily used to determine the diffusion parameters of biomolecules [1, 2]. To a much lower extent, some studies used FCS to characterize the photophysics and photochemistry of dyes [3-6] and to investigate structural fluctuations in biomolecules [7-12]. The imprints of these processes in correlation curves are exponential decays and we refer to them as dark states. There are several approaches to study the conformational dynamics by FCS.

In one approach, a biomolecule is labeled with a FRET pair and therefore the fluorescence emission varies between different FRET states. A FRET-FCS fitting model, even though only for a two state system, has been developed and implemented experimentally for this configuration [10, 11, 13]. However, this model is based on the assumption of having the same diffusion parameters for both states. This assumption implies the same hydrodynamic radii for both states, e.g., the folded and the unfolded states of a protein. It also neglects the possible differences in detection volume parameters of the donor and the acceptor channels.

In another approach, the acceptor is substituted with a quencher to perform photoinduced electron transfer in combination with FCS (PET-FCS). Here the fluorescence fluctuates between a bright and a less bright (or a dark) state and result in exponential decay terms in the correlation curves. PET-FCS is used in a DNA hairpin, to investigate the kinetics of conformational fluctuation between its open and closed states [8]. Alternatively, the extrinsic quencher can be replaced by an intrinsic and therefore non-disruptive quencher, e.g., the aromatic amino acids tryptophan and tyrosine. This technique is used to show the long-range modulation of chain motions within an intrinsically disordered domain of a tumor suppressor protein [14] as well as to measure the folding-unfolding kinetics of several proteins [12, 15]. PET-FCS employs simple pseudo autocorrelation analysis (see chapter 2). This can be accounted as an advantage compared to the two-color cross correlation design of the FRET-FCS; as the pseudo autocorrelation analysis has been based on fewer assumptions which are experimentally verified, and encounters simpler sample preparation (single labeling vs. FRET pair labeling).
Analysis of PET-FCS measurements requires carefully-designed control measurements, which, unfortunately, are not properly described in the literature. The precise fingerprint of the photophysics of the labeling dye in the measured correlation curve needs to be known first in order to distinguish it from the (possible) signal fluctuation induced by intra- or inter-molecular dynamics of the labeled biomolecule. For example, it is well known that typical fluorescent probes such as rhodamine dyes can have several triplet states and other nonradiative decay channels that can be activated under various experimental conditions [16]. However, their presence is either fully neglected or misunderstood in FCS applications to biology [12, 14].

Sasmal et al. [12] observed 3 dark states in Human Serum Albumin (HSA) labeled with 7-dimethylamino-3-(4-maleimidophenyl)-4-methylcoumarin (CPM) dye. These dark states were ascribed to chain dynamics, interchain reaction and concerted chain motion of the protein. Surprisingly, the authors did not consider the possibility for any of these components to simply arise from the intrinsic photophysical processes of the dye itself. PET-FCS measurements were recently performed on an intrinsically disordered protein labeled with an oxazine fluorophore, AttoOxa11, in order to study the loop closure kinetics of its segments [14]. Interestingly, the authors also observed microsecond relaxation lifetimes in the correlation curves, which they claim to originate entirely from the formation of nonrandom chain configurations. Furthermore, Chattopadhyay et al. used tetramethylrhodamine (TMR) to study the kinetics of conformational fluctuations in an unfolded protein. They have observed a ~2 μs decay component in their FCS data which was viscosity dependent [15]. They conclude that this is an unknown conformational relaxation process, assuming that electronic quenching cannot be viscosity dependent.

However, not only that typically reported lifetimes for triplet relaxation of rhodamine dyes is in the order of a few microseconds [3, 4, 17], but also the dependence of nonradiative decay processes on viscosity in rhodamine dyes have been reported in several studies [18, 19]. More specifically, the triplet state decay was shown to be viscosity sensitive and could be described by a Stern-Volmer equation [20, 21]. In addition, not only quenching but also other non-fluorescence de-excitation channels such as triplet, charge-transfer or photo-isomeration pathways are often sensitive to environment changes [22, 23]. Hence the presence of a sub-diffusion decay component in FCS data may correspond to a newly activated dark state and not the direct modulation of fluorescence intensity by the movement of biomolecular quenching groups in and out of the proximity of the dye. Consequently, (sub) microsecond kinetics observed
by FCS may not be assigned straightforwardly to local conformational dynamics of the biological molecule and a better understanding of the mechanism of the modulation of fluorophore’s emission by conformational dynamics is required to correctly interpret the data.

The present chapter is focused on addressing the ambiguities related to the influence of dye photophysics versus intermolecular dynamics of biomolecules in FCS data. We measured, characterized and assigned a number of dark states of rhodamine dyes observed by FCS measurements under various conditions. As an example for application, the information obtained was used in the FCS analysis of proteins, single- and double-stranded DNA oligonucleotides.

3.2 Experimental

3.2.1 Materials

All samples were diluted in TRIS buffer (50mM Tris, 150mM NaCl, pH 7.5). Glycerol (4750, VWR International, Mississauga, Canada) solutions were also prepared in TRIS buffer. Oxygenated buffer was made by bubbling O₂ through TRIS buffer for 60 minutes prior to experiment. The fluorescent dyes used in the experiments are: 5-carboxy-tetramethylrhodamine N-succinimidyl ester (TMR-NHS) (81125, AnaSpec, Fremont, CA, USA), tetramethylrhodamine-5-maleimide (TMR-maleimide) (T-6027, Life Technologies Inc., Burlington, Canada), Rhodamine 6G (Rh6G) (R4127, Sigma Aldrich, Oakville, Canada), Rhodamine 101 (Rh101) (8369, Sigma Aldrich, Oakville, Canada), Rhodamine 110 (Rh110) (80103, Biotium, Inc. Hayward, CA), Bodipy Fluorescein-maleimide (Bodipy FL-maleimide) (B-10250, Life Technologies Inc., Burlington, Canada). Fluorescent polystyrene beads (F8809, Life Technologies Inc., Burlington, Canada) were used for the alignment of the excitation and detection paths in the confocal microscope.

Protein denaturation assays were performed in 6M of Guanidinium hydrochloride (GdmCl) (G9284, Sigma Aldrich, Oakville, Canada). All dye samples were diluted to concentrations of 1-10 nM, which are most suitable for FCS measurements. For experiments, sample solutions of ca. 50 μL were dropped on plasma-cleaned coverslips. When necessary, the clean coverslips were coated with bovine serum albumin (BSA) (15260-037, Life Technologies Inc., Burlington, Canada) in order to prevent nonspecific adsorption of the dyes to the glass surface. To prevent the evaporation of the solvent, all samples were sealed between two coverlips using silicon spacers.
3.2 Experimental

(P17175, Life Technologies Inc., Burlington, Canada) and/or adhesive spacers (S24735, Life Technologies Inc., Burlington, Canada).

3.2.2 Protocols

DNA hybridization:

Custom-designed complementary DNA oligonucleotides, 3'-ATT TAG ATT TCA /6-TMR/ TG TAT TCC TTT GTA TTG CCA TTC AGG T-5' and 5'-TAA ATC TAA AGT TAC ATA AGG AAA CAT AAC GGT AAG TCC A-3', were purchased from Integrated DNA Technologies (Coralville, IA, USA). The two sequences were hybridized together using a protocol provided by the manufacturer. Briefly, the DNA oligos were dissolved in STE Buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA) at high concentration (~200 μM). The TMR-labeled strand and the complementary strands were mixed in 1:10 molar ratios. The mixture was heated to 94°C and gradually cooled on the bench top overnight. The resulting product was desiccated and stored in 4°C.

Oxygen scavenger assay:

The oxygen scavenger assay was performed according to protocols given elsewhere [24]. The oxygen-scavenging solution contained: TRIS buffer (pH 7.5), 5% (w/v) glucose, 5% (v/v) glycerine, 25 μg/ml glucose-oxidase (G2133, Sigma Aldrich, Oakville, Canada), 100 μg/ml catalase (C40, Sigma Aldrich, Oakville, Canada), and 0.05 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, C47.6, Sigma Aldrich, Oakville, Canada). The pH of the solution was monitored upon mixing all ingredients. It was found that the pH reduced from 7.5 to 7 during the first hour, and further to 6 by the end of the second hour. Therefore, FCS experiments on samples in the oxygen scavenger buffer were performed within the first hour upon mixing to prevent significant pH influence on our data.

Protein labeling:

Wild type (WT) and T22G mutant of N-terminal Src-homology-3 of Drosophila (SH3 DrkN) were labeled non-specifically with TMR-NHS. The NHS-amine coupling reaction was performed in PBS buffer with pH adjusted to 8. The reaction was started by adding the NHS-ester-activated fluorophores to a 50 μL solution of 200 μM proteins at a molar ratio dye:protein of 1:2. This low labelling ratio ensured that most proteins had only one dye label attached. The sample was gently
shaken for 3 hours in the dark at room temperature. The excess dye was removed by size-
exclusion chromatography using Sephadex G-25 gels (G2580, Sigma Aldrich, Oakville, Canada)
in a BioLogic LP system (731-8300, Bio-Rad, Mississauga, Canada).

The cysteine-mutated caseinolytic protease (ClpP E150C) from *E. coli* was labeled by TMR-
amaleimide as discussed in detail in section 1.3.2.2. TCEP was added to the protein solution (50
μL, 2 μM protein, TRIS buffer) at 20:1 molar ratio excess compared to the protein in order to
reduce the possible disulfide bonds [25]. The labeling reaction was initiated upon addition of
TMR at 10:1 molar ratio with the protein. The reaction proceeded by gentle vortexing for 3 hours
at room temperature and overnight in 4 °C. The remaining excess dye was removed by size-
exclusion chromatography [26] using Sephadex G-50 gels (G5080, Sigma Aldrich, Oakville,
Canada).

### 3.2.3 Procedures

Under some simplifying assumptions, such as a 3D Gaussian shape of the detection volume,
the fluorescence correlation function for free Brownian diffusion of a single molecular species
with dark state blinking is given by [27]:

\[
G(t) = \frac{1}{N_{\text{eff}}^{0.5}} \left( 1 + \frac{\tau}{\tau_d} \right)^{-1} \left( 1 + \frac{\tau_s^2 \tau_d}{s^2 \tau_d} \right)^{-0.5} \prod_i \left( 1 + \frac{f_{DS,i}}{1 - f_{DS,i}} \exp \left( -\frac{\tau}{t_{DS,i}} \right) \right) \tag{3-1}
\]

In equation (3-1), \(N_{\text{eff}}\) is the average number of molecules in the detection volume, \(V_{\text{eff}} = \pi^{3/2} z_0 w_0^2\),
and it is given by the inverse of the diffusion-related amplitude, \(G_0 = N_{\text{eff}}^{-1}\). In addition, \(s\) is the
structure parameter, namely the ratio between the axial and the lateral radii of the detection
ellipsoid \(s = z_0 / \omega_0\), \(\omega_0\) is the diffusion time, which is related to the diffusion coefficient \((\omega_0^2 = 4D \cdot \tau_d)\), \(t_{DS,i}\) is the dark state relaxation time and \(f_{DS,i}\) is the fraction of molecules of the \(i^{th}\)
dark state. In a 3-level system such as that shown in figure 3-1, \(f_{DS}\) and \(t_{DS}\) are given by [28]:

\[
t_{DS} = \left( k_{31} + \frac{k_{12}k_{23}}{k_{12} + k_{21}} \right)^{-1},
\]

\[
f_{DS} = \frac{k_{23}k_{12}}{k_{12}(k_{23} + k_{31}) + k_{31}(k_{21} + k_{23})}. \tag{3-2}
\]
3.2 Experimental

Under low excitation intensity condition that we are using in our experiments $k_{21}$ is by far the largest transition rate and $t_{DS} \sim k_{31}^{-1}$.

![Simplified 3-level scheme of the electronic states of a fluorophore](image)

**Figure 3-1.** Simplified 3-level scheme of the electronic states of a fluorophore: ground-state ($S_0$) and excited-state ($S_1$) singlets, the triplet state ($T_1$) and the transition rate constants between them. This system can be viewed as a two state model of a bright state and a dark state manifold.

Similarly, any transition from a fluorescent to a non-fluorescent (dark) state will give rise to a decay of the fluorescence intensity autocorrelation with a rate matching the relaxation rate of the dark state. In general, $t_{DS}$ and $f_{DS}$ can be expressed in terms of population ($k_{bd}$) and depopulation ($k_{db}$) rates, i.e., the rates of transition between the bright (fluorescent) state BS, and the dark (nonfluorescent) state DS [29]:

$$
BS \xrightarrow{k_{bd}} DS, \quad k_{db} \quad \text{and} \quad t_{DS} = \frac{1}{k_{bd} + k_{db}} \quad \text{and} \quad f_{DS} = \frac{k_{bd}}{k_{bd} + k_{db}}.
$$

From equations 3-2 and 3-3, it is possible to define effective $k_{bd}$ and $k_{db}$ in terms of transition rates between electronic levels or vice versa. Again assuming $k_{21}$ is by far the largest transition rate we can find $k_{bd} \sim k_{23}k_{12}/k_{21}$ and $k_{db} \sim k_{31}$.

Rhodamine 6G (Rh6G) and Rhodamine 110 (Rh110) are the standard dyes to determine the geometric parameters of the FCS detection volume, $s$ and $w_0$, for the two laser lines, 532 nm, and 480 nm, used in this study. However, instead of the commonly used values of $D_{Rh6G} = 2.8 \times 10^{-10}$ m$^2$ s$^{-1}$ [30] and $D_{Rh110} = 2.7 \times 10^{-10}$ m$^2$ s$^{-1}$ [31]], we applied the recently updated values of $D_{Rh6G} = 
3.3 Observation of multiple dark states

In Chapter 2 of this thesis, we focused on the accuracy of determining the diffusion constant and the local concentration by FCS, and we showed that, for reliable FCS measurements, the excitation intensity should be less than 20% of the saturation intensity. However, in addition to introducing saturation and photobleaching, high excitation intensity can alter the correlation curve by changing the manifold of dark states of the fluorophore. Here we can mention for example the increase of the triplet state population and lifetime with excitation intensity [17], the transition to the photo-induced transient states [6], and the broadening of the DS lifetime distribution due to non-uniform excitation across the detection volume. Therefore, in order to study the dark states and molecular dynamics with FCS, we performed measurements under low intensity conditions, namely $I_{ex} < 0.1 I_s$, unless stated otherwise.

In order to satisfactorily fit the correlation curve obtained for freely diffusing Rh6G in aqueous buffer only one DS term was sufficient in the classic model (eq. 3.1). This is in agreement with previous FCS reports on organic dyes [3, 4, 17]. However, the DS decay was fitted to a value of 17 $\mu$s, which is an order of magnitude larger than the reported triplet state lifetime of $\sim 2$ $\mu$s [4, 17]. We further investigated this disagreement by performing FCS at different excitation intensities between 1 to 50 kW/cm$^2$, the latter being the typical excitation intensity in published FCS studies. Interestingly, fitting the correlation curve measured at the highest intensity required an extra exponential decay term with a lifetime of 1.8 $\mu$s (see figure 3-2), which closely matches the reported triplet lifetime value in FCS literature [4, 17]. In addition, the relative population of this component, 0.22, is similar to literature values ($\sim 0.2$, figure 4 in [4]). The absence of the fast decay component under low power excitation is probably due to reduced transition rate to this DS under these conditions (see also below).
3.3 Observation of multiple dark states

Figure 3-2. FCS analysis of Rh6G at different excitation intensities. At the lowest intensity, a model with one dark state fitted the data. At higher intensities, two dark states were required for good data fitting. The weighted residuals of the fit with both models at the highest excitation intensity are shown in the lower panel.

Figure 3-3. FCS data (symbols) and fit (lines) for SH3-TMR in 6M GdmCl (pH 6) and TMR in 50% w/w glycerol in TRIS (pH 7.5). Four dark states were needed to fit the SH3 data and two dark states were needed for TMR in glycerol. The weighted residuals of the SH3 data fitting with three and four dark states (DS) are shown in the lower panel.
3.3 Observation of multiple dark states

Additional DS decay terms in the FCS fitting model were required not only for free dyes in aqueous buffers in the intense excitation regime, but also for the same dyes under low power excitation in viscous solutions or in biological samples. The maximum numbers of exponential decays, four, were required for fitting the correlation curve measured at low intensity for the DrkN-SH3 protein labeled with the amine-reactive TMR-NHS dye (see figure 3-3). We named these dark states DS₁, DS₂, DS₃ and DS₄, from the fastest ($t_{DS1} < 1 \mu s$) to the slowest ($t_{DS4} > 60 \mu s$). The focus of this chapter is to identify the processes corresponding to these FCS decays and evaluate which of them are intrinsic to rhodamine dyes and which are modulated by the protein environment.

We used the glycerol buffer (50%w/w) in order to increase the diffusion time ($\tau_d$) of the free dye from ~50 $\mu$s to ~300 $\mu$s, which is the value estimated for SH3. This allows for the same time window for measuring the photophysics of TMR by FCS in the two cases: free dye vs. dye attached to protein. Under these "equivalent" conditions, quite surprisingly, the correlation curve of the dye alone shows only two dark states instead of four. This may suggest that the other two decays are induced by intramolecular conformational dynamics in the protein. However, other explanations such as activation of a nonradiative decay channel upon labeling should not be neglected. Therefore, further investigations are required to assign the dark states in FCS curves and understand which ones can be used to track molecular dynamics.

Rhodamine dyes are known to exhibit complex photophysics behavior consisting of multiple dark states with lifetimes varying from nanoseconds to seconds [16, 23, 34, 35]. Although one can question the capability of the fluorescence correlation analysis to measure all photophysical processes, it is striking that most FCS studies report one triplet state or no triplet at all, even when the triplet decay is expected to be faster than the translational diffusion across the confocal volume [34]. Consequently, one can question the accuracy of the claims made in those reports that DS decays correspond to molecular dynamics of the labeled biomolecule.

For all FCS data presented in this chapter, the fitting was performed for a 6-orders of magnitude time scale, from 100 ns to 100 ms. Our raw data extends down to 4 ps, but we started at 100 ns to eliminate antibunching and rotational kinematics from analysis. Typical diffusion times of the biomolecules used in this study were in the order of 0.5 ms. Therefore the observed DS exponential decays spanned over 3 orders of magnitude on the time axis. The amplitudes of the DS decays estimated at low excitation power are typically an order of magnitude smaller than
the amplitude of the diffusion curve itself. In addition to the expected broadening of the decay times from non-uniform irradiance at the focal spot, the criteria for resolving multiple decays by FCS also applies. According to Meseth et al. [36], in the case of two independently diffusive species, their diffusion coefficients should be at least a factor of 1.6 times different to be resolved by FCS. Furthermore, the resolving power depends strongly on their relative quantum yield as well as the concentration of each component in solution. The aforementioned maximum number of observed DS decay terms of four is probably due to this limitation, and each decay component may correspond to several processes with lifetimes too close to be resolved by FCS.

3.4 Dark states of rhodamine dyes

In order to investigate the nature of dark states in the correlation curves, a brief review of the photophysics of the corresponding dyes is useful. Rhodamine is a family of fluorescence dyes which are used extensively in single-molecule studies due to their high fluorescence quantum yield (90-100%) and photostability [37]. Figure 3-4 shows the chemical structure of four representative rhodamine dyes and a non-rhodamine dye, Bodipy fluorescein (Bodipy FL), that were used in this study.

![Figure 3-4](image)

**Figure 3-4.** The dyes investigated in this chapter: the first four are rhodamine dyes, while the last one is a Bodipy dye.

As discussed above, the fast, sub-diffusion correlation decay arises from fluctuations of the fluorescence intensity due to reversible transitions of the fluorophore to a dimmer or non-fluorescent state. These states are collectively named dark states (DS) even though they may not be fully dark or represent a new excited state. For example, **emission shifts** upon changes in local pH and/or polarity that can be induced by conformational dynamics of the host biomolecule, will also appear as an exponential decay in FCS curve. This is clearly not a new electronic (dark) state of the fluorophore, but is very useful as a probe for the time scale of the underlying molecular dynamics of the host. **Intermolecular energy transfer** is another example of a process that can
lead to a DS-like decay in FCS, in which the dye fluorescence is quenched by a molecule in close proximity, such as an aromatic amino-acid in proteins or a base in nucleic acids.

3.4.1 Triplet state

Rhodamine dyes are rich in non-fluorescence deactivation pathways. After excitation, deexcitation continues by returning to the ground state either through photon emission, or intersystem crossing to a triplet state followed by non-radiative relaxation or phosphorescence [38].

3.4.2 Twisted-intramolecular charge-transfer state

To explain the nature of nonradiative decay in the rhodamine dye family several rhodamines with different substitution patterns at the amino nitrogen are investigated in several studies [16]. The enhancement of the intersystem crossing rate by temperature and its reduction by viscosity [39] has been explained by the assumption of an energy barrier in the torsional motion of the amino groups in the excited state [16]. Interestingly, dyes such as Rh110, Rh6G and Rh101 which have non-substituted, partially alkylated or rigidly locked amino groups, have quantum yields of close to unity. Furthermore their quantum yields are almost temperature independent [16, 39]. Contrary, the quantum yield of RhB that has a fully alkylated amino groups strongly depends on solvent viscosity and temperature [39].

In addition, based on the very high quantum yield of fluorescence ~100% observed for RhB in glycerol (high viscosity) and for Rh101, it is postulated that the transition from the excited planar state (S1) to a twisted-intramolecular charge-transfer (TICT) state through the torsional motion of the amino groups forms a route of nonradiative relaxation in rhodamines [19].

3.4.3 The lactone form

Rhodamines dyes like Rh101 can exist in solution in Zwitterionic (Z), Lactonic (L) or Cationic (C) forms (figure 3-5) [40-42]. The Z and C forms have a similar molar extinction coefficient, ~95,000 M⁻¹ cm⁻¹ for Rh101 at 568 nm [16]. The C form only exists in polar solvents under acidic conditions (pH<4), therefore there is no need to consider it in this study. At higher pH in polar protic solvents, the Z form is dominant and present in equilibrium with the L form (82% vs 18% in water) [40]. Due to the disruption of the π-electronic system of the xanthene
group in the L form, it does not absorb or fluoresce in the visible region [16]. Therefore, the fluorescence is due entirely to the presence of dye in the Z form. The L form behaves as a dark ground state which can be populated both by the Z-L conversion in the ground state or by the branching of the zwitterionic excited state as a result of lactone ring formation.

![Diagram](image)

**Figure 3-5** Lactonic, zwitterionic and cationic forms of TMR. The zwitterion is favored by polar protic solvents. Lactone dominates in nonpolar and polar aprotic media. Cations occur under acidic conditions.

### 3.4.4 Butterfly motion

Even though the butterfly motion of the xanthene moiety does not represent any specific dark state, it can play a role in the nonraditive processes by coupling the excited states to the other dark states or by enhancing internal conversion [16]. For example, the quenching of zwitterionic excited state of Rh101 in nonpolar solvents is explained by deactivation through a higher triplet state mediated by the bending vibration of xanthene skeleton [16].

### 3.4.5 Photoionization

In addition to these structural associated non-radiative decay pathways, photochemical dark states like photoionization followed by electron trapping (redox blinking) have been observed in Rh6G. To explain the excited-state kinetic of the Rh6G in Poly(vinyl alcohol), Zondervan et al. [23] considered a four-level system where a long-lived dark state was populated through an intermediate triplet state. They further proved its nature as an anionic radical. A similar dark state, although with a higher stability of the radical, was observed in ethanol [23]. Interestingly, oxygen was effective in destabilizing this dark state, probably by assisting the recovery as an electron carrier.
3.5 Dark states in FCS

Given the complexity of excitation relaxation in organic dyes, it is essential to characterize their dark state contribution to the correlation curve prior to labeling biomolecules for FCS experiments. To this end, we performed FCS on free TMR, Rh6G, Rh101, Rh110 and Bodipy Fl under various experimental conditions. Rh6G is an esterified rhodamine and lacks the lactonic dark state, while Rh101 does not have a TICT state because of its rigid structure. On the other hand, based on structural considerations, TMR and Rh110 can in principle display all dark states. Bodipy FL is a non-rhodamine dye and exhibits a different excited state manifold. We investigated the changes of the dark states with viscosity, oxygen, irradiance and detection geometry for the aforementioned dyes.

3.5.1 Viscosity

In order to increase the time window of FCS analysis of the dark states to a situation comparable to that of a labeled protein, we performed FCS measurements of the free dyes in TRIS buffer with 50% w/w glycerol (v = 6 cP). In our confocal setup, this solvent viscosity results in a diffusion time of ~300 μs for the free dyes, which is similar to that found for a ~40
kDa protein in aqueous buffer. The FCS measurement on Rh6G sample in glycerol buffer solution was also repeated using a 75-µm instead of the regular 50-µm pinhole.

Table 3-1. FCS fitting results for free dyes in glycerol and in water. All data were acquired at “low” excitation intensity, 1 kW/cm². The uncertainties for $\tau_d$, $t_{DS2}$, $t_{DS4}$, $f_{DS2}$ and $f_{DS4}$ are less than 3%, 35%, 10%, 30% and 5%, respectively. Even though Bodipy and Rhodamine are different families of dyes with different photophysics, the dark states of Bodipy FL are given in the DS2 and DS4 columns for the sake of comparison.

<table>
<thead>
<tr>
<th>Dye</th>
<th>$\tau_d$ (µs)</th>
<th>DS2</th>
<th>DS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rh6G (in buffer)</td>
<td>56</td>
<td>2.4</td>
<td>1.7</td>
</tr>
<tr>
<td>in 50% w/w glycerol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rh6G 50 µm</td>
<td>340</td>
<td>7.8</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>533</td>
<td>13</td>
<td>6.1</td>
</tr>
<tr>
<td>Rh110</td>
<td>346</td>
<td>16</td>
<td>2.6</td>
</tr>
<tr>
<td>TMR</td>
<td>351</td>
<td>3.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Rh101</td>
<td>299</td>
<td>4.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Bodipy FL</td>
<td>315</td>
<td>37</td>
<td>10</td>
</tr>
</tbody>
</table>

The FCS results are given in the Table 3-1. Bodipy FL is the only dye that showed a single decay component, while all rhodamine dyes showed two dark states, DS2 and DS4 (for notation details, see above). For comparison, FCS results obtained for Rh6G in aqueous buffer at low excitation intensity are also included. Note that this data was fitted with a two-dark-state model even though one component fit was sufficient. Overall, the fastest component (DS2) is similar in water and glycerol and it is only populated at a fraction of a few percent. The slowest component (DS4) is about 7.5 times slower in glycerol than in water, while its fractional population is nearly the same, around 15%. The DS4 lifetime ratio almost matches the ratio between the viscosities of the two solutions (6). Also, $t_{DS2}$ is somewhat larger in glycerol, ca. 3 times, but this could be due to (over)fitting artifacts. For measurements done in water the decay times of translational diffusion and dark state photophysics are quite closely spaced, and this could affect the accuracy of DS lifetimes estimated by FCS fitting. Furthermore, we acquired FCS data for TMR and Rh6G at different viscosities and a linear increase of $t_{DS4}$ was clearly observed, while $t_{DS2}$ was nearly constant between 1 and 20 cP (figure 3-6).
Figure 3-6. Dark state life times versus solvent viscosity for TMR and Rh6G, as estimated by FCS.

In addition, the FCS data for Rh101 could be satisfactorily fitted by only one dark state, DS4 (70 μs, 11%). For confidence, the measurements were repeated at a different time and the analysis produced nearly identical results. We also repeated the measurements at higher excitation intensities (7 and 21 kW/cm²), for which the faster dark state, DS2, was needed to fit the data. The fraction of this component, f_{DS2}, increased with the excitation intensity (11% and 26%, respectively), while the population of the original state, f_{DS4}, was invariant (11% in all cases). A similar behavior was also observed for Rh110. Therefore, we presented the two-dark-state fit for these dyes in Table 3-1.

3.5.2 Oxygen

The triplet-state quenching by molecular oxygen works through either energy transfer or the formation of an oxidized intermediate state (charge transfer) [46]. In a parallel process, the dye can undergo irreversible photobleaching mediated by singlet oxygen; therefore, it is desirable to remove the oxygen from the buffer to increase the photobleaching lifetime. The concentration of oxygen in water at 20°C is 7.6 mg/L (~ 1.25 mM) [47]. By the oxygenation of the buffer this value can increase to 20 mg/L and even higher. Oxygen can effectively be removed to achieve few μM concentrations by enzymatic oxygen-scavenging systems, such as Gloxy (a mixture of glucose, glucose oxidase and catalase, see section 3.2.2) [46].

To explore the effect of oxygen on a long time scale, we labeled a large protein, the cysteine-mutated, 300 kDa caseinolytic protease (E150C ClpP) from *E.coli*, with TMR maleimide. ClpP is
a proteolic chamber consisting of two stacked heptameric rings [48], and to prevent multiple labeling with TMR, the protein was labeled at low labeling efficiency (< 25%). This approach was preferred to the free dye in high-viscosity solvent, because the oxygenation of glycerol buffers produced persistent bubbles. As a result, we were able to investigate by FCS the effect of O₂ and Gloxy on the dark states of TMR on a time scale up to the diffusion time of the protein, ~0.5 ms, (Table 3-2).

Table 3-2. The effect of oxygen and Gloxy on the dark states of TMR attached to ClpP E150C. The uncertainties of the $t_{DS}$ and $f_{DS}$ values for DS₁-₄ were less than (80%, 45%, 15%, 10%) and (20%, 20%, 10%, 6%), respectively.

<table>
<thead>
<tr>
<th></th>
<th>DS₁</th>
<th>DS₂</th>
<th>DS₃</th>
<th>DS₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloxy</td>
<td>$t_{DS₁}$ (μs)</td>
<td>$f_{DS₁}$ (%)</td>
<td>$t_{DS₂}$ (μs)</td>
<td>$f_{DS₂}$ (%)</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>8</td>
<td>11.7</td>
<td>14</td>
</tr>
<tr>
<td>TRIS buffer</td>
<td>5.2</td>
<td>8</td>
<td>34</td>
<td>22</td>
</tr>
<tr>
<td>Oxygenated buffer</td>
<td>1.2</td>
<td>6</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>

Figure 3-7 FCS analysis of ClpP E150C-TMR in oxygenated and de-oxygenated buffers. In oxygenated buffer, besides the diffusion term, four exponential decays were required to fit the correlation curve. The dotted curve shows the diffusion term only. The double dot-dashed, the dot-dashed, and the dashed curves show the contribution of the diffusion term plus 1, 2 and 3 slowest dark states to the final fit (solid curve), respectively. Three dark states were sufficient to fit the data for the oxygen-depleted buffer (blue curve, fitting not shown).
Interestingly, in addition to DS₂ and DS₄, two more dark states (DS₁ and DS₃) were observed, of which DS₁ appeared exclusively in the oxygen-depleted buffer. The experimental FCS curves and the fits for the sample in oxygenated and oxygen-depleted buffers are shown in the figure 3-7.

3.5.3 Intramolecular dynamics in proteins and DNA

In order to investigate the capability of FCS to measure intramolecular dynamics, we performed correlation experiments on a marginally stable protein and on DNA samples. These systems were chosen because they can exist in two different phases: a highly fluctuating structure and a more rigid arrangement.

Drk (Downstream of receptor kinase) is an adaptor protein essential in the development of the Drosophila eye. The Drk N-terminal 59-residue sequence forms a β-barrel SH3 (Src homology-3) domain that is marginally stable under non-denaturing conditions: the folded and unfolded states are quasi-equally populated, with a very slow exchange rate of around 2 s⁻¹ [49]. A single point threonylation mutation (T22G) fully stabilizes the folded state [50]. These characteristics make it a useful model system for our purpose.

Table 3-3. FCS analysis of WT-SH3 and the T22G mutant in TRIS and 6 M GdmCl. The uncertainty of τᵋ values are less than 2% and the uncertainties of the tₒₛ and fₒₛ values for DS₁-₄ were less than (60%, 45%, 15%, 10%) and (25%, 25% 10%, 5%), respectively.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Solvent</th>
<th>τᵋ (μs)</th>
<th>tₒₛ₁ (μs)</th>
<th>fₒₛ₁ (%)</th>
<th>tₒₛ₂ (μs)</th>
<th>fₒₛ₂ (%)</th>
<th>tₒₛ₃ (μs)</th>
<th>fₒₛ₃ (%)</th>
<th>tₒₛ₄ (μs)</th>
<th>fₒₛ₄ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T22G</td>
<td>TRIS</td>
<td>207</td>
<td>3</td>
<td>4</td>
<td>25</td>
<td>6</td>
<td>85</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 M GdmCl</td>
<td>426</td>
<td>6</td>
<td>10</td>
<td>38</td>
<td>4</td>
<td>160</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>TRIS</td>
<td>231</td>
<td>0.3</td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>28</td>
<td>7</td>
<td>96</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>6 M GdmCl</td>
<td>412</td>
<td>0.9</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>38</td>
<td>5</td>
<td>147</td>
<td>14</td>
</tr>
</tbody>
</table>
In addition to the N-terminus the DrkN SH3 amino-acid sequence has five lysines, which are the typical target of dye labeling by N-HydroxySuccinimide (NHS) chemistry. By labeling at low dye-protein ratio, we ensured that most proteins had a single TMR label, attached at any of the six different locations. FCS experiments were performed for wild-type and mutant (T22G) SH3 samples both in TRIS (pH = 7.5) and in 6 M GdmCl (pH = 7.5) buffer. The data obtained was analyzed and the results are presented in Table 3-3.

![Figure 3-8. The measured FCS curves of WT-SH3 (red) and SH3 T22G (blue) in TRIS buffer and in 6 M GmdCl. In both conditions, the WT-SH3 showed an additional fast-decaying dark state on the sub-μs time scale.](image)

Unlike T22G, the WT-SH3 consists of two subpopulations, the folded and the unfolded states. The ratio of hydrodynamic radii is \( \frac{R_{h\text{, unfolded}}}{R_{f\text{, folded}}} = 1.3 \) [51], therefore, according to Meseth criteria [36], the correlation curve of the WT-SH3 cannot be fitted with two separate diffusive components. In the chapter 4 of this thesis, we will show how this restriction can be overcome, however here we present the FCS analysis with only 1 diffusive component for the WT protein. The correlation curve of T22G was satisfactory fitted with three dark states but the WT-SH3 required four dark states (figure 3-8).

Double-stranded (ds) DNA has a rigid structure with persistence length of about 50 nm, or ~150 base pairs (bp). Therefore the 40-bp dsDNA that we used in this study can be regarded as a solid rod ~13 nm long. In contrast, the single-stranded (ss) DNA is a very flexible chain, with a persistence length of just 1-3 nm [52]. Therefore, we chose this system, containing a TMR label
at position 13, in order to separate the chain dynamics from the dye photophysics in fluorescence correlation curves.

![Correlation curves of the ssDNA (red) and dsDNA (blue), fitted to a 3-dark-state model. The residuals of the fit of dsDNA data with 2- and 3-dark-state models are shown in the lower panel.](image)

**Figure 3-9.** Correlation curves of the ssDNA (red) and dsDNA (blue), fitted to a 3-dark-state model. The residuals of the fit of dsDNA data with 2- and 3-dark-state models are shown in the lower panel.

The correlation curve of ssDNA was satisfactorily fit with 3 dark states with $t_{DS} \approx 1.9$, 14 and 77 $\mu$s (Figure 3-9 and Table 3.4). However, fitting dsDNA data returned unexpected results, as models with 2, 3 or 4 dark states were tested. The 2-dark-state model fit was not satisfactory (figure 3-9) and produced illogical results: the $\tau_d = 214$ $\mu$s obtained was smaller than that of ssDNA (222 $\mu$s), and the slow-dark state with $t_{DS} = 32$ $\mu$s matches neither DS$_3$ nor DS$_4$ of the ssDNA with $t_{DS3} \approx 14$ $\mu$s and $t_{DS4} = 77$ $\mu$s. The 3-dark-state model was satisfactory and it resulted in a reasonable $\tau_d = 280$ $\mu$s, and $t_{DS3} = 20$ $\mu$s and $t_{DS4} = 98$ $\mu$s, corresponding to DS$_3$ and DS$_4$. In addition, it surprisingly resulted in a slow, $\sim 2.4$ ms, decay which is put in the column labeled DS$_2$, and will be discussed in section 3-6-2. The 4-dark-state model fit produces exactly the same values plus an additional microsecond-lived dark state resembling DS$_2$, but with a virtually zero population, 0.6% (not shown).
Table 3-4. FCS fitting parameters for 40-bp ssDNA and dsDNA labeled with TMR. The uncertainty of $\tau_d$ values are less than 2%. The uncertainties of the $t_{DS}$ and $f_{DS}$ for DS$_{2,4}$ were less than (45%, 25%, 12%) and (35% 10%, 5%), respectively, except for the fitting dsDNA data with 3-dark-state model, which have $\Delta t_{DS2}$~12% and $\Delta f_{DS2}$~3%.

<table>
<thead>
<tr>
<th># of DS in the fitting model</th>
<th># of DS in the fitting model</th>
<th># of DS in the fitting model</th>
<th># of DS in the fitting model</th>
<th># of DS in the fitting model</th>
<th># of DS in the fitting model</th>
<th># of DS in the fitting model</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssDNA</td>
<td>ssDNA</td>
<td>ssDNA</td>
<td>ssDNA</td>
<td>ssDNA</td>
<td>ssDNA</td>
<td>ssDNA</td>
</tr>
<tr>
<td>3</td>
<td>222</td>
<td>1.9</td>
<td>9</td>
<td>14</td>
<td>12</td>
<td>77</td>
</tr>
<tr>
<td>2</td>
<td>222</td>
<td>6.6</td>
<td>15</td>
<td>62</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>dsDNA</td>
<td>dsDNA</td>
<td>dsDNA</td>
<td>dsDNA</td>
<td>dsDNA</td>
<td>dsDNA</td>
<td>dsDNA</td>
</tr>
<tr>
<td>3</td>
<td>280</td>
<td>2393</td>
<td>22</td>
<td>20</td>
<td>7</td>
<td>98</td>
</tr>
<tr>
<td>2</td>
<td>214</td>
<td>3.8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.6 Discussion: the assignment of dark states

According to our data analysis, two dark states (DS$_2$ and DS$_4$) were resolved by FCS in the pure rhodamine solutions. One or two additional dark states (DS$_1$ and DS$_3$) were observed when a rhodamine dye (TMR) was attached to proteins and a DNA. DNA hybridization, protein mutation and oxygen influenced the significance of these dark states. For further discussion, it is useful to calculate blinking rates by implementing $t_{DS}$ and $f_{DS}$ estimated by correlation curve fitting (Table 3-2) in the equation 3-3. The results, $k_{bd}$ and $k_{db}$ for DS$_{1-4}$, are given in Table 3-5.

Table 3-5. DS population (shaded) and de-population rates (in ms$^{-1}$) based on Table 3-2 data and Eq. 3-3.
3.6 Discussion: the assignment of dark states

3.6.1 Assignment of dark state 1

In ClpP E150C, DS\textsubscript{1} was observed only in the deoxygenated buffer (Table 3-2), which suggests that this is in fact a TMR triplet state (figure 3-10 A). Molecular oxygen is a well-known triplet quencher, and, at normal O\textsubscript{2} concentrations in the buffer solution, it seems to cause DS\textsubscript{1} to relax much faster than its population rate. In the oxygen-depleted buffer, the depopulation rate slows down enough to increase the occupancy of this state to ~10% and make it detectable in the FCS curve. Note that, since measurements are typically done in the presence of oxygen, DS\textsubscript{2} instead of DS\textsubscript{1} was reported in literature as the fastest-decaying triplet state \[3, 4, 7\].

A fast, DS\textsubscript{1}-like decay was also observed in an unstable protein (WT-SH3) under normal O\textsubscript{2} concentrations (Table 3-3). This phase was absent in the stable form (SH3 T22G), and thus could be attributed to protein conformational dynamics, such as segment rotation. This can arise, for instance, by quenching of TMR fluorescence by aromatic residues moving in and out of the fluorophore proximity. If this is the case, one may expect that the correlation curves of the two samples would be similar under denaturing conditions. However, even at 6 M GdmCl, the T22G curve does not show the sub-\(\mu\)s decay component (Figure 3-8 and Table 3-3). This suggests that the local environment of the dye is different in the WT protein \textit{versus} the T22G mutant, which is reasonable considering the non-specific dye labeling of the protein. Both proteins have 5 lysines (LYS6, 21, 26, 44, 56), all being potential sites of amine labeling with the TMR-NHS dye. These lysines will be labeled based on their relative solvent accessibilities. A lysine that is less exposed in the T22G mutant could become the dominant labeling site in the WT protein.

An alternative scenario is that the fast decay does not reflect the SH3 protein dynamics, but it is in fact a triplet state which is quenched in T22G and active in the WT protein. Assuming that the dye label is at different sites in the two cases, upon labeling the WT protein and protein refolding, the dye will be trapped in the protein matrix and protected from oxygen. What seems unlikely is the ability of this local matrix to survive, although with a lower probability (lower \(f_{DS1}\)), in high denaturant concentration. However, there is evidence that some SH3 residues remain in close proximity to each other even under severe denaturing conditions (chapter 4). In the WT protein, the DS\textsubscript{1}-like decay becomes 3 times slower at high [GdmCl], which in this scenario can be attributed to the low oxygen solubility in high salt solutions \[53\] and therefore a reduced triplet depopulation rate.
In conclusion, the fast $\text{DS}_1$-like decay in SH3 could arise either from local conformational dynamics or from induced photophysics, both due to the different labeling sites of TMR in WT SH3 versus the T22G mutant. Further investigations are needed to elucidate the underlying mechanism. The appearance of $\text{DS}_1$ in correlation curve of ClpP protein in deoxygenated buffer supports the induced photophysics scenario. However, still a more solid proof is required, such as the development of $\text{DS}_1$ in the correlation analysis of free dye and T22G mutant in an oxygen depleted buffer. On the other hand, site-specific labeling can help to clarify the role of conformational dynamics in $\text{DS}_1$. To this end, the lysines should be substituted one by one by a cysteine for site-specific labeling. If a mutant shows a fast decay behavior in its correlation curve, the corresponding mobile region contains that mutation site. According to the molecular dynamics simulations of Sarah Rauscher (unpublished data) the LYS 21 and LYS 26 are the least solvent-accessible lysine in the folded state, and therefore likely candidates for mutation in future studies.

3.6.2 Assignment of dark state 2

Neither Rh101, nor Rh110 correlation curves show a noticeable $\text{DS}_2$ component at low excitation intensities (table 3-1). Rh101 is structurally restricted and Rh110 is non-alkylated, therefore these two dyes are not expected to show the TICT state. This strongly suggests that $\text{DS}_2$ could correspond to the TICT state. However, $\text{DS}_2$ is “activated” for both dyes at high excitation intensities, and its population becomes quite significant compared to the other dark states. This power dependence suggests that $\text{DS}_2$ is a hybrid of at least two closely spaced dark states, one of which is a TICT state, which are indistinguishable by FCS.

A similar picture arises from the response of the $\text{DS}_2$ (de)population rates to oxygen, which shows a hybrid behavior. Consider the two steps of increase of oxygen concentration, from deoxygenated Gloxy buffer (few μM) to the TRIS buffer (~1 mM), then to the oxygenated buffer (~3-6 mM). In the first step, the $\text{DS}_2$ depopulation rate $k_{db}$ increases by ~2.5 times, while the population rate $k_{bd}$ stays roughly the same (compare Gloxy and TRIS rows in table 3-5). A further increase of the oxygen concentration in the second phase leads to an increase of both rates by 3 to 4 times. We can understand the two-phase behavior by assuming the existence of parallel competitive deactivation channels, including the oxygen-sensitive TICT state and a second state
which is independent of oxygen in the μM range (figure 3-10). The measured DS$_2$ parameters are averaged values of these two states.

The DS which is less sensitive to oxygen is most likely the lactonic form of rhodamines. We argue that this is the lactone form because twisting of amino group requires breakdown of the nitrogen double bond. Consequently, the positive charge will transfer to the pyran group which can mediate the lactone ring formation (figure 3-5). Secondly, oxygen is a hole-type charge carrier, promoting the $Z\rightarrow L$ transition and not the reverse. In addition this dark state is most probably a ground state because in the lactonic form the absorption peak is shifted to the shorter wavelengths (<350 nm) [16, 54], consequently resulting in higher energy of the excited levels compared to those for the zwitterionic form. Therefore the population of a lactonic dark excited state from the TICT state is unlikely. However, still a lactonic dark excited state is possible to be populated directly from the zwitterionic singlet excited state (see the discussion of DS$_3$ below).

![Figure 3.10.](image)

**Figure 3.10.** The proposed relaxation scheme for DS$_2$. The solid lines represent the excitation and the deactivation pathways corresponding to each observed dark state. Fluorescence is shown by the dashed line.

The results of DNA experiments (table 3-4) can be adequately explained by assigning DS$_2$ to the TICT state. The backbone of the DNA is highly negatively charged. The insertion of the positively charged amine moiety of TMR into the rigid dsDNA grooves will lock it in place. Consequently, the energy barrier between in-plane and tilted states will increase, slowing down the associated relaxation to the millisecond time scale. In ssDNA there are no rigid grooves and the relaxation will be faster (2 μs).

### 3.6.3 Assignment of dark state 3

The DS$_3$ phase, which is only observed in FCS experiments on labeled biomolecules, is most likely induced by intramolecular dynamics. As we mentioned above, intramolecular dynamics can appear in fluorescence correlation curves due to activation/population of a (nearly) inactive
dark state, modulation of an active dark state and/or direct modulation of the fluorescence rate induced, for instance, by quenching by (mobile) aromatic residues.

The presence of DS₃ for all labeled biomolecules and the response of DS₃ to oxygen (tables 3-2 and 3-5) support the fact that this phase is not caused by conformational dynamics and direct modulation of the fluorescence (e.g. by fluorescence quenching or enhancement or by shifting the fluorescence spectra) as previously reported [12]. Instead, we propose that DS₃ is a state with very low population under normal conditions, such that it is invisible in the FCS curves of free dyes. An indication to the nature of this “hidden” dark state can be found in Table 3-1 by comparing $t_{DS₄}$ values for different rhodamine dyes. In terms of diffusion time, Rh110, TMR and Rh101 have a very similar DS₄ relaxation time, $t_{DS₄} \approx 0.25 \tau_d$, while for Rh6G this is significantly larger (~0.35 $\tau_d$). Interestingly, Rh6G is the only dye of the series lacking the lactone form. Let’s assume that DS₃ is due to the population of the lactonic dark state, which is “hidden” in nearly all free rhodamines (figure 3-11). A very low population of DS₃, with a decay time close to that of DS₄, may result in fitting artifacts of the correlation curves of the free dyes. Consequently, assigning a single dark state to the two unresolvable states (DS₃ and DS₄) will result in dark state parameters close to that of DS₄ but with a reduced $t_{DS₃}$.

Figure 3.11. The proposed relaxation scheme for DS₃. The solid lines represent the excitation and the deactivation pathways of the dark state. Dashed lines represent other processes in the system.

The reported 18% population of the lactonic form in water [40] is not inconsistent with the assignment of DS₃ to the lactonic dark state (not the ground state which previously was assigned to DS₂) even though DS₃ was not observed in free dyes. It is because the 82%-18% population of zwitterions-lactones is achieved from the equilibrium in the ground state. The zero fraction of DS₃ in free dyes indicates that the ground-state conversion of lactone to zwitterion is much faster
than the decay rate of the singlet excited state to the lactonic excited dark state (maybe a lactonic triplet state).

The clear appearance of DS₃ in SH3 and ClpP proteins indicate a higher chance of lactone formation in the proximity of nonpolar residues of proteins. Interestingly, based on the \( f_{DS3} \), the population of the lactonic form is 2-3 times higher in ClpP than in SH3, indicating that the fluorophore is located in a more nonpolar environment. In DNA, the phosphate negative charge has the potential to destabilize the charged zwitterionic form and increase the conversion rate to lactone, therefore reducing the \( t_{DS3} \) compared to proteins. In fact, DS₃ lifetime is 20-50% shorter in DNA compared to the SH3 protein and even more (60%) compared to the ClpP protein (table 3-2). Finally, the DS₃ lifetime increases by ~40% in 6 M GdmCl, which is most likely due to the higher viscosity (1.67 cP vs 1.00 cP) and to the fact that the zwitterion-lactone conversion rates are viscosity dependent.

3.6.4 Assignment of dark state 4

The presence of DS₄ in all rhodamine dyes studied strongly suggests that this is not a TICT or lactonic dark state. The observed linear viscosity dependence of DS₄ implies an underlying viscosity dependant mechanism like butterfly motion, photo-ionization or non-gaussian artifacts. Molecular vibrations are known to be viscosity dependent [55], however the relaxation of DS₄ is several orders of magnitude slower than the time scale of vibrational motions (ps-ns). Therefore butterfly motion is not a dark state itself, but it can help to couple electronic states of the excited fluorophore. On the other hand, slow, millisecond scale photo-ionization processes were recently reported in rhodamine solutions [23].

The ratio \( t_{DS4}/\tau_d \) for free rhodamine dyes in glycerol solution falls in the 25-35% range (table 3-1). This fraction resembles the reported fictitious dark state related to the non-gaussian detection volume. The 50-µm pinhole used in our setup rejects 30% of the fluorescence signal from 200 nm beads located in the centre of focal spot, implying an aperture size of 1 optical unit. Therefore, we do not expect a fraction greater than 2% for the fictitious decay in our measurements [43]. In fact, upon increasing the pinhole size to 75 µm, \( \tau_d \) and \( t_{DS4} \) increased by different amounts, 57% and 21%, respectively (table 3-1). This implies that \( t_{DS4} \), while it is increasing with viscosity, is not proportional to the diffusion time and it is therefore not fictitious.
$t_{DS4}/\tau_d$ was changed between 25 and 35\% for free dyes (table 3-1) and between 25 and 47\% for TMR-labeled biomolecules (table 4-2,4-3 and 4-4). This indicates that DS$_4$ is a dark state which is sensitive to the dye structure and its environment. For example, for ClpP the maximum ratio of 47\% was found in the Gloxy buffer and the minimum ratio of 25\% was observed, just oppositely, in the oxygenated buffer. This shows the sensitivity of the $t_{DS4}$ to oxygen and hence to the viscosity, according to the Stern-Volmer equation [20, 21]. This sensitivity to oxygen is as one would expect for photo-ionization followed by electron trapping (figure 3-12). Remarkably, the $t_{DS4}/\tau_d$ found for Bodipy FL was significantly (by a factor of 2-3) lower than those of the dyes of the rhodamine family.

![Figure 3.12](image)

Figure 3.12 The proposed relaxation scheme for DS$_4$. The solid lines represent the excitation and the deactivation pathways of the dark state. Dashed lines represent other processes in the system. The assumption of an intermediate triplet state is made based on [23] (see section 3.4.5).

The variations of the DS$_4$ population, $f_{DS4}$, among the rhodamine are more significant than the $t_{DS4}$ variations. For free dyes at low excitation intensity, TMR showed the highest fraction at 19\%, while at the other end, Rh101 only showed about 10\%. High oxygen concentration reduced the DS$_4$ population of the ClpP-TMR from 26\% to 20\% (table 3-2).

Based on these observations we are inclined to accept DS$_4$ as a real dark state; although at this point we cannot completely reject the possibility that this decay has a contribution from optical artifacts as the 3D gaussian profile of the detection volume which is assumed in the FCS model is just an approximation. The linear viscosity dependence of DS$_4$ makes it hard to separate from the upper-limit diffusion component and investigate it in detail by FCS. Based on our data so far, we speculate that photoionization is the main underlying process of the slowest phase observed in the correlation curves.
3.7 Summary

We observed 2 dark states, with decay times of 1-5 μs and at least 60 μs, respectively, upon FCS analysis of free rhodamine dyes in viscous solutions. After labeling biomolecules (proteins and DNA) with one of the rhodamines (TMR), 1 or 2 additional dark states were observed in the correlation curves. One of these states is the fastest one of all (DS₁, decay time ~0.5 μs), and we attributed it to a rhodamine triplet state which is heavily quenched by oxygen in normal aqueous buffers. By comparative analysis, we concluded that the other 3 dark states are also due to rhodamine photophysics: a twisted-intramolecular-charge-transfer state (TICT) state (DS₂, decay time 1-5 μs), a lactonic dark state (DS₃, decay time 20-40 μs), and photo-ionization followed by electron trapping (DS₄, decay time >60 μs). In the FCS literature, the commonly reported dark state with a lifetime of a few μs is usually attributed to the triplet state of the dyes. Our results show that this is a hybrid dark state consisting of a TICT and another oxygen insensitive (most likely, lactone) dark state. A summary of the assignment of dark states in this study is given in figure 3-13.

Organic dyes like rhodamines have a very rich excited state photophysics which is distributed over the whole observation time window in FCS measurements. Accordingly, the application of PET-FCS to measuring the intramolecular dynamics of biomolecules requires extra considerations. Our experiments and analysis suggests that intramolecular dynamics mainly appeared in the correlation curve via modulation of the dark states of the fluorophore. This is of critical importance, as the common approach in literature is to directly interpret the relaxation times in FCS measurements as the time constant of conformational dynamics of biomolecules.
without a comprehensive analysis and solid justification. In this context, our results could be used to pinpoint the mechanism of modulation of correlation curve by conformational fluctuation, to identify appropriate kinetic components that most likely reflect this dynamics and to screen and identify fluorophores that are most sensitive to intramolecular dynamics.

3.8 Bibliography


Chapter 4

Size, shape and folding of the SH3 Domain of Drk Drosophila Adaptor Protein

Abstract

Src-homology-3 (SH3) is a commonly found modular binding domain that mediates protein-protein interactions. The N-terminal SH3 domain (6.8 kDa) of the Drosophila adaptor protein Drk is marginally stable under physiological conditions and exists in nearly 50/50 equilibrium between the folded and unfolded states with slow interconversion rates which make it an excellent model system for the study of unfolded states of proteins.

Fluorescence correlation spectroscopy (FCS) experiments performed on the wild-type and on a single-site mutant that stabilizes the folded state provided the hydrodynamic radii of the folded \( (R_{h,F}) \), unfolded \( (R_{h,U}) \) and denatured \( (R_{h,D}) \) states at 20°C. Quite surprisingly, the results indicate that the unfolded state under non-denaturing conditions is hydrodynamically less compact than the chemically-denatured state. The ratio of the \( R_{h,D} \) and previously reported radius of gyration of the denatured state \( (R_{g,D}) \), led to a shape factor predicted for circular polymers.

A dual cysteine mutant (E2C-61C-G62) was labeled with a FRET pair. Single molecule FRET experiments were performed to quantify the end-to-end size distribution and thus the heterogeneity of the unfolded and denatured state ensemble. Strikingly, even at very high denaturant concentration a heterogeneous high FRET cluster persist in the FRET histogram indicating presence of long-lived non-local interactions that brings the ends close to each other. The nature of high FRET scaffold was investigated by several osmolytes and it is concluded that at least two different conformations contribute in the high FRET cluster. Cooperative nonpolar contacts, hydrogen bonds and salt bridges are discussed as possible mechanisms of loop formation in the denatured state.
4.1 Introduction

Historically the structure–function paradigm was developed long before the determination of the first protein structure [1]. In 1894, Fischer explained the specificity of enzymes by the lock and key model [2]. Numerous theories and experiments supported the link between structure and activity of proteins thereafter. For example the loss of a protein’s function in denaturants is commonly associated with the loss of 3D structure [3]. This description was widely accepted until NMR studies demonstrated physiologically relevant protein disorder under non-denaturing buffer conditions [4, 5]. Therefore, the focus of the structural biologist was mainly on the well-defined folded states. Today there are 85582 solved protein structures in the Protein Data Bank.

Until the 1980’s, a dominant view of protein folding was as a highly cooperative, all-or-none process, essentially described by a two-state model [6]. This included a highly structured folded state, also known as the native state, and a highly disordered state with almost no secondary or tertiary structure, also known as the random-coil or the denatured state. In 1983 Ohgushi et al. found a folding intermediate in cytochrome c [7]. This state, which conserves a native-like secondary structure with a dynamic tertiary structure, was referred to as the molten globule (MG) state. Pre-molten globule (pre-MG) is a transient intermediate state which has been observed in the early stage of folding (within the first few msec) of several proteins in stopped-flow experiments [8]. Pre-MG is defined to have non-native secondary structure and no long-range interaction. Unfolded state is a general name used for both compact unfolded states under non-denaturing conditions (U\text{exch}) and the denatured states (U\text{denatured}). However, in this study we use the term only for the first case.

However, with the discovery of intrinsically disordered proteins (IDPs), a consensus emerged in the 1990s that the native functional states of many proteins are disordered. The disordered state is defined as lacking a stable tertiary structure, although different than a statistical random coil. It therefore resembles a (pre-)MG with some secondary and transient tertiary structures [9] and, more generally, the unfolded ensemble of proteins [10]. It is estimated from bioinformatics approaches that 33 % of proteins contain long disordered regions (> 30 residues) [11]. Many of these proteins are involved in cell signaling, molecular recognition and transcriptional regulation [11-13]. In addition, the majority of oncogenic proteins have significant disordered regions [14].
In spite of the biological importance of disordered states of proteins, there was little success in their characterization either by computational and conventional experimental approaches. This is because describing a dynamic ensemble of interconverting structures is significantly more challenging than a well-defined singular structure of a folded state and it cannot be characterized only by few parameters [15, 16]. However, single-molecule fluorescence (SMF) techniques are capable of resolving molecular heterogeneities and can provide detailed information about the conformational populations and interconversion dynamics of proteins. SM measurements are performed in the very low concentration range of pico-nanomolar and therefore they can be considered free of aggregates. Protein aggregates can cause artifacts at standard conditions used in conventional techniques like nuclear magnetic resonance (NMR) or small angle X-ray scattering (SAXS).

In addition and probably more relevant for the protein research field, the information obtained from SMF techniques such as the end-to-end ($R_{\text{EE}}$) or intra-chain distance distributions, the hydrodynamic radii ($R_h$) and the interconversion rates can be used as new experimental restraints to molecular dynamic (MD) simulations and hybrid experimental-simulation approaches [15, 17]. The ENSEMBLE algorithm is an example of a hybrid approach [15, 18] that determines which sets of randomly generated structures together generate calculated ensemble-averaged parameters such as J-couplings, chemical shifts, hydrodynamic radius and solvent accessibility most consistent with experimental data. The ENSEMBLE algorithm has been applied to characterize the unfolded state of a marginally stable protein, the N-terminal Src-homology-3 domain of Drosophila Drk (Drk N SH3) [15, 19].

The sets of structures determined to represent the experimental disordered state, however, are not unique. It is always possible to find different structure sets that can produce the same average experimental values. In addition, calculations of experimental observables from structures rely on approximations that may not hold precisely. For example, the $R_h$ values calculated using the HYDROPRO program [20] and the crystal structure of globular proteins are typically smaller than the corresponding measured values [21]. In contrast, the intrachain distance distributions, including $R_{\text{EE}}$, are directly obtained by SM(F) techniques such as FRET can significantly enhance the accuracy of the structural refinement.
In this chapter we are using DrkN SH3 as a model system for the study of unfolded states of proteins [17, 19, 22-25]. First, we discuss a disagreement between computational and experimental methods for the determination of the size of the unfolded state under non-denaturing conditions and the chemically denatured state of the protein. Our fluorescence correlation spectroscopy (FCS) experiments confirm the results obtained previously by NMR and SAXS, but at a much lower sample concentration (nM instead of mM). We performed single-molecule Förster resonance energy transfer (smFRET) measurements between the two ends of the protein in various osmolyte solutions (e.g. sodium sulfate, glycerol, Guanidinium hydrochloride (GdmCl), urea and formamide) in order to disentangle the conformational landscape of DrkN SH3. Surprisingly, a heterogeneous looped state (with small R_{EE}) was observed even under high denaturant concentration. The presence of these looped but denatured states in the FRET histogram has been attributed to a looping mechanism possibly associated with cooperative nonpolar contact, salt bridge and hydrogen bond formation.

### 4.2 Experimental

#### 4.2.1 Materials

The fluorescent probes used in the experiments were: 5-carboxy-tetramethylrhodamine N-succinimidyl ester (TMR-NHS) (81125, AnaSpec, Fremont, CA, USA), Alexa 555-NHS (A555-NHS) (A20009, Life Technologies Inc., Burlington, Canada), Bodipy Fluorescein maleimide (BFL-maleimide) (B-10250, Life Technologies Inc., Burlington, Canada) and Alexa 647-maleimide (A647-maleimide) (A-20347, Life Technologies Inc., Burlington, Canada). Rhodamine 6G (Rh6G) (R4127, Sigma Aldrich, Oakville, Canada) was used for the calibration of detection volume in FCS measurements. Fluorescent polystyrene beads (F8809, Life Technologies Inc., Burlington, Canada) were used for the alignment of the excitation and detection paths in the confocal microscope.

Guanidinium chloride (GdmCl) (G9284, Sigma Aldrich, Oakville, Canada), urea (9510, VWR International, Mississauga, Canada) and formamide (47671, Sigma Aldrich, Oakville, Canada) were used for protein denaturation assays. GdmCl and urea solutions are prepared in MilliQ (MQ) water, resulting in pH values of 6.0 and 8.2, respectively, with less than 2% variation. All other samples unless clearly stated were diluted in TRIS buffer (50mM Tris,
150 mM NaCl, pH 7.5). Formamide, glycerol (4750, VWR International, Mississauga, Canada) and Na₂SO₄ (238597, Sigma Aldrich, Oakville, Canada) solutions were also prepared in TRIS buffer. L-Arginine (11009, Sigma Aldrich, Oakville, Canada) is used as an intermediate state suppressor [26, 27].

4.2.2 Protocols

4.2.2.1 Purification of Urea

Cyanate is a decomposition product of urea and irreversibly interacts with amino groups to form carbamoyl derivatives [28]. The hydrolysis products and ionic impurities, which exist even in the best grades of urea, should be removed for background-free smFRET experiments. The purification protocol for urea is described in detail elsewhere [28]. Briefly, a 7 M solution of urea was prepared in MilliQ water and passed 3 times through Amberlite MB-1 resin (M8032, Sigma Aldrich, Oakville, Canada) to achieve a deionized urea with a conductivity similar to that of the MilliQ water.

4.2.2.2 Protein expression and purification

The protein expression and purification was carried out in the Forman-Kay Lab (Sick Kids Hospital, UofT). A plasmid coding for the terminal SH3 domain of Drk (DrkN SH3), residues 1 through 59 of the protein Drk [25], was transfected into Escherichia coli HMS 174 cells under the control of the T7 promoter. The expression of the protein was induced for 4 h at an OD600 of 0.6 by addition of 250 mg/liter IPTG to bacterial growths at 37 °C in M9 minimal medium, supplemented with 0.3% ¹³C labeled glucose, 0.1% ¹⁵NH₄Cl, 100 mg/L ampicillin, 10 mg/L thiamine, 10 mg/L biotin and 1 mM of each MgSO₄ and CaCl₂. Cells were lysed by sonication in 50 mM Tris, 2 mM EDTA, 5mM benzamidine HCl and 7 mM β-mercaptoethanol. The DrkN SH3 domain was purified on a DE 52 ion-exchange column with a linear gradient of NaCl (0-1 M) followed by Superdex 75 gel filtration column in 0.15 M NaCl, 50 mM Tris, 2 mM EDTA, 5mM benzamidine HCl and 7 mM β-mercaptoethanol, followed by a Mono Q ion-exchange column with a linear gradient of NaCl (0-0.3 M). The T22G [29] and the C2 C61 G62 mutants were expressed and purified in a similar manner to the wild-type (WT) protein. The identity and purity of the three protein samples were confirmed by mass spectrometry.
4.2.2.3 Protein labeling

WT and T22G proteins were labeled non-specifically with TMR-NHS and A555-NHS dyes. The NHS-amine coupling reaction was performed in PBS buffer with pH adjusted to 8. The reaction was started by adding the NHS-ester-activated fluorophores to a 50 μL solution of 200 μM proteins at a molar ratio dye:protein of 1:2. This low labelling ratio ensured that most proteins had only one dye label attached. The sample was gently shaken for 3 hours in the dark at room temperature. The excess dye was removed by size-exclusion chromatography using Sephadex G-25 gels (G2580, Sigma Aldrich, Oakville, Canada) in a BioLogic LP system (731-8300, Bio-Rad, Mississauga, Canada). A typical labeling efficiency of ~5% was calculated based on the absorption spectra of the product, which was further confirmed by mass spectrometry.

Thiol-reactive dyes, BFL1-maleimide and A647-maleimide, are used for the site-specific labeling of the SH3 C2 C61 G62 protein. 10 μL DMSO was added to a 50 μL of 100 μM protein in TRIS buffer to improve the solubility, and hence the labeling efficiency of the BFL. TCEP was added to the protein at a 20:1 molar ratio to reduce the disulfide bonds. Then BFL was added to the mixture at ~5:1 molar ratio with the protein. Oxygen was removed by flushing the sample with argon gas in a desiccator for ca. 5 min. Then the vial was capped tightly and the reaction was allowed to continue by gentle vortexing for ca. 12 hours at room temperature. Afterwards, A647 was added at a 20:1 molar excess to the protein, the solution was flushed with argon in a vacuum desiccator, and the reaction continued for 3 days at room temperature in the sealed vial. The remaining free dye was removed by size-exclusion chromatography. A labeling efficiency of ~4% and ~70% for BFL and A647, respectively, were estimated from the absorption spectra of the labeled proteins using the protocol provided by the supplier.

For spectral calibration purposes, donor-only and acceptor-only proteins were prepared using a similar protocol. Using these control samples, the quantum yields of the fluorophores attached to the SH3 protein were found to be 55% for BFL and 37% for A647.

4.2.2.4 Sample preparation

All samples were diluted to donor-protein concentrations of 1-10 nM and 50-100 pM, which are most suitable for FCS and smFRET burst experiments, respectively. For a typical experiment, a sample solution of ca. 50-100 μL was dropped on the surface of plasma-cleaned
coverslips. The nonspecific protein adsorption to the coverslip was prevented by adding 0.005% (v/v) Tween-20 (P2287, Sigma-Aldrich, Canada) to the sample [30]. This worked well for the denaturation assay, while for the other experiments the plasma-cleaned coverslips were coated with bovine serum albumin (BSA) (15260-037, Life Technologies Inc., Burlington, Canada) [31].

4.2.3 Instrumentation

Time-resolved fluorescence measurements were performed on a multiparameter confocal microscope that was described in detail in Chapter 1. CW laser excitation at 532 nm and 473 nm was used for all FCS and smFRET experiments and the measurements were done in the T2 mode of PicoHarp photon counting module (PicoQuant), as described in chapter 1. For the lifetime analysis, the femtosecond laser (Tsunami HP, Spectra Physics) was tuned to 1055 nm to obtain 527.5 nm pulses after frequency doubling and data was acquired on PicoHarp in the T3 mode (see chapter 1).

For FCS measurements, the fluorescence from the sample went through a 50 μm pinhole and then it was split by a broadband polarizer cube into its polarization components. Each component was focused onto a separate single-photon avalanche diodes (SPADs) (PD5CTC, MPD, Italy). In the FRET-FCS experiment, the fluorescence was split in four different channels. The signal was split by a dichroic filter (FF560-Di01, Semrock, Rochester, NY, USA) into the donor and acceptor contributions, each of them being further divided into their polarization components. The same type of SPAD detectors was used for all four detection channels.

To perform FRET-FCS measurements, the 50 μm pinhole was substituted by a 100 μm one, the polarizer cubes were removed and a red-sensitive SPAD (COUNT, Laser Components, USA) was installed in the acceptor channel. The donor and acceptor signals were cleaned spectrally using bandpass filters (BP530/50m and HQ690/70, Chroma, Bellows Falls, VT, USA). A custom-written LabView code was used to construct the correlation curves using logarithmic time binning and 24 bins per temporal decade (see Chapter 1). A LabView program developed in the Gradinaru Lab by Baoxu Liu was used for the smFRET burst analysis to construct smFRET histograms. The burst recognition is based on the sliding window burst search algorithm [32].
4.2 Experimental

4.2.4 Procedures

4.2.4.1 Fluorescence correlation spectroscopy

Under some simplifying assumptions, such as a 3D Gaussian shape of the detection volume, the fluorescence correlation function for free Brownian diffusion of a single molecular species with dark-state (DS) blinking is given by [33]:

\[ G(\tau) = \frac{1}{N_{\text{eff}}} \left( 1 + \frac{\tau}{\tau_d} \right)^{-1} \left( 1 + \frac{\tau}{s^2 \tau_d} \right)^{-0.5} \sum f_{\text{DS},i} \left( 1 + \frac{f_{\text{DS},i}}{1 - f_{\text{DS},i}} \exp\left( -\frac{\tau}{t_{\text{DS},i}} \right) \right) \]  

(4-1)

In equation (4-1), \( N_{\text{eff}} \) is the average number of molecules in the detection volume, \( V_{\text{eff}} = \pi^{3/2} z_0 w_0^2 \), and it is given by the inverse of the diffusion-related amplitude, \( G_0 = N_{\text{eff}}^{-1} \). In addition, \( s \) is the structure parameter, namely the ratio between the axial and the lateral radii of the detection ellipsoid (\( s = z_0 / w_0 \)), \( \tau_d \) is the diffusion time, which is related to the diffusion coefficient (\( \omega_0^2 = 4D \cdot \tau_d \)), \( t_{\text{DS},i} \) is the dark state relaxation time and \( f_{\text{DS},i} \) is the fraction of molecules of the \( i^{th} \) dark state.

For samples consisting of a mixture of non-interactive diffusive species, the net correlation function is given by the brightness-weighted sum of the individual correlation functions [34]:

\[ G(\tau) = \sum N_i^2 \eta_i^2 G_i(\tau) / \left( \sum N_i \eta_i \right)^2 \]  

(4-2)

where \( N_i \) is the number of molecules of species \( i \) in the detection volume and \( \eta_i \) is their specific molecular brightness.

4.2.4.2 smFRET-burst

Single-molecule fluorescence bursts were identified using the sliding burst search scheme [32]. For each burst identified, the FRET efficiency can be calculated based on the number of detected photons in the donor \( (n_D) \) and acceptor \( (n_A) \) channels [35]:

\[ E = \frac{n_A}{n_A + \gamma n_D} \]  

(4-3)
where $\gamma$ is the ratio of the detection efficiencies ($\zeta$) and the quantum yields ($\Phi$), $\gamma = \zeta_A \Phi_A / \zeta_D \Phi_D$. In addition, corrections were applied on both $n_D$ and $n_A$ to subtract the background and the spectral cross talk [36]. The histogram of the calculated FRET efficiency values of individual burst events were used for the smFRET burst analysis.

The distance between the donor and the acceptor probes, $R$, can be calculated from the FRET efficiency by the following equation [35]:

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{4-4}$$

where $R_0$ is the Förster radius and is the interdye distance that results in 50% FRET efficiency. The dependence of $R_0$ to the spectral characteristics of FRET dye pair is described by [35]:

$$R_0^6 = \frac{9000 \ln 10 \kappa^2 \Phi_D J}{128\pi^5 n^4 N_A} \tag{4-5}$$

where $n$ is the refractive index of medium, $N_A$ is the Avogadro number, $\kappa^2$ is the dipole-dipole orientation factor, often assumed to be $2/3$, as calculated for freely rotating dipoles. $J$ is the overlap integral of donor emission and acceptor absorption spectra. $R_0 = 44.4$ Å was estimated for BFL and A647 dyes conjugated to the DrkN SH3 protein in TRIS buffer (pH 7.5).

To calculate the smFRET efficiency under high denaturant concentration, the fluorescence spectra and the quantum yield of the donor and acceptor dyes were measured in similar solutions. The shift of fluorescence spectra will affect the analysis in two ways. Firstly the spectral shift relative to the spectra of the used band pass filters in donor and acceptor channels will change the detection efficiencies. The maximum calculated change of detection efficiency ($\sim 21\%$) was in the donor channel due to a $\sim 5$-nm red shift of the emission spectrum of BFL in 8M GdmCl. Secondly, due to the shift of the emission spectrum, the spectral overlap, $J$, in equation 4-5 must be adjusted, thus changing the value for $R_0$.

The upper bound for the shot noise standard deviation ($\sigma_{sn}$) of a FRET peak is given by [36]:

$$\sigma_{sn}^2 \leq \frac{\langle E \rangle(1 - \langle E \rangle)}{N_T} \tag{4-6}$$
where $\langle E \rangle$ is the mean FRET efficiency and $N_T$ is a threshold imposed on the total number of photons in a bin. For the folded conformation of a protein, the standard deviation (SD) in the peaks of the FRET efficiency is expected to approach the shot noise limit. On the other hand, the heterogeneous ensemble of the unfolded conformations may result in a broader FRET peak.

### 4.3 DrkN SH3

Drk, Downstream of receptor kinase, is an adaptor protein essential in the development of the Drosophila eye. It consists of a single Src homology 2 (SH2) domain and two flanking Src homology 3 (SH3) domains [37]. The SH2 domain binds phosphotyrosine-containing peptides while the SH3 domains bind short proline-rich sequence motifs [38].

The N terminal SH3 domain (figure 4-1) is a 59 residue-β-barrel consisting of five anti-parallel β-strands (PDB 2A36). It is marginally stable, with the folded ($F_{exch}$) and unfolded ($U_{exch}$) states almost equally populated under normal conditions [25, 39-41], and interconverting with a slow exchange rate, $\sim 2 \text{ s}^{-1}$ [40]. Exposing the protein to 2 M guanidinium chloride (GdmCl) will significantly denature the wild type (WT) protein [22, 25, 42]. On the contrary, the addition of 0.4 M Na$_2$SO$_4$, or a single Threonine mutation to a Glycine, T22G (PDB 2A37), fully stabilizes the folded state of DrkN SH3 [39, 43].

![Figure 4-1](image.png)  
**Figure 4-1** The solution structure of the DrkN SH3 domain (PDB ID: 2A36).
4.3 DrkN SH3

4.3.1 Motivation: simulations vs. experiments

The gyration radii ($R_g$) of the $F_{exch}$, $U_{exch}$ and the chemically denatured state ($U_{GdmCl}$) of SH3 have been determined by SAXS experiments to be $11.7\pm0.5$, $16.7\pm1.4$ and $21.9\pm0.5$ Å, respectively [41]. Pulsed-field-gradient NMR (PFG-NMR) measured the diffusion coefficients, and therefore the hydrodynamic radii, of the two states coexisting at physiological conditions. For the $F_{exch}$ state, the reported $R_{h,F} = 15.6\pm0.2$ Å results in a shape factor $R_g/R_h \approx 0.75$, in agreement with what is expected for globular folded proteins ($R_g/R_h = \sqrt{3}/5 \approx 0.775$) [21]. For the $U_{exch}$ state, $R_{h,U} = 20.3\pm0.2$ Å, resulting in a slightly larger shape factor $R_g/R_h \approx 0.82$. Based on these studies, the average size of the native unfolded state, $U_{exch}$, was estimated to be 30-40% larger than the native folded conformation, $F_{exch}$.

![Figure 4-2](image.png)

**Figure 4-2.** The $R_{EE}$ distribution of the folded ensemble of WT SH3 calculated from a canonical MD simulation starting from the NMR structure. Three examples of simulated conformations are shown. (Sarah Rauscher & R. Pomès, reproduced with permission).

Molecular Dynamic (MD) simulations of WT SH3 were carried out by the Pomès group (Biochemistry, University of Toronto, data not published). These simulations perform both canonical ensemble (constant temperature) and generalized-ensemble sampling (random walk in temperature) with replica-exchange methods [44]. The canonical MD simulations of the WT SH3 starting from NMR structure of the folded state shows a relatively heterogeneous $R_{EE}$ distribution of the folded state (Figure 4-2). The $R_{EE}$ values found by the generalized-ensemble algorithm shows a very compact distribution at 300 K, with a sharp peak around 5 Å, a broad peak around 20 Å and almost no population above 40 Å (figure 4-3). The $R_{g,U} = 11.11\pm0.03$ Å from of the
4.3 DrkN SH3

The unfolded ensemble is only 10% larger than the folded $R_{g,f} = 10.09 \pm 0.03$ Å. The hydrodynamic radii were calculated by HYDROPRO and also gave a very similar compactness ratio, $R_{h,U}/R_{h,F} = 1.09$, considerably less than the 30-40% derived from SAXS and NMR experiments. In addition, the denatured state was simulated in 2M GdmCl and $R_{g,D} = 10.92 \pm 0.04$ Å, which deviates even more from the value found experimentally, 21.9±0.5 Å [41]. NMR and SAXS experiments were performed in relatively high concentrations, on the order of 1 mM, and therefore they might be prone to aggregation effects. In addition, a time-dependent increase of $R_g$ was observed in SAXS experiments, probably due to radiation damage of the protein [41].

![Figure 4-3](image) The end-to-end distance (REE) distribution of WT SH3 at 300 and 802 K. (S. Rauscher & R. Pomès, reproduced with permission).

The force field is a technical term for a potential energy that describes a molecular structure and is a function of parameters such as partial bond lengths, bond angles and partial charges of atoms. In MD simulations of proteins, the force field parameterization of different atoms is typically calculated from small organic molecules. In fact, there are a large number of force fields used for proteins with considerably different potential terms [45]. Therefore, the accuracy of the underlying potential function and the solvent model are critical for capturing the essential features of the molecular structure in liquid solvent [45, 46]. In recent years, MD simulations of the disordered states of protein have been published [47-51]. Consequently, even more than before, we need a careful assessment and comparison of the computational methods with experiments.
In the 60’s, Tanford showed that, at a high concentration of chemical denaturant (6M GdmCl), proteins obey a scaling law expected from a random coil [52]. Since then, the random-coil model became very popular as the basis of our understanding of the denatured state of proteins. A closely related question to the characteristics of the disordered protein conformations is the validity of the random-coil model for the denatured state. Recent observations of substantial residual helical and even native-like structures by NMR at a high denaturant concentration (6-8 M GdmCl or urea) challenge the random-coil model [53-55]. Figure 4-4 shows the measured or computed gyration radius of the disordered/denatured state relative to that of the folded state for a number of proteins. The blue symbols are experimental values [21, 41, 56, 57] and the red ones are found from MD simulations [58-60]. In the simulation of α-synuclein and protein L the radii are reported for their disordered states. The $R_g$ of apomyoglobin from simulation is for its thermally denatured state. All the other proteins were chemically denatured in GdmCl or urea solutions. For four proteins (apomyoglobin, α-synuclein, Protein L and DrkN SH3), both experimental and theoretical values are available. For each of them, there is a clear trend of underestimating the size of the denatured or unfolded state in the MD simulations.

Figure 4-4. Normalized $R_g$ of the disordered/denatured state by the $R_g$ of the folded state vs the number of residues (N) for a number of proteins. The data for SH3 are shown by stars. The data are from 1: [56], 2: [57], 3: [21], 4: [41], 5 and 8: Sarah Rauscher (communication, unpublished), 6: [58], 7: [59], and 9: [60].

In this study we are using single-molecule fluorescence techniques to investigate the aforementioned disagreement. Single-molecule techniques, such as FRET and FCS, are very suitable for protein research because they avoid the high-concentration artifacts that may occur in
NMR and SAXS measurements and, also, they provide additional restraints (i.e. $R_{EE}$) to refine the MD simulations. In addition, they have an immense potential to address the current confusion about the nature of denatured state, which is difficult to interrogate with other techniques.

### 4.4 FCS: Size and shape of $F_{\text{exch}}$, $U_{\text{exch}}$ and $U_{\text{den}}$ states

FCS experiments were performed on a homebuilt customized setup in order to measure the $R_h$ ratio and the fraction of unfolded/folded protein at 20°C. Without prior information about diffusion properties or relative populations of the diffusive components, Meseth et al.[61] showed that the discrimination between two diffusive components with a $R_h$ ratio of less than 1.6 is impossible by standard one-color FCS on a mixed sample. The estimated $R_h$ ratio of 1.3 (1.1) determined by NMR (MD simulation) is too small to fulfill this criteria. One way to improve the sensitivity of the FCS measurement in these conditions is to determine the diffusion parameter of one of the species by performing FCS on a pure sample and then use it in the data analysis of the mixed sample. Here, the hydrodynamic radius of the folded state was measured by using the T22G mutant, which is fully folded at room temperature. However, instead of fixing the diffusion parameter of the folded state, we fitted the FCS curves of both the T22G mutant and the WT SH3 with linked parameters.

A series of FCS experiments were carried out upon labelling the proteins at their free amine groups with two different dyes, either A555 or Tetramethylrhodamine (TMR). A555 is heavier (~1250 Da) and negatively charged and is known for not being sensitive to environment. One the other hand, TMR is a small (527 D), neutral and environment-sensitive dye. TMR fluorescence lifetime is 3.6 ns in nonpolar environments and 2.4 ns in aqueous (polar) solutions. In addition, aromatic residues (tyrosine and tryptophan) are known to quench TMR and reduce its fluorescence lifetime to ~1.5 ns) [62].

Figure 4-5 presents the FCS data and fitted curves for WT and T22G SH3. As expected, the correlation curve of the WT protein is slightly shifted to longer decay times because of the unfolded contribution. The FCS-measured unfolded/folded hydrodynamic radius ratios of $R_{h,u}/R_{h,f} = 1.29\pm0.09$ from A555-labeled proteins and $R_{h,u}/R_{h,f} = 1.31\pm0.05$ from the TMR-labeled proteins are in an excellent agreement with the reported value of 1.30±0.01 from PFG- NMR experiments. The fluorescence lifetime analysis of A555 in both WT and T22G samples gave the same lifetime
components (0.80 ns, 1.58 ns) with the same fractions (0.80, 0.20). This suggests that the photophysics of the dye is not influenced by intramolecular dynamics of the protein. Therefore, it can be assumed that the dye has the same dark-state parameters in both folded and unfolded states. Consequently, the FCS-measured fractions of the diffusive components can be directly attributed to the relative population of the states (see eq. 4-2). The FCS-measured folded fraction of 0.52 for A555 labeled protein is very close to the literature values of ~0.50 at 30 °C [25, 39], 0.66 at 20 °C [40] and 0.58 at 5 °C [41].

![Figure 4-5](image)

**Figure 4-5** Intensity correlation data of WT SH3 (red dots) and SH3 T22G (blue dots) labeled with A555. The solid red and blue lines represent the fitting results for the WT and T22G samples, respectively. Two representative protein configurations for the folded and unfolded states are shown for illustration purposes only (provided by Sarah Rauscher, Pomès group)

The lifetime histogram of the TMR-labeled T22G mutant was well fitted by a single exponential component (2.73 ns). However the WT SH3 shows a bi-exponential behavior with lifetimes of 2.1 ns and 3.3 ns, with relative fractions of 0.49 and 0.51, respectively. These fractions resemble the unfolded and folded populations (0.48, 0.52) found by FCS analysis of the A555-labeled proteins. Because of the change in the photophysics of the TMR, the dark state parameters may differ for the two proteins, and they cannot be shared in the fitting analysis. In fact, an additional dark state decaying in ~0.2 μs appeared in the WT SH3 data (figure 4-6), while for \( \tau > 1 \) μs, the dark states parameters are almost the same for the two states. The fast sub-μs decay was absent in the correlation curves of the free dyes and A555-labeled proteins (data not
shown). Thus, we applied global fitting of FCS data for $\tau > 6 \mu s$ when the proteins labeled with TMR and for $\tau > 0.2 \mu s$ when the proteins labeled with A555.

![Figure 4-6](image.png)

**Figure 4-6** Correlation curves of TMR-labeled SH3 (WT and T22G) in Tris buffer (pH 7.5) and in 6M GdmCl. The solid lines are fits from non-global analysis and assumption of a single diffusive component for all samples.

The appearance of the sub-$\mu$s term only in the WT SH3-TMR could be attributed to protein structure fluctuations such as segmental motions of protein subdomains during (local) unfolding. To further investigate the nature of the $0.2 \mu s$ decay, we performed FCS measurements of both WT and T22G in a 6M GdmCl solution (pH 7.5) (figure 4-6). Interestingly, the sub-micron second decay appears only in WT SH3-TMR but with a slower time constant ($\sim 0.9 \mu s$ at 6M GdmCl). This suggests that the local environment of the dye is different in the WT protein versus the T22G mutant. Therefore, the fast decay could arise either from local conformational dynamics or from induced photophysics, both due to the different labeling sites of TMR in WT SH3 versus the T22G mutant. Both proteins have 5 lysines, all being potential sites of amine labeling with the TMR-NHS dye. These lysines will be labeled based on their relative solvent accessibilities. A lysine that is buried in the T22G mutant could become the dominant labeling site in the WT protein. For further discussion please refer to section 3.6.6.1.

A summary of $R_h$ and $R_g$ of WT SH3 measured by NMR and SAXS experiments [41] and by our FCS analysis is provided in table 4-1. The radii of the unfolded and the denatured states are normalized with respect to the corresponding folded state radius. The NMR and SAXS measurement of the denatured protein is reported in 2M GdmCl. FCS-measured hydrodynamic
radii ($R_{h,d}$) of the denatured states of WT SH3 in 2 M and 6 M GdmCl are almost identical (less than 0.5% difference), implying the protein is fully denatured in 2 M GdmCl. The viscosities of the 2 and 6 M GdmCl solutions (1.13 and 1.67 cP, respectively) were determined by FCS measurements of Rhodamine 6G in these solutions and used to correct the $R_h$ values estimated for the denatured state. The normalized $R_h$ of the T22G mutant in 2 M GdmCl was 1.18±0.03 implying this mutant is not completely denatured, in agreement with its higher stability. The denatured proteins data were fitted using a single diffusive species. The obtained denatured radius is thus the average radius of the denatured ensemble.

Table 4-1 Hydrodynamic ($R_h$) and gyration ($R_g$) radii of the folded, unfolded and denatured$^1$ states of SH3. Unfolded and denatured state parameters are normalized to those of the folded state.

<table>
<thead>
<tr>
<th></th>
<th>Folded (Å$^3$)</th>
<th>Unfolded (normalized)</th>
<th>Denatured (normalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCS</td>
<td>$R_h$</td>
<td>14.4±0.2</td>
<td>1.31±0.05</td>
</tr>
<tr>
<td></td>
<td>$R_g$</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>NMR$^2$</td>
<td>$R_h$</td>
<td>15.6</td>
<td>1.30±0.01</td>
</tr>
<tr>
<td>SAXS$^2$</td>
<td>$R_g$</td>
<td>11.9±0.5</td>
<td>1.40±0.14</td>
</tr>
</tbody>
</table>

1. Measured in 2 M GdmCl.
2. Data from [41].

Although the measured compactness of unfolded state ($R_{h,u}/R_{h,f}$=1.31) is in agreement with NMR experiments, quite surprisingly, FCS measurements indicate that hydrodynamic radius of the unfolded state under non-denaturing conditions is smaller than that of the chemically-denatured state ($R_{h,u}/R_{h,f}$ = 1.31±0.05 versus $R_{h,d}/R_{h,f}$ = 1.20±0.03). To better understand this trend, we calculated the gyration radii from the FCS-measured hydrodynamic radii using shape factors ($\rho$). The shape factor, also known as characteristic ratio or dynamic structure factor, is defined as the ratio of gyration and hydrodynamic radii ($\rho = R_g/R_h$) and depends on the polymer configuration. In case of a globular protein (folded state) the shape factor is predicted to be $\rho = 0.775$ [21, 63], in agreement with experiments (i.e. $\rho = 0.75$ for HIV-1 Capsid Protein [64]). For the native unfolded state of SH3, $\rho$ was found to be 0.82 [41]. The estimated gyration radii for the
folded and unfolded states are in a very good agreement with SAXS data (rows 2 and 4 in Table 4-1).

High concentration denaturants are regularly treated as “good solvents” [65]. From renormalization group theory, $\rho \approx 1.56$ was calculated for an excluded-volume linear chain (a polymer in a good solvent). However for a linear Gaussian chain (a polymer in theta solvent) $\rho = 8/3\pi^2 = 1.508$ is predicted by Tanford following the Zimm model [66, 67]. Experimentally, shape factors around 1.5 [68, 69] and 1.35 – 1.61 [70] for several synthetic linear polymers in good solvents and shape factors around 1.35-1.51 in theta solvents [71] are reported.

Based on the FCS-measured $R_h$ values (Table 4-1) and the shape factors of 1.56 and 1.508, the $R_{g,d}/R_{g,f}$ was estimated to be 2.42 and 2.33, respectively, for the protein in good and theta solvent. Both of these values are substantially larger than the SAXS-measured value of 1.84. Actually, based on the SAXS-measured $R_{g,d}$ and the FCS-measured $R_h,d$, we calculated a shape factor $\rho = 1.19$. A very similar shape factor ($\rho = 1.2$) was found for IgG-domain of protein L using single molecule fluorescence methods [72]. In addition, $\rho = 1.06$ was found for a range of highly denatured polypeptides (50-400 residues) by PFG-NMR and SAXS measurements [63].

Interestingly, for a circular polymer in theta solvent $\rho = \left(\frac{\pi}{2}\right)^{1/2} \approx 1.25$ was estimated [67, 73], which is also expected to be insensitive to the nature of the solvent [73, 74]. This shape factor was used in table 4-1 to calculate the normalized $R_{g,d} \approx 1.92$, which is in agreement with the measured SAXS value of 1.84±0.09. We will discuss the circular chain nature of the denatured protein in more detail in sections 4-7 and 4-8. However, we can conclude that, while the denatured state has a smaller hydrodynamic radius compared to the unfolded state, the gyration radius of the denatured state is considerably larger, in agreement with the common sense.

### 4.5 smFRET: the heterogeneity of disordered conformations

The distance between the ends of a freely-joint chain is related to its gyration radius by $R_{EE} = \sqrt{6} R_g$ [54]. Denatured proteins are often approximated using this model, so we can estimate the $R_{EE}$ of denatured SH3 from the SAXS-measured and FCS-estimated $R_g$ values, to be 52.7 Å. smFRET is an alternative, more direct technique that not only measures $R_{EE}$ of any
4.5 smFRET: the heterogeneity of disordered conformations

protein conformation, including the unfolded and denatured states, but also describes their heterogeneity by quantifying the spread of $R_{EE}$ within each conformation.

4.5.1 Direct observation of folded and unfolded states

In order to perform smFRET experiments, a double cysteine mutant (SH3 C2 C61 G62) was labeled by a dye pair, BFL and A647, with a Förster radius of 44.4 Å. The low labeling efficiency of the donor dye (~4%) ensures that double-donor-labeled proteins are very rare. However, single-donor-labeled proteins are still present in the sample and they will result in a zero-efficiency peak in the FRET histogram [75].

A 1-second time-trajectory of a ~10-hour measurement on 50 pM WT SH3 in TRIS buffer (pH 7.5) is given in the inset of Figure 4-7. The burst events are identified using sliding window burst search algorithm [76]. Generally, the bursts detected in the donor or the acceptor channel are due to the low or high FRET efficiency populated at the two ends of the FRET histogram, while the hybrid burst events belong to the intermediate FRET efficiency distribution. The low FRET values are due to either a very extended protein conformation or a protein with inactive acceptor. The high FRET values results from conformations with small inter-dye distances such as the folded state of SH3.

![Figure 4-7. smFRET histogram of SH3 labeled with BFL and A647 in TRIS buffer (pH 7.5). The green line is the fit of the data with three Gaussian terms. The smFRET histogram of a donor-only SH3 sample is shown in black for comparison. The blue- and red-colored parts of the histogram represent the folded and the unfolded populations, respectively. The inset shows a corresponding 1-second fluorescence time trajectory.](image)
The histogram of calculated FRET efficiencies of the identified bursts is given in Figure 4-7. The folded and unfolded states of the protein correspond to the 99% and 55% FRET efficiency peaks, which are colored in blue and red, respectively in Fig. 4-7. If we exclude the zero-FRET peak (see also below), the relative populations of unfolded vs. folded states can be calculated by integrating the area under the peaks to be 0.54 and 0.46, respectively. These values closely resemble the FCS-measured values of 0.52 and 0.48, respectively. The high FRET population results in a $R_{EE} < 25 \, \text{Å}$, which is close to the expected value from the folded ensemble (Figure 4-2). The FRET peak centered around 55% leads to an average $R_{EE}$ of the unfolded ensemble of around 43 Å, which is significantly larger than that found from MD simulations at the corresponding temperature (Figure 4-3). In addition the 15.3% FRET efficiency standard deviation found for this peak is almost two times larger than the calculated shot-noise standard deviation (eq. 4-9, $N_T=50$). This implies that the unfolded ensemble of the SH3 is considerably heterogeneous. The low FRET peak centred at 0% efficiency does not represent any conformational state and we will discuss it next.

4.5.1.1 The nature of low FRET efficiency peak

Nearly half of the recorded fluorescence intensity bursts contribute to the zero-FRET peak. It is usually assumed that this peak corresponds to proteins with donor as the only active label [77], such as donor-only labeled proteins and proteins with an inactive acceptor due to bleaching or transition to a long-lived dark state. smFRET experiments on pure buffer solutions (data not shown) demonstrate that the fluorescent impurities in the buffer contribute less than 1% to this peak and show no contribution at all to the other FRET peaks.

The peaks of the FRET histogram are typically approximated by Gaussian distributions and less commonly by beta distributions, which are more suitable to describe the asymmetric shapes at the two ends of the FRET histogram (0 and 100%). However, the choices of fitting functions were found not to be critical for the interpretation of smFRET histograms [36]. For the sake of simplicity, we will use Gaussian terms to fit our data. The zero-FRET peak is about one third broader than that of the donor-only sample and 2.5 times broader than the 100%-FRET peak (Figure 4-7). This evidence suggests that the zero-FRET population it is not entirely due to donor-only proteins.
In particular, the transition of the acceptor to a long-lived dark state may explain the broadening of zero-FRET efficiency peak. To be identified as a zero-FRET event in the burst experiments, the lifetime of the dark state should be at least 100 μs, i.e., similar or larger than the diffusion time of the protein through the focal spot. Such a slow blinking process, occurring on the ms time scale for synthetic dyes, has been reported previously in smFRET-time trajectory data [78].

We further investigated the long-lived dark state hypothesis by performing FRET-FCS experiments on a sample of 2 nM SH3-BFL-A647 in a TRIS buffer solution (pH 7.5). The pseudo-autocorrelation and cross-correlation curves and their ratios are shown in Figure 4-8. The two-state model given by equations 1-13 did not yield a satisfactory fit of the correlation ratios over the entire time range, implying that a more complex model is needed. This is also clear from the non-monotonic behavior of the AA/DD ratio, which reaches a maximum around $\tau = 300$ μs. When restricting the data analysis to $\tau > 1$ ms, the fits improved considerably and resulted in $k_{12}=68\pm11$ s$^{-1}$, $k_{21}=30\pm8$ s$^{-1}$, $E_1=90\pm3$ % and $E_2=7\pm3$ %.

Figure 4-8 FRET-FCS analysis. The inset shows the pseudo-autocorrelation and cross correlation curves. The ratios of the correlation curves (open symbols) were fitted simultaneously for $\tau > 1$ ms, using the model given by equations 4-4 (red solid lines).

The obtained FRET efficiencies of the two states are close to 0% and 100% and they are consistent with the existence of acceptor blinking. The high (90%) FRET efficiency is significantly larger than the average value expected for the folded/unfolded mixture (77%) and it
4.5 smFRET: the heterogeneity of disordered conformations

is closer to the folded state value (99%). This is consistent with increased blinking rates of the acceptor dye in the unfolded state. Slow, (sub)ms dark states of organic dyes are most likely the result of charge transfer processes to a local trap in the proximity of the dye. Previous FRET studies on surface-immobilized molecules have shown that the blinking of organic dyes are environment sensitive and follow a power-law statistics [79, 80]. We therefore argue that the differences between the correlation ratio and the model fitting, even \( \tau > 1 \) ms, are related to the non-exponential blinking statistics and to the heterogeneity of the nano-environment of the dyes in the unfolded and folded SH3 states.

The low-FRET peak centered on 1% (Figure 4-7) corresponds to a \( R_{EE} > 9 \) nm. This is similar to the maximum extension \( (R_{EE} \approx 9 \) nm) of the polypeptide chain assumed to be a rigid, rod-like \( \alpha \)-helix, and therefore the low FRET population cannot be attributed to a conformational state of SH3 (see also below). We conclude that the low FRET efficiency peak is entirely due to proteins with only donor active labels and does not correspond to a meaningful end-to-end distance in the SH3 protein.

4.5.2 The charge and pH effect on the Unfolded state

A relatively high net charge of -5.9 at pH 7 and an isoelectric point (pI) of 4.4 were estimated for the DrkN SH3 sequence using Innovagen's Peptide Property Calculator [81]. To further investigate the effect of pH and the net charge on the stability of SH3 we performed smFRET measurements under high (9.0) and low (4.0) pH conditions. The donor and acceptor dyes are relatively photostable and insensitive in this pH range according to the manufacturer. The obtained FRET histograms are shown in figure 4-9 and the fitting results are provided in table 4-2.

Upon decreasing the pH of the sample solution from 9 to 4 (below pI=4.4), \( R_{EE} \) decreased for both folded and unfolded conformations due to less electrostatic repulsion, while the unfolded fraction increased from about 0.38 to about 0.5. This implies that the folded state is destabilized at low pH. Note that even around the pI a substantial fraction of expected protein remains unfolded, thus suggesting that the electrostatic repulsion does not playing an essential role in the destabilization of the folded state. This is also in agreement with the aforementioned complete
The effect of a variety of compounds on SH3 conformations

stabilization of the folded state by a single mutation (T22G) without affecting the net charge of the protein.

Figure 4-9 smFRET histogram of DrkN SH3 at different pH.

smFRET measurements on SH3 were also performed in MilliQ water (pH 5.5). The changes observed compared to the pH neutral sample (larger unfolded fraction and end-to-end distance) can be understood in terms of the reduced charge screening in the absence of salt in the solution.

Table 4-2. Average FRET efficiency, width (Half-Width at Half-Maximum, HWHM), $R_{EE}$ and fraction ($f$) of the folded and unfolded SH3 protein estimated by fitting the smFRET histograms at several pH values and in MilliQ water. The uncertainty of $E$ for the folded and unfolded states are <0.2% and <7% respectively.

<table>
<thead>
<tr>
<th>pH</th>
<th>Folded</th>
<th>Unfolded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E \pm$ HWHM (%) $R_{EE}$ (Å) $f$</td>
<td>$E \pm$ HWHM (%) $R_{EE}$ (Å) $f$</td>
</tr>
<tr>
<td>9</td>
<td>98.6±2.6     &lt; 26.1 0.62±0.03</td>
<td>52.0±13.8     43.8±4.3 0.38±0.03</td>
</tr>
<tr>
<td>7.5</td>
<td>99.1±2.1     &lt; 24.9 0.54±0.02</td>
<td>55.0±15.3     42.9±4.7 0.46±0.02</td>
</tr>
<tr>
<td>4</td>
<td>99.5±2.0     &lt; 24.1 0.50±0.02</td>
<td>62.6±13.5     40.7±4.1 0.50±0.02</td>
</tr>
<tr>
<td>MilliQ</td>
<td>99.4±2.1   &lt; 24.3 0.48±0.07</td>
<td>44.4±15.5     46.1±5.5 0.52±0.07</td>
</tr>
</tbody>
</table>

4.6 The effect of a variety of compounds on SH3 conformations

Osmolytes are small molecules that are ubiquitously used by living organisms to control macromolecular stability [82]. In order to investigate the heterogeneity within the observed
conformational states of DrkN SH3, we used various osmolytes and other compounds such as protein stabilizers (sodium sulfate and glycerol) and protein denaturants (GdmCl, urea and formamide), which modulate the $R_{EE}$ distribution and the interconversion rate of the conformational states.

4.6.1 Folded state stabilizers

It is well known that, at the extreme end of the Hofmeister series of anions, Na$_2$SO$_4$ stabilizes the folded state by promoting hydrophobic interactions [39]. On the other hand, glycerol is known to stabilize proteins and prevent aggregation not only by shifting the native protein ensemble to more compact states but also by destabilizing aggregation-prone partially unfolded intermediates [83]. smFRET measurements in 0.4 M Na$_2$SO$_4$ and in 20% glycerol show the SH3 protein only in its high-FRET conformation, while the lower FRET, unfolded population is fully suppressed (Figure 4-10). In addition, in the presence of stabilizers the zero-FRET peak is either slightly narrower (in glycerol) or slightly broader (in Na$_2$SO$_4$) than the one found in their absence. This further confirms that the zero-FRET peak in our data is most likely caused by donor-only molecules and it does not correspond to an extremely extended conformational state.

![Figure 4-10](image)

**Figure 4-10** smFRET histogram of SH3 in the presence of protein stabilizers, 0.4 M Na2SO4 (grey bars) and 20% glycerol (red line).
4.6.2 Denaturants: resolving a compact scaffold

![Figure 4-11 smFRET histograms of SH3 at various concentrations of GdmCl and urea. Each histogram is calculated from ~10000 bursts.](image)

Using smFRET and FCS, we performed a systematic study of SH3 conformations under titration of denaturants (Figure 4-11). In other smFRET studies of protein denaturation, e.g., the β cold shock protein [77], upon addition of GdmCl, an extended state emerges. This becomes more populated and extended as the denaturant concentration increases. Surprisingly, in DrkN SH3 the high-FRET (~100%) peak does not shift to the lower FRET efficiency nor does it disappear for the high denaturant concentrations (figure 4-11; 8M GdmCl and 7 M Urea). Instead, upon addition of denaturants, it becomes a broad population. For example beyond 2M GdmCl four different Gaussians were required to fit satisfactorily the high FRET values in the histograms.
4.6 The effect of a variety of compounds on SH3 conformations

Also the high FRET population was less heterogeneous when urea is used instead of GdmCl, and three Gaussian terms were sufficient to fit the high-FRET population. The full width at half maximum of the high-FRET population increases from 4.6%, in the absence of denaturants, to a maximum of 12% in 6M GdmCl, and 10.4% in 7M urea. Consequently, the fraction of population with $E > 95\%$ reduces from 0.36 in TRIS to 0.14 in 6M GdmCl, and to 0.13 in 7M urea. In addition, the FRET peak corresponding to the unfolded state gradually shifts from $E = 55\%$ to $E = 20\%$ upon denaturation.

Figure 4-12 shows a high-quality smFRET histogram of SH3 in 6M GdmCl, which was obtained from an overnight experiment (ca. 10 hours, $\sim 10^5$ bursts). The histogram was fitted with six Gaussians: one that can be assigned to the chemically-denatured state (blue), one resembling the unfolded state (green), three of them adding up to a heterogeneous high-FRET scaffold (red) and one accounting for the donor-only molecules (dashed cyan line). The fitting values are provided in the table 4-3. The Förster radius calculated in 6M GdmCl, $R_0 = 4.27$ nm, was used to derive the average $R_{EE}$ for each population.

![smFRET histogram of SH3 in 6M GdmCl. The histogram was fitted with six Gaussian terms.](image)

<table>
<thead>
<tr>
<th>$E \pm$ HWHM (%)</th>
<th>$R_{EE}$ ($\text{A}^\circ$)</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.9±2.2</td>
<td>&lt; 22.9</td>
<td>0.10±0.02</td>
</tr>
<tr>
<td>95.7±3.8</td>
<td>25.5±2.8</td>
<td>0.09±0.02</td>
</tr>
<tr>
<td>86.6±5.8</td>
<td>30.3±3.4</td>
<td>0.10±0.03</td>
</tr>
<tr>
<td>68.2±13.9</td>
<td>37.6±4.1</td>
<td>0.11±0.01</td>
</tr>
<tr>
<td>23.1±12.4</td>
<td>52.1±6.1</td>
<td>0.60±0.03</td>
</tr>
</tbody>
</table>

Table 4-3. Fitting parameters obtained by fitting the smFRET histogram from figure 4-12.

The gyration radius of the random coil model is given by: $R_g = R_{EE}/\sqrt{6}$. Applying this model to the most expanded SH3 conformation (centered at 23% FRET, blue line in figure 4-12), we obtain $R_g = 21.4 \pm 0.5$ $\text{A}^\circ$ which is in an excellent agreement with $R_g = 21.5$ $\text{A}^\circ$ and $21.9 \pm 1.1$
4.6 The effect of a variety of compounds on SH3 conformations

A° estimated by FCS and measured by SAXS, respectively, for denatured SH3 (table 4-1). Therefore we assign this FRET population to the chemically-denatured conformation of SH3.

According to table 4-3, in 6 M GdmCl, close to 40% of the molecules studied by smFRET show significantly smaller $R_{EE}$ compared to the denatured conformation. This is inconsistent with previous NMR and fluorescence spectroscopy studies that showed that this SH3 domain is almost completely denatured at 2 M GdmCl [22, 25, 42]. In addition, if a more compact shape is assumed and therefore a smaller average $R_g$ for the non-denatured conformers (based on their smaller $R_{EE}$), we would expect a smaller average $R_g$ to be measured in SAXS experiments. SAXS is a bulk method and can only estimate the $R_g$ for the entire ensemble and cannot section different populations like smFRET. By assuming that 40% of molecules remain folded in 6 M GdmCl and by making use of $R_{g,D} = 21.4$ Å (table 4-3) and $R_{g,F} = 11.9$ Å [41], an ensemble average $R_g$ of ~17.6 Å is expected. Clearly, that is not the case in the actual SAXS measurements ($R_g = 21.9$ Å) [41].

**Figure 4-13** The high-FRET efficiency scaffold of SH3 in GdmCl, urea and TRIS buffer. The data was normalized to the peak maximum for comparison purposes.

As mentioned above, a similar trend was observed upon urea denaturation, even though urea seems to be a better denaturant (figure 4-11). In 1 M urea the fully denatured fraction reached its maximum population ($f = 0.67$). Interestingly the high-FRET efficiency scaffold in urea seems less heterogeneous than in GdmCl, although they are both considerably broader than the FRET population corresponding to the folded conformation in TRIS buffer (figure 4-13). This
broadening suggests that the high-FRET scaffold observed for DrkN SH3 domain in denaturing conditions is distinct from the folded conformation.

The underlying mechanism of protein destabilization by denaturants such as guanidinium salts and urea is not unique and it is still controversial [84, 85]. In general, these compounds disrupt the hydrogen (H-) bonds that hold the protein’s unique structure. Typically, this effect explains satisfactorily the denaturation of proteins by urea. Guanidium, unlike urea, is charged and partially operates through electrostatic interference with polar or charge-charge interactions within the polypeptide chain, without perturbing the hydration shell [85]. In addition, denaturants are thought to disrupt hydrophobic interactions by changing the solvent structure and promoting its hydrophobic character [85, 86].

The different mechanism of denaturation by GdmCl and urea can explain the aformentioned difference in the heterogeneity of their high-FRET scaffold. The lower heterogeneity in urea may be due to the fact that urea is more efficient in forming H-bonds with peptides. Therefore, we also investigated the denaturation of SH3 by formamide, a neutral H-bond disruptor that is commonly used to stabilize single-stranded DNA.

Figure 4-14 shows the smFRET histogram of SH3 in 20% (v/v) aqueous formamide solution in the absence and in the presence of 4 M GdmCl. Also, the smFRET histogram of the protein in 4 M GdmCl is shown for comparison. Interestingly, formamide alone suppressed the high-FRET scaffold nearly five times more efficiently than either GdmCl or Urea. Upon addition of
4M GdmCl the high-FRET scaffold population increased to higher levels than that of the protein in 4 M GdmCl alone. This suggests that these denaturants are not cooperative and probably the interaction between guanidinium and formamide molecules reduces the effective amount of denaturants in the sample.

### 4.6.3 The high-FRET scaffold is not due to aggregates

As shown in the previous section, the high-FRET population is highly heterogeneous and is not compact so it can not be the folded state. In addition, the appearance of any distinguishable peak in the smFRET burst histogram implies a long-lived (lifetime > 0.5 ms) conformation. The presence of stable and heterogeneous protein aggregates could explain both observations. Nevertheless, all smFRET experiments were performed on sub-nanomolar concentration of proteins. In addition, denaturants are known to be relatively effective in monomer recovery of pre-formed aggregates specially at higher denaturant concentrations and smaller initial aggregate size [87]. Disulfide-bonded aggregates are the exception [88], but in our case both cysteines are labeled.

![Figure 4-15](image)

**Figure 4-15** The histogram of photon burst brightness for SH3 in 6M GdmCl. Bursts from aggregates (dimers, trimers, etc) are expected to fall in the shaded region.

We further investigated the existence of SH3 oligomers in smFRET experiments using burst brightness analysis. The fluorescence burst brightness is defined as the number of detected photons (sum of both donor and acceptor channels) in a burst divided by the burst duration. Figure 4-15 shows the histogram of the brightness of the bursts detected in the smFRET experiment of SH3 in 6M GdmCl. The analysis is performed on the same data set used to
construct the FRET histogram in figure 4-13. The average burst brightness of 108.5 kHz corresponds to about 40 photons per burst. Because aggregates contain more fluorescent labels they are expected to produce brighter bursts, i.e. 200 kHz and higher. If the high-FRET population was caused by protein aggregates, bright bursts would be detected quite frequently, as ~40% of the SH3 proteins show high-FRET (figure 4.12 and table 4.3). Clearly, these bright photon bursts are absent in figure 4-15 implying a complete monomeric SH3 sample.

4.6.4 Conformational clusters

The presence of distinct clusters in the smFRET histogram of SH3 in denaturants implies a slow interconversion rate (relative to the diffusion timescale) between the completely denatured and the partially denatured conformations. If the interconversion between conformers was faster than diffusion, each detected burst would sample several protein conformations and the FRET peaks would be broad. At the extreme, when the interconversion rate is very fast, both peaks merge into a broad single peak with intermediate FRET [36]. Therefore, not all Gaussian terms used in the fitting of the FRET histogram can be easily assigned to different protein conformations. At least some of these Gaussians may arise from the broadening of FRET peaks due to interconversion between different protein conformations.

Aggregation inhibitors like arginine mask hydrophobic surface of proteins and suppress aggregation-prone conformations [26, 27]. We investigated the partially denatured SH3 populations in the presence of arginine in order to assess if they correspond to aggregation-prone conformations. Figure 4-16 shows the FRET histogram of SH3 in 6M GdmCl and 60 mM arginine. The zero-FRET peak is considerably less populated than in the absence of arginine. This is most likely due to a reduced acceptor blinking in the presence of arginine. On the other hand, while a considerable contribution of high-FRET events remained in the data acquired for this sample, the high-FRET population is less heterogeneous than in the absence of arginine. For instance, the high FRET scaffold present in 6M GdmCl (figure 4-13 and table 4-3) becomes narrower and shifts to the higher FRET efficiencies upon addition of 60 mM arginine. Therefore instead of three only two Gaussians centered on $E = 89\%$ and 96% were sufficient to fit this scaffold. As a result, the broad peak centered on $E = 68\%$ stands out clearly as a distinct conformation in the arginine data (figure 4-16).
In conclusion, the combined smFRET data of SH3 in various denaturants shows at least three distinct conformation clusters, with the average FRET efficiencies of 23%, 68% and >95%. The 87% FRET peak could either be a distinct conformation or a broadening effect of the fast interconversion between the 68% and >95% FRET clusters, which is hindered by arginine.

![Figure 4-16 smFRET histogram of SH3 in 6M GdmCl and 60 mM arginine.](image)

### 4.7 Looping and knotting

As discussed above, the shape factor $\rho \approx 1.19$ found from FCS and SAXS measurements of $R_h$ and $R_g$ is reasonably close to the value expected from a ring polymer (~1.25 [67, 73]). In contrast to Nuclear Overhauser Effect (NOE) spectroscopy and NMR chemical shift experiments that suggests that DrkN SH3 is almost completely denatured at 2 M GdmCl [22, 25, 42], smFRET data at even higher denaturant concentration indicates the presence of a heterogeneous population of conformations with an unexpectedly high FRET efficiency. However, further analysis suggested that this scaffold is considerably more extended than the folded SH3 structure. A natural explanation of these observations is that the high-FRET population in the smFRET histogram data belongs to a denatured SH3 conformation with close ends. This can be understood by considering loop formation and knotting mechanism which can be quite effective even at high denaturant concentrations.

Knots as a topological form of biomolecules were observed for the first time in long single-stranded DNA in the 70s [89] and their behavior and biological relevance have been studied
extensively thereafter [90, 91]. However, the knots in proteins were discovered in the native state in 1994 [92] and up to now more than 273 knotted proteins are identified and collected in the protein knot server (http://knots.mit.edu) [93].

The folding pathway of knotted proteins and the underlying mechanism in efficient knotting are challenging and subject of several studies [94, 95]. Mallam et al. [96] made two natively knotted proteins into circular loops using disulfide chemistry under high urea concentration and nonreducing conditions. Both proteins refold to their native knotted structure upon renaturation. This suggests that natively knotted proteins can remain knotted even when denatured. However, to the best of our knowledge, no knot in the denatured state of natively unknotted proteins has been reported yet. This is reasonable, as in these proteins knotting may lead to misfolding [96].

Recently, coarse-grained folding simulations of two structurally similar enzymes, one natively knotted and the other unknotted, was reported [97]. The knotting event was only observed for the natively-knotted protein and only when non-native interactions between amino acids are introduced [97]. In another words, the first step in the knotting process may involve the formation of a transient loop in the polypeptide chain.

### 4.7.1 Fluctuating structure in high denaturant concentration

The presence of loops in the denatured state is consistent with the observation of significant fractionally populated structure in highly denatured proteins such as apomyoglobin, staphylococcal nuclease and DrkN SH3 [53, 98]. Some good reviews of this topic can be found here [10, 53, 99, 100]. Backbone NOE, relaxation and chemical shift experiment results provide evidence of fluctuating structure in SH3, mainly turn-like, under chemically denatured condition [23, 98]. This data was explained by a rapid equilibrium between the random coil and the turn-like structure [98]. As mentioned above, these and other studies (long range NOE, chemical shift, NMR and fluorescence denaturation) show a substantial difference between the unfolded state under non-denaturing conditions and the chemically denatured state [22, 25, 42] and suggest that SH3 is almost fully denatured in 2 M GdmCl.

A more complex picture arises from NMR studies of apomyoglobin and the staphylococcal nuclease. The Shortle group conducted hybrid spin label experiments and MD simulations on denatured staphylococcal nuclease, and later residual dipolar coupling experiments in oriented
gels. They argued that, even in 8M urea, the mean structure of the denatured ensemble is similar to the structure of native state, despite the absence of tertiary structure [55, 101]. This suggested that the native-like structure persists even in strong denaturation conditions. Furthermore, the authors suggest that hydrophobic interactions alone are able to encode the protein’s fold topology in the denatured state despite the absence of cooperative hydrogen bonding.

In NMR chemical shift experiments on denatured apomyoglobin in 8 M urea, no considerable secondary structure was observed [102]. Strikingly similar results to those obtained for the Staphylococcal nuclease were obtained in residual dipolar coupling measurements of denatured apomyoglobin in 8 M Urea at pH 2.3 [53]. However, the authors proposed a different explanation, by taking into account the persistence length and considering the protein as a polymer of statistical segments. The observed dipolar couplings were attributed to the transient local alignment of statistical segments instead of the persistence of long-range interactions. This has been further supported by MD simulations of various polypeptides, which revealed that, even though none of the denatured conformers had native-like features, the mean ensemble structure had a native-like geometry [103]. In fact, this mean-structure hypothesis, also known as the "reconciliation problem" [104], is not yet fully understood and the role of long-range interactions in the denatured state of proteins remains controversial.

Charge-charge is a long range interaction which is found to have a significant effect for folding stabilization in several protein engineering studies [105]. This is somewhat surprising, since its contribution to the folding free energy is estimated to be two orders of magnitude smaller than the hydrophobic effect [105]. pKₐ values calculated from NMR chemical shifts as a function of pH were used to probe the electrostatic interactions in the denatured state of a 56-residue protein, the N-terminal domain of L9 from Bacillus stearothermophilus [106]. Based on the discrepancy of the pKₐ values of the folded, denatured and small protein segments, Kuhlman et al. concluded that both local and nonlocal electrostatic interactions exist in the denatured state [106]. In addition, further analysis using the Gaussian-chain model indicated that the residual electrostatic effect in the denatured protein may be attributed to nonspecific nonlocal charge-charge interactions. Both local and long-range structures in the denatured states have also been produced by hybrid experimental-simulation studies [10].
4.7 Looping and knotting

4.7.1.1 Evidence of loop formation from the size scaling and intrinsic viscosity

In solution, a polymer chain experiences three types of interactions: the monomer-monomer, monomer-solvent, and steric repulsions. Flory studied the effect of steric exclusion on the size of polymers in solutions which led to the definition of good, theta and poor solvents. The theta point in a solvent is achieved when the nonsteric forces cancel the steric repulsion to the first order while in the good (bad) solvent the net interaction is repulsive (attractive) [107]. Flory showed that all solvents can be assigned to one of these categories. According to the Flory’s model the size of polymer in a solution as a function of its number of monomers is given by

\[
R = R_0 N^\nu. \quad \nu = \begin{cases} 
3/5 & \text{good solvent} \\
1/2 & \text{theta solvent} \\
1/3 & \text{poor solvent} 
\end{cases}
\]

(4-7)

where \( R \) is a polymer size characteristic size of the polymer, such as \( R_g \), \( R_h \) or \( R_{EE} \), and \( R_0 \) is a prefactor corresponding to the effective segment length [107]. A better approximation revised the scaling factor for a good solvent to \( \nu \approx 0.588 \) [107]. For globular proteins, water is considered to be a poor solvent, and denaturants such as GdmCl and urea are generally assumed to be good solvents.

The experimental values of the scaling exponents have been determined by fitting either the calculated \( R_g \) of set of proteins from protein data bank [108] or the measured \( R_g \) and \( R_h \) by SAXS, PFG NMR and single-molecule techniques [63, 108]. The obtained \( \nu = 0.34 \) and 0.29 [63, 108] for water closely matches the theoretical value for the poor solvent. Based on \( R_g \) data, the scaling exponent of denatured proteins was found to be \( \nu = 0.59 \) [109], \( \nu = 0.59 - 0.67 \) in 6 M GdmCl, with the larger values in this range for the highly charged intrinsically disordered proteins [77]. In general, these values are in agreement with the good solvent conditions.

On the other hand, the \( R_h \) data provided \( \nu = 0.57 \) [63], \( \nu = 0.52 \) in 6 M urea and 0.55 in 6 M GdmCl [110]. These values are somewhat smaller than those found from \( R_g \) studies and are distributed between the theta and the good solvent values. From theory, both radii (\( R_g \) and \( R_h \)) are expected to follow the same scaling behavior [21, 107], but the scaling exponents were found for different sets of proteins and one must be cautious when comparing them.
Burry addressed the observation of residual structure in denatured proteins in NMR studies by analyzing and reevaluating published $R_g$, $R_h$ and intrinsic viscosities ($[\eta]$) obtained for dilute protein solutions in strong denaturant [111]. The resulting scaling factors were $\nu = 0.60$ using $R_g$, resembling a good solvent, and $\nu = 0.52$ using $R_h$, resembling a theta solvent, for the same set of proteins under the same conditions (5-6 M GdmCl). In addition because of the different scaling behavior of the $R_g$ and $R_h$, the shape factor $\rho$, which according to the Zimm model is expected to be $\geq 1.5$ [66, 67], for small proteins ($N < 30$) is considerably decreased and even became smaller than unity.

Burry also investigated the $K$ parameter, which is calculated from $R_g$, $R_h$, $[\eta]$ and the molecular weight $M$ according to the following equation:

$$K = \frac{M[\eta]}{\pi N_{av} R_g^2 R_h},$$

(4-8)

where $N_{av}$ is the Avogadro number. For linear polystyrene polymers (PSs) in a theta solvent, a semi-constant experimental value $K_{linear} \approx 2.6$ was obtained over three orders of magnitude of molecular weights ($M = 1 \text{ kDa} - 3 \times 10^3 \text{ kDa}$) [111]. $K$ was smaller in a good solvent but approached the theta-solvent value for small linear PSs. For long, ring-shaped polymers without intramolecular excluded volume, $K_{ring} = 1.519K_{linear}$ was estimated by theory [111]. For several ring-shaped PSs, $K_{ring}$ values (~3-3.4) much smaller than the predicted value of ~3.9 were measured, which was attributed to the impurity of the samples. However, for small proteins (<100 kDa) in GdmCl, $K \approx 3.4$. For larger proteins, the $K$ value decreased and approached 2.6, similar to linear PSs. The observation of larger $K$ values for the lower range of $M$ was interpreted as evidence of transient loop formation by association of widely separated amino acids along the chain backbone. Furthermore, the author showed that these ambiguities can be partially resolved if, for example, denatured proteins are assumed to be random chain polymers, with ca. 25-30% of them containing a loop with two dangling ends, each one-fourth the total length of the chain [111].
4.8 Loop formation: Cooperative salt bridges and H-bonds

Even though little is known about residual structure in the denatured state, our results and the literature studies discussed in the previous section come together into a coherent picture by assuming a loop formation mechanism in DrkN SH3 domain, which may also initiate a knotting event [97].

The DrkN SH3 has 26 ionizable groups (~44% of its composition) [112], which is about 40% higher than the average content in proteins [113]. In the light of the aforementioned observation of nonlocal charge-charge interaction in denatured state [106] and residual turn-like structure observed in NMR, we propose that cooperative salt bridges and hydrogen bonds (side chain and backbone) cause the long range amino acid association and hence loop formation in SH3. In addition, a looped denatured state may also lead to nonpolar contacts and reduction of exposed hydrophobic side chains.

Figure 4-18 shows the contact map along the protein backbone for half-turn (180°) looping events, depicted in the same figure. The nNC-axis, the lower horizontal axis, is the separation between the N-terminus and the projection of C-terminus on the protein backbone when the

**Figure 4-17** The $K$ parameter as a function of the molecular weight (logarithmic scale) for unfolded proteins (filled circles), linear polystyrene chains under Flory theta conditions (diamonds) and good solvent conditions (filled diamonds), and ring-shaped polystyrene chains (circles). The graph is reproduced from [111] with permission.
dangling segment in the C-terminus is shorter than the one in the N-terminus. For example assume a loop formed by a set of contacts including the one between residues 8 and 60, which is the closest contact to the end-termini. Because the residue-index of the N- and the C- termini are 1 and 62, the two dangling segments are $8 - 1 = 7$-residue-long, and $62 - 60 = 2$-residue-long, respectively. The projection of the C-terminus on the protein backbone is the residue number $6$ ($8$ minus $2$) and therefore the separation between the N-terminus and the projection of the C-terminus on the protein backbone is $n_{NC} = 5$ ($6$ minus $1$). In this contact map, the color of a cell with coordinate $(n_{NC}, y)$ shows the nature of interaction between residues with indices

$$(N - y + 1) + n_{NC} \quad \text{and} \quad y$$

where $N$ (for our protein equals $62$) is the protein length. Similarly the $n_{CN}$ axis, the upper horizontal axis, is the separation between the C-terminus and the projection of N-terminus on the protein backbone when the dangling segment in the N-terminus is the shorter one. In addition a cell with coordinate $(n_{CN}, y)$ belongs to residues with indices

$$(N - y + 1) - n_{CN} \quad \text{and} \quad y$$

The contact map is shown for only one branch of each loop, because the complementary branch is symmetrically colored as shown for a small portion of the map in the centre region of figure 4-18.

A similar contact map for full-turn ($360^\circ$) loops is shown in figure 4-19. In this case, the horizontal axis ($n$-axis) is the separation of the N-terminus and its projection ($i$) on the protein backbone. For example assume a loop formed by a set of contacts including the one between residues 3 and 47, which is the closest contact to the N-terminus. In this case, the dangling segment in the N-terminus is $(3$ minus $1)\ 2$-residues-long. The projection of the N-terminus on the protein backbone is the residue number $45$ ($47$ minus $2$), and therefore, the separation between the N-terminus and its projection on the protein backbone is $n = 44$ ($45$ minus $1$). For each separation value the contacts are given from the projection of the N-terminus on the protein backbone (a.k.a. $i$) to the C-terminus. In this contact map, the color of a cell with coordinate $(n, y)$ shows the nature of interaction between the residues $n + y$ and $y$. 
Figure 4-18  Contact map of half turn loops for DrkN SH3. $n_{CN}$ and $n_{NC}$ are the separations between the C- (or N-) terminus and the projection of N- (or C-) terminus on the protein backbone. Blue, green and yellow pixels represent salt bridges, side chain H-bonds and non-polar contacts, respectively. Black pixels represent contacts between charged amino acids of the same polarity, which are not favorable for loop formation. The regions that are more stable against sliding of the contact segments are shown by purple ellipses. Adjacent favorable contacts are highlighted by red boxes.
Figure 4-19 Contact map of full-turn loops. $n$ is the separation of N-terminus and its projection (i) on the protein backbone. Blue, green and yellow pixels represent salt bridges, side chain H-bonds and non-polar contacts, respectively. Black pixels represent contacts between charged amino acids of the same polarity, which are not favorable for loop formation. The regions that are more stable against sliding of the contact segments are shown by purple ellipses.

We consider a looping event successful when several adjacent blue, green and yellow contacts are formed, with minimum white and black interspaces, like the red-boxed segments highlighted in figures 4-18 and 4-19. In addition, a successful loop should be stable against sliding of the contact residues. The purple ellipses show areas of the contact map where the
numbers of contacts are minimally affected by sliding. The residues responsible for these highly favorable and conserved contacts are highlighted along the protein backbone (vertical axis of figure 4-18 and 4-19) by purple boxes and these SH3 segments are likely the ones involved in the looping process. Seven different active segments are recognized from the contact maps. The closest such segment to the N-terminus, formed by residues 3-9, is in fact the longest. The association of this segment and the other two active segments close to the C-terminus, formed by residues 51-53 and 57-60, is most likely responsible for the high-FRET population ($E \geq 90\%$, $R_{EE} < 30$ Å) in the denatured protein.

For the half-turn loops, the $R_{EE}$ can be easily estimated by assuming that the two dangling segments at the ends are random coils. Then in the Flory model, we equate $N$ to the sum of the lengths of the two dangling segments. The effective segment length was calculated to be 4.7 Å from equation 4-7 by substituting $\nu = 0.588$ for a good solvent and the $R_{EE} = 52.1$ Å for the denatured state (table 4-3). We calculated $R_{EE}$ for the best 7 loops in the contact map using this approach (figure 4-18). Note that the $R_{EE}$ estimated here represents the distance between the two cysteine labeling sites. It is very interesting that the obtained $R_{EE}$ values (12, 15, 19, 23, 25, 39 and 41 Å) are very similar to three $R_{EE}$ values (37.6, 25.5 and <15.2 Å) found by smFRET of SH3 in 6 M GdmCl (table 4-3). On the other hand, the estimation of $R_{EE}$ for full-turn loops is not straightforward and requires detailed information of the 3D association of the active blocks.

Summarizing these findings, figure 4-20 shows three examples of possible loop formations through cooperative salt bridges (dashed blue links), side chain H-bonds (solid green links) and nonpolar contacts (dashed black links) in the DrkN SH3 protein. Backbone hydrogen bonds are not shown in this scheme. Positively and negatively charged residues are shown by blue and red color, respectively. The remaining green and black colored residues are polar and nonpolar, respectively. The first scheme (figure 4-20a) corresponds to the half turn loop shown in figure 4-18 with $R_{EE} = 15$ Å. Figure 4-20b shows a full-turn loop with $n=44$ and in figure 4-20c a hybrid scenario with both half-turn ($n_{CN}=39$) and full-turn ($n=17$) loops is shown. Hybrid configurations made of two or more half and/or full-turns are potentially very important because they may led to knotting events.
4.9 Summary and outlook

The size and shape of the folded, unfolded and chemically-denatured states of the DrkN SH3 was investigated by FCS and smFRET burst experiments. The folded and unfolded hydrodynamic radii were determined with FCS measurements and are in excellent agreement with previously reported values by PFG NMR. In addition, using appropriate shape factors, they closely reproduce the gyration radii determined in SAXS experiments. However, we found that the chemically-denatured ensemble, U_{GdmCl}, is hydrodynamically more compact than the natively disordered state, U_{exch}. In addition, a shape factor of \( \rho \approx 1.19 \) was obtained from the FCS-measured \( R_{h,D} \) and the SAXS-measured \( R_{g,D} \) [41], which is considerably smaller than the value.
expected for a linear polymer in good solvent or theta solvent, i.e., 1.56 and 1.5 respectively [66, 67] and resembles the shape factor of a circular polymer $\rho \approx 1.25$ [67, 73].

The SH3 conformational distribution was investigated by smFRET burst measurements under various experimental conditions, including pH, denaturants and other osmolytes. The FRET histogram obtained from experiments in normal conditions resolved two separate peaks representing the classical $F_\text{exch}$ and $U_\text{exch}$ states of SH3. The $R_{\text{EE}}$ of the $U_\text{exch}$ state has reduced by about nearly 10% by lowering the pH from 9 to 4. This is explained by the high net charge of the protein which results in an expansion of its overall size by charge repulsion at pH values higher than $pI \approx 4.4$. On the contrary, the $U_\text{exch} R_{\text{EE}}$ increased in MilliQ water in the absence of charge screening.

The $U_\text{exch}$ state completely disappeared upon addition of protein stabilizers such as 0.4 M Na$_2$SO$_4$ or 20% glycerol. Very interestingly, a high-FRET efficiency cluster (>90%) persisted at high denaturant concentrations, such as 8M GdmCl, 7M urea and 20% formamide. The FRET histogram of SH3 in 6 M GdmCl was fitted to 5 Gaussian populations, including a completely denatured state ($E \approx 25\%$), three components of the high-FRET scaffold and one conformation with intermediate FRET efficiency, $\approx 68\%$. A random Gaussian chain model was assumed to describe the completely denatured population. The obtained $R_{g,D} = 21.4 \ \text{Å}$ is in an excellent agreement with the SAXS-measured $R_{g,D} = 21.9 \ \text{Å}$ and the FCS-estimated $R_{g,D}=21.5 \ \text{Å}$. The broad population with intermediate FRET efficiency was better resolved upon addition of 60 mM arginine, thus implying that it does indeed represent a distinct conformational state of SH3.

Although the energy transfer of the high-FRET scaffold may resemble that of the folded state, based on the high heterogeneity of this population and on previous NMR studies that suggest that the SH3 protein is fully denatured in 2 M GdmCl [22, 25, 42], we conclude that it does not correspond to the folded protein. Both the fraction and the heterogeneity of this scaffold changed in different denaturants. The lowest fraction was achieved in 20% formamide and is almost five times lower than in 6M GdmCl.

Based on these results, we suggest that the high-FRET population in the FRET histogram obtained under denaturing conditions represents looped states of the denatured protein, which is fully denatured otherwise. Connections through cooperative nonpolar contacts, H-bonds and salt-bridges were discussed in detail as possible explanation of the underlying mechanism.
This interpretation of our data is in agreement with observations of residual helical structure [53, 98], local and long-range electrostatic interactions [106] in NMR studies of denatured proteins and evidences of loop formation from denatured state size scaling [111].

In this study of DrkN SH3 we highlighted the possibility of looped conformations in protein folding pathways. The underlying mechanism and the interconversion rates between the contributing looped states can be further studied by smFRET-time trajectory experiments on various mutants of the protein. A wider range of osmolytes can also be used to perturb the intramolecular interactions and to investigate the contribution of salt bridges, H-bonds and non-polar contacts in the association of specific protein segments.

In conclusion, our data suggests that SH3 continues to be an excellent model system for the study of the protein folding. The finding of this study, in conjunction with the extensive body of research on SH3 over the last two decades, provide a solid ground work for the future investigations. We foresee that this system will provide a great insight for the protein folding mechanism.

4.10 Bibliography


Chapter 5

Direct Analysis of Single-Cell miRNA by Capillary Electrophoresis with Confocal MultiParameter Fluorescence Detection

Abstract

For the purpose of direct quantitative analysis of multiple miRNAs without enrichment or modification, we developed a simple method to efficiently couple on-column, standard Capillary Electrophoresis with Confocal MultiParameter Fluorescence detection (CE-CMPF) using only commercially available components. A molecular counting efficiency of 13% and a concentration limit of detection of 1.5 pM of fluorescein were achieved in our instrument by gating the arrival time of individual photons to reduce the scattering contribution. The proposed scheme allows for amplification-free detection and separation of three different micro-RNAs from MCF-7 cell lysate. The detection sensitivity is approximately 500 times better and the separation time is ca. 3 times faster compared to protocols on commercial CE instruments. Although the optical design can be further improved, it is shown that the current CE-CMPF prototype is already capable of analyzing the lysate content of single cells. In addition, all CE protocols previously developed on commercial instruments can be performed with our CE-CMPF without modification but with nearly three orders of magnitude higher sensitivity.

This chapter is reproduced with permission from:


5.1 Introduction

MicroRNAs (miRNAs) are short, non-coding RNA molecules (18-25 nucleotides) that participate in the regulation of the majority of cellular processes [1]. Various disease processes as well as every cancer type studied so far have been associated with altered expression patterns of multiple miRNAs [2-4]. miRNAs are so exceptionally stable in plasma as well as serum that they are detectable even in the serum of cancer patients [5]. Therefore, expression profiling of miRNAs as non-invasive biomarkers is useful in clinical diagnosis, estimation of prognosis, prediction of therapeutic efficacy and in cancer classification.

The efficient use of miRNA-biomarker panels requires accurate quantitative analyses of multiple miRNAs. However, the direct detection methods of miRNA are either semi-quantitative (e.g. Northern blotting, signal amplifying ribozymes, and in situ hybridization) or unuseful for multiple miRNAs (bioluminescence detection, and two-probe single-molecule fluorescence) [6-10]. In addition, indirect methods (e.g. quantitative reverse-transcriptase polymerase chain reaction, microarrays, surface plasmon resonance, next generation sequencing, etc.) are not only cumbersome but also suffer from reduced accuracy due to different efficiencies of modifications for different miRNAs [11-13].

To address this problem, we used an embedded capillary interface (ECI) to efficiently couple on-column, standard Capillary Electrophoresis with Confocal Multi Parameter Fluorescence (CE-CMPF) detection. We compared the sensitivity of our instrument with a commercial CE machine and by employment of CE-CMPF in a recently developed method of Direct Quantitative Analysis of Multiple miRNAs (DQAMmiR) [14] we aimed to achieve a sufficiently high sensitivity to analyze miRNA in fine-needle-biopsy samples (~100 cells) and even in single cells without enrichment or modification of miRNA. In addition, we designed the instrument so that all standard commercial CE protocols can be transferred to our platform.

5.2 CE review and our setup design

Capillary Electrophoresis with Laser-Induced Fluorescence (CE-LIF) detection is a highly versatile analytical tool, beneficial for various pharmaceutical, genetic, and biomedical
investigations [15-25]. Commercial CE-LIF instruments are easy-to-use as they provide fully automated injection, separation, and collection of samples and on-column fluorescence detection through the capillary wall. Therefore, the majority of CE protocols have been developed using commercially available instruments. However, due to the high background level [26] these instruments have a poor detection limit, on the order of \(10^4\text{-}10^5\) injected fluorophore molecules.

### 5.2.1 Sensitivity definitions

Several CE-LIF methods were developed to improve the detection sensitivity. However, the obtained sensitivities were not reported in consistent terms. Therefore, it is helpful to define some necessary terms before reviewing these methods and discussing our CE-CMPF apparatus.

The mass Limit of Detection (mLOD) is the minimum number of molecules (not necessary all pass through the detection volume) in a plug such that the plug is detectable; the Molecular Counting Efficiency (MCE) is the fraction of injected molecules that pass through the laser probe volume; the absolute Limit of Detection (LOD) is the minimum number of molecules required to pass through the laser probe volume such that the plug is detectable (approximately equal to \(\text{MCE} \times \text{mLOD}\)); and the concentration LOD (cLOD) is the minimum concentration of a plug such that it is detectable. For all terms mentioned above the phrase “such that is detectable” refers to a CE peak with a signal to noise ratio of \(S/N = 3\).

### 5.2.2 CE-LIF methods

**Off-column detection.** The first CE-LIF approach that was developed to address fluorescence sensitivity employed post-column detection using a sophisticated sheath-flow cuvette to detect analytes after they exited the capillary [26, 27]. This technique was adopted from flow cytometry and reduces background signal by eliminating scattering and autofluorescence originating from the capillary walls. Using this design mLODs of less than 100 molecules were reported nearly 20 years ago [28, 29]. Despite the effectiveness of this technique, there are no current commercial implementations of this approach, probably because inherent difficulties with sample collection, capillary exchange and alignment procedures.
5.2 CE review and our setup design

**Tight channels.** Another approach to sensitive CE-LIF was the use of (sub)micrometer-sized separation channels. Such designs force all analytes to flow through a tightly focused excitation beam, thus improving MCE. Examples of micro-channel designs include: custom-made pinched capillaries [30], microfluidic chips [31-34], etched glass channels [35], patch-clamp pipettes [36, 37], and quartz capillaries [36]. The use of micro-channels often hinders diffusion in 1 or 2 dimensions, thus beneficially causing molecules to reside in the detection volume longer and emit more photons [36, 37]. Single molecule fluorescence bursts with S/N ≈ 50 were recorded [34]; however, the drawback of the method is that small-size channels are more prone to contamination and clogging [33, 38, 39]. Consequently, there is only one micro-channel study involving a complex biological sample (cell lysate) which avoids the aforementioned troubles by using a relatively wide channel (1.8 µm × 60 µm height/width) [33]. In addition, the shape and material of micro-channels often differ from that of standard fused silica capillaries. This difference may introduce changes to analyte migration speeds and retention times [30, 36], analyte velocity profiles [33, 40], and electro-osmotic flows compared to a standard round capillary. The variations in these parameters can make transferring protocols from commercial instruments to micro-channel devices non-trivial. Also, the media that electrophoresis is carried out on is known to degrade over time and must be replaced; further making the case for the use of inexpensive and easy-to-replace commercial round capillaries rather than more complicated chips and custom-made capillaries.

5.2.3 CE-CMPF

In order to improve the detection limit, the problem of high background due to, i.e., (i) the reflections of excitation beam at the capillary walls, (ii) the Raman scattering of water, and (iii) the autofluorescence of the capillary glass, must be addressed without altering the on-column detection scheme. Axial resolution in the direction of beam propagation is achievable with a confocal detection scheme, which suppresses out-of-focus background photons originating from Rayleigh and Raman scattering and from fluorescence impurities. This technique can, therefore, be used to address points i) and iii) by limiting the axial size of the focal volume to within the inner bore of the capillary, and point ii) through limiting the amount of buffer that is optically sampled. Under typical experimental conditions a well-designed confocal microscope allows for highly sensitive fluorescence detection with up to single molecule resolution in a detection
5.2 CE review and our setup design

volume of about 1 fl. Our goal in this work was to effectively couple our ultrasensitive confocal fluorescence detection with a standard commercial round capillary.

To efficiently couple the confocal detection system with the CE-LIF platform we had to avoid the pitfalls of previous confocal-CE studies: poor MCE (< 0.1%) in capillaries much larger than the size of the detection volume [3, 41, 42] and clogging in sub-micrometer channels [33]. With these studies in mind, we chose a capillary with a 10 µm inner diameter (ID) and a standard confocal detection volume. For this ID value, commercially available capillaries have outer diameters (OD) greater than 150 µm resulting in an OD/ID > 15. At such large ratios distortions of the detection volume by lensing effects and spherical aberrations due to the curvature and wall thickness of capillaries can be so severe that even bright samples like 20-nm fluorescent beads may pass undetected in a confocal-CE arrangement [30]. Therefore, careful choice of capillary geometry and mounting media is essential to limit the distortion of the detection volume.

![Figure 5-1](image)

**Figure 5-1.** Left: schematic showing geometry and refractive indexes of ECI. Centre: actual photograph of the ECI. Right: photographs of a laser beam incident on a 10-µm inner diameter capillary in air, and one embedded in a transparent matrix of adhesive (the sketch of the capillary is guide for the eye).

Our microscope objective is compensated against spherical aberrations when used with 170 µm thick coverslips and because of its high numerical aperture (NA=1.4), it is very sensitive to deviations from this thickness. Therefore, the interface between the objective and the capillary inner bore was designed to be close to 170 µm and to have a uniform refractive index near that of glass. To achieve these requirements, we proposed the use of an ECI, whereby the capillary was fixed to a support surface by means of a transparent matrix (Figure 5-1, left). The thickness of the support surface had to be less than 170 µm, so we chose commercially available 110 µm coverslips. The use of a commercially available capillary with an inner diameter of 10 µm, an
outer diameter of 150 µm and wall thickness of 58 µm (after the opaque 12 µm polyimide coating was removed under a flame) ensured a final thickness of 168 µm, thus mimicking the optimal coverslip. The 2 µm deviation between the ECI and the optimum coverslip thickness, although not negligible, is within the 3 µm uncertainty of the capillary thickness provided by the manufacturer and the <5µm measured uncertainty of our coverslip batch. For the purposes of this study this deviation is acceptable.

The capillary and support surface were fixed together using an optical transparent adhesive with a refractive index of 1.56 (Figure 5-1, centre). By closely matching the refractive index of fused silica (1.46), borosilicate glass (1.52), and immersion oil (1.53) the extent of scattering and aberrations is reduced (Figure 5-1, right). In addition, back-scattering was further reduced by mounting the ECI along the y-axis parallel to the laser polarization [43].

![Figure 5-2. Backscattering from a dry capillary. Images were acquired by scanning the stage in the transverse plane (x and y). a) focal plane located at the centre of the capillary; the central bright line is the capillary channel and the dark regions around it are the capillary wall. b) focal plane located at the bottom glass-air interface inside the capillary.](image)

The ECI was mounted on the 3-D piezo stage along the y-axis in the confocal microscope for imaging through the high NA objective. A flashlight was used to illuminate the ECI from the top and the transmission image of the capillary through the objective was projected on a white paper. This magnified image aided the placement of the capillary at the centre of the scanning range of the x-axis. The axial (z-) position was adjusted to focus on the inner bore and sharpen the capillary transmission image. Fine lateral and axial alignments of the ECI were done by maximizing the backscattering signal from the glass-air interface. We used this method to find
the bottom of the inner bore (Figure 5-2), with the fluorescence filters removed in the detection line to make the scattering more prominent.

![Figure 5-3. Backscattering (red curve) from a dry capillary and fluorescence (black curve) from a capillary filled with 100 nM Rhodamine 110 (black) as functions of the distance from the inner wall of the capillary. In this representation, the excitation beam comes from the left of the image and the signal is detected on the same side.](image)

Once the capillary inner bore was located, a sample of 100 nM of Rhodamine 110 was passed through the capillary. Figure 5-3 shows the backscattering and the fluorescence signals as functions of focal point distance, as measured from the wall of the inner wall closest to the microscope objective. The fluorescence intensity has a maximum at a distance of ~4.5 µm from the wall, which closely matches the optimum depth into aqueous solutions mounted on regular cover slips which produces minimal aberration in FCS measurements (see chapter 1). All the subsequent CE measurements were carried out with the laser beam focused at 4.5 µm inside the capillary inner bore.

5.3 ECI detection volume and sensitivity

In principle, matching the detection volume to the capillary size would be desirable to maximize the molecular counting efficiency (MCE) of the instrument. However, the detection point spread function, and therefore the detection volume, is determined by both excitation and detection optics. For example, the combination of the area size of the avalanche photodiode detector and the magnification of the objective determines the maximum possible lateral diameter.
of the detection volume (~0.5 μm for a magnification of 100x and chip diameter of 50 μm). In addition, the deformation of the detection volume by cylindrical lensing effect of the curved capillary walls will reduce the photon collection efficiency (PCE) and thus the mLOD. Therefore, in order to determine the MCE and investigate the ECI performance in bioanalytical applications, it is imperative to quantitatively measure the detection volume of the instrument and to estimate and control the optical aberrations introduced by the ECI approach.

In order to estimate the aberrations introduced by the ECI, we immobilized fluorescent polystyrene beads on the inner wall of the capillary and imaged them on the confocal microscope (Figure 5-4 (a)). A series of images taken at different axial (z) positions showed two focal planes separated by ca. 4 μm along the z-axis (Figure 5-4 (c-g)). In these focal planes the bead image was elliptical, elongated about 5 times along either the x- or the y-direction. This astigmatism is likely caused by the cylindrical lensing effect of the curved capillary wall. Upon filling the capillary with water, the astigmatism decreased considerably such that the two separate focal points were no longer separated. In this case, the shape of the image was almost round (Figure 5-4 (b)), similar to those obtained from immobilized beads on standard cover slips. However, the image of the ECI-immobilized bead is ca. 50% wider in both x- and y-directions compared to the coverslip-immobilized bead (~350 nm full width half max diameter).

**Figure 5-4.** (a) An image of immobilized beads in a dry capillary (image is rotated 90 degree). The centre line is the backscattering from the capillary (long-pass and band-pass filters are removed). (c-g) Images of a fluorescent bead in a dry capillary at different axial positions of the focal point. Z is the distance between centre of the focal point and the glass-air interface. (b) An image of a fluorescent bead when the capillary is filled with MQ water.
5.3 ECI detection volume and sensitivity

We compared the fluorescence intensity from a dye sample (100 nM Rhodamine 110 (Rh110)), as measured in the ECI and on a standard coverslip. Based on FCS analysis, a 1.02 nM solution on a coverslip is equivalent to having on average 1 molecule the detection volume. Each Rh110 molecule gives on an average fluorescence count rate of 0.75 kHz at an excitation power of 10 μW (measured at the sample), so the signal measured on the 100 nM sample is 74 kHz. Under similar sample/excitation conditions, the signal measured in the ECI was found to be 23 kHz, about a third of that of the coverslip sample. Since Rh110 does not adsorb to either borosilicate coverslips, fused silica capillaries or sample vials [44], the fluorescence reduction in ECI is most likely due to optical aberrations. For instance, the aberrations lead to lower excitation intensity because of the larger focal spot (Figure 5-4 (b)) and to lower photon collection efficiency.

![Experimental FCS curve of 10 nM Rhodamine dyes measured on the ECI platform (black squares) fitted by 2D (blue) and 3D (red) diffusion models.](image)

Figure 5-5. Experimental FCS curve of 10 nM Rhodamine dyes measured on the ECI platform (black squares) fitted by 2D (blue) and 3D (red) diffusion models.

To determine the individual contributions of the expansion of these aberration-induced effects to the overall 3-fold reduction of the fluorescence signal, we estimated the size of the detection volume in the ECI using FCS data from a solution of 10 nM Rhodamine 6G diffusing freely inside the capillary. The correlation curve obtained was best fitted to the 2D diffusion model (Figure 5-5), implying that the detection volume spans (almost) entirely across the inner capillary bore. Thus, the axial size of the detection volume is estimated to be close to 10 μm. In addition, a lateral radius of \( r \sim 430 \text{ nm} \) was found from the fitting of the correlation curve using the 2D diffusion model. This represents a \( \sim 1.3 \) times expansion of the detection volume in the
transverse directions compared to the same measurement on coverslips and it leads to a MCE of \(~13\%\) in our ECI (for details see section 5.5). The expanded axial and transverse radii lead to a \(~3.3\) times larger volume compared to the coverslip experiment.

The relative reduction in collection efficiency, as compared to a measurement made on a coverslip, can be estimated from a simple relation \( F \propto N I C_{\text{ph}} \), where \( F \) is the measured fluorescence intensity, \( N \) is the average number of molecules in the detection volume, \( I \) is the beam intensity at focal spot, and \( C_{\text{ph}} \) is the photon collection efficiency. Experimentally, we found that \( F \) was reduced by a factor of 3, \( N \), which is proportional to the detection volume, increased by a factor of 3.3, and \( I \) decreased by a factor of 1.7 (1.3×1.3). By implementing these values in the expression above, \( C_{\text{ph}} \) for the ECI mode was found to be \(~5\) times lower than the typical value for the coverslip mode, e.g., 2.7\% [see appendix 1]. In future iterations of the design, the photon collection efficiency can be improved by choosing Single-Photon Avalanche Diode (SPAD) detectors with larger photoactive area and higher quantum efficiency and by employing corrective lenses that minimize the optical aberrations.

### 5.4 Signal-to-noise optimization

While the preceding discussion identified the theoretical limit of ECI-based measurements as set by the optics of the system, in the next section, we seek improvements to the detection limit through increasing the S/N using different experimental conditions and data analysis techniques. To calculate the S/N, the signal was found from the peak amplitude minus the baseline and the noise is considered to be the standard deviation of the baseline.

#### 5.4.1 Data binning time

The advantage of time-tagged photon detection is that the fluorescence signal can be binned post-acquisition at any desired dwell time. To explore the influence of data binning on S/N, we performed electrophoresis of a 60 pL sample containing 100 pM of fluorescein, 5 nM of Alexa-488 labelled DNA, and 100 pM of the unlabeled complementary strand of RNA (Figure 5-6). A maximum S/N for fluorescein was obtained using a bin size of 800 ms, while the maximum S/N for the duplex peak occurred when we chose a bin size of 400 ms. This is consistent with the fact
that the duplex peak is narrower than the fluorescein peak probably due to electrofocusing. To ensure that the S/N is optimal both for fluorescent dyes and for labeled biomolecular complexes, we used a compromise value of 600 ms for the time bin of the data presented here.

**Figure 5-6.** Normalized S/N as a function of data binning time for the electrophoretic separation of fluorescein (100 pM), DNA probe (5 nM) and DNA-miRNA hybrid (100 pM). The measured data is shown in the insert.

### 5.4.2 Excitation intensity

**Figure 5-7.** S/N versus the laser excitation power measured after the objective lens. The sample used is a long plug of 100 pM fluorescein.

Long plugs of 100 pM fluorescein (peak width > 10 s) were injected into the capillary by negative pressure to study the effect of the excitation intensity on the S/N. All fluorescence
5.4 Signal-to-noise optimization

Photons were detected in a single SPAD. The excitation intensity was optimized by varying the input laser power and measuring the resulting S/N from the sample. Our data indicate that the S/N is maximal when the sample is illuminated at 200 µW (Figure 5-7). This leads to an estimated focal plane intensity of 35 kW/cm², based on the FCS-measured focal point radius of 0.43 µm. This optimum excitation intensity was used in all the subsequent measurements.

5.4.3 Polarization and lifetime filtering

By placing a polarizer cube in the detection path, the fluorescence signal was divided into two polarization channels – one parallel and one perpendicular to the direction of polarization of the excitation source. The background signal in the ECI was measured using a sample containing only tetraborate buffer and it was found to be split in a ratio 2:1 between the parallel and perpendicular channels, respectively. The start-stop time (lifetime) histogram features a sharp peak near time zero, indicating that a considerable amount of scattering contributed to the background (Figure 5-8 (a)). Post-acquisition time-gating was applied in software to suppress the background by excluding photons that were detected within a short time window after the excitation pulse.

Figure 5-8. (a) Lifetime histograms of tetraborate buffer (black) and 100 pM fluorescein (red) measured in the ECI configuration. (b) S/N for 100 pM fluorescein versus time-gating window in different polarization channels.

Figure 5-8 (b) shows the S/N as a function of the width of the time-gate applied to the lifetime histogram obtained from the electropherogram of a 10-second plug of 100 pM
5.5 Detection sensitivity

Highly sensitive single-molecule measurements are typically performed by detecting fluorescence bursts as single molecules pass through the detection volume. Even though this is the best achievable sensitivity, biologically-relevant cellular concentrations produce at least 1000 molecules in the focal volume in ECI (~5.8 fL) and the burst analysis is not applicable unless the detection volume is reduced by the same ratio. Therefore, this mode of detection was not used for the work presented in this chapter; however the performance of CE-CMPF-ECI in the burst mode is evaluated in appendix B.

To characterize the sensitivity of our CE-CMPF-ECI setup, 60 pL samples of serially diluted fluorescein were injected into the capillary and propagated to the detector using electrophoresis. Samples containing roughly 3200, 1600, 800, 400 and 200 molecules were studied. The resulting S/N values of the fluorescein peaks, plotted as a function of the number of injected molecules, follow a linear trend ($R^2=0.998$) and give S/N = 3 at 50 molecules (mLOD) (Figure 5-9). This value is a theoretical limit, which is calculated by assuming a linear dependence of the signal detected on the number of molecules injected, which may not be valid as the number of molecules becomes smaller. However, as seen in Fig. 5-9 (a), for the most diluted sample, which contains only 200 molecules, the signal is well above the fluctuations of the background and the calculated limit of detection of 50 molecules seems very plausible. The concentration of this sample was 1.5 pM, giving a cLOD comparable to those achieved using burst analysis in tight
capillaries (250 fM [30] and 2 pM [35]). The error bars indicated in Figure 5-9 (b) represent a 15% variation and were determined by measuring the variation in peak area between 10 separate 60 pL injections of 100 pM of fluorescein. This error is presumably due to some irreproducibility in the volume of sample injected by the automated syringe pump.

![Figure 5-9](image)

**Figure 5-9.** (a) Stacked electropherograms showing the peak size of 60 pL injections of fluorescein of different concentrations: 25 pM, 12 pM and 6 pM, respectively. (b) The S/N ratio of the electropherograms was used to determine a linear calibration curve which indicates a S/N = 3 for ~50 molecules of fluorescein.

To determine the MCE and LOD of our setup we estimated the fraction of injected molecules passing through the detection volume by dividing the cross section of the detection volume, as measured perpendicular to the capillary’s axis, by the cross section of the capillary’s inner bore. Based on the results of the FCS measurements (see section 5.3), we estimated the MCE to be about 13%, which, combined with the mLOD value of 50, yields the LOD = 6.5 molecules of fluorescein for the CE-CMPF-ECI instrument. This LOD is comparable to results achieved using off-column designs and the mLOD is three orders of magnitude better than those achieved using a commercial CE instrument.

### 5.6 Analysis of miRNA biomarkers

Upon optimizing and characterizing the CE-CMPF-ECI instrument, we used it to perform the DQAMmiR protocol previously developed on a commercial CE instrument [14]. DQAMmiR is used to directly detect native miRNA species from cell lysate samples. Using fluorescently labeled ssDNA probes designed as complements to specific miRNA species, the probe-miRNA
hybrids are separated from the unbound probes using CE and the miRNA content is quantified. The method is applicable to multiple miRNA species by using different probes with different drag tags (i.e. biotin, hairpin-forming extensions on the probes and short peptides [46]), which shift the probe migration times from one another. In addition, the running buffer contains excess of single-stranded DNA binding protein (SSB), which shifts the migration times of all unbound probes from those that are bound.

Figure 5-10. Three miRNA species (from left to right, miR145, miR125b, miR21) spiked into a sample containing an internal standard (IS) and a 5-fold excess of probes. The sample is injected and separated by CE in a running buffer containing SSB. Electropherograms for CE-CMPF-ECI (a) and a commercial CE instrument (b) are shown.

Samples containing three DNA probes mixed with known, and equal, concentrations of three miRNA species were analyzed using CE-CMPF-ECI and the commercial CE instrument. The results are shown in Figure 5-10 and demonstrate similar S/N ratios for similar peaks, despite the fact that the CE-CMPF-ECI samples contained 600-times fewer miRNAs. Also notable was the fact that baseline separation was achieved by CE-CMPF-ECI in a third of the separation time required by the commercial CE instrument. By extrapolating the average S/N ratios of the three hybrid peaks down to S/N = 3, the miRNA mLOD of each setup can be determined. This gives
values of $10^5$ and 200 miRNA copies for commercial CE and CE-CMPF-ECI instruments, respectively, i.e., a 500 times higher sensitivity on our setup.

![Figure 5-11](image)

**Figure 5-11.** Three DNA probes are incubated in a sample of MCF-7 cell lysate along with an internal standard. The entire sample is injected and separated by CE in a running buffer containing SSB. Results obtained with CE-CMPF-ECI (a) and a commercial CE instrument (b) are shown. Only one miRNA species was detected (miR21) using both instruments. Around 1200× more cell lysate (and hence miRNA) was injected in the lower trace.

To show the performance of our setup in the analysis of crude biological samples, we performed the DQAMmiR protocol on a sample of MCF-7 breast cancer cell lysate. Though the use of cell lysate changes the conductivity of the sample and affects the migration times of analytes (with respect to Figure 5-10), the use of an internal standard allows the time axis to be scaled for inter-run comparisons. Figure 5-11 shows the data from similar samples measured on the CE-CMPF-ECI setup and the commercial CE instrument. The two electropherograms are identical except for some fine-structure detail on the SSB-DNA probe peak and internal standard peak, and a significant difference in miRNA-DNA peak height. The first two discrepancies are understood to be due to the high sensitivity of the CE-CMPF-ECI, which picks up small details that may have otherwise gone unnoticed. The difference in miRNA-DNA hybrid peak height is because 1,200-times more lysate was used for the commercial CE than for the CE-CMPF-ECI. Only one miRNA species is detectable (miR21) using both instruments, indicating the expression
of the other two miRNAs is suppressed in the cancerous cell. This is expected, since the overexpression of miR21 in MCF-7 cells has been reported previously [47].

Remarkably, the small injection volumes required for analysis by CE-CMPF-ECI allows for the possibility of analyzing the contents of a single cell. In order to test this capability, we applied DQAMmiR performed with CE-CMPF-ECI to the analysis of the 3 miRNAs in MCF-7 breast cancer cells (Figure 5-12, lower trace).

![Figure 5-12. Analysis of 3 miRNAs (miRNA 145, 125b, and 21) by DQAMmiR using the CE-CMPF-ECI setup. The upper trace corresponds to sampling a mixture of 3 synthetic miRNAs in MCF-7 lysate (positive control). The middle trace corresponds to a sample with no miRNAs (negative control). The bottom trace corresponds to sampling the MCF-7 cell lysate; the inset shows the zoomed-in peak corresponding to 12 thousand copies of the miRNA 21-probe hybrid. Note that despite the difference in peak heights on the top trace, the area of the three miRNA hybrids are comparable. The 3 traces are offset along the Y axis for clarity of presentation.](image)

The hybridization mixture was prepared with a crude cell lysate and 30 pL of this mixture was sampled for CE-CTRF-ECI. The volume sampled was estimated to contain the cellular content of approximately 1 cell (0.8). The probe-target hybrid for miRNA 21, which is up-regulated in these cells, was detected and corresponds to 12,000 copies of miRNA 21 per cell. The signals from miRNA 125b and 145, which are down-regulated in these cells [3], were below the 200-copy detection limit. The results obtained are in agreement with the previous estimates by other methods [9] and thus opens up interesting possibilities of using the CE-CMPF-ECI approach for single cell analytics.
5.7 Summary

In this study, we described the development and characterization of ultrasensitive capillary electrophoresis on a confocal setup. An embedded capillary interface was designed using a commercially available capillary and a simple protocol. This interface allowed highly sensitive measurements to be carried out in-capillary using multiparameter fluorescence detection at photon level which advantageously combines spectral, spatial, and temporal photon filtering. The signal-to-noise ratio was investigated by optimizing the excitation power, signal binning time, lifetime and polarization filtering after data acquisition. Sensitive CE experiments were carried out using this setup and a molecular collection of 13% and a concentration limit of detection of 1.5 pM of fluorescein molecules were achieved, comparable to those achieved by off-column and custom capillary designs. By performing on-column detection in a commercially available capillary we showed our design to be able to robustly perform: i) DQAMmiRNA with a 500× better detection limit compared to a commercial instrument, and ii) the analysis of samples of cell lysate down to the contents of a single cell. This makes the method applicable to biological samples containing a small number of cells such as fine-needle aspiration biopsies. In addition, our design was constructed entirely from commercially available parts. This makes the experimental setup not only feasible for production but also robust in operation. Importantly, this approach has a potential for further improvement in the detection limit without complicating the interface, by optimizing the excitation and collection optics for in-capillary detection. We foresee that CE-CTRF-ECI will become a practical instrumental platform for highly sensitive analyses of miRNAs and other disease biomarkers.

5.8 Experimental

Sample preparation. To calibrate the optical system, 200 nm fluorescent micro-spheres (F8809, Life Technologies, USA) were chemically immobilized to the surface of the inner bore of the capillary. Polyethyleneimide was injected and incubated in the capillary, followed by rinsing with water, then the injection and incubation of a solution of fluorescent micro-spheres, followed by another rinsing step with water and finally drying by air flowing (14 psi) through it. The duration of each step was 5 minutes.
Two fluorescent dyes, Rhodamine110 (Rh110) (20310, Biotium, USA) and fluorescein (F6377, Sigma Aldrich, USA), were prepared in a buffer of 50 mM Tris-acetate, 50 mM NaCl, 10 mM EDTA, pH 7.8. The limit of detection of the CE-CMPF instrument was determined by analyzing the signal from injections of fluorescein. Separate samples of 100, 50, 25, 12.5, and 6 pM of fluorescein were injected into a capillary pre-filled with 25 mM sodium tetraborate (pH 9.3). The fluorescein samples were injected by sealing the outlet vial to the syringe and programming the syringe pump to draw the syringe back, hold for 10 s, and then return to its initial position. The volume of sample injected was determined by finding propagation speed of fluorescein under the pressure of the syringe with no voltage applied. By drawing the syringe back 2 mL, with an initial dead volume of 10 mL, the speed of propagation was measured as 0.09 mm/s, and the pressure of injection was determined to be approximately 2.4 psi using the Hagen-Poiseuille equation. Under these conditions, 70 pL of sample was injected in 10 s.

The CE-based DQAMmiR method was applied for detecting different micro(mi)RNA species (miR21, miR145 and miR145 (synthesized by IDT, Coralville, IA, USA) using a procedure described elsewhere [14]. Briefly, the miRNAs were hybridized to single-stranded (ss) DNA probes labeled with Alexa-488 in the incubation buffer (50 mM Tris-acetate, 50 mM NaCl, 10 mM EDTA, pH 7.8). Each miRNA species was diluted to a concentration of 100 pM for CE-CMPF-ECI or 2 nM for commercial CE, with a 5-fold excess of probe ssDNA in each case, while the samples of MCF-7 cell lysate were used at $2.72 \times 10^7$ cells/mL and $118 \times 10^7$ cells/mL, respectively. Also included in the mixtures for commercial CE were 10 nM fluorescein, 5 µM masking DNA (20-nucleotide DNA strand and 2 µM masking RNA (tRNA library), while for CE-CMPF-ECI they were all reduced by a factor of 20. The fluorescein was added as an internal standard to help quantify variations in the volume of injected sample and in peak migration times. The masking DNA and RNA were added to prevent the degradation of DNA and RNA, respectively, as well as to prevent adsorption of the probe or the miRNA to the walls of the vial during incubation. Sample injection volumes were 6 nL for the commercial CE instrument, 200 pL for CE-CMPF-ECI instrument, and 30 pL for the injection of the equivalent lysate of a single cell using CE-CMPF-ECI.

Pre-existing DNA or RNA structures were denatured by raising the temperature of the mixture to 80°C, and hybridization was promoted by cooling the mixture to 37°C at a rate of
20°C/min. The mixture was incubated at 37°C overnight to ensure complete hybridization without optimizing the hybridization time. To determine the miRNA detection limit, a single DNA probe-miRNA pair was used and SSB was removed from the run buffer. The uncertainty on the limit of detection of microRNA was ±20% (one standard deviation) as determined by measuring the run-to-run variations in peak area from 30 pL injections of the various concentrations of microRNA, which were performed in triplicates and normalized by the area of the fluorescein internal standard.

Cell lysates were prepared in the following way: MCF-7 cells were purchased from ATCC and grown in incubator at 37°C in the atmosphere of 5% CO₂. Cells were grown in DMEM/F12 media (Invitrogen, Carlsbad, CA, USA) with 20 ng/mL hEGF, 0.5 µg/mL hydrocortisone, 10 µg/mL insulin, FBS and 10,000 µg/mL penicillin, streptomycin in a 100 mm Petri dish. When cells covered roughly 90% of the plate they were washed with PBS, trypsinized to be detached from bottom of dish and centrifuged at 150 ×g for 5 min. Pellet was washed twice with PBS. The cells were counted using haemocytometer and lysed with 1% Triton in 50 mM Tris-acetate, 50 mM KCl, 10 µM masking RNA, 0.1 mM EDTA, pH 8.16. Cell lysates were aliquoted and stored in liquid nitrogen for further use.

**Capillary electrophoresis in the ECI.** To construct the ECI, we first created a 2-cm detection window in the capillary (TSP010150, Polymicro Tech., Phoenix, AZ, USA) by removing the capillary’s polyimide coating under an open flame for several seconds. The detection window and the surface of the microscope coverslip (Gold Seal #3223, Electron Microscopy Sciences, Hatfield, PA, USA) were cleaned using ethanol and the detection window was then carefully pressed to the cover slip using the plastic prongs of tweezers, which were held roughly 1 cm apart. A small bead of adhesive (NOA61, Thorlabs, Newton, NJ, USA) was placed directly in between the two prongs and the bead was cured under a UV lamp for 1 minute. The prongs were then removed and the rest of the bare capillary resting on the cover slip was coated in adhesive (to reduce fragility) and cured for 2-3 minutes.

Fluorescence detection was performed by placing the ECI device on top of our multiparameter confocal microscope. Samples were injected into the capillary by negatively pressurizing an airtight chamber sealed to the outlet end of the capillary. The chamber was
pressurized by drawing back the plunger on a 60 mL syringe by a pre-defined distance for a pre-defined duration of time. The exact distance and duration was controlled using an automated syringe pump (NE-1010, New Era Pump Systems Inc., Farmingdale, NY, USA). The total amount of sample injected was determined by experimentally measuring the flow rate of a fluorescent dye under the pre-defined injection pressure then using this flow rate to determine injection volumes for fixed injection durations. 200 pL injections were performed by beginning with the syringe plunger at the 0 mL mark, drawing the plunger back 5 mL, holding for 5 s, then returning the plunger to the 0 mL mark and opening the chamber to the atmosphere. 60 and 30 pL injections were performed by beginning with the syringe plunger at the 10 mL mark, then drawing the plunger back to 12 mL, holding for 10 and 5 s, respectively, then returning the plunger back to the 10 mL mark and opening the chamber to the atmosphere. Electrophoresis was carried out by placing a platinum electrode in the inlet and outlet vials and using a high voltage power supply (CZE1000r Spellman, Hauppauge, NY, USA) to provide an electric field of 300 V/cm across the capillary’s length. The running buffer was 25 mM sodium tetraborate, pH 9.3, with 50 nM of single-stranded DNA binding protein (SSB) from E. coli (SSB02200, Epicentre Biotechnologies, Madison, WI, USA) added to the buffer for miRNA detection experiments. The capillary was flushed prior to every CE run with 0.1 M HCl, 0.1 M NaOH, deionized H2O, and running buffer for 1 min each. The total length of the capillary was ~50 cm and the length from the injection end to the detection point was ~7 cm.

**Capillary electrophoresis in a commercial instrument.** We used a P/ACE MDQ capillary electrophoresis system (Beckman-Coulter, Fullerton, CA, USA) with laser-induced fluorescence detection. Fluorescence was excited with a 488 nm continuous wave solid-state laser (JDSU, Santa Rosa, CA, USA). We used bare fused-silica capillaries with an outer diameter of 365 μm, an inner diameter of 75 μm, and a total length of 50 cm. The distance from the injection end of the capillary to the detector was 39 cm. The running buffer was 25 mM sodium tetraborate, pH 9.3, with 50 nM SSB. The capillary was flushed prior to every CE run with 0.1 M HCl, 0.1 M NaOH, deionized H2O, and running buffer for 1 min each. Samples were injected by a pressure pulse of 0.5 psi for 5 s, the volume of the injected sample was ~6 nL. Electrophoresis was driven by an electric field of 500 V/cm. Electropherograms were analyzed using 32Karat Software. Peak areas were divided by the corresponding migration times to compensate for the dependence of the residence time in the detector on the electrophoretic velocity of species. For electropherograms
from both instruments all areas were normalized by dividing them by the area of internal standard, fluorescein.

**CE-CMPF instrument.** Time- and polarization-resolved laser-induced fluorescence measurements were performed on a multiparameter confocal microscope described in detail in chapter 1. Briefly, a Ti:sapphire laser (Tsunami HP, Spectra Physics, Santa Clara, CA, USA) was used as the excitation source, producing laser pulses of ~ 100 fs duration and 10-12 nm spectral width. The output was tuned to a center wavelength of 956 nm and frequency-doubled to 478 nm by focusing the laser beam into a β-BBO crystal. A dichroic mirror directed laser light to a 1.4 NA/100× oil immersion microscope objective (PlanApochromat, Carl Zeiss, Toronto, Canada). The embedded capillary interface (ECI) was mounted on a three-axis piezo scanner (T225, MadCity Labs, Madison, WI, USA), which controls the position and depth of the focus into the capillary.

The fluorescence was collected by the same objective and the excitation laser scattering was removed by a series of long-pass and band-pass spectral filters. Light was passed through a 150-μm pinhole for confocal detection and then it was directed either fully to a single-photon avalanche diode (SPAD) detector (PD5CTC, Optoelectronic Components, Kirkland, Canada) or split by a broadband polarizer cube into components and focused onto two SPADs.

Measurements were performed in the Time-Tagged Time-resolved (TTTR) mode. In this mode, the output of the SPAD and a small fraction of the laser excitation pulse, as detected by a fast photodiode (PHD-400-N, Becker & Hickl, Berlin, Germany), were used as inputs for the Time-Correlated Single-Photon Counting (TCSPC) module (PicoHarp300, PicoQuant, Berlin, Germany). The delay between the excitation pulse and the detected photon (start–stop time), and the absolute arrival time of each photon are recorded with 4 ps resolution and up to a maximum count rate of 10 MHz. Custom LabView codes previously developed in the Gradinaru Lab were used to control photon data acquisition, process the data and visualize intensity time-trajectories and start–stop (lifetime) histograms.
5.9 Bibliography


Appendix A

Photon detection efficiency

Under pulsed excitation conditions, the brightness of a molecule situated at the center of the detection volume can be written as [1]:

$$
\eta = \varepsilon_0 \cdot \frac{T_0 \cdot \Phi F}{T_0} \cdot \frac{1 - \exp\left(-\langle I_0 \rangle / I_S \right)}{\kappa \cdot \frac{T_0}{\Phi \text{ISC}}} \left[1 - \exp\left(-\langle I_0 \rangle / I_S \right)\right]
$$  \hspace{1cm} (A.1)

The parameters of this expression, including the time-averaged excitation intensity \(\langle I_0 \rangle\), are explained in detail in the article. To calculate the brightness for a dye solution with \(N_{eff} = 1\), one needs to substitute \(\langle I_0 \rangle\) and \(\varepsilon_0\) with \(\langle I(x, y, z) \rangle\) and \(\epsilon(x, y, z)\) and integrate over the volume. Instead of this cumbersome calculation, we used an approximation by substituting \(\langle I_0 \rangle\) and \(\varepsilon_0\) with their spatial-averaged values \(I_{avg}\) and \(\epsilon_{avg}\) in the expression above. Our purpose is to find out the single-molecule detection efficiency \(\epsilon_0\) from the average value \(\epsilon_{avg}\), which can be determined by fitting the experimental saturation data to the expression for molecular brightness.

We can assume a Gaussian excitation beam given by:

$$
\langle I(x, y, z) \rangle = \frac{2P}{\pi \Omega_0^2 \left(1 + \left(\frac{z}{z_0}\right)^2\right)} g(x, \omega(z)/2) g(y, \omega(z)/2)
$$  \hspace{1cm} (A.2)

where \(P\) is the measured time-averaged excitation power after the microscope objective and 
\(g(x, \sigma) = \exp\left(-\frac{x^2}{2\sigma^2}\right).\) The spot size of the beam varies as \(\omega(z) = \Omega_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2},\) where \(\Omega_0\) is the beam waist and \(z_0\) is the Rayleigh range, \(\pi \Omega_0^2 / \lambda.\) By integration, we obtain the average excitation intensity over the beam waist:
\[
I_{\text{avg}} = \frac{P}{\pi \Omega_0^2} \approx \frac{P}{2\pi \omega_0^2}
\]  

(A.3)

where \( \omega_0 = \Omega_0 / \sqrt{2} \) is waist of the detection volume as determined by FCS calibration measurements (see below). The derivation of the expression for \( \eta \) was done using the intensity \( \bar{I} = P / \pi \omega_0^3 \). In order to account for correct averaging across the waist of the beam, this must be divided by a factor of 2. Note that this correction leads to a value for saturation intensity of EGFP that is in excellent agreement with the value reported in literature (see chapter 2).

The average collection efficiency \( \varepsilon_{\text{avg}} \) was calculated assuming that the confocal pinhole was sufficiently small so that its projection in the object plane is smaller than the excitation beam waist \( \Omega_0 \). This is supported by the fact that the lateral size of the detection volume measured by FCS with and without pinhole was 230 nm and 370 nm, respectively. It follows that \( \varepsilon(x, y, z) \) can be approximated, at least in the lateral dimensions, by the point spread function:

\[
\varepsilon(x, y, z) = \varepsilon_0 g(x, \Omega_0 / 2) g(y, \Omega_0 / 2) g(z, s \omega_0 / 2)
\]  

(A.4)

The typical dimensions of the detection ellipsoid measured by FCS are \( \omega_0 = 230 \text{ nm} \) and \( \omega_z = s \omega_0 = 1.26 \mu m \), the latter considerably larger than the Rayleigh range \( z_0 = 0.68 \mu m \). This discrepancy is a common feature in FCS experiments with high numerical aperture (NA) objectives, for which measured \( s \) values are around 6. Spherical aberrations of these objectives are considered the main source of the axial elongation of the detection volume [2, 3]. For the purpose of estimating \( \varepsilon_{\text{avg}} \), we will assume that \( \omega_z \) is mainly determined by the size of the pinhole and that the excitation intensity is axially uniform over the detection volume.

The average collection efficiency \( \varepsilon_{\text{avg}} \) depends on both the collection efficiency function and the probability of fluorescence emission:
At low excitation intensity, the probability distribution of fluorescence is given by the normalized excitation intensity distribution. Therefore, as explained above, we can assume that the probability of fluorescence does not depend on \( z \). We can further assume that photons emitted outside the range \(-\omega_z < z < \omega_z\) are blocked by the confocal pinhole, so that a step function of width of \( 2\omega_z = 2s\omega_0\), \( U(2s\omega_0)\), can be used to approximate the \( z \)-component of the fluorescence probability function. With these simplifications, \( \epsilon_{avg} \) can be calculated as:

\[
\epsilon_{avg} = \int \epsilon(x, y, z)P(x, y, z)dv
\]

(A.5)

\[
\epsilon_{avg} = \frac{2}{\pi\Omega_0^2} \frac{1}{2s\omega_0} \int g(x, \Omega_0/2)g(y, \Omega_0/2)U(2s\omega_0)dv
\]

\[
= \epsilon_0 \frac{2}{\pi\Omega_0^2} \frac{1}{2s\omega_0} \int g(x, \Omega_0/2)^2 dx \int g(y, \Omega_0/2)^2 dy \int g(z, s\omega_0/2)dz
\]

(A.6)

\[
= \epsilon_0 \frac{1}{\pi\omega_0^2} \frac{1}{2s\omega_0} \int g(x, \omega_0/2)dx \int g(y, \omega_0/2)dy \int g(z, s\omega_0/2)dz
\]

\[
= \epsilon_0 \frac{1}{\pi\omega_0^2} \frac{1}{2s\omega_0} \frac{\pi s\omega_0^2}{\sqrt{2}} = \epsilon_0 \frac{\sqrt{\pi}}{4\sqrt{2}} \approx 0.31 \epsilon_0
\]

With an average value \( \epsilon_{avg} = 0.73\% \) (main text), the photon detection efficiency of the setup is calculated to be \( \epsilon_0 = 2.3\% \).

Alternatively, we assumed that the axial size of the detection volume is \( \sqrt{2} \omega_z \), which is determined by the beam length and point spread function and it is not confined by pinhole. In this case, both the collection efficiency and the excitation intensity profiles are assumed to be identical and the \( z \)-component of the collection function is replaced by another Gaussian function:
\[ \varepsilon_{\text{avg}} = \int e_0 g(x, \Omega_0 / 2)g(y, \Omega_0 / 2)g(z, s\Omega_0 / 2) \]
\[ = \varepsilon_0 \frac{2}{\pi \Omega_0^2} \frac{\sqrt{2}}{\sqrt{\pi s\Omega_0}} g(x, \Omega_0 / 2) \frac{g(y, \Omega_0 / 2)g(z, s\Omega_0 / 2)dv}{2} \]
\[ = \varepsilon_0 \frac{2}{\pi \Omega_0^2} \frac{\sqrt{2}}{\sqrt{\pi s\Omega_0}} \int g(x, \Omega_0 / 2)^2 dx \int g(y, \Omega_0 / 2)^2 dy \int g(z, s\Omega_0 / 2)^2 dz \]
\[ = \varepsilon_0 \frac{1}{\pi \omega_0^2} \frac{1}{\sqrt{\pi s\omega_0}} \int g(x, \omega_0 / 2)dx \int g(y, \omega_0 / 2)dy \int g(z, s\omega_0 / 2)dz \]
\[ = \varepsilon_0 \frac{1}{\pi \omega_0^2} \frac{1}{\sqrt{\pi s\omega_0}} \frac{\pi \omega_0^2}{2} \frac{\sqrt{\pi s\omega_0}}{2} = \varepsilon_0 \frac{1}{2\sqrt{2}} \quad (\text{if } s = 5.5) \]
\[ \approx 0.35 \varepsilon_0 \]

Using again the experimental value for \( \varepsilon_{\text{avg}} \) (0.73%) we obtain \( \varepsilon_0 = 2.1\% \), thus implying that the estimation of the photon detection efficiency does not critically depend on the details of the model used for the axial dependence of the collection efficiency function.

References


Appendix B

Single-molecule bursts in the capillary

A capillary was filled by electrophoresis buffer (tetraborate, pH 9.3). Single-molecule bursts were recorded from a short plug of 100 pM fluorescein propagated through the capillary by pressure (Figure A.1 (a)). A 200-μs binning time was applied to the fluorescence intensity trajectory. This binning time resembles the diffusion time of a free dye through the detection volume found from FCS (107 μs). In order to determine the threshold level for the burst recognition, a 0.5 s data slice was analyzed in an interval without a clear peak (Figure A.1 (b)). Three standard deviations above the average number of photons/bin resulted in the threshold level of 4 photons/bin, which was used in subsequent analysis. The 0.5 s time trajectory from tetraborate buffer resulted in 13 bursts (Figure A.1 (b)).

In the centre of the plug (at 169 s in the full trajectory) the bursts are not distinguishable due to the high concentration (Figure A.1 (d)). In fact, based on the capillary and detection region size we estimate to have an average of 1 molecule in the detection volume at a concentration of ~300 pM. Far enough off the centre of the plug the concentration reduces due to the broadening of the plug during propagation through capillary and therefore burst analysis becomes feasible. Figure A.1 (c) shows a 0.5 s slice of the intensity-time trajectory at 167 s. The average concentration in this slice is ~5× lower than the concentration in the center of the plug, i.e., ~ 20 pM, and it contains a total of 117 bursts. An average S/N = 4.5 was calculated for these bursts. Here signal (S) is calculated by subtracting the average number of photons/bin found from the buffer slice (Figure A.1 (b)) from the average number of photons in these bursts. Noise (N) is taken as the standard deviation of the buffer slice. This low S/N ratio can be explained by the 6-fold reduction
in the fluorescence intensity of the bursts in ECI compared to the sub-micrometer capillaries due to excitation beam expansion and reduction of the photon collection efficiency, as discussed in the main text. In addition, regular capillaries are prone to higher background fluorescence and backscattering because of curvature of capillaries, index mismatch, extended optical path in the glass and a larger detection volume.

**Figure B.1** 100 pM of fluorescein propagated through the capillary by pressure. In (a) the full passage of the plug is shown, in (b), (c), (d) 0.5 intervals at t = 16, 167, 169 s are shown, respectively.
Statement of Contributions

Chapter 1:

I was involved in the building, the development and the daily maintenance of the confocal multiparameter fluorescence (CMPF) microscope. I developed the core program for FCS analysis in LabView. Baoxu Liu, developed the main data acquisition and analysis program. I developed several sample protocols based on the existing literature, and finalized them with the assistance of other lab members. These protocols include coverslip surface modification, protein labelling and purification, etc.

Chapter 2:

I investigated the experimental conditions should be fulfilled in order to obtain reliable physical parameters via FCS. I showed the performance of FCS in several in vivo and in vitro applications. The fluorescent Drosophila S2 cells were prepared and provided by Dr. Xinping Qiu from Prof. Bryan Stewart group (Biology, UTM). The STAT3 protein and the F*-peptide were provided by Joel Drewry from Prof. Patrick Gunning group (Chemistry, UTM).

Chapter 3:

I investigated the nature of dark states in the correlation curves. Abdullah Bahram prepared the dyes and oxygen scavenger solutions. Abdullah also performed some of the experiments under my supervision. ClpP protein was provided by Angela Yu from Professor Walid Houry group (Biochemistry, UofT). The SH3 proteins were provided by Hong Lin from Prof. Julie Forman-Kay group (Biochemistry, UofT).
Chapter 4:

I investigated the size and shape of the folded, unfolded and chemically-denatured states of DrkN SH3 by FCS and smFRET burst experiments. The LabView code for the smFRET burst analysis was developed by Baoxu Liu. The MD simulations were performed by Sarah Rauscher and Grace Li from prof. Regis Pomès group (Biochemistry, UofT). Abdullah Bahram performed some of the experiments under my supervision. The Amberlite MB-1 resin and the basic protocol for urea purification were provided kindly by Prof. Voula Kanelis. Abdullah Bahram purified urea and helped me in SH3 labeling. The SH3 proteins were prepared and purified by Hong Lin from Prof. Julie Forman-Kay group (Biochemistry, UofT).

Chapter 5:

I and Bryan Dodgson developed an ultrasensitive capillary electrophoresis on our CMPF microscope to perform DQAMmiR experiments. Bryan constructed the ECI based on my prototype. Bryan also involved in all measurements. I characterized the setup and I myself analyzed all the data acquired in our setup except for detection sensitivity estimation and miRNA biomarkers measurements which mainly analyzed by Bryan. The cell lysates and miRNA samples were prepared by David Wegman. The experiments on the commercial CE instrument were performed in the Prof. Sergey Krylov lab (Chemistry, York University) by Bryan and David.
List of Publications


* The authors contributed equally to the work


* The authors contributed equally to the work


