Single-Molecule Spectroscopy of Disordered States and Dynamics in Proteins

by

Zhenfu Zhang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

© Copyright 2017 by Zhenfu Zhang
Single-Molecule Spectroscopy of Disordered States and Dynamics in Proteins

Zhenfu Zhang

Doctor of Philosophy, 2017
Graduate Department of Physics
University of Toronto

Abstract

Single-molecule fluorescence techniques, such as Förster resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCS), were applied to investigate the conformations, dynamics and interactions of disordered protein systems using custom-built confocal and total internal reflection microscopes.

Conformational states of the N-terminal Src-homology-3 domain of downstream of receptor kinases (drkN SH3) were characterized. Both folded and unfolded states of drkN SH3 were detected under non-denaturing conditions. When exposed to high concentrations of urea and GdmCl denaturants, the protein still exhibits two distinct smFRET populations. We propose that the high-FRET population corresponds to denaturation-induced looped conformations. FRET experiments in formamide and DMSO suggest that interactions between hydrophobic amino-acid groups in the distal regions are involved in the formation of the looped state.

To gain more insight into the determinants of unfolded protein structures, we studied an intrinsically disordered protein (IDP), the eukaryotic initiation factor 4E (eIF4E) binding protein 2 (4E-BP2). Nanosecond-scale dynamics was observed by FCS and was tentatively
assigned to local peptide chain contact formation. Our data suggest that multi-site phosphorylation of the protein slows down the proximal chain motions and modulates the kinetics of distal regions. Segmental rotational correlation times and wobbling cone angles provided a rigidity map of the protein at different sites and were used to evaluate the binding mode to eIF4E. smFRET analysis reveals changes in the conformational ensemble responding to phosphorylation, denaturation, salt and pH. It is shown that both hydrophobic and electrostatic interactions play vital roles in determining the conformations of 4E-BP2.

Encapsulating proteins into lipid vesicles has been widely used in studying their structural properties and the kinetics of protein-protein interactions. False signals will be introduced in these studies if the fluorescent probe itself interacts with the lipid wall. We employed the FCS method to systematically quantify the interaction between commonly used fluorophores and lipids and proposed mechanisms for the underlying interactions. This study offers a baseline correction for non-specific interactions and can be used to guide the pairing of dyes and lipids in the single-molecule fluorescence studies.
To my beloved parents

献给我挚爱的父母
Acknowledgments

I would like to express my sincere thanks to all the people who supported and helped me in the past six years at the University of Toronto, not only in research and teaching, but also in daily life. Even if you are not on the page, you are surely in my heart.

I am particularly indebted to my supervisor Prof. Claudiu Grădinaru for bringing me into this fascinating single molecule biophysics field and for his endless supervising, support and help in my research. My thesis projects would not be accomplished without his vision, guidance and insightful discussions. Here, I really want to appreciate his efforts during these years and say “Thank you” to him.

Next, I would like to thank my committee members, Prof. Joshua Milstein and Prof. David McMillen for their insightful inputs, helps and discussions throughout the years. I also want to thank all my defense committees, Prof. Elizabeth Rhoades, Prof. Régis Pomès, and Prof. William Ryu to review my Ph.D thesis and provide valuable feedback.

I would like to thank my former lab members, Dr. Baoxu Liu, Dr. Amir Mazouchi and Abdullah Bahram. They patiently answered my technical questions and offered meaningful discussions at the beginning of my Ph.D program. Because of their support and encouragement, I stumbled less in my journey.

I would like to express my distinctive gratitude and appreciations to our research collaborators: Prof. Julie Forman-Kay, Prof. Peter Macdonald, Prof. Ulrich Krull, Prof. Hue Sun Chan, Dr. Qasim Saleem, Dr. Alaji Bah, Hong Lin, Dr. Andrew Chong, Dr. Rhea Hudson, Dr. Jennifer Dawson, Dr. Uvaraj Udayasankar and Dr. Jianhui Song. Without their constructive suggestions and valuable samples provided, the content of my thesis would not be possible.

I would also like to thank other graduate students in the lab, Yuchong Li, Gregory Gomes, Dennis Fernandes. They provided valuable support and discussion in my research projects. My projects would not go smooth without them.
Following that I would like to acknowledge other members worked in the Gradinaru group for creating a nice working atmosphere in the lab and being friends with me. They are Anthony Pugliese, Daniel Jacobs, Jordan Rebelo, Darius Chia, Amanda Leung, Jessica Kun, Spencer Smyth, Sam Siu, Dan Yomo, Mehik Rizvi, Jasbir Bamrah, John Castroverde, Juan Castaneda, Charlie Huston and Florent Mercier. It is my great pleasure meeting you all.

I would like to thank Department of Physics at the University of Toronto and the “CIHR Training Program in Protein Folding and Interaction Dynamics: Principles and Diseases” for financial support within these years.

Last but not least, this thesis could not have been written without the absolute loving support of my dear family. I would like to thank my parents Honggen Zhang and Nianxia Zhu for giving birth to me and their insight, guidance and encouragement will always in my heart. Your love makes this thesis come true.
Contents

List of Tables xiii
List of Figures xiv
List of Appendices xvi
List of Abbreviations xvii

Chapter 1 Introduction 1

1.1 Molecule of Life ................................................................................................... 2
  1.1.1 Proteins ............................................................................................................ 2
    1.1.1.1 Structural organization in proteins ..................................................... 2
    1.1.1.2 Protein folding ....................................................................................... 3
    1.1.1.3 Intrinsically disordered proteins ......................................................... 5
  1.1.2 Lipids .............................................................................................................. 8
    1.1.2.1 Lipids and the lipid bilayers ............................................................... 8
    1.1.2.2 Liposomes and bicelles ................................................................. 10
    1.1.2.3 Drug delivery applications .............................................................. 12
  1.2 Single Molecule Fluorescence ...................................................................... 13
    1.2.1 Basics of fluorescence ............................................................................ 13
      1.2.1.1 Organic fluorophores and Jablonski diagram .................................. 13
      1.2.1.2 Fluorescence quenching ................................................................. 16
      1.2.1.3 Förster resonance energy transfer ................................................. 19
      1.2.1.4 Fluorescence correlation spectroscopy .......................................... 22
      1.2.1.5 Fluorescence anisotropy decay ...................................................... 24
      1.2.1.6 Fluorescence photobleaching ....................................................... 26
      1.2.1.7 Fluorescence recovery after photobleaching .................................. 27
    1.2.2 Single molecule fluorescence setups .................................................... 33
      1.2.2.1 Total internal reflection fluorescence microscope ...................... 33
1.2.2.1.1 Principles of total internal reflection fluorescence microscope ................................................................. 33
1.2.2.1.2 TIRFM data acquisition .............................................. 34
1.2.2.1.3 TIRFM alternating laser excitation ............................ 39
1.2.2.1.4 TIRFM data analysis ................................................. 41
1.2.2.2 Multi-parameter fluorescence confocal microscope ....... 50
1.2.2.3 Dual-color fluorescence correlation microscope ............ 52

Chapter 2 Resolving Conformational Heterogeneity of an SH3 Domain Using Single-Molecule Fluorescence Spectroscopy and an Excluded-Volume Polymer Model 55

2.1 Abstract ............................................................................................................... 55
2.2 Introduction ........................................................................................................ 56
2.3 Materials and Methods ....................................................................................... 58
  2.3.1 Materials ...................................................................................................... 58
  2.3.2 Protein expression and purification ........................................................ 58
  2.3.3 Sample preparation .................................................................................. 59
  2.3.4 Instrumentation ..................................................................................... 61
  2.3.5 FCS analysis ........................................................................................... 62
  2.3.6 smFRET analysis .................................................................................. 62
  2.3.7 Polymer model for R_g inference from smFRET data ...................... 63
2.4 Results ................................................................................................................ 65
  2.4.1 Chemical denaturants lead to a heterogeneous set of conformations that include a high FRET cluster .......................................................... 67
  2.4.2 The effect of different dye pairs ............................................................. 73
  2.4.3 The high FRET cluster in denaturants is not caused by aggregation...... 75
2.5 Discussion ........................................................................................................... 79
2.6 Conclusions ........................................................................................................ 87

Chapter 3 Single-Molecule Dissection of the Conformations, Dynamics and Binding of the Disordered 4E-BP2 Protein 91

3.1 Abstract ............................................................................................................... 91
3.2 Introduction ........................................................................................................ 91
3.3 Materials and Methods ................................................................. 95
  3.3.1 Materials .................................................................................. 95
  3.3.2 Protein expression and purification .......................................... 95
  3.3.3 Sample preparation ................................................................... 97
  3.3.4 Instrumentation .......................................................................... 98
  3.3.5 FCS analysis ............................................................................. 99
  3.3.6 FAD analysis ............................................................................ 99
  3.3.7 smFRET analysis ........................................................................ 100
  3.3.8 Polymer model for $R_G$ and $\langle R_{EE}^2 \rangle$ inference from smFRET data .......................................................... 102

3.4 Results and Discussion ................................................................. 103
  3.4.1 Towards a rigidity map of 4E-BP2: probing local chain flexibility ..... 103
  3.4.2 Local chain flexibility highlights the difference between intrinsically and denatured disordered states ........................................ 106
  3.4.3 Segmental motion analysis reveals the protein-protein interaction interface ................................................................. 108
  3.4.4 Intra-chain kinetics for 4E-BP2 .................................................. 111
    3.4.4.1 The effect of phosphorylation .............................................. 113
    3.4.4.2 The effect of denaturation .................................................. 115
  3.4.5 Inter-chain kinetics for 4E-BP2 bound to eIF4E ........................... 117
  3.4.6 The linear size(s) of 4E-BP2 ........................................................ 119
    3.4.6.1 The effect of phosphorylation .............................................. 119
    3.4.6.2 The effect of denaturation .................................................. 122
    3.4.6.3 Conformation exchange dynamics: the width of smFRET peaks ................................................................. 123
    3.4.6.4 Non-cooperative conformational transition ...................... 124
    3.4.6.5 The effect of salt .................................................................. 125
    3.4.6.6 The effect of pH ................................................................. 128
    3.4.6.7 $R_G$ and shape factor inference from smFRET data ........... 131

3.5 Conclusions .................................................................................. 133

Chapter 4 Lipids and Fluorophores --- Friends or Foes? 137
  4.1 Abstract ....................................................................................... 137
  4.2 Introduction .................................................................................. 137
4.3 Materials and Methods ................................................................. 139
  4.3.1 Materials ...................................................................................... 139
  4.3.2 Sample preparation ....................................................................... 140
    4.3.2.1 Small and large unilamellar vesicles preparation ............... 140
    4.3.2.2 Giant unilamellar vesicle preparation ................................ 141
    4.3.2.3 Supported lipid bilayer preparation .................................. 141
    4.3.2.4 Organic fluorophore preparation ....................................... 142
    4.3.2.5 Polyethylene glycol surface .............................................. 142
  4.3.3 Instrumentation ........................................................................... 143
    4.3.3.1 Dynamic light scattering ................................................... 143
    4.3.3.2 Fluorescence emission spectrum ....................................... 143
    4.3.3.3 FCS experiments .............................................................. 143
    4.3.3.4 Fluorescence lifetime experiments .................................. 144
    4.3.3.5 Fluorescence Imaging ....................................................... 145
  4.3.4 LogD calculations ....................................................................... 145
  4.3.5 FCS analysis ............................................................................... 146
  4.3.6 Theory of partition coefficient and free energy of partitioning ...... 148
  4.4 Simulations ..................................................................................... 149
  4.5 Results ........................................................................................... 151
    4.5.1 The brightness control ............................................................. 151
    4.5.2 Partition coefficient and free energy of partitioning ............... 153
    4.5.3 The electrostatic interaction .................................................... 155
    4.5.4 The effect of lipid head groups .............................................. 156
    4.5.5 The effect of acyl chain saturation level ............................... 158
    4.5.6 The effect of cholesterol ......................................................... 159
    4.5.7 Fluorescence imaging on GUV .............................................. 160
    4.5.8 Single-molecule imaging of fluorophores on supported lipid bilayers 161
    4.5.9 The shift of emission spectra ................................................. 162
  4.6 Discussion ...................................................................................... 164
  4.7 Conclusions .................................................................................... 168

Concluding Remarks and Future Directions .......................... 172
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>176</td>
</tr>
<tr>
<td>Appendix A</td>
<td>204</td>
</tr>
<tr>
<td>Appendix B</td>
<td>206</td>
</tr>
<tr>
<td>Appendix C</td>
<td>212</td>
</tr>
<tr>
<td>Appendix D</td>
<td>213</td>
</tr>
<tr>
<td>Statement of Contributions</td>
<td>219</td>
</tr>
<tr>
<td>List of Publications</td>
<td>223</td>
</tr>
</tbody>
</table>
## List of Tables

1-1 Results of Monte Carlo simulations of FRAP curves ........................................ 32
1-2 Output of histograms fittings ........................................................................... 49

2-1 Fitting parameters from the smFRET histogram .............................................. 67
2-2 FRET efficiencies and fractional populations ................................................... 69
2-3 $R_H$, $R_G$ and $\rho$ of different conformation states of drkN SH3 ...................... 77
2-4 $R_H$ of drkN SH3 in GdmCl, formamide, and DMSO solutions ....................... 77
2-5 $R_G$ of different conformations of drkN SH3 ................................................... 80

3-1 $\rho_{seg}$ of 4E-BP2 in different conditions .......................................................... 107
3-2 $\rho_{seg}$ and dye half-cone angles in the apo and bound states ......................... 110
3-3 $\langle E \rangle$ and $\sigma$ of NP and HP 4E-BP2 at different GdmCl and urea concentrations. ................................................................. 122
3-4 $\langle E \rangle$ of NP and HP 4E-BP2 at different KCl concentrations ....................... 126
3-5 Net charge and $\langle E \rangle$ of the 4E-BP2 at different pH .................................... 130
3-6 $R_G$ and $\sqrt{\langle R_{EE}^2 \rangle}$ values under different solvent environment .......... 132

4-1 $K_p$ of different fluorophores with DOTAP, DOPC, and DOPG lipids .............. 154
4-2 Density of fluorescent molecules observed on DOPC SLB .............................. 162
4-3 Shifts of the fluorescence emission spectra .................................................... 163
List of Figures

1-1 The solution structure of the drkN SH3 domain ............................................. 2
1-2 Schematic representation of the energy landscape ........................................... 3
1-3 Structures of DOPC lipid and cholesterol ....................................................... 10
1-4 Scheme of a liposome and a bicelle ................................................................. 12
1-5 Jablonski diagram ............................................................................................ 15
1-6 The quenching rates of At488 by amino acids ................................................. 17
1-7 Sodium dithionite induced quenching of NBD-PE fluorescence ....................... 19
1-8 (A) FRAP LabVIEW interface (B) Time-lapse FRAP data acquisition ............ 28
1-9 Fluorescence intensity reduction in different excitation conditions .................. 29
1-10 FRAP images ..................................................................................................... 30
1-11 FRAP recovery curves ..................................................................................... 31
1-12 Optical layout of the TIRFM setup ................................................................. 35
1-13 The interface of TIRFM LabVIEW data acquisition program ......................... 38
1-14 TIRFM ALEX .................................................................................................... 40
1-15 Scheme of encapsulation of single molecules within a liposome .................... 41
1-16 Flow chart of the custom-written Matlab program for TIRFM data analysis .... 42
1-17 Single-molecule intensity-time trajectories ..................................................... 46
1-18 smFRET histogram fitting results ................................................................. 49
1-19 Optical layout of the multi-parameter fluorescence confocal microscope ....... 52

2-1 Trytophan fluorescence of the drkN SH3 domain ............................................ 65
2-2 smFRET histograms of drkN SH3 end-labeled with BFL/A647 ......................... 66
2-3 smFRET histograms at various concentrations of GdmCl and urea ................. 68
2-4 smFRET histogram of drkN SH3 in formamide and DMSO ............................ 71
2-5 smFRET denaturation series of drkN SH3 measured with A555/A647 .......... 74
2-6 FCS of WT SH3 in Tris buffer, T22G SH3, WT SH3 in 6M GdmCl ................. 76
2-7 Auto-correlation and cross-correlation functions of SH3 ................................. 78
2-8 $R_{EE}$ of unfolded state in the native drkN SH3 ................................................. 81

3-1 Simplified interaction and phosphorylation scheme of 4E-BP2 ....................... 93
3-2 Half-cone angles of dye rotation on 4E-BP2 in different states................. 101
3-3 The sequence of wild-type 4E-BP2............................................................. 104
3-4 FAD curves of 4E-BP2 at different locations in the apo and bound states..... 105
3-5 $\rho_{\text{seg}}$ of 4E-BP2 in different conditions ........................................... 106
3-6 FCS curves of Atto 488 and 4E-BP2 with eIF4E and $R_H$ values ............... 109
3-7 $\rho_{\text{seg}}$ and half-cone angles of the At488 in the apo and bound state........ 110
3-8 FCS data of 4E-BP2 in NP and HP conditions ............................................. 112
3-9 At488-amino acid quenching rates of 4E-BP2 in NP and HP conditions...... 114
3-10 At488-amino acid quenching rates of NP and HP 4E-BP2 in PBS and 6 M GdmCl ............................................................................................................. 116
3-11 Interaction on rates between 4E-BP2 and eIF4E ........................................... 118
3-12 Interaction off rates between 4E-BP2 and eIF4E ......................................... 119
3-13 smFRET of NP and HP 4E-BP2 at different GdmCl and urea concentrations ............................................................................................................. 121
3-14 smFRET of NP and HP 4E-BP2 at different KCl concentrations................. 125
3-15 Force ratio for two phosphate groups screened by 150 mM and 3 M KCl... 127
3-16 pH dependent smFRET of NP and HP 4E-BP2 ............................................ 129
3-17 Charge and hydrophobicity distributions of 4E-BP2 at different pH .......... 131

4-1 Simulations of FCS using average values and normal distributions.......... 147
4-2 Simulated average number of lipid-bound fluorophores per vesicle .......... 150
4-3 Simulated vesicle-bound fraction of fluorophore at different $K_p$ values.... 151
4-4 Fluorescence intensity decays curves of dyes in the presence of LUV ........ 152
4-5 FCS curves of Atto 488, Atto 594, and Atto 647N with DOPC ................. 153
4-6 FCS of Atto 647N with DOTAP at different NaCl concentrations. ............... 156
4-7 (A) FCS of Atto 565 with DOPS and DOPG; (B) Fractions of fluorophores in the presence of DOPS or DOPG ......................................................... 157
4-8 (A) FCS of Atto 647N with DOPC, POPC and DPPC; (B) Fractions of fluorophores in the presence of DOPC or POPC ........................................... 158
4-9 FCS of Atto488, Atto565 and Atto 647N with different Chol % DOPC ....... 159
4-10 Wide-field fluorescence images of GUVs .................................................. 160
4-11 TIRF images of DOPC SLB with Atto 488, Atto 565, and Atto 647N ....... 161
List of Appendices

A  Lipid structures, logD values, charges and phase transition temperatures  204
B  Fluorophore structures and their charge and logD values  206
C  Free energy of partitioning  212
D  TIRFM user operation and diagnosis manual  213
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>4E-BP2</td>
<td>Eukaryotic initiation factor 4E binding protein 2</td>
</tr>
<tr>
<td>A488</td>
<td>Alexa 488</td>
</tr>
<tr>
<td>A555</td>
<td>Alexa 555</td>
</tr>
<tr>
<td>A647</td>
<td>Alexa 647</td>
</tr>
<tr>
<td>ACF</td>
<td>Auto-correlation function</td>
</tr>
<tr>
<td>ALEX</td>
<td>Alternating laser excitation</td>
</tr>
<tr>
<td>At488</td>
<td>Atto 488</td>
</tr>
<tr>
<td>BFL</td>
<td>Bodipy fluorescein</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>CCF</td>
<td>Cross-correlation function</td>
</tr>
<tr>
<td>Chol</td>
<td>Cholesterol</td>
</tr>
<tr>
<td>CW</td>
<td>Continuous-wave</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>dcFCS</td>
<td>Dual-color FCS</td>
</tr>
<tr>
<td>DHPC</td>
<td>1,2-hexanoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>DMPC</td>
<td>1,2-dimyristoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DOPC</td>
<td>1,2-dioleoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>DOPG</td>
<td>1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)</td>
</tr>
<tr>
<td>DOPS</td>
<td>1,2-dioleoyl-sn-glycero-3-phospho-L-serine</td>
</tr>
<tr>
<td>DOTAP</td>
<td>1,2-dioleoyl-3-trimethylammonium-propane</td>
</tr>
<tr>
<td>DPH</td>
<td>1,6-diphenyl-1,3,4-hexatriene</td>
</tr>
<tr>
<td>DPPC</td>
<td>1,2-dipalmitoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>drkN SH3</td>
<td>N-terminal Src Homology 3 domain of downstream of receptor kinase</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded DNA</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>eIF4E</td>
<td>Eukaryotic initiation factor 4E</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>EMCCD</td>
<td>Electron-multiplied charge-coupled device</td>
</tr>
<tr>
<td>FA</td>
<td>Fluorescence anisotropy</td>
</tr>
<tr>
<td>FAD</td>
<td>Fluorescence anisotropy decay</td>
</tr>
<tr>
<td>FCS</td>
<td>Fluorescence correlation spectroscopy</td>
</tr>
<tr>
<td>FCCS</td>
<td>Fluorescence cross-correlation spectroscopy</td>
</tr>
<tr>
<td>FL</td>
<td>Fluorescein</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>GdmCl</td>
<td>Guanidinium hydrochloride</td>
</tr>
<tr>
<td>GUV</td>
<td>Giant unilamellar vesicle</td>
</tr>
<tr>
<td>HbA</td>
<td>Hemoglobin alpha</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>HP</td>
<td>Hyper-phosphorylated</td>
</tr>
<tr>
<td>HWHM</td>
<td>Half-width at half-maximum</td>
</tr>
<tr>
<td>IDP</td>
<td>Intrinsically disordered protein</td>
</tr>
<tr>
<td>IDR</td>
<td>Intrinsically disordered region</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropylthio-β-galactoside</td>
</tr>
<tr>
<td>IRF</td>
<td>Instrument response function</td>
</tr>
<tr>
<td>ITO</td>
<td>Indium-Tin-Oxide</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>LUV</td>
<td>Large unilamellar vesicle</td>
</tr>
<tr>
<td>MAPK</td>
<td>Mitogen-activated protein kinases</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamic</td>
</tr>
<tr>
<td>Met</td>
<td>Methionine</td>
</tr>
<tr>
<td>MG</td>
<td>Molten globule</td>
</tr>
<tr>
<td>MLV</td>
<td>Multilamellar vesicle</td>
</tr>
<tr>
<td>MPD</td>
<td>Micro Photon Devices</td>
</tr>
<tr>
<td>MQ</td>
<td>MilliQ water</td>
</tr>
<tr>
<td>MSD</td>
<td>Mean square displacement</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>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>NP</td>
<td>Non-phosphorylated</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>PA</td>
<td>Phosphatidic acid</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidylcholine</td>
</tr>
<tr>
<td>PCA</td>
<td>Protocatechuic acid</td>
</tr>
<tr>
<td>PCD</td>
<td>Protocatechuate-3,4-dioxygenase</td>
</tr>
<tr>
<td>PE</td>
<td>Phosphatidylethanolamine</td>
</tr>
<tr>
<td>PET</td>
<td>Photo-induced electron transfer</td>
</tr>
<tr>
<td>PG</td>
<td>Phosphatidylglycerol</td>
</tr>
<tr>
<td>PI</td>
<td>Phosphatidylinositol</td>
</tr>
<tr>
<td>POPC</td>
<td>1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>PRE</td>
<td>Paramagnetic relaxation enhancement</td>
</tr>
<tr>
<td>PS</td>
<td>Phosphatidylserine</td>
</tr>
<tr>
<td>PS</td>
<td>Polystyrene</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>R_{EE}</td>
<td>End-to-end distance</td>
</tr>
<tr>
<td>RG</td>
<td>Radius of Gyration</td>
</tr>
<tr>
<td>RH</td>
<td>Hydrodynamic radius</td>
</tr>
<tr>
<td>RNaseA</td>
<td>Ribonuclease A</td>
</tr>
<tr>
<td>SAW</td>
<td>Self-avoiding walk</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small angle X-ray scattering</td>
</tr>
<tr>
<td>SHG</td>
<td>Second-harmonic generation</td>
</tr>
<tr>
<td>SLB</td>
<td>Supported lipid bilayers</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>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>STAT3</td>
<td>Signal transducer and activator of transcription 3</td>
</tr>
<tr>
<td>SUV</td>
<td>Small unilamellar vesicle</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl)phosphine hydrochloride</td>
</tr>
<tr>
<td>TCSPC</td>
<td>Time-correlated single-photon counting</td>
</tr>
<tr>
<td>TIRFM</td>
<td>Total internal reflection fluorescence microscope</td>
</tr>
<tr>
<td>TMR</td>
<td>Tetramethylrhodamine</td>
</tr>
<tr>
<td>Trolox</td>
<td>6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid</td>
</tr>
<tr>
<td>Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

This thesis first introduces basic biophysical concepts of proteins and lipids that are related to the topics in the following chapters. Then the readers will be informed about the single-molecule fluorescence spectroscopy and microscopy techniques, the instrumentation and data analysis methods that were applied throughout the biophysical studies described in the following chapters. Next, the thesis focuses on investigating structural and dynamic properties of intrinsically unfolded proteins, such as the N-terminal Src Homology 3 domain of the Drosophila adapter protein Drk (drkN SH3) and the eukaryotic initiation factor 4E binding protein 2 (4E-BP2). Finally, we address the basis for dye and lipid selection when studying proteins within lipid vesicles using single-molecule fluorescence tools.

In this chapter, bio-macromolecules, such as proteins and lipids, are highlighted. In the protein section; the four levels of protein structure, the protein folding problem, and an introduction to intrinsically disordered proteins (IDPs) are presented. In the lipids section; lipids, liposome and bicelle models are discussed and their application to drug delivery are introduced. Following that, biophysical techniques especially the single molecule fluorescence spectroscopy modalities that were used throughout the thesis are introduced. Experimental setups that were designed and built for single-photon, single-molecule sensitivity are described, including confocal and total internal reflection fluorescence microscopes (TIRFM). The design of the TIRFM instrument and the data acquisition and data analysis programs that I wrote are discussed in larger detail.
1.1 Molecules of Life

1.1.1 Proteins

1.1.1.1 Structural organization in proteins

There are four levels of protein structure: primary, secondary, tertiary, and quaternary [1]. There are different amino acids that make up the building blocks of proteins, and their linear arrangement (sequence) is the primary structure of a protein. The secondary structure of a protein consists of regions exhibiting structural symmetry as a result of the formation of strings of amino acids with a specific pattern of hydrogen bonds between backbone atoms. There are two main types of secondary structures: the α-helix and the β-strand. The tertiary structure is the complete three-dimensional structure of a polypeptide chain, which typically comprises several secondary structures connected by loops. The tertiary structure is sometimes called a motif or a domain. For example, the N-terminal Src Homology 3 domain of the Drosophila adapter protein Drk (drkN SH3) is made of five anti-parallel β-strands and arranged into two β-sheets packed against each other [2] (See Figure 1-1). The tertiary structure is stabilized through several interactions between the side groups of the amino acids. There are four main interactions that maintain the spatial conformation of the proteins: hydrophobic interactions, hydrogen bonding, electrostatic interactions and disulfide bonding. Incorporation of multiple domains from different polypeptide chains into a unified structure forms a quaternary structure. For example, hemoglobin alpha (HbA) consists of four polypeptide chains, two α-helices, and two β-strands arranged in space in a well-organized manner [3].
1.1.1.2 Protein folding

Our understanding of the numerous biological activities comes from the elucidation of the protein’s three-dimensional structure. In order to connect the protein structures with their biological functions, and link the folding pathways with the misfolding mechanisms, it is critical to comprehend how proteins fold into their native states from unfolded states [4]. The fundamental problem of protein folding is how the protein primary sequence encodes for the three-dimensional structures. It has been investigated by protein science community for several decades, but still remains one of the most challenging problems in biology.

One of the pioneering studies in protein folding occurred in the 1950s. Nobel Laureate Christian B. Anfinsen [5] showed that the ribonuclease A (RNaseA) is an unstructured random polymer and loses it enzymatic activity when the protein is denatured by urea. Subsequently, when urea is removed, RNaseA folds back into its native structure and regains its biological function. The Anfinsen experiment argued that the native structure is determined only by the protein's amino acid sequence and unfolded proteins can spontaneously refold.

Figure 1-2. Schematic representation of the energy landscape of (A) a globular protein and (B) a disordered protein.
From the thermodynamic point of view, the energy landscape and the folding funnel is a general approach to understand the protein folding pathway [6, 7]. Imagining the free energy surface of a protein folding process as a free energy funnel, the depth of the funnel represents the free energy while the width of the funnel reflects the configurational entropy of the system (see Figure 1-2). Each protein has its unique folding funnel and the shape of the funnel is determined by the amino acid sequence. The unfolded states with high free energies and high degrees of configuration entropy are located at the top of the funnel, which allows larger amount of distinct conformational states to be sampled. On the contrary, the folded state has the lowest free energy and samples a smaller number of conformations at the bottom of the funnel. The surface outside the funnel is relatively flat representing the heterogeneity of the random coil states. Partially folded proteins are located at the local minima along the energy landscape. The folding funnel theory helps visualize the fact that a protein can proceed from the unfolded state to the native state via many different folding pathways instead of a fixed route [6, 7].

The wealth of biophysical information for folded proteins has shed light on the physical basis for protein folding and stability. On the other hand, studies of unfolded, denatured or partially folded states are limited owing to the difficulty in capturing short-lived states and to characterize their structure and energetics. The unfolded state of a protein serves as a reference and a starting point for protein folding studies [8]. Thus, understanding the natively unfolded and chemically denatured states is critical for attaining a complete picture of protein folding and stability [8-11]. Random coil models have been treated as the standard model of the denatured state of proteins for a couple of decades. It describes the backbone of the denatured state ensemble as a random sampling of all possible conformations whereby the amino acids lack any specific interaction [12, 13]. However, recent experimental and theoretical studies showed that denatured state ensembles are not classic random coils [10, 14-17] and there is growing evidence for relatively nonspecific hydrophobic clustering and other inter-chain interactions in denatured states – even in the presence of high concentration of guanidinium chloride (GdmCl) or urea [9, 15, 18, 19]. In addition, significant residual structures in the denatured state appear to be at variance with the observation of random coil behavior [14, 20].
Thus, more studies on protein denaturation are needed to lead to more physically realistic models of the unfolded state and to better understand the determining factors of protein folding.

The Gradinaru lab is collaborating with the Forman-Kay lab (Biochemistry, UofT and The Hospital for Sick Children) on several projects focused on characterizing the structure and the dynamics of proteins involved in signal transduction. One of these systems is the N-terminal drk SH3 domain, a modular protein studied in Chapter 2 of this thesis. SH3 domains are found in proteins involved in regulating tyrosine kinase and Src kinase signaling, cytoskeletal reorganization and mediating protein-protein interactions [21, 22]. The drkN SH3 domain comprises 59 amino acid residues and binds the repeating sequence PxxP at C-terminal of the guanine nucleotide exchange factor Sos, which is important in coupling receptor tyrosine kinases to Ras activation [23]. The drkN SH3 domain is marginally stable and exists in equilibrium between folded and unfolded states under non-denaturing conditions [24] whereas upon binding to its Sos-derived target peptide, it becomes fully folded [25]. The interconversion rate between the two states is 2 sec⁻¹, giving rise to distinct sets of solution nuclear magnetic resonance (NMR) chemical shifts [25]. This makes the drkN SH3 protein an excellent model for comparing conformations and dynamics of unfolded states under near-physiological conditions with unfolded states under chemical denaturing conditions. Chapter 2 of this thesis focuses on characterization of folded, unfolded and denatured states of drkN SH3 using single-molecule techniques and new polymer model.

1.1.1.3 Intrinsically disordered proteins

Several decades ago, it was believed that a well-defined three-dimensional structure was a prerequisite for the biological function of a protein and deviation from the native structure would result in a decrease or loss of biological activities. This argument is referred to as the structure-function paradigm. However, a growing number of proteins which lack stable secondary and tertiary structures while still being capable of carrying out biological function were found. They are termed intrinsically disordered proteins (IDPs) [26-28]. Statistics of multiple sequenced genomes shows that up to 50% of amino acids in eukaryotic proteins are
disordered, with about 35% of proteins containing stretches of greater than 30 residues of intrinsic disorder and 15% begin completely disordered [29-32].

The structure and function of IDPs can be attributed to their characteristic amino acid sequence composition. Amino acid content of IDPs has a higher proportion of charged and polar residues and a lower percentage of hydrophobic and bulky aromatic residues. These properties allow IDPs to be prone to more extended conformations owing to electrostatic repulsion and prevent the formation of a hydrophobic interior [33].

The degree of disorder among IDPs ranges from completely extended random coils to highly collapsed molten globules [27]. The shape of the energy landscape of an IDP could be significantly different from that of an ordered protein (see Fig. 1-2). For an ordered protein, as discussed in section 1.1.1.2, the energy landscape could be described as having relatively few shallow wells with one deeper well representing native conformation. In addition, the local energy minima could have substantial deviation from the global energy minima [7]. On the contrary, the energy landscape of an IDP would have numerous but relatively shallow wells. This configuration allows IDPs to populate multiple conformations and fluctuate amongst them [34].

The lack of a persistent folded structure and hydrophobic core observed in IDPs leads to the naive argument that IDPs are indistinguishable from random coils. However, the transient secondary and tertiary structures, electrostatic interactions, and backbone torsion angle propensities turn IDPs into different degrees of compactness [29], thus IDPs exhibit a broad range of compactness with respect to polymer length when compared to chemically denatured proteins [26, 35, 36]. Comparison between the native disordered 4E-BP2 and its chemically denatured states is discussed in Chapter 3.

Determination of structure and dynamics of IDPs is a challenging task because fluctuation among all possible conformations and multiple timescales of their dynamic behavior makes them hard or even impossible to crystalize. There are only a few structural ensembles of IDPs which have been characterized [37-41]. However, the protein community still doubts how accurately these structure ensembles represent the true equilibrium ensemble
In addition, NMR and/or X-ray diffraction analysis only characterize the ensemble averages of these behaviors; thus structures and dynamics of IDPs at a molecular level still remains poorly understood [43-45].

IDPs play a crucial role in mediating interactions with multiple partners and often function as protein interaction hubs [46, 47]. A binding partner can select an existing conformation from the IDP ensemble before the interaction. This is called conformational selection. An interaction partner can also induce a novel conformation on binding. This is termed induced fit. Both mechanisms play a critical role in IDP complexes formation [48]. The resulting protein complexes can undergo disorder to order transitions or remain dynamic [29]. In addition, it was shown that disorder-to-order transitions allow proteins to bind with high specificity coupled with relatively low affinity due to the entropic cost of folding upon binding. This provides a thermodynamic switch that is appropriate for signal transduction, where a protein must not only bind to its target specifically but also be able to dissociate when the signaling process is complete [49].

Phosphorylation, in which an amino acid residue is appended with an additional covalently bound phosphate group by a protein kinase, is one of the most common post-translational modifications (PTM). Phosphorylation has been shown to induce significant structural reorganization of the IDPs and may induce the protein to undergo disorder to order transitions and vice versa [29, 50, 51]. The change in the structure preference also regulates protein flexibility and protein-protein interactions [51-53].

IDPs are involved in a number of neurodegenerative diseases including Alzheimer’s, Parkinson’s disease, Huntington's disease and Prion diseases [54-56]. Approximately 80% of human cancer-associated proteins are predicted to have intrinsically disordered regions [57]. Studies have shown that oligomerization of α-synuclein, Tau, and Aβ42, not only form fibrils, but also turns to be toxic. Discovering small molecules to bind to monomers to inhibit their aggregation is shown to be a viable treatment strategy [58-60]. Besides neurodegenerative diseases and cancer, IDPs have been shown to be highly represented in proteins associated with cardiovascular diseases [61], diabetes [54], HIV [62], and cystic fibrosis [63].
In the nervous system, translational control is obligatory for neurodegeneration [64], metabolic disorders, cellular transformation and cancer [65]. Cap-dependent translation initiation is regulated by the interaction of eukaryotic initiation factor 4E (eIF4E) with disordered eIF4E binding proteins (4E-BPs) in a phosphorylation dependent manner. In particular, the neural IDP 4E-BP2 functions in regulating synaptic plasticity, essential for learning and memory [66]. Recent studies show that 4E-BP2 has significant number of transient structure. Binding of IDPs often leads to ordering, yet the eIF4E:4E-BP2 complex has been shown to be highly dynamic with an exchanging bipartite interface [67]. However, detailed structural pictures of free 4E-BP2 and its dynamic complex are not available. Phosphorylation of 4E-BP2 reduces eIF4E binding affinity by inducing a folded structure for residues 18-62 with long-range stabilizing interactions to residues outside the folded region not well-characterized [53]. The conformational distributions and dynamics of 4E-BP2, with and without eIF4E, under different phosphorylation conditions are discussed in Chapter 3. This work will help design small molecules that stabilize or destabilize the folded conformation for therapeutics to treat autism spectrum diseases.

1.1.2 Lipids

1.1.2.1 Lipids and the lipid bilayer

Lipids are amphiphilic small molecules that consist of a hydrophilic head group and hydrophobic tail. They are soluble in organic solvents such as chloroform, methanol or ethanol. Structural properties of lipids are characterized by the type of polar head group, backbone structure, as well as the length and degree of saturation of the hydrocarbon chains. In aqueous solution, the amphiphilic lipids self-assemble into a variety of organized structures, including spherical or irregular shaped liposomes, micelles, bilayers, and inverted structures [1, 68]. The final assembled structure depends on several factors. A high head group diameter to tail length ratio leads to a “conical” lipid shape, which is more amenable to induced curvature. When the ratio approaches one, the lipids become “cylindrical” and favor flat structures [68]. Two competing forces drive the self-assembly. Hydrophilic head groups are in favor of exposing to
aqueous environment while hydrophobic interaction between the non-polar acyl chains drives the tails to form a hydrophobic core. The steric and electrostatic interactions also make hydrophilic head groups repel each other, thus preventing membranes from spontaneous fusion or aggregation [68].

Phospholipids are a major class of lipids. A phospholipid with either one or two hydrophobic chains connects a glycerol backbone and a polar head group that contains a phosphate group. Common phosphate head groups include phosphatidylethanolamine (PE), phosphatidylinositol (PI), phosphatidylserine (PS), phosphatidylglycerol (PG), and phosphatidic acid (PA). The 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) structure is shown in Figure 1-3 (A) and some phospholipid structures used in the following chapters are illustrated in Appendix A, with charge, transition temperature and calculated hydrophobicity values.

After Gorter and Grendel first identified the lipid bilayer in 1925 [69], Langmuir and Blodgett pioneered demonstrations of the lipid bilayer self-assembly on a synthetic substrate [70, 71]. Lipid bilayers are only a few nanometers in width but impermeable to most water-soluble molecules and ions. The lipid bilayer have been produced using various techniques under various conditions. One of the simplest and straightforward methods to form lipid bilayer is by liposome fusion, which was first discovered by Brian and McConnell in 1984 [72]. Since then, supported bilayers have become an important biophysical platforms to study the physical properties of lipids, such as diffusion and mobility [73, 74], phase separation [75] and interactions between the lipid bilayer and membrane proteins [76, 77].

All lipids have a characteristic temperature at which they undergo a phase transition from the gel phase to the liquid crystalline phase. Lipid bilayers in the liquid crystalline phase have relatively high in-plane mobility and exhibit a random-walk-like behavior. The time scale for one lipid molecule to exchange positions with its neighbors is on the order of nanoseconds. In contrast, lipid molecules in a gel phase diffuse relatively slow [68]. In general, the phase transition temperature increases with increasing length and saturation of the hydrophobic chains.
Cholesterol is a lipid molecule that is essential to cell membranes. This molecule plays a crucial role for maintaining the structural integrity of the membrane and largely determines its mechanical properties [78, 79]. The structure of cholesterol is quite different from that of phospholipids (see Figure 1-3). The hydrophilic domain consists of a single hydroxyl group, several fused aromatic rings and a short single chain tail. The presence of cholesterol increases membrane packing and rigidity while remaining membrane fluidity and its diffusion coefficient via the interaction with phospholipid fatty-acid chains when in the liquid crystalline phase [68]. The mechanism responsible involves intercalation between lipid molecules followed by an increase of the ordering of the nearby lipids [80]. On the contrary, introducing cholesterol to lipid bilayers in the gel phase disrupts local packing order and increases the diffusion coefficient [81].

Figure 1-3. Structures of (A) DOPC lipid and (B) Cholesterol. Figures are adapted from Avanti Polar Lipids, Inc.

Cell membranes are complex systems that consist of lipids membranes and membrane-spanning proteins. Using model lipid systems to mimic cell membranes is an important biophysical approach towards advancing our understanding of the cell morphology and its signaling functionality.

1.1.2.2 Liposomes and bicelles

Liposomes were discovered by Alec D. Bangham more than half a century ago [82]. Since then, they have been extensively studied and used by chemists and biologists, in part due to their promise for medicinal drug coating and delivery. A liposome is a spherical vesicle having at least one lipid bilayer containing mostly phospholipids and cholesterol. Because of the hydrophobic effect, amphiphilic lipids spontaneously bend into a spherical shape forming
a continuous bilayer and self-assemble to vesicular structure when exposed to an aqueous solutions. Liposomes can be produced in several ways and can be formed in a wide range of size from tens of micrometers to tens of nanometers. The three common methods to produce liposomes are extrusion, sonication, and electroformation. The four major types of liposomes are the multilamellar vesicle (MLV), the giant unilamellar vesicle (GUV) (diameter size >1 µm), the large unilamellar vesicle (LUV) (diameter size: 100 nm ~ 1 µm) and the small unilamellar vesicle (SUV) (diameter size: 20 nm ~ 50 nm) [68]. Multilamellar liposomes have more than one lamellar phase lipid bilayers and unilamellar vesicles have single lamellar bilayer. Extrusion is a common tool to produce LUVs and SUVs [83]. Sonication is mostly used to make SUVs [84] and electroformation is the preferred technique to form GUVs [85, 86].

Bicelles, or bilayered micelles, are disk-like structures with long chain amphiphiles, usually 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), assembling into a planar lipid bilayer that is stabilized at its edges by short chain amphiphiles, usually 1,2-hexanoyl-sn-glycero-3-phosphocholine (DHPC) [83]. The planar bilayer region of a bicelle provides an excellent approximation of the natural membrane environment, so that bicelles have become increasingly popular in membrane protein structural studies using a variety of techniques. Bicelles are also used to incorporate membrane protein [87] and are served as pharmaceutical delivery vehicles [88]. Bicelles have two advantages for lipid bilayer coatings compared to liposomes. First, bicelle fabrication is simple and carried out under mild conditions which is conducive to retention of native membrane proteins structure and function. Second, bicelles in the form of small planar discs can adapt readily to rough or irregular surfaces [83]. The schemes of a liposome and a bicelle are shown in Figure 1-4.
Interaction of lipid membranes with proteins, peptides, oligonucleotides and nanoparticles plays a critical role in cellular signaling and in drug delivery. Fluorescence techniques are one of the most commonly used methods to study these interactions. One point of concern is that researchers need to introduce organic fluorophores to conjugate lipids, proteins, peptides, DNA, RNA or nanoparticles. Possible artifacts caused by the presence of fluorescent labels have not been studied in detail. Chapter 4 is an extensive quantitative discussion on interactions between commonly used fluorophores and lipids. This study provides valuable feedback on how to wisely pair lipids and dyes for fluorescence-based studies.

1.1.2.3 Drug delivery applications

As discussed above, the structure, composition and physical properties of the lipid bilayer are similar to those of the biological cell membrane. In addition to their non-toxic nature and high drug encapsulation efficiency, lipids and lipid bilayers have attracted enormous attention as ideal models for drug delivery systems and for signal amplification in immune systems. By varying the composition, charge and phase transition temperature of the lipid component, liposome-based delivery systems can carry encapsulated drugs in a spatially and temporally controlled manner. These systems have been shown to modulate release kinetics, circulation times, and specificity using electric and magnetic fields, light and ultrasound [89].
The drugs can be quite versatile including both hydrophobic and hydrophilic molecules. Polar molecules are encapsulated in liposome’s aqueous interior, while hydrophobic drugs are anchored in the lipid bilayer milieu. One example studied in Gradinaru lab is the signal transducer and activator of transcription 3 (STAT3) protein, which was anchored to the membrane of liposomes and cells as an inhibitory approach to cancer [90]. A number of liposome-based drug delivery systems have been commercially available and new formulations that are undergoing clinical trials [91, 92]. In 2013, the Nobel Prize in physiology and medicine was awarded to R. Schekman, J. Rothman and T. Sudhof for their discoveries of machinery regulating lipid vesicle traffic in cells.

A hydrogel is a network of hydrophilic polymer chains which can trap water in its interior. Hydrogel systems show promise in a number of areas, such as devices, drug delivery and regenerative medicine [93]. Hydrogels are already widely used as three-dimensional cell and tissue culture environments due to their tissue compatibility [94].

The performance of drug delivery systems is highly dependent on the physical and chemical characteristics of their coatings. Surface properties, such as charge, hydrophobicity, permeability, diffusion rate, etc., are critical. For example, a surface coating that regulates permeability can be applied to the hydrogel matrix to manipulate a variety of controls over the rate of drug release [95]. The study of liposome-coated hydrogel spheres used for drug delivery vehicles has been published as a 2nd author [96]. However, the content is not shown in the thesis.

1.2 Single Molecule Fluorescence
1.2.1 Basics of fluorescence
1.2.1.1 Organic fluorophores and Jablonski diagram

Fluorophores are chemical compounds that emit visible light upon excitation in the UV-VIS range. Fluorophores are typically organic dyes which contain several aromatic groups
and π bonds, the details of which determine their photophysical properties, such as absorption and emission wavelength range, molar extinction coefficient, fluorescence quantum yield, lifetime, and photostability, etc. The molar extinction coefficient reflects how strongly a fluorophore attenuates light at a certain wavelength. The fluorescence quantum yield is defined as the number of emitted photons relative to the number of absorbed photons. Higher quantum yield means that the fluorophore displays brighter emission. The lifetime is a measure of the average time that the fluorophore spends in the lowest singlet excited state. The quantum yield and the lifetime are related quantities and both depend on the local environment of the fluorophore. The emission spectrum and the fluorescence quantum yield are typically independent of the excitation wavelength.

With the rapid development of fluorescence imaging applications in life sciences, high performance fluorophores have been synthesized in the past two decades, with a wide range of colors (and other properties) available. Typically, absorption and emission wavelength in the visible range are preferred, although some UV and IR dyes are also available. For multicolor microscopy and spectroscopy applications in biology, narrow spectra of absorption and emission are preferred to eliminate or at least reduce cross-talk effects.

Single molecule approaches are now established in biophysics when there is a need to study heterogeneous systems or dynamics [97, 98]. For single molecule fluorescence experiments, selecting fluorescent labels with small molecular weight, high photostability, molar extinction coefficient and quantum yield is critical for ensuring non-perturbative conditions and a good signal-to-noise ratio for detection. Fluorophores with long fluorescence lifetimes and/or high fundamental anisotropies are used to study the interaction and the conformational dynamics of proteins by monitoring lifetime and anisotropy changes. The anisotropy is defined as the ratio of the intensity difference between parallel and perpendicular polarization to the total intensity [99]. Environment sensitive fluorophores are typically selected to probe the local conformational dynamics of the labelled biomolecule [100] and report on biomolecular reaction [101].
Structures of common fluorescent dyes used in single-molecule studies are shown in Appendix B with calculated charge and hydrophobicity at physiological condition. Studies on the interactions between these commonly used fluorophores with various lipids are discussed in Chapter 4.

A Jablonski diagram is a schematic representation of the possible consequences of a particular molecule responding to the upcoming photons. Figure 1-5 shows a typical Jablonski diagram.

![Jablonski Diagram](image)

**Figure 1-5.** Jablonski diagram. Black lines are the discrete energy levels. $S_0$, $S_1$, and $S_2$ are singlet electronic states and $T_1$ is the triplet state. $V_1$, $V_2$, $V_3$, ... $V_n$ are vibrational states. Blue: photon absorption; Orange: internal conversion; Red: photon emission; Green: intersystem crossing; Cyan: vibrational relaxation; Purple: phosphorescence.

The energy representation of a molecule system can be simplified as the combination of a ground state, some excited singlet and triplet electronic states coupled with a number of vibrational states at each of these electronic states. A fluorophore is first excited to some higher vibrational states of the excited states $S_1$ or $S_2$, which takes place in $\sim 10^{-15}$ seconds and the molecules rapidly relax to the lowest vibrational state of $S_1$. This process is called internal
conversion, which occur in less than $10^{-12}$ seconds. Following that, the molecules can either complete radiative decay to the vibrational states of the ground state during the lifetime of the fluorophore or undergo intersystem crossing to the vibrational states of the triplet state $T_1$. Here, a number of interactions, such as quenching and energy transfer, are ignored for the sake of simplicity. The lifetime of the fluorophore is on the order of nanoseconds and the intersystem crossing happens in the range of $10^{-8}$-$10^{-3}$ seconds. The molecule then undergoes internal conversion again to the lowest vibrational states of either $S_0$ or $T_1$ depending on the previous decay pathway. Intersystem crossing from triplet state to singlet ground state is called phosphorescence, the time scale of which ranges from milliseconds to seconds [99].

### 1.2.1.2 Fluorescence quenching

Quenching is a phenomenon in which a fluorophore in the excited state is de-excited upon interacting with quenchers in the solution. Quenching reduces the fluorescence intensity of the fluorophore. There are three types of quenching: static quenching, dynamical (collisional) quenching, and energy transfer [99]. Energy transfer is discussed in the next section. In the case of the static quenching, the fluorophore and the quencher bind to each other and form dark (non-fluorescent) complexes. This process happens in the ground state of both molecules and does not rely on collision. Conversely, in the case of dynamical quenching, the quencher collides with the fluorophore in the excited state and it causes the energy to be transferred to excite the quencher molecule from ground state to the excited state.

The Stern-Volmer equation [99] describes the fluorescence quenching process: $I_0/I = 1 + K_{SV}[Q]$, where $I_0$ and $I$ are the fluorescence intensities in the absence and in the presence of quencher, respectively, $[Q]$ is the quencher concentration and $K_{SV}$ is the Stern-Volmer quenching constant. For pure collisional quenching, we have $\tau_0/\tau = 1 + k_q \tau_0[Q]$, where $k_q$ is the bimolecular quenching rate constant and $\tau$ and $\tau_0$ are the fluorophore lifetimes in the presence and in the absence of quencher, respectively. If both static and collisional quenching occur, $I_0/I = (1 + k_q \tau_0[Q]) (1 + K_a[Q])$, where $K_a$ is the association constant of the complex.

Figure 1-6 illustrates the quenching curves of the Atto 488 dye by four different amino acids,
tryptophan (Trp), tyrosine (Tyr), methionine (Met), and histidine (His). Steady state fluorescence experiments were performed using a PTI spectrofluorimeter (Photon Technology International, Canada). The emission spectra of the Atto 488 were taken in the presence and absence of each of the four amino acids. The lifetime experiments were performed under custom-built multi-parameter confocal setup (See Chapter 1, Section 2.2.2). Fluorescence intensities and lifetimes relative to the ones in the absence of the amino acid were calculated and plotted in Figure 1-6. All experiments were performed on L-amino acids and dissolved in phosphate-buffered saline (PBS) buffer at pH 7.4.

**Figure 1-6.** The quenching of Atto 488 by 4 amino acids: Trp (red), Tyr (green), Met (blue), and His (black). Ratio of the fluorescence intensities (A) and of the fluorescence lifetimes (B) recorded in the absence vs. the presence of amino acids, respectively. The magenta dash line in (A) represents the best fit to the static quenching model for the Trp data. The insets are the Tyr data at low concentrations up to 1.8 mM due to its low solubility.

The data indicates that the quenching of Atto 488 by Tyr, Met and His is entirely collisional. However, the intensity data shown in Figure 1-6 (A) for the Trp quenching is quadratic, not linear, which suggests that quenching of Atto 488 by tryptophan involves not only collisional quenching but also the formation of dark complexes between the two molecules. From the fit, the Stern-Volmer quenching constants, $K_{SV}$, and the collisional quenching rates, $k_q$, were obtained: $K_{SV,Tyr} = 9.3 \pm 0.3 \text{ M}^{-1}$, $K_{SV,Met} = 6.6 \pm 0.1 \text{ M}^{-1}$, $K_{SV,His} = 1.7 \pm 0.1 \text{ M}^{-1}$. $k_{q,Tyr} = 3.25 \pm 0.03 \times 10^9 \text{ s}^{-1}\text{M}^{-1}$, $k_{q,Tyr} = 2.38 \pm 0.15 \times 10^9 \text{ s}^{-1}\text{M}^{-1}$, $k_{q,Met} = 1.67 \pm 0.03 \times 10^9 \text{ s}^{-1}\text{M}^{-1}$, $k_{q,His} = 0.25 \pm 0.05 \times 10^9 \text{ s}^{-1}\text{M}^{-1}$. Also, the association constant of the Trp-Atto 488 complex, $K_{a,Tyr} = 7.71 \pm 0.15 \text{ M}^{-1}$ was determined. This is the first time, to our knowledge, that quenching of Atto 488 by natural amino acids were measured. Compared to the results of
quenching of Alexa 488 studied by Webb group [102], it is shown that Trp, Tyr and Met have similar collisional quenching effect to both dyes, while Alexa 488 was shown to be quenched more significantly by His than Atto 488 whereas His has limited effect on it.

Quenching of fluorophores by aromatic amino acids allows researchers to study the conformational dynamics of biomacromolecules, such as proteins, peptides and oligonucleotides [103]. Fluorescence quenching via photo-induced electron transfer (PET) requires van der Waals contact and can be used as reporter for structural changes on the spatial scale of angstroms to nanometers and on the temporal scale of nanoseconds to milliseconds [104, 105].

Fluorescence quenching method can also be applied to examine the integrity of the lipid bilayer. In the study of lipid bilayer deposition on polystyrene (PS) bead surfaces using bicelles and liposomes, fluorescence quenching measurements were performed to prove that the coating was a single lipid bilayer. Detailed description can be found in the publication [83]. Briefly, 0.10 mol% NBD-PE lipid was incorporated in the DMPC/DMTAP/CHOLOA lipid bilayer and sodium dithionite was used to quench the NBD fluorescence. Ideally, dithionite ions added externally will not permeate to the interior compartment of an intact unilamellar spherical lipid bilayer, and the NBD-PE fluorescence would be reduced to 50% of the initial value, which was found in the liposome coated PS bead, as shown in Figure 1-7. When PS bead lipid coatings formed from bicelles were interrogated in this fashion, NBD-PE fluorescence was reduced to roughly 30%, indicating significant permeation of dithionite to the interior side of the lipid bilayer coating.
Figure 1-7. Sodium dithionite induced quenching of NBD-PE fluorescence from PS bead lipid coatings formed using bicelles (black curve) and liposomes (gray curve). Dithionite was added at time point “D” and the detergent Triton X100 was added at time point “T”. Reprinted with permission from reference [83]. Copyright (2015) American Chemical Society.

1.2.1.3 Föster resonance energy transfer

Förster resonance energy transfer (FRET) describes the transfer of excitation energy between two molecules. A donor molecule in its lowest singlet excited state may transfer energy to an acceptor molecule through non-radiative dipole–dipole coupling. The dipole–dipole interaction potential is [106]

\[ V_{DA} = \frac{1}{4\pi\varepsilon_0} \left[ \frac{r^2 \mu_D^* \cdot \mu_A^* - 3(\mu_D^* \cdot \vec{r})(\mu_A^* \cdot \vec{r})}{\vec{r}^5} \right], \]  

(1-1)

where \( \vec{r} = \vec{r}_D - \vec{r}_A \) is the distance between two fluorophores and \( \vec{r}_D, \vec{r}_A \) are the center of mass coordinates of donor and acceptor. \( \mu_D^* \) and \( \mu_A^* \) are the transient dipole moments of donor and acceptor, respectively. The FRET efficiency is proportional to ratio of the donor’s energy absorbed by the acceptor to the energy emitted by the donor in the absence of the acceptor. The rate of acceptor absorbance is proportional to the square of the donor electric field surrounding the acceptor [107]. The donor electric field is inversely proportional to the \( r^3 \) based on eq. (1-1). Therefore, FRET is very sensitive to small changes in the distance, \( r \), because the energy
transfer efficiency, \( E \), is inversely proportional to the sixth power of the distance between donor and acceptor fluorophores, as described in eq. (1-2) [108]

\[
E = \frac{1}{1 + (r/R_0)^6}.
\]

(1-2)

\( R_0 \) is the Förster radius of the donor-acceptor pair and, as eq. (1-2) shows, it represents the distance at which the FRET efficiency is 50%. The Förster radius depends on the relative orientation of the donor emission dipole moment and the acceptor absorption dipole moment, \( \kappa^2 \), the donor quantum yield in the absence of acceptor, \( \text{Q}_D \), the refractive index of the sample, \( n \), and the spectral overlap of the donor emission spectrum and the acceptor absorption spectrum \( J \):

\[
R_0^6 = \frac{9(\ln 10)}{128\pi^5 N_A} \frac{\kappa^2 Q_D}{n^4} J,
\]

(1-3)

where \( J = \frac{\int f_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int f_D(\lambda) d\lambda} \), \( \kappa = \mu_A \mu_D - 3(\hat{\mu}_D \cdot \hat{\mu}_A) \), \( N_A \) is Avogadro’s number, \( f_D(\lambda) \) is the donor emission spectrum, \( \varepsilon_A \) is the acceptor molar extinction coefficient, \( \mu_D \) and \( \mu_A \) are the normalized transition dipole moments of the donor and acceptor fluorophores and \( \hat{r} \) denotes the normalized inter-fluorophore displacement.

Single-molecule FRET (smFRET) is a powerful tool to measure protein conformations, especially proteins with disordered regions or those entirely disordered (IDPs). Since such proteins lack stable secondary and tertiary structures, they are not amenable to traditional structural biophysics methods, such as NMR and X-ray crystallography. smFRET is able to measure the heterogeneity of the conformations within the conformational ensemble and thus has the advantage of exploring rare sub-populations that may be hidden in ensemble measurements. It is also a powerful tool in the studies of conformational dynamics [109-111] and binding associations [112, 113]. Importantly, smFRET is able to distinguish the contribution from doubly labeled species from singly labeled or photo-physically damaged samples by employing alternating laser excitation (ALEX) [114, 115]. Some biomolecules are prone to aggregation at the concentrations used for ensemble measurements (\( \mu \text{M} – \text{mM} \)), while
single-molecule methods operate at low concentrations (pM – nM) and allow the properties of monomeric species to be probed.

The FRET efficiency is measured experimentally using either ratiometric intensity or ratiometric lifetime methods. Intensity measurements do not require pulsed femto-pico-second lasers, and therefore it is easier and more popular. The measurements can be realized by illumination with a continuous-wave (CW) laser at a donor excitation wavelength, followed by collecting spectrally-separated donor and acceptor emission signals. FRET efficiency is then calculated as a ratio of the corrected emission signals, as

\[ E = \frac{I_a}{\gamma I_d + I_a}, \]  

(1-4)

where \( I_d \) and \( I_a \) are the donor and acceptor emission signals, respectively, and \( \gamma \) is a correction factor, which accounts for differences in the quantum yield ratios of donor and acceptor and in the detection efficiencies in the two color channels.

There are typically two sorts of smFRET measurements: the molecules are immobilized or they are freely diffusing. The immobilized smFRET experiments were pioneered by R. M. Hochstrasser and co-workers to investigate protein folding [116]. They immobilized the GCN-4 peptide on a glass surface and observed anti-correlated changes in donor and acceptor signals due to conformational changes. The inevitable non-specific interaction between proteins and surface may cause flaws in the data interpretation. To reduce these artefacts, Haran group [117] encapsulated the adenylate kinase protein inside liposomes and immobilized those onto the surface. This encapsulation protocol has been optimized by Ha group [118-120] and Gradinaru group [121] and has been widely used to study protein folding at single-molecule level. The advantage of immobilized method is that it allows one to study longer dynamics of individual molecules because of the longer observation time.

Another simpler approach is to measure freely diffusing molecules in diluted solutions using a confocal microscope. When a labeled molecule passes through the small, femtoliter focal volume, conjugated fluorophores will emit a burst of photons and the donor and acceptor photon numbers are used to infer the conformation of the molecule. The freely diffusing
method is limited by the time scale of translational diffusion through the detection volume, which is typically between 100 µs and 10 ms. However, this scheme offers easy sample preparation and provides information about sub-millisecond fluctuations for a statistically large number of single molecules. The other merit for this approach is that it is easy to change solvent conditions in the solution rather than inside the liposome. It should be noted that common lipid vesicles cannot support harsh solvent conditions [122], such as denaturants and organic solvents, and it is difficult to do sample or buffer mixing due to ion impermeability and osmotic pressure of the lipid bilayer.

smFRET can not only probe the protein structural properties, but also investigate the binding, assembly of molecular complexes and their resulting function. One of the challenging biological questions that has been addressed in single molecule studies is how protein conformational changes couple with binding and interaction. For example, Deniz and co-workers used smFRET to study how binding of α-synuclein to amphiphilic small molecules modulates conformational transitions between a natively unfolded state and multiple α-helical structures [123].

1.2.1.4 Fluorescence correlation spectroscopy

Fluorescence correlation spectroscopy (FCS) is a technique that is based on the fluorescence intensity fluctuations with time occurring in a femtoliter volume. This small observation volume can be realized via confocal optics or multi-photon excitation [99]. The fluctuation of the fluorescence intensity is due to molecular Brownian motion. The correlation function provides the average number of fluorescent molecules and diffusion time when the molecules pass through the focal volume.

On the nanosecond time scale, photon antibunching can be observed. Photon antibunching is the average time after a photon emission event in which a molecule needs to wait to be re-excited again to emit the next photon. If the emission is split into horizontal and vertical polarizations, the correlation function will reflect information about the average
molecular rotational diffusion time, which is typically on the nanosecond time scale. Photophysics of the fluorophore and conformational fluctuations range from (sub-)microsecond to (sub-)millisecond time scale. On the (sub-)millisecond time scale, the correlation function represents the translational diffusion time of the molecule and the hydrodynamic radius, $R_H$, can be calculated based on the Stokes-Einstein equation,

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

(1-5)

where the diffusion coefficient $D$ is related to diffusion time $\tau_D$ by $\omega_0^2 = 4D\tau_D$, and $\omega_0$ is lateral radius of the detection volume.

The fluorescence intensity in a typical FCS experiment is split into two or more channels to construct pseudo-autocorrelation and cross-correlation curves. The use of more than one photon detector for autocorrelation curves is required in order to eliminate after-pulsing effects [124].

The normalized fluorescence intensity correlation function is defined as:

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

(1-6)

where $i$ and $j$ are detection channels, $\delta F(t) = F(t) - \langle F(t) \rangle$, and brackets denote time averaging. With some simplified assumptions, an analytical formula for the fluorescence intensity correlation function was derived for a mixture of molecular species [125] :

$$G(t) = \frac{1}{\sum_i (Q_iN_i)^2} \sum_i Q_i^2 N_i \left(1 + \frac{t}{\tau_{D,i}}\right)^{-1} \frac{1}{\frac{t}{s^2\tau_{D,i}} + \frac{1}{2}} \left(1 + K_i e^{-t/\tau_{k,i}}\right).$$

(1-7)

In eq. (1-7), $Q_i, N_i, \tau_{D,i}, K_i, \tau_{k,i}$ are the brightness, the average number of molecules in the detection volume, the diffusion time, the triplet fraction, and the triplet lifetime of $i^{th}$ species, respectively; $s$ is the aspect ratio of the axial and the lateral radii of the detection volume, $s = z_0/\omega_0$. In the case of two species of equal brightness and each exhibiting anomalous diffusion, eq. (1-7) becomes:
\[ G(t) = \frac{1}{N_{\text{tot}}} \left[ f_1 \left( 1 + \left( \frac{t}{\tau_{D1}} \right)^{\alpha_1} \right)^{-1} \left( 1 + \frac{1}{s} \left( \frac{t}{\tau_{D1}} \right)^{\alpha_1} \right)^{-\frac{1}{2}} \left( 1 + K_1 e^{-t/\tau_{k1}} \right) + (1 - f_1) \left( 1 + \left( \frac{t}{\tau_{D2}} \right)^{\alpha_2} \right)^{-1} \left( 1 + \frac{1}{s} \left( \frac{t}{\tau_{D2}} \right)^{\alpha_2} \right)^{-\frac{1}{2}} \left( 1 + K_2 e^{-t/\tau_{k2}} \right) \right] , \quad (1-8) \]

where \( N_{\text{tot}} = C N_A V_{\text{eff}} \) is the total number of molecules in the effective detection volume \( V_{\text{eff}} = \pi^{3/2} z_0 w_0^2 \). \( \alpha_{1,2} \) are the anomalous factors for each diffusion species; \( \alpha > 1 \) corresponds to super-diffusion, where mean square displacement (MSD) is faster than linearly growing with time and for \( \alpha < 1 \), the particle undergoes sub-diffusion, where MSD grows slower than linearly with time [126, 127].

FCS is versatile biophysical tool with applications ranging from triplet-state dynamics [128] to polymer physics [129, 130] and to living cell studies [131].

Fluorescence cross-correlation spectroscopy (FCCS) is a method to detect interactions between two different species of molecules that have fluorophores with separated emission spectra [132]. The dual-color fluorescence intensities emitted from the overlapping observation volume are separated by a dichroic mirror and fed to different detectors. If the two species diffuse independently, their fluctuations are uncorrelated, and thus the amplitude of the cross-correlation function is zero. However, if these two species interact with each other, they diffuse into and out of the detection volume together, and thus provides non-zero amplitude in the cross-correlation function. The amplitude of the cross-correlation function informs on the fraction of co-diffusing molecules [132].

1.2.1.5 Fluorescence anisotropy decay

Fluorescence anisotropy decay (FAD) measurements are based on the principle of photoselective excitation of fluorophores by polarized light. The polarized light preferentially excites those fluorophores at the ground state having their dipole transition moments aligned with the excitation field. After the excitation, molecules diffuse and rotate causing emissive polarization changes. FAD monitors how the emission polarization changes during the excited-
state lifetime. Anisotropy will decay to zero with time and the rate of decay is a direct measure of the rate of rotational motion. FAD can be used to study the rotational diffusion of molecules, and to infer information about their shape, size and local rigidity. Fluorescence anisotropy is given by [99]:

\[
\text{r}(t) = \frac{I_{\parallel}(t) - GI_{\perp}(t)}{I_{\parallel}(t) + 2GI_{\perp}(t)}.
\]  

(1-9)

Here the G factor, G, is a parameter which defines the detection efficiency ratio of the parallel channel to that of the perpendicular channel. If the fluorophore tumbles fast enough, the polarization orientation has already randomized before the emission occurs. In this case, \( I_{\parallel}(t) = GI_{\perp}(t) \) and the anisotropy value equals to zero.

Kinosita, et al., [133] decoupled the local dye motion from the protein dynamics and proposed a “wobble-in-a-cone” model to analyze FAD curves. The model function is:

\[
\text{r}(t) = r_0[(1 - A)e^{-t/\rho_1} + A]e^{-t/\rho_2}.
\]  

(1-10)

Fundamental anisotropy \( r_0 = 0.4p_2(\cos(\lambda)) \) refers to the anisotropy observed in the absence of other depolarizing processes, such as rotational diffusion or energy transfer [99]. \( p_2(x) = (3x^2 - 1)/2 \) is the second-order Legendre polynomial, \( \lambda \) is the angle between the absorption and emission dipole moment fixed in the coordinate frame of the dye. \( A = \frac{1}{\xi^2}(1 + \cos\theta)\cos\theta \) is a parameter describing the degree of motional restriction and is related to the semicone angle \( \theta \). \( \rho_1, \rho_2 \) are the rotational correlation times of the dye and the protein, respectively.

The FAD measurements not only probe the overall size and shape of the labelled protein but also provide information about local chain dynamics in the vicinity of interaction sites. This suggests that we can map the binding affinities within a relatively large protein complex using segmental motion analysis [134]. The FAD technique was applied to study the structural flexibility of disordered 4E-BP2 and the binding strength of eIF4E/4E-BP2 at segmental level in Chapter 3.
1.2.1.6 Fluorescence photobleaching

Photobleaching is a terminology to describe the situation when fluorophores are converted to chemically non-fluorescent compounds mostly due to high light intensity exposure. The photobleaching also occurs when the fluorophores react with oxygen and become non-fluorescent. This reaction typically occurs after the fluorophores transit from singlet excited state to the triplet state, as shown in Figure 1-5.

In general, photobleaching is avoided in order to extend the observation time of biological events. This is especially important in single-molecule studies, in which relatively high excitation intensities are used. Several oxygen scavenging solutions based on chemical and enzymatic reactions have been proposed to increase the photobleaching time of different fluorophores for single molecule studies [135, 136]. For example, a combination of glucose oxidase, catalase and glucose can reduce the concentration of oxygen in an aqueous solution from 270 µM to 14 µM and thus increase the photobleaching time by a few orders of magnitude for some fluorophores [135, 136]. The downside of this protocol is that the pH value of the solution drops over time (from pH 8.0 to below pH 6.0 in 10 mM Tris buffer within 30 min) [137], which poses difficulties for studying pH-dependent biological phenomena. Another enzymatic system consisting of protocatechuic acid (PCA) and protocatechuate-3,4-dioxygenase (PCD) can also significantly decrease the oxygen concentration in solution from 270 µM to ~ 3 µM, but does not induce a pH drop over time [136]. The side effect of these oxygen scavengers is that they introduce blinking due to the increase of the triplet state lifetime. In single molecule experiments, triplet quenchers, such as 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox), are added to these enzymatic oxygen scavenger solutions to reduce photobleaching and blinking [138].

Photobleaching has drawbacks but it can also be employed as a quantitative tool to study biological systems. Single molecule photobleaching has been used to study the size distribution of small molecular conjugates, with applications such as biosensor design [139].
and protein oligomerization [140, 141] studies. Another wide-spread application of photobleaching kinetics used to study diffusion of biological molecules is described below.

1.2.1.7 Fluorescence recovery after photobleaching

Fluorescence recovery after photobleaching (FRAP) is a method to quantify the mobility of the fluorescently-tagged molecules in artificial lipid membrane and in live cells [83, 142]. FRAP can also be used to investigate protein binding in various environments [143].

A FRAP experiment starts by collecting the fluorescence signal within a specific area or an entire fluorescence image before photobleaching. Following that, the specific area is illuminated by light with appropriate wavelength, focal size and intensity for a certain time period. The fluorophores in this region are photobleached to a large degree. If the rescanning speed is much faster than the recovery rate, a noticeable dark spot will be observed and fluorescence signal will drop to a dip point. As fluorescent molecules undergo Brownian motion, they will diffuse randomly throughout the sample and more and more bright molecules will diffuse into the photobleached area and replace the dark ones. The consequence is that the fluorescence intensity in this area increases until an equilibrium value is reached. Sometimes the fluorescence signal doesn’t recover to the pre-bleaching level, which is indicative of an immobile fraction in the specific area.

There are a number of pre-requisites for a FRAP measurement. First of all, the illumination intensity in the photobleaching period should be relatively high but the duration of such strong illumination should be short. These conditions ensure that the majority of fluorescent molecules will be photobleached in the specific area but number of molecules that randomly diffuse into this area is minimized during the photobleaching period. Secondly, the samples should not be photobleached during the pre-photobleaching and the recovery stages. This can be realized by using highly photostable fluorophores and relatively low laser excitation intensity during these stages. With the assistance of other hardware devices, such as a computer controllable fast shutter [83], we can reduce the unnecessary photobleaching events
to the minimum level. Thirdly, the fluorescence signals in the pre-photobleaching and recovery phases should not cause saturation of the detector. The saturation will cause an offset to the original and recovery signals. Fourthly, it is crucial to acquire fluorescent signals immediately after photobleaching in order to capture the fast diffusion processes.

I developed a time-lapsed FRAP data acquisition modality by writing a LabVIEW program which controls the shutter and OD filters on the automatic filter wheel (Pacific Scientific, Model No. 5240) to prevent the sample from being continuously exposed to the laser light and to switch between photobleaching and recovery phases automatically. The advantage of this method is that it minimizes unwanted photobleaching during data acquisition. The interface of the FRAP LabVIEW program and a time-lapse data acquisition module are shown in Figure 1-8.

![Figure 1-8](image)

**Figure 1-8.** (A) FRAP LabVIEW Interface, including shutter, OD controls and time-lapsed mode. (B) Time-lapse FRAP data acquisition module. Reprinted with permission from reference [83]. Copyright (2015) American Chemical Society.

Figure 1-9 compares the intensity loss observed using continuous versus time-lapse data acquisition methods. With continuous excitation over 10 min. approximately 10% of the fluorescence was lost due to photobleaching. In contrast, using the time-lapse excitation
scheme with a 1:4 shutter open/close ratio, virtually no photobleaching was observed even after 40 min. In both cases the excitation laser intensity was approximately 1 W/cm².

**Figure 1-9.** Fluorescence intensity reduction in (A) continuous laser excitation and (B) time-lapse laser excitation case. Reprinted with permission from reference [83]. Copyright (2015) American Chemical Society.

FRAP technique was employed in the Gradinaru group to address the question of whether or not bicelles and liposomes bound at the polystyrene (PS) bead surface fuse into a continuous lipid bilayer and to quantify the lateral diffusion of lipids bound to the PS bead surface [83]. The lipid components of the liposomes are DMPC, DMTAP and CHOLOA and an additional short chain lipid DHPC was added to form the bicelles. The lipids used here are purchased from Avanti Lipids Polar, Inc. and CHOLOA was synthesized by Dr. Qasim Saleem from Prof. P. M. Macdonald group (Chemistry, University of Toronto). Both bicelles and liposomes contain 0.05 mol% RhB-PE (cat. # 810157, Avanti Lipids Polar, Inc., AL, USA). Detailed information on the chemical composition ratio can be found in [83]. The laser intensity was set to 3 kW/cm² for 0.5 seconds in order to photobleach the lipids within a diffraction-limited spot on the surface of PS beads. The photobleached area constituted around 4% of the total surface area of a lipid coated PS bead. Subsequent images were obtained by re-scanning the entire bead with a laser intensity of 1 W/cm². To obtain high-resolution fluorescence recovery curves, the photobleached spot was monitored immediately after the photobleaching using a laser intensity of a 1 W/cm² and 10 ms binning. The acquisition scheme in Figure 1-8 (B) was applied. However, continuous data acquisition was still used for the first minute in order to properly capture the fast recovery component.
Figure 1-10. FRAP images (measured at 20 °C) of 0.05 mol% RhB-PE-labelled PS bead bicelle lipid coatings formed at 30 °C: (A) prior to photobleaching, (B) immediately after photobleaching, and (C) 20 minutes subsequent to photobleaching. Reprinted with permission from reference [83]. Copyright (2015) American Chemical Society.

Figure 1-10 shows the results obtained for the case of bicelles allowed to bind to PS beads at 30 °C, i.e., above the transition temperature of DMPC. Prior to photobleaching (Figure 1-10 A), the RhB-PE fluorescence is a continuous ring arising from lipid bound at the PS bead surface. Immediately after photobleaching a diffraction-limited spot, the ring of fluorescence is discontinuous (Fig. 1-10 B), but it reacquires the continuous shape on a time scale of 10-20 minutes (Fig. 1-10 C). This can only occur if individual bicelles have fused and formed a continuous lipid bilayer (or bilayers) on the surface of the PS bead.

FRAP curves combined with Monte Carlo simulations were used to extract lateral diffusion coefficients because no analytical expression for extracting the diffusion coefficient(s) from the FRAP recovery curve for a micrometer-sized spherical lipid bilayer photobleached in this fashion was found. Figure 1-11 shows the experimental FRAP curves obtained from (A) a single PS bead bicelle lipid coating and (B) a single PS bead liposome lipid coating, along with the corresponding best-fit Monte Carlo simulation. Lateral diffusion coefficients and relative populations obtained from the Monte Carlo simulations are listed in Table 1-1. The simulation indicates the presence of two populations, virtually equal in size, but differing by several orders of magnitude in their lateral diffusivity.
Figure 1-11. FRAP recovery curves (black), Monte Carlo simulations (red) and 2D analytical plot (blue) for 0.05 mol% RhB-PE-labelled PS bead (A) bicelle lipid coatings (B) liposome lipid coatings. Simulations assumed two different lateral diffusion coefficients and population weightings. Results of best-fits for the lateral diffusion coefficients and populations are listed in Table 1-1. Parameters used in 2D analytical plot are identical to those used in the simulation. Reprinted with permission from reference [83]. Copyright (2015) American Chemical Society.

The lateral diffusion coefficient of the “fast” fraction \( D = 1 \times 10^{-14} \text{ m}^2\text{s}^{-1} \) is in accord with values for DMPC lipid bilayers at temperatures below \( T_M = 24^\circ\text{C} \) [144]. As for the “slow” population, supported bilayer membranes generally exhibit two roughly equal lipid populations with differing lateral diffusion properties and this is usually attributed to friction between the inner bilayer leaflet and the underlying support that is absent from the outer leaflet [145, 146]. Even if the bilayer is separated from the support by spacer groups intended to alleviate such friction, the anchors are immobile and can introduce a “picket fence”-type frictional barrier within the inner leaflet [147, 148].

The mobile fraction of lipids in the supported bilayer \( m \) (Table 1-1) is extracted by applying eqs. (1-11) and (1-12) to the raw FRAP data:

\[
(1 - m)(1 - B) + m(1 - B) = F(0), \tag{1-11}
\]

\[
(1 - m)(1 - B) + m[(1 - f_a) + f_a(1 - B)] = F(+\infty), \tag{1-12}
\]

where \( 0 \leq B \leq 1 \) is the degree of photobleaching in the photobleached area, and \( f_a \) is the fraction of total fluorophores photobleached. \( F(0) \) and \( F(+\infty) \) are the intensities measured immediately after the photobleaching event (average of first 10 data points) and at the end of
the recovery (average of last 30 data points), respectively, both normalized relative to the pre-photobleaching intensity $F(-\infty)$.

For bicelle-coated PS beads, the mobile lipid fraction is on the order of 70% (Table 1-1). This must be considered as the lower limit, since an inspection of Figure 5-10 shows that the fluorescence continues to recovery slowly out to long times.

Table 1-1. Results of Monte Carlo simulations of FRAP curves of lipids bound to PS beads. Reprinted with permission from reference [83]. Copyright (2015) American Chemical Society.

<table>
<thead>
<tr>
<th>Lipid Morphology</th>
<th>$D_1$ ($\times 10^{-14}$m$^2$/s)</th>
<th>$D_2$ ($\times 10^{-16}$m$^2$/s)</th>
<th>$P(D_1):P(D_2)$</th>
<th>Mobile Fraction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicelles</td>
<td>1.0</td>
<td>4.6</td>
<td>50:50</td>
<td>68 ± 3</td>
</tr>
<tr>
<td>Liposomes</td>
<td>0.40</td>
<td>1.4</td>
<td>40:60</td>
<td>46 ± 7</td>
</tr>
</tbody>
</table>

In order to examine the possibility that the biphasic FRAP curves were merely the result of a single population diffusing across the three-dimensional geometry of the PS beads, as opposed to two distinct populations, a two-dimensional system was modelled, consisting of a disk of finite radius with a small photobleached spot. The disk has the radius $R$ and contains fluorophores at a uniform concentration $C_0$ prior to photobleaching. The photobleached region is a small circular area of radius $a$ located at the center of the disk. Immediately after photobleaching, the fluorophore concentration in this area decreases to $C_1$ and then changes with time according to the equation [106]:

$$\frac{\partial C}{\partial t} = D \nabla^2 C,$$

where $D$ represents the diffusion coefficient of the fluorophores. The general solution of this differential equation is:

$$C = \sum_m A_m e^{-\left(\frac{x_m}{R}\right)^2} D t \int_0^\infty \left(\frac{x_m}{R}\right) \,,$$
where $A_m (m=0,1,2\ldots)$ are the coefficients, $f_0(x)$ is the zero-order Bessel function, and $x_m$ is the zero points of the first-order Bessel function $f_1(x)$. By applying the initial conditions, the normalized number of fluorophores in the photobleached area can be written as:

$$n(t) = \frac{C_1 \sigma^2 + C_0 (R^2 - \sigma^2)}{R^2} + 4(C_1 - C_0) \sum_{m=1}^{\infty} e^{-\frac{(x_m)^2}{R^2} \sigma^2} \frac{f_1^2(x_m)}{x_m f_1^2(x_m)},$$  \hspace{1cm} (1-15)

The total surface area, the photobleached area, the initial fluorescence intensity and the mobile fraction of the 2D disk are set to match those estimated for the 3D sphere. The recovery curves obtained using eq. (1-15) are shown in Figure 1-11 (blue line). The comparison shows that while FRAP in the “fast” phase is similar in the two cases, recovery in the “slow” phase is enhanced in the three-dimensional case. This is likely due to the more complex geometry of the photobleached region in the three-dimensional case [148], relative to the strictly circular photobleached spot assumed for the two-dimensional case. Nevertheless, this comparison makes clear that the dimensionality of the PS bead is not the origin of the biphasic FRAP curves, but rather, that there are indeed two different diffusing populations.

### 1.2.2 Single molecule fluorescence setups

#### 1.2.2.1 Total internal reflection fluorescence microscope

##### 1.2.2.1.1 Principle of total internal reflection fluorescence microscope

Total internal reflection is an optical phenomenon that happens when light reaches the interface of two density-different media. When the light is incident from a higher refractive index medium to a lower refractive index one at an angle greater than the critical angle $\theta_c$, the light is totally reflected from the interface. The value of the critical angle is defined as $\theta_c = \sin^{-1}(n_2/n_1)$, where $n_1$ and $n_2$ are the refractive indices of the two media and $n_2 < n_1$ is the prerequisite condition for the occurrence of total internal reflection.

D. Axelrod developed total internal reflection fluorescence microscopy (TIRFM) in 1980s, by employing the total internal reflection at the interface between a glass coverslip and
the buffer solution [149]. TIRFM uses the evanescent field to illuminate fluorophores in a restricted area of the specimen immediately adjacent to the glass-water interface. The evanescent field decays exponentially from the interface into the lower refractive index material, and thus penetrates to a depth of approximately 100 nm into the sample medium. The evanescent field intensity can be written as \( I(z) = I(0) \exp \left( -\frac{z}{d} \right) \), where \( I \) is the evanescent field intensity, \( z \) is the penetrating depth, \( d \) is the exponentially decay distance \( \lambda = \frac{\lambda}{4\pi n_1^2 \sin^2 \theta - n_2^2} \), \( \lambda \) is the wavelength of the excitation light in vacuum, \( n_1 \) and \( n_2 \) are the indices of refraction of the liquid and glass, and \( \theta \) is the angle of incidence. The penetration depth does not depend on polarization, although the polarization of incident light and the incident angle affect the intensity of the evanescent wave at the interface [150].

The purpose of TIRF is to eliminate background fluorescence from outside the focal plane, which can significantly improve the signal-to-noise ratio, and consequently, the spatial resolution. This is true in live cell systems where cells are immobilized at the solid-liquid interface and only fluorophores located within or close to the lower cell membrane can be excited, which considerably reduces background fluorescence inside the cells.

TIRFM has become one of the most widely used single-molecule microscopic techniques. It is used in combination with other techniques, such as FRET [151], FCS [152], FA [153] and FRAP [154] to measure conformational and orientational distributions, translational diffusion, binding kinetics and local concentrations [150]. Another advantage of TIRFM is that it enables temporal and spatial tracking of the all the molecules in the illumination area in parallel, thus allowing the observation of reaction dynamics. TIRFM also provides the platform for buffer exchange to monitor chemical reactions in situ using microfluidics chambers [155].

1.2.2.1.2 TIRFM data acquisition

Typically, there are two types of TIRFM configurations. One is the prism-type configuration. The prism is mounted on the stage of an inverted fluorescence microscope and is optically coupled to the surface of interest by using a refractive index matching liquid such
as glycerol. The laser beam is focused at the point of internal reflection and evanescent field is generated along the surface [149]. The second configuration is to use through-objective TIRFM. This type is implemented in our lab and the optical layout of TIRFM is shown in Figure 1-12. The total internal reflection is created by focusing a laser beam through the periphery of a high numerical aperture objective (1.45NA/60X Plan-Apochromat, Olympus, USA). The fluorescence emitted by the sample is collected through the same oil-immersion objective and it passes through a dichroic mirror, and a series of long-pass and band-pass filters to remove scattering light and unwanted fluorescence signals.

**Figure 1-12.** Optical layout of the TIRFM setup. In the excitation paths (solid lines), 4 collimated lasers (via L1-L8) with different excitation wavelengths, 405nm, 473nm, 532nm, and 633nm, are combined in AOTF and magnified by a telescope (L9, L10). Excitation beams are reflected by quadruple band dichroic mirrors and pass through the oil-immersion objective. 1D motorized stage shifts the incident angle, 1D axial nano-positioning stage controls the image focus and 2D motorized stage monitors illumination area. In the emission paths (dash lines), fluorescence signals pass the identical objective and dichroic and then are filtered by proper long-pass and band-pass filters. Flipping mirror 1 (FM1) and FM2 convert the measurement modality from single channel imaging or dual channel FRET assay.
realized by a dichroic pair (DM3, DM4) to dual channel FA assay accomplished by polarized cubes (PC1, PC2). Filtered signals are collected by the EMCCD camera for imaging and post data process. Single molecule image of Alexa 555 encapsulated in 100 nm vesicle and an intensity-time trajectory are demonstrated on the upper right corner.

Figure 1-12 shows a diagram of the current TIRFM setup, which I developed in the Gradinaru lab from a version developed previously by Dr. Amir Mazouchi and Dr. Baoxu Liu. In the excitation paths (solid lines), there are four lasers with various emission wavelengths, 405nm, 473nm, 532nm and 633nm. Lenses L1-L8 are used for collimating the beams and for controlling the size of the laser beam spot. All laser beams are merged together by passing through or by reflecting off dichroic mirrors and reach the acousto-optical tunable filter (AOTF) to have similar waist sizes.

The AOTF (TF-525-250-6-3-GH18A, Gooch & Housego, USA) can select a specific wavelength from the incoming laser sources. It uses the acousto-optic effect to diffract the light using radio frequency (RF). Basically, a piezoelectric transducer is attached to a birefringent crystal. When the transducer is stimulated by the applied RF signal, acoustic waves are generated and they produce a periodic modulation of the refraction index. This modulation provides a moving phase grating that diffracts the incident beam. For a certain RF frequency, only a narrow range of wavelength can meet the phase matching condition and thus are able to diffract. By tuning the RF frequency the transmitted wavelength changes, and the rate of change is ca. 0.1 μs. The intensity of the diffracted beam is determined by the amplitude of the RF signal applied to the birefringent crystal. The AOTF can control up to 8 different wavelengths simultaneously.

After the AOTF, the selected laser beam passes through a pair of telescopes, sandwiched by a pinhole, to collimate the beam, remove aberrations, produce a smooth transverse Gaussian intensity profile and magnify the beam to the desired size. The following two irises are used for beam alignment and to control the size of the beam if necessary. The lens FL1 can be electronically flipped in and out of the beam path to help users check the backscattering and to provide photobleaching in FRAP measurements. The laser beam is
reflected up by a quadruple band dichroic mirror (405/488/532/635nm), passes through an oil-immersion objective (1.45NA/60X, Plan-Apochromat, Olympus, USA) and excites the samples deposited on the glass surface. An axial piezo element (Nano-F, Mad City Labs Inc., USA) moves the objective to focus the image on the camera and a motorized stage (LX-XYB-Z5A-FW, Applied Scientific Instrumentation, Inc., OR, USA) controls the area of the sample to be imaged.

I wrote the LabVIEW program to control all the hardware in the TIRFM setup. The interface of TIRFM LabVIEW program is shown in Figure 1-13. The left panel controls the hardware operation. It includes adjusting the position of the piezo stage for focusing, controlling the positions of the 2D motorized stage for sample scanning in different areas, and moving the position of the motorized actuator for changing the incident angle and image mode to perform TIRF and wide field imaging. The program controls the wavelength and power of light that reaches the objective (by using the AOTF), the TTL setting (which is synced to camera) and the ALEX imaging modality. The middle panel sets various camera parameters, such as the acquisition mode, exposure time, accumulation scan, electron multiplying (EM) gain, number of frames, readout and shift speeds, region of interest, bin size and baseline level. The right panel is the on-the-fly data processing. The user can obtain the positions and intensities of any point in an image, attain an intensity profile as a function of position over the selected region of interest and also save the acquired data.
**Figure 1-13.** The interface of TIRFM LabVIEW data acquisition program. The middle part is the controls and settings of the EMCCD camera. The left side is the controls and settings of the lateral and axial position, TIRF angle, laser excitation wavelength and intensity, and control the ALEX modality. The right side is dedicated to on-the-fly position and intensity diagnosis, line plot and save functions of this data.

The fluorescent signals pass through the same objective and quadruple band dichroic mirror and go through the emission path (dashed lines in Figure 1-12). In our setup, users can perform three kinds of measurements. Flipping mirrors 1 and 2 (FM1 and FM2) are used for switching between them. With FM1 and FM2 flipped down, single-color imaging can be performed in the absence of dichroic mirrors 3 and 4 (DM3 and DM4), dual-color imaging, e.g. FRET, can be performed in the presence of DM3 and DM4. With FM1 and FM2 flipped up, dual-channel FA measurements can be performed by employing two polarized cubes (PC1, PC2). The fluorescence signals eventually project on the EMCCD camera for collection.
I designed a sliding blade using Solid Edge (Student Edition, Siemens, TX, USA). It was mounted immediately in front of the camera to control the fraction of the camera chip used for imaging. This mode of operation is used for fast kinetics studies to ensure that only small number of pixel lines can be illuminated, for instance 20-30 out of a total of 512. The EMCCD camera (DU-897BV, Andor Technology, USA) in the setup is capable of 33 full frames per second. The blade effectively reduces the size of the frame, so it can be read faster and ensure that kinetics on the (sub-) millisecond time scale can be measured.

All TIRF measurements were performed using long-pass filters and band-pass filters in each color or polarization channel as shown in Figure 1-12. Examples of a single-molecule image of Alexa 555 encapsulated in the 100 nm lipid vesicle and one of its intensity-time trajectories are shown in Figure 1-12. The TIRFM operation and diagnosis manual is written in Appendix D.

1.2.2.1.3 TIRFM alternating laser excitation

Alternating laser excitation (ALEX) is a technique that complements smFRET measurements by providing a wide dynamic range of distance information and confirms the presence of both donor and acceptor fluorophores [114]. Common donor-acceptor FRET pairs have a Förster radius, $R_0$, around 50 Å. In practice, it is difficult to differentiate between low FRET species for a dye separation that is larger than 70 Å, and species labelled with only donor or with a photobleached acceptor. The advantage of ALEX is that fluorescent molecules can be sorted by separating singly labelled species from species that contain both fluorophores via alternating the excitation wavelength to selectively excite the donor and the acceptor, respectively [114].

Combining TIRFM with ALEX is a powerful way to extend the distance range of surface immobilized molecules in real time [156]. The top panel of Figure 1-14 shows a TIRFM ALEX scheme with periodic 2/20 seconds-ON/OFF alternation for donor and acceptor excitations. Spin-coated 20 nm fluorescent bead was used to prove the concept. The fluorescence intensity-time trajectories in the donor and acceptor channels corresponding to the excitation scheme above were shown in the bottom panel of Figure 1-14.
Figure 1-14. (Top) Illustration of a TIRFM ALEX sequence with periodic 2 seconds ON and 20 seconds OFF for the acceptor excitation and 20 seconds ON and 2 seconds OFF for the donor excitation. Blue denotes the donor excitation condition and red represents the acceptor excitation case. (Bottom) Fluorescence signals of 20 nm bead in donor channel (blue) and acceptor channel (red) detected as a result of the ALEX excitation sequence.

A variant of the ALEX scheme is the ADA scheme, in which the excitation sequence comprises 3 instead of 2 sequential excitation periods. Acceptor excitation first, followed by donor excitation and then again acceptor excitation at the end. This excitation scheme is particularly useful for smFRET measurements of molecules encapsulated in liposomes [157].

Figure 1-15 shows schematically how proteins can be encapsulated inside a liposome, which is then tethered to a surface coated with polyethylene glycol (PEG) via biotin streptavidin interaction. The number of molecules encapsulated into a lipid vesicle follows the Poisson distribution [121], which means that it can be more than one molecule inside the liposome container. This complicates smFRET measurements and may provide false FRET signals. The ADA scheme can effectively eliminate these artefacts and only select one molecule labelled with both donor and acceptor fluorophores in the liposome when users are interested in studying intramolecular FRET or only select two molecules with one donor on one molecule and one acceptor on the other molecule to study intermolecular FRET.
Let’s consider the example of intramolecular FRET. The initial acceptor excitation is to make sure that there is only one acceptor in the liposome. This can be realized by using single acceptor control measurement to determine the single acceptor fluorescence intensity range under same excitation and emission conditions. Next, donor excitation can help to eliminate single acceptor and more than one donor cases. The last acceptor excitation will help discard independent donors and acceptors trapped in the same liposome. After these three steps, only the case of one molecule with both donor and acceptor fluorophores attached will be retained for further analysis.

![Figure 1-15](image)

**Figure 1-15.** Scheme of encapsulation of single molecules within a liposome immobilized on a PEG-PEG biotinylated surface via biotin streptavidin interaction.

### 1.2.2.1.4 TIRFM data analysis

As discussed in the above section 2.2.1.2, TIRFM can perform both single molecule spectroscopy and microscopy imaging routines, including FRET, FA and FRAP. I wrote the TIRFM data analysis Matlab program to post-process those spectroscopy and microscopy images. The brief flow chart of the Matlab data analysis program is shown in Figure 1-16.
Figure 1-16. Flow chart of the custom-written Matlab program for TIRFM data analysis.

The TIRFM analysis program consists of two main parts. In the microscopy imaging part, the Matlab program can output a FRET, a FA or a co-localization image. In the single-molecule spectroscopy part, the program can build and analyze photobleaching data, smFRET traces and anisotropy traces. The sequence of steps of the program is the following:

1. **Choose the analysis mode.** The options are spectroscopy data and microscopy image. In the spectroscopy data category, single molecule photobleaching (smPB), smFRET and smFA can be selected. In the microscopy image category, FA, FRET and co-localization modes can be chosen.

2. **Load control and calibration files.** There are several calibration files are required to be loaded before sample images are processed and analyzed. These calibration files are the background image, flatness calibration image, G factor calibration image, channel overlapping image and channel leakage images. All these control images should be taken at the same data acquisition condition, such as setup configuration, alignment, laser excitation intensity, and all other hardware settings. The background image is referred to as the image with buffer solution. Non-uniform illumination is produced via a Gaussian excitation beam directed through a high-NA objective to create an evanescent field at the sample. To account for this spatial inhomogeneity, TIRF images of a solution with fluorophores (the fluorophore could be the same as the one used in the ongoing sample measurements) need to be collected before each
series of single-particle acquisitions. After background subtraction, a flatness matrix is calculated by normalizing the intensity of the image. Applying the flatness matrix to the sample images could help to homogenize the intensity across the entire TIRF image of immobilized fluorophores. The G factor calibration file can also be performed using a solution of fluorophore. The purpose of the G factor calibration is to un-bias the detection sensitivities of the two channels. The channel overlapping calibration is done by imaging spin-coated bead. The image transformation matrix is obtained from the coordinates of the bead image in both channels. The transformation matrix can then be applied to correct each channel’s shift relative to one another in the sample images. Channel leakage information is obtained by imaging donor and acceptor fluorophore solutions and calculating the intensity ratio of each corresponding pixel in both channels. For smPB analysis, users need to upload the background and flatness images. For smFRET analysis, users need to load the background file, channel overlapping, donor leakage and acceptor leakage files. Users also need to input the maximum single-donor and single-acceptor intensity counts under given experimental conditions, which can be obtained from the brightness histograms of donor-only and acceptor-only samples. In the smFA analysis, users need to load the background file, channel overlapping image and the G factor calibration file.

3. **Load single-molecule images and assign program parameters.** After users obtain all the calibration information stated above, the program will prompt users to load the sample images for analysis. Users need to input the cycle time of the measurement, pixel per spot, brightness threshold for identifying single molecules, and the number of frames that users are interested in analyzing. The cycle time value is the time it takes for one frame. The program converts the output trajectory from ‘frame unit’ to units of time. The pixel number per spot depends on the magnification of the objective and image magnification of the lens (L11, L12 or L13 shown in Fig. 1-8) before the EMCCD. This value is obtained by visualizing the number of pixels one single molecule occupies. The pixel number per spot value could help to exclude aggregated particles in images. The brightness threshold is set to keep all points with intensity above the threshold. For instance, a brightness threshold of 0.9 means that the top 10% of pixel intensity in the image will be selected. The program will automatically print the total number of frame
in the stack image. If the user inputs frame numbers and warning will be triggered if the number of frame the user sets is larger than the total number of frames in the stack image.

4. **Identify the position of single molecules.** By default, the program will identify all the particles positions in each successive frame in order not to exclude particles that were initially in the dark state (i.e., blinking). However, users can have the option to choose a range of frames of interest. This option can be used in the ALEX modality or for the purpose of reducing the computation time. The molecule identification is based on the brightness threshold that user inputs in step 3. The program searches all the pixels whose brightness value are above the threshold. A 2D Gaussian maximum-likelihood estimation (MLE) fitting model [158] was used to obtain the center, width and amplitude of the identified molecules. The x and y locations of the identified spots were indexed and then the program removes the overlapping spots within a distance of certain pixels to avoid analyzing separate particles tethered too close to each other. To avoid analyzing aggregated molecules, spots that have width larger than two times the number of pixel per spot are removed. The final molecule coordinates are saved for later use. Depending on the specific application, the program provides users freedom to adjust the input parameters based on their own interest.

5. **Generate intensity-time trajectories and change-point analysis.** Based on the positions of the identified spots, the program outputs the intensity-time trajectories for each identified spot. At the same time, the program uses the change-point analysis method to identify each change point occurring in the intensity-time trace. A detailed description of change-point algorithms can be found elsewhere [159-161]. Briefly, the algorithm first calculates cumulative sum (CS) of the intensity of each individual time series $I(t)$ starting from the origin to the final data point, according to the formula:

$$CS(t) = \sum_{t'=1}^{t} (I(t') - \langle I(t) \rangle).$$  \hspace{1cm} (1-16)

The CS fluctuates is around zero if there is no transition in the raw intensity. However, if a transition occurs, CS(t) would increase or decrease depending on the direction of the change in the original signal. The putative change-point was assigned to the location where the
$|CS(t)|$ is maximal. The algorithm then evaluates the significance of this putative change-point based on the user-defined criterion. The criterion basically compares with the p value. The p value is obtained by performing a statistical t test of two subgroups of data that are separated by the putative change-point. The p value tells us that there is less than p chance that the change is due to random chance. If the p-value is less than or equal to the chosen significance level, it suggests that the observed data is inconsistent with the null hypothesis, so the null hypothesis must be rejected [162]. Thus any deviation above the user-defined criterion is designated as a statistically significant change in the intensity-time trajectory and the program records and truncates the data at that point. Following that, the program takes this point as a new origin and continue recursively until no true change points are found. Akaike information criterion (AIC) is incorporated in the program to choose the statistically optimal criterion value for the change point analysis [163, 164].

$$AIC = nln\left(\frac{\chi^2}{n}\right) + 2k + \frac{2k(k+1)}{n-k-1},$$

where n is the number of data points, k is the number of free parameter, $\chi^2 = \sum l \frac{(data_i - fit_i)^2}{\sigma^2}$ and $\sigma^2$ is the variance of the data. The preferred model is the one with the minimum AIC value [164]. It should be noted that the choice of criterion value sometimes needs to be validated with the physical meaning of the occurrence of the change points.

After the change-point analysis is done for all the identified molecules, the raw data are fitted by a trajectory with discrete intensity levels interspersed with stepwise jumps. The data smoothing option [165] can also be chosen. Two examples of fluorescence intensity-time trajectories, including raw and fit, are shown in Figure 1-17. The first is a smPB trace of streptavidin labelled with Alexa555 (Molecular Probes, Thermo Fisher Scientific, Canada); The second is a smFRET trace of 6bp dsDNA labelled with TMR (donor) and Alexa647 (acceptor) (Integrated DNA Technologies, Inc., Iowa, USA).
Figure 1-17. (A) An example of smPB trace of streptavidin-Alexa555 immobilized on the PEG-PEG biotinylated surface. The black data points are the raw data and the red lines are the fitted intensity levels. (B) An example of FRET-time trace of 6bp dsDNA labelled with TMR and Alexa647 immobilized on PEG-PEG biotinylated surface. Red and green lines denote the raw acceptor and donor intensities and the black and blue lines represent the fitted intensity levels.

6. **Build histograms.** Depending on what program modality is used, the program outputs global histograms of different fitted parameters from all molecules analyzed. If the user analyzes smPB data, a histogram of the number of smPB transitions per molecule will be produced, as well as the histogram of step sizes. If the user analyzes smFRET (or smFA) data, FRET traces and a FRET efficiency histogram, (or anisotropy traces and a histogram of anisotropy values) will be generated. In addition, a FRET/anisotropy transition map and dwell time histograms will be produced.

Some constraints can be added to make the data analysis more meaningful. For instance, fluorophores sometimes blink and sometimes enter the dim state upon excitation [166]. These states could be misinterpreted as photobleaching events in the smPB analysis. Users can set an intensity drop range and if the intensity drop at the change point is not within the range specified, it will not count as a smPB step. Similarly, in the smFRET analysis, anti-correlated donor and acceptor intensity fluctuations are observed before either of them photobleaches. The program calculates the correlation functions of each smFRET trace and discard those trajectories that are not anti-correlated before either donor or acceptor photobleaches, thus removing false FRET information from the overall histogram.
7. **Statistical analysis of single-molecule histograms**. Figure 1-18 (A) is the simulated FRET histogram with three normally distributed sub-populations. The mean efficiencies are set to be 0, 0.66 and 0.97 with a width of 0.15 for all three clusters. The corresponding fractions are 20%, 40% and 40%. The program employs statistical optimization algorithms, such as K-means [167], MLE of Gaussian mixture models (GMM) [168], AIC [164] and bootstrap methods [169] to analyze the histogram data without human bias. The main advantage of the program is that users do not need to input the expected number of clusters (sub-populations) or to guess the starting values of the mean and the width of each cluster. The core of the bootstrap algorithm is to estimate the sampling distribution using random permutation methods [169]. The data is replicated N times, and random permutation are performed to construct a new set of bootstrapped data sets (see Fig. 1-14 (C-E)). Each bootstrapped data set is regarded as an independent data set and from which the global mean and standard deviation can be obtained.

The objective of K-means algorithm is to minimize the total intra-cluster variance, which is expressed as [167]

\[
\text{var} = \sum_{i=1}^{k} \sum_{x \in D_i} |x - \mu_i|^2 ,
\]  

(1-18)

where \(\mu_i\) is the mean of the data points \(D_i\) and \(k\) is the cluster number.

The purpose of GMM is to define a Gaussian mixture distribution that consists of a mixture of one or more Gaussian distribution components and calculate the maximum-likelihood of the mixture distribution. For example, one Gaussian distribution with mean \(\mu\) and width \(\sigma\) is

\[
p_i(x|\mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right] ,
\]  

(1-19)

and the Gaussian mixture distribution of \(M\) components is

\[
p(x|\theta) = \sum_{j=1}^{M} \alpha_j p_j(x|\theta_j) ,
\]  

(1-20)
where $\theta_j = (\mu_j, \sigma_j)$, $\Theta = (\alpha_1, ..., \alpha_M, \theta_1, ..., \theta_M)$, and the fraction $\alpha_j$ has $\sum_{j=1}^{M} \alpha_j = 1$. The log-likelihood can be expressed as [168]

$$
\log L(\Theta) = \sum_{i=1}^{N} \log p(x_i | \Theta) - \sum_{i=1}^{N} \left( \log \sum_{j=1}^{M} \alpha_j p_j(x_i | \theta_j) \right).
$$

Based on log-likelihood value, the AIC value can be calculated. The preferred fitting model is the one with the minimum AIC value. The AIC values of each fitting model are shown in Figure 1-18 (B). It demonstrates that the fitting model with three clusters is the optimal one because it has the minimum AIC value. This result is in agreement with simulated parameters. The default fitting functions are Gaussians (Fig. 3-14 (C-E)), but the user can change the distribution function depending on specific prior information. After loading the data file, the program will automatically output the mean, the width and the fraction of each cluster of the histogram, together with their standard deviation (Table 1-2). The smFRET/smFA transition map could also be outputted by the program. Figure 1-18 (F) illustrates the FRET transition map for the data shown in Figure 1-18 (A). The $k_{on}$ and $k_{off}$ rates are critical when studying binding kinetics. The program is capable of analyzing the on and off dwell time of single molecule intensity-time trajectories and outputs the on and off dwell time histograms (Fig. 1-14 (G, H)).
Figure 1-18. smFRET histogram fitting results. (A) Simulated FRET efficiency histogram; (B) AIC values of different number of clusters used in the histogram fitting models; (C-E) Examples of bootstrap histogram based on original FRET efficiency histogram of (A). Green curves are the Gaussian fits for each FRET cluster and the red curve is the overall fit to the FRET histogram; (F) FRET transition map of initial FRET efficiency and final FRET efficiency; (G) Dwell on-time histogram; (H) Dwell off-time histogram. The red decay curves are the fitting curves.

Table 1-2. Example of histograms fitting output

<table>
<thead>
<tr>
<th>Cluster No</th>
<th>E_{center}</th>
<th>E_{center std}</th>
<th>E_{width}</th>
<th>E_{width std}</th>
<th>Area (%)</th>
<th>Area_{std} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0103</td>
<td>0.0129</td>
<td>0.1553</td>
<td>0.0125</td>
<td>19.57</td>
<td>1.30</td>
</tr>
</tbody>
</table>
This versatile TIRFM data analysis program has been used in the Gradinaru lab on various projects, i.e., to measure the distribution of DNA oligos conjugated to quantum dot biosensors [139], the size of DNA ladder constructs for a new diagnostic assay, and the oligomeric size of the M2 muscarinic receptor and its associated G-protein [141]. Currently, the setup and the analysis program are being used to study protein conformations and interactions in disordered protein and G protein coupled receptors (GPCRs) via smFRET.

1.2.2.2 Multi-parameter fluorescence confocal microscope

The confocal microscope increases the spatial resolution and contrast of fluorescence imaging through the insertion of a pinhole in the detection path to prevent out-of-focus emission from reaching the detector [170]. Multi-parameter fluorescence confocal microscope (MPFCM) allows users to obtain multiple parameters from an experiment, including emission color, intensity, lifetime, anisotropy, and fluorophore stoichiometry.

A simplified optical layout of the multi-parameter fluorescence confocal setup is shown in Figure 1-19. The excitation light can be chosen from three CW lasers 473 nm, 532 nm and 635nm and a Ti-sapphire laser (Tsunami HP, Spectra Physics, Santa Clara, CA, USA) which is continuously tunable from 700 nm to 1100 nm and outputs approximately 100-fs pulses at a repetition rate of 80 MHz. Nonlinear β-BBO crystals are utilized for second harmonic generation (SHG) to obtain pulses in the visible spectrum range. Flipping mirrors are used to switch between different excitation paths. A TTL-modulated CW laser at 635 nm is used for ALEX measurements.

The excitation light bounces off the dichroic mirror, passes through a 1.4NA/100x Plan-Apochromat oil objective (420790-9900, Carl Zeiss, Canada) and then focuses onto the sample mounted or freely diffusing in a solution on top of a glass coverslip. The coverslip is mounted
on a three-axis piezo nanopositioning stage (T225, Mad City Labs Inc., USA), which controls the 2D lateral position and axial focus. The fluorescence emission is collected through the same objective and is transmitted through the excitation dichroic mirror. Rayleigh and Raman scattering, impurities and out-of-focus fluorescence signals are removed via a series of long-pass and band-pass filters, followed by a confocal pinhole. The “clean” fluorescence signal is split by an emission dichroic into two different spectral channels. Polarization cubes further divide each spectral channel and direct it onto separate single-photon avalanche diodes (SPADs) for fluorescence anisotropy analysis.

Two types of SPADs are installed in the MPFCM setup. The first type is a Micro-Photon Device (MPD) module (PD5CTC, MPD, Italy), which has fast-response, ca. 50 ps, a detection efficiency of ~50% at 550nm and a low dark count, 20-50 cps. The other type is a red-sensitive COUNT module (COUNT, Laser Components, USA), with ~70% photon detection efficiency at 660 nm, and with a slower time response (~1000 ps). PicoHarp300 (PicoQuant, Germany) is a time-resolved single-photon counting (TCSPC) system with a 4 ps resolution. The SPADs are connected to this system via a router module (PHR800), which has 4 input channels and 4 “marker” channels to receive synchronization signals from other hardware [171]. These markers provide the chance to sort out photons coming from different triggers and implement the ALEX modality. Detailed description of MPFCM setup can be found in the thesis of Dr. Baoxu Liu from the Gradinaru Lab. The confocal ALEX operation was designed and implemented by Gregory Gomes; I used the ALEX modality on this setup for the smFRET study on 4E-BP2 protein (Chapter 3).
Figure 1-19. Simplified optical layout of the multi-parameter fluorescence confocal microscope. The laser sources include CW lasers 473nm, 532nm and 635nm and femto-second pulse laser Tsunami 700-1100nm. After SHG, laser frequency gets doubled and visible light is used to illuminate the sample. FM switches between CW lasers and femto laser. 635nm laser is modulated for ALEX experiments. Fluorescence signals pass through the same objective and dichroic mirror Di1 and separated by Di2 based on different fluorescence colors. PC1 and PC2 further divide the fluorescence signals into two SPADs for polarization analysis. The SPADs connect PicoHarp single photon counting module.

1.2.2.3 Dual-color fluorescence correlation microscope

Dual-color fluorescence correlation microscope (dcFCM) is another fluorescence confocal microscope in the lab. It was built by Yuchong Li and Jordan Rebelo in the Gradinaru Lab. The main differences between dcFCM and MPFCM is that dcFCM does not use a TCSPC system, thus it does not provide photon-by-photon information and it is not suitable for picosecond-nanosecond time-resolved measurements. The dcFCM employs a digital correlator (Flex02-01D, correlator.com) to provide fast fluorescence correlation calculation, so that FCS and FCCS measurements can be efficiently and reliably performed on this setup. The excitation and emission optical path is similar to MPFCM as described in detail in the previous section.
The setup has been used to study the aggregation status of drkN SH3 domain (Chapter 2) and the lipid-dye interactions (Chapter 4).
Chapter 2
Resolving Conformational Heterogeneity of an SH3 Domain Using Single-Molecule Fluorescence Spectroscopy and an Excluded-Volume Polymer Model*

2.1 Abstract

Conformational states of the metastable drkN SH3 domain were characterized using single molecule fluorescence techniques. Under non-denaturing conditions, two Förster Resonance Energy Transfer (FRET) populations were observed, corresponding to the folded and an unfolded state. FRET-estimated radii of gyration and the hydrodynamic radii estimated by Fluorescence Correlation Spectroscopy (FCS) of the two coexisting conformations are in agreement with previous ensemble X-ray scattering and nuclear magnetic resonance measurements. Surprisingly, when exposed to high concentrations of urea and GdmCl denaturants, the protein still exhibits two distinct FRET populations. The dominant conformation is expanded, shows a low FRET efficiency and is consistent with the expected behavior of a random chain with excluded volume. However, approximately one third of drkN SH3 conformations showed high, nearly 100% FRET efficiency, which is shown to correspond to denaturation-induced looped conformations that remain stable on a timescale of at least 100 μs. These loops may contain interconverting conformations that are more globally collapsed, hairpin-like or circular, giving rise to the observed heterogeneous broadening of this population. Although the underlying mechanism of chain looping remains elusive, FRET experiments in formamide and DMSO suggest that interactions between hydrophobic groups in the distal regions may play a significant role in the formation of the looped state.

* The results described in this chapter have been published in Biophysical Journal, 2016. 110(7): p.1510-1522. Reprint with permission.
2.2 Introduction

There is growing awareness that the biologically functional state of many proteins is not folded in a stable configuration, but instead contains one or more disordered regions. Although they lack a stable 3D fold, these intrinsically disordered proteins (IDPs) and regions generally exhibit fluctuating secondary and tertiary structures [172]. IDPs are involved in cell signaling, molecular recognition and transcriptional regulation, with a majority of oncogenic proteins having significant disordered regions [29, 47, 173]. Many similarities exist between IDPs and the unfolded state ensemble of globular proteins, as may be expected given that some IDPs undergo a disorder-to-order transition upon interaction with other proteins or upon post-translational modification [29, 53]. A more quantitative description of unfolded state ensembles is much needed, not only from a protein folding perspective, but also as a model or reference for IDPs. Since the 1960s, when Tanford showed that proteins in high concentrations of chemical denaturant obey a random-coil scaling law [12], the random-coil model has been widely used as the basis for understanding the physical properties of unfolded proteins. However, the validity of this model for describing the unfolded ensemble is limited, since significant residual structure was reported in several denatured proteins, such as apomyoglobin and staphylococcal nuclease [174], and local hydrophobic interactions restricting backbone motions have been shown to persist even under the harshest denaturing conditions [11].

Despite the biological importance of disordered/unfolded states of proteins, conventional experimental approaches for structural and thermodynamic characterization provide limited information and the nature of their transient structure remains elusive [175, 176]. Computational methods have recently focused on the study of unfolded and disordered proteins [177-179], although the selection of the force field in the molecular simulations is critical and the most popular options seem to be biased towards describing the folded state. Recent computational studies point to an overestimation of the helical structure in small proteins and to an underestimation of the size of denatured proteins compared to experiments [180]. Hybrid computational-experimental approaches in which sets of conformers are selected to fit nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS) experimental data were developed [181], although powerful single-molecule fluorescence data
are still to be incorporated to help provide a better evaluation and understanding of structural ensembles.

The N-terminal Src homology 3 domain of Drosophila downstream of receptor kinase (drkN SH3) is a 59-residue \( \beta \)-barrel consisting of five anti-parallel \( \beta \)-strands (PDB 2A36). It has a marginally stable structure, with the folded (\( F_{exch} \)) and unfolded (\( U_{exch} \)) states almost equally populated under non-denaturing conditions and interconverting at a slow exchange rate, \( \sim 2 \text{ s}^{-1} \) [182]. NMR chemical shift, J coupling, residual dipolar coupling, \( O_2 \)-induced paramagnetic shift, \( ^{15}N \) relaxation, nuclear Overhauser effect (NOEs), paramagnetic relaxation enhancement (PREs) and hydrodynamic radius (\( R_H \)) measurements, as well as SAXS data, have been used to calculate structural ensembles for the \( U_{exch} \) state [183]. The data and the derived ensembles provide evidence for fluctuating structure in the \( U_{exch} \) state, including significant non-native \( \alpha \)-helical secondary structure as well as both native-like and non-native tertiary structure, in particular in the region of the central \( \beta \)-sheet of the folded state showing multiple contacts to the tryptophan (Trp) at position 36. NMR and Trp fluorescence data point to significant differences between the \( U_{exch} \) unfolded state under non-denaturing conditions and the chemically denatured state and demonstrate that drkN SH3 is virtually unfolded in 2 M GdmCl [25, 184, 185].

Single-molecule fluorescence (SMF) spectroscopy provides an elegant quantitative characterization of co-existing protein conformations [186], and it has become a powerful tool for studying the physical mechanism of protein folding [187-189]. In order to describe and compare the conformations associated with the unfolded state and the chemically denatured state of the drkN SH3 domain, SMF spectroscopy experiments were performed, such as fluorescence correlation spectroscopy (FCS) and single-molecule Förster resonance energy transfer (smFRET). The protein was found to preserve its conformational heterogeneity in the denatured state. Based on the single-molecule data, the hydrodynamic radius (\( R_H \)), the end-to-end distance (\( R_{EE} \)), and the radius of gyration (\( R_G \)) were estimated for different conformational states resolved for drkN SH3. The linear size and the topology of the denatured protein were interpreted using a coarse-grained, sub-ensemble based polymer model that accounts for excluded volume [190]. The single-molecule results obtained in this study are compared with
previous drkN SH3 ensemble data obtained by NMR and SAXS, and interpretations are proposed in terms of denaturant-induced intra-chain interactions.

2.3 Materials and Methods

2.3.1 Materials

The fluorescent probes used for labelling the SH3 domain for SMF experiments were: 5-carboxy-tetramethylrhodamine N-succinimidyl ester (TMR-NHS) (AnaSpec, Fremont, USA), Bodipy Fluorescein (BFL) maleimide, Alexa 647 (A647) maleimide and Alexa 555 (A555) maleimide (ThermoFisher Scientific, Canada). 40-base pair-long oligonucleotides labeled with Cy5 at 5’ and 6-FAM at 3’ were purchased from Integrated DNA Technologies, Inc. (Coralville, Iowa, USA). The sequence was chosen as follows to prevent FRET between two conjugated dyes: Cy5-5’-TAA GCC TCG TCC TGC GTC GGA GCC CGT CTG CCA GCG GAAT-3’- 6-FAM. DNA sample was usually used as a detection volume calibration sample. Guanidinium chloride (GdmCl) (G9284, Sigma Aldrich) and urea (EM-9510, EMD Millipore) were used for protein denaturation. Immediately before use, urea was purified from hydrolysis products and ionic impurities as described previously [191]. GdmCl and urea solutions were adjusted to pH 7.5 for all the experiments unless stated otherwise. Each GdmCl and urea concentration was also confirmed by measuring the solution viscosity by FCS. All other samples were diluted in Tris buffer (50 mM Tris, 150 mM NaCl, pH 7.5). Formamide (47671, Sigma Aldrich), dimethyl sulfoxide (DMSO) (472301, Sigma Aldrich), glycerol (EM-4750, EMD Millipore), sodium sulfate (238597, Sigma Aldrich) and L-arginine (11009, Sigma Aldrich) solutions were prepared in the Tris buffer.

2.3.2 Protein expression and purification

Detailed protein expression and purification procedures can be found in our previous papers (16, 22). Briefly, plasmid encoding the N-terminal SH3 domain of the drk protein
(residues 1 through 59) with ampicillin resistance was transformed into BL21-codonplus (DE3)-RIL competent E. coli cells. Bacterial cultures were grown overnight on Lennox Lysogeny broth (LB) (L7658, Sigma-Aldrich) –ampicillin Agar plate in a 37°C incubator. A single colony was inoculated into LB medium with 100 mg/L final concentration of ampicillin (AB0028, Bio Basic Inc., Canada) and shaken in a 37°C incubator for ~5 hours until cell density reached an OD_{600} of 0.6. A 1 mM final concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG) (IB0168, Bio Basic Inc.) was then added and the medium was transferred to a 16°C incubator for overnight growth up to a cell density having an OD_{600} of around 1.3. Cells were lysed by sonication in a buffer containing 50 mM Tris-HCl, 2 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (646547, Sigma-Aldrich), 1 mM Ethylenediaminetetraacetic acid (EDTA) (EDT111, BioShop, Canada), 2 mM Benzamidine-HCl (BD0076, Bio Basic Inc., Canada), pH 7.5. The drkN SH3 domain was purified on an ion-exchange column (17-1153-01 HiTrap Q HP, GE Healthcare) with a linear gradient of NaCl (0-1 M) followed by HiLoad Superdex 75 PG gel filtration column (28-9893-34, GE Healthcare) in 50 mM Tris, 150 mM NaCl, 2 mM TCEP, 1 mM EDTA, 2 mM Benzamidine-HCl, pH 7.5. The T22G, C2 (Cys inserted at N-terminus position 2) and the C2/C61/G62 (Cys inserted at N-terminus position 2, Cys and Gly added at C-terminus, positions 61 and 62) mutants were expressed and purified in a similar manner to the wild-type (WT) protein. The identity and the purity of the three protein samples were assessed by mass spectrometry.

2.3.3 Sample preparation

For FCS experiments, free amines in the WT and the T22G drkN SH3 proteins were labeled using TMR-NHS. The NHS-amine coupling reaction was performed in a PBS buffer at pH 8.0, by adding the NHS-ester-activated fluorophores to a 50 μL solution of 200 μM protein at a dye:protein molar ratio of 1:2. This ratio was chosen in order to limit the fraction of multi-labelled proteins. For dual-color FCS (dcFCS) control experiments, the SH3 C2 protein was singly labeled using BFL-maleimide, and A647-maleimide. The samples were 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) in a BioLogic LP system (731-8300, Bio-Rad).

For smFRET experiments, the double-cysteine mutant (C2/C61/G62) was labeled in a site-specific manner by a pair of thiol-reactive dyes, i.e. BFL-maleimide (donor, neutral charge) and A647-maleimide (acceptor, negative 3 charge). 10 μL DMSO was added to a 50 μL solution of 100 μM protein in Tris buffer to improve the solubility and the labeling efficiency of BFL. TCEP was added at a 20x molar excess to the protein in order to reduce the disulfide bonds, then BFL was added at a 5:1 molar ratio to the protein. Oxygen was removed by flushing the sample with argon gas in a desiccator for 5-10 min. The vial was capped tightly and shaken gently for 3 hours at room temperature. Then A647 was added at a 20x molar excess to the protein, the solution was flushed with argon in a vacuum desiccator and kept at 4 °C for 24 hours in a sealed vial in the dark.

The free dyes were removed by size-exclusion chromatography. For some SMF experiments, a different donor dye was used: A555 (negative 4 charge) instead of BFL (neutral charge). In order to estimate the Förster radius and the intensity corrections factors for smFRET analysis, donor-only and acceptor-only proteins were prepared using a similar protocol in which the protein was incubated for 3 hours with only one species of thiol-reactive dye. Using these control samples, the fluorescence quantum yields of the dyes attached to drkN SH3 were found to be 0.55 for BFL, 0.39 for A555 and 0.37 for A647 in Tris buffer (pH 7.5), 0.38 for BFL, 0.40 for A555 and 0.36 for A647 in 6M GdmCl, and 0.28 for BFL and 0.34 for A647 in 7M urea. These represent ensemble-averaged and labelling-site-averaged values.

All samples were diluted to concentrations of 1-20 nM and 20-100 pM, which are most suitable for (dc)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. Non-specific protein adsorption to the coverslip was prevented by adding 0.005% (v/v) Tween-20 (P2287, Sigma-Aldrich) to the solution, and bovine serum albumin (BSA) (15260-037, ThermoFisher Scientific) was used to coat the clean coverslips. All experiments were performed at 20 °C.
2.3.4 Instrumentation

Quantum yield and tryptophan fluorescence measurements were performed using a QuantaMaster PTI spectrofluorimeter (Photon Technology International, Canada) equipped with a red-sensitive photomultiplier tube (R928P, Hamamatsu). Single-molecule measurements were performed on a custom-built multiparameter fluorescence microscope that was described in detail elsewhere [121, 163]. Laser excitation at 532 nm (LRS-0532, Laserglow Technologies, Canada) and 473 nm (04-01 series, Cobolt Blues, Sweden) was used in the experiments and the fluorescence data was acquired using a single photon-counting system (PicoHarp300, PicoQuant, Germany).

In FCS measurements, the fluorescence from the sample was focused through a 50-μm pinhole and then divided into two channels by a non-polarizing cube beam-splitter. Each beam was focused onto a separate single-photon avalanche diode (SPAD) (PD5CTC, MPD, Italy). A pseudo-autocorrelation curve with logarithmic time binning and 24 bins per temporal decade was calculated from these two signals using a photon-by-photon algorithm [163].

In smFRET measurements, a larger confocal pinhole was used (75 μm vs. 50 μm) and the beamsplitter cube was replaced by a dichroic mirror (FF560-Di01 or FF640-Di01, Semrock, Rochester, NY, USA), which splits the fluorescence into donor and acceptor dye contributions. In addition, a red-sensitive SPAD detector (COUNT-100, Laser Components, USA) was used to increase the signal in the acceptor channel. The donor and acceptor signals were spectrally filtered using band-pass filters (Chroma, Bellows Falls, VT, USA): BP530/50m and HQ690/70 (for BFL-A647), and D575/70 and D705/80 (for A555-A647). A custom-written LabVIEW code was used to identify and analyze dual-color single-molecule fluorescence bursts, from which a smFRET histogram was constructed [192].

dcFCS measurements were performed on a different confocal microscope, on which the samples were excited with 488-nm (TECBL-488, World Star Tech, Canada) and 633-nm (TECRL-633, World Star Tech, Canada) lasers simultaneously. Fluorescence from blue (BFL
or 6-FAM) and red (A647/Cy5) probes was passed through a 30-µm pinhole and a combination of long-pass and bandpass filters and focused on single photon-counting avalanche photodiodes (SPCM-CD 3017, Perkin Elmer Optoelectronics). BLP01-488 (Semrock, USA) and HQ 520/66 (Chroma, USA) were used in BFL/6-FAM emission channel and BLP01-647 (Semrock, USA) and HQ 685/80 (Chroma, USA) were used in A647/Cy5 emission channel. The signal output of the detectors was fed into a 4-channel hardware correlator (Flex02-01D, Correlator.com), which provided autocorrelation and cross-correlation curves.

### 2.3.5 FCS analysis

Under some simplifying assumptions, the fluorescence intensity correlation function for the free Brownian diffusion of a single molecular species with triplet-state blinking is given by [193]:

\[
G(\tau) = \frac{1}{N_{\text{eff}}} \left(1 + \frac{\tau}{\tau_d}\right)^{-1} \left(1 + \frac{\tau}{s^2 \tau_d}\right)^{-0.5} \left(1 + \frac{f_T}{1-f_T} e^{-\tau/\tau_T}\right). \tag{2-1}
\]

Prior to each set of measurements, a dilute solution of a dye with known diffusion coefficient was used to characterize the detection volume [163]. Then, by fitting the measured correlation curve to the model described by eq. (2-1), one obtains the local concentration, \(N_{\text{eff}}\), and the diffusion time, \(\tau_d\), which is used to estimate the diffusion coefficient and then the hydrodynamic radius of the molecule via the Stokes-Einstein equation [163].

### 2.3.6 smFRET analysis

The FRET efficiency was calculated using the number of detected photons in the donor \((I_D)\) and the acceptor \((I_A)\) channels in each single-molecule burst [194]:

\[
E = \frac{I_A}{I_A + \gamma I_D}, \tag{2-2}
\]
where \( \gamma \) is the ratio of the detection efficiencies (\( \zeta \)) and the quantum yields (\( \Phi \)), i.e., \( \gamma = \zeta_A \Phi_A / \zeta_D \Phi_D \) (0.72 for BFL-A647 and 0.91 for A555-A647 in Tris buffer, 0.9 and 1.17 for BFL-A647 in 6M GdmCl and 7M urea, respectively). In addition, corrections were applied on both \( I_D \) and \( I_A \) to subtract the background and the spectral cross talk. An \( E \) value was estimated for each single-molecule intensity burst, and for each sample approximately 10,000 bursts were processed and a FRET histogram was produced.

The distance between the FRET probes, \( R \), can be estimated from the measured energy transfer efficiency using the Förster equation:

\[
E = \left[ 1 + \left( \frac{R}{R_0} \right)^6 \right]^{-1}.
\]  

(2-3)

The Förster radius, \( R_0 \), was estimated to be 4.4 ± 0.2 nm for the BFL-A647 pair and 5.9 ± 0.2 nm for the A555-A647 pair. These estimates were made using the fluorophore spectra and quantum yields measured upon labelling drkN SH3. At each concentration of denaturant, the spectra and quantum yields of the donor and the acceptor were measured and used in the FRET calculations. For instance, at 6 M GdmCl, the estimated \( R_0 \) values were 4.3 ± 0.2 nm for the BFL-A647 and 5.6 ± 0.2 nm for A555-A647.

### 2.3.7 Polymer model for \( R_G \) inference from smFRET data

The drkN SH3 domain with dyes and linkers was modelled as a coarse-grained self-avoiding walk (SAW) [190] with \( n = 66 \) residues, i.e., 60 residues separating the dyes plus 6 equivalent residues (3 at each end) accounting for flexible dye linkers. Each amino-acid residue is modelled by a single spherical bead with excluded volume defined by a hard core radius \( R_{hc} = 0.4 \) nm. The SAW is sampled without energetic bias in order to populate conformations with equal \textit{a priori} probabilities. From this unbiased full ensemble, an inferred sub-ensemble is constructed by re-weighting the conformations in accordance with \textit{a posteriori} information from the smFRET histograms.
Briefly, this proceeds by seeking a distribution of end-to-end distances $P(R_{EE}|R_G)$, which minimizes the deviation $\Delta E(R_G) = |\langle E \rangle_{exp} - \langle E \rangle_{sim}(R_G)|$ where,

$$\langle E \rangle_{sim}(R_G) = \int_0^{l_c} dR_{EE} \frac{R_0^6}{R_0^6 + R_{EE}^6} P(R_{EE}|R_G) . \quad (2-4)$$

To construct $P(R_{EE}|R_G)$, conformations sampled in the unbiased full ensemble SAW are sorted into sub-ensembles conditioned on a narrow range of $R_G$. The inferred $R_G$ is the one for which the deviation $\Delta E(R_G)$ is minimized, and the uncertainty in $R_G$ is reported such that $R_G \pm \sigma_{R_G}$ is in the range for which $\Delta E(R_G) < \sigma_{(E)exp}$. The uncertainty in the mean experimental FRET efficiency, $\sigma_{(E)exp}$, was estimated to be $\sim 0.02$ using the bootstrap technique by fitting the FRET efficiency histograms of 1000 replicate data sets generated by random sampling with replacement from the original set of fluorescence bursts. More details about the use of coarse-grained protein chains in the interpretation of smFRET data, and the comparison with conventional procedures which use either the Gaussian chain model or the mean-field Sanchez polymer theory can be found in ref. [190].

The population-weighted root mean squared (RMS) radius of gyration for a mixture of conformation clusters was calculated using the following equation:

$$< R_{G,ens}^2 >^{1/2} = \sqrt{f_1 < R_{G1}^2 >} + (1 - f_1) < R_{G2}^2 > , \quad (2-5)$$

where $\langle R_{G1,2}^2 \rangle^{1/2}$ are the RMS $R_G$ values calculated or measured for each FRET sub-population, and $f_1$ and $f_2 = 1 - f_1$ are their respective fractions obtained by fitting the smFRET histograms. 5000 in silico data sets were created by adding Gaussian random noise to $f_1 R_{G1}$ and $R_{G2}$ with standard deviation according to their respective uncertainties. $\langle R_{G,ens}^2 \rangle^{1/2}$ was calculated for each of the 5000 in silico data sets, and the standard deviation of these values was reported.
2.4 Results

Fluorescence of the Tryptophan residue at position 36 in the sequence of wild-type and C2C61G62 of drkN SH3 samples was monitored upon increasing the concentration of GdmCl in solution. The fluorescence of Trp increases when exposed to an aqueous solution and it is used as an indicator in protein denaturation studies [195]. The results are shown in Figure 2-1 and indicate that the C2C61G62 mutant, which was used for smFRET studies, did not alter the denaturation mid-point and the overall denaturation behavior of the wild-type protein.

![Normalized Trp fluorescence of the drkN SH3 domain, WT (black) and C2C61G62 (blue), measured at different GdmCl concentrations. The emission intensity was recorded at 358 nm upon excitation at 280 nm.](image)

**Figure 2-1.** Normalized Trp fluorescence of the drkN SH3 domain, WT (black) and C2C61G62 (blue), measured at different GdmCl concentrations. The emission intensity was recorded at 358 nm upon excitation at 280 nm.

The distribution of protein end-to-end separations was measured by smFRET using the double-cysteine protein labeled with the BFL-A647 donor-acceptor dye pair. Figure 2-2 shows the smFRET histograms obtained under non-denaturing conditions and in the presence of protein folding stabilizers.
Figure 2-2. smFRET histograms of drkN SH3 end-labeled with BFL and A647. (A) Measurements in Tris buffer (pH 7.5). The blue- and red-colored parts of the histogram correspond to the folded ($F_{exch}$) and the unfolded ($U_{exch}$) sub-populations, respectively; the black curve was obtained from a donor-only sample. The inset shows representative structures for the two states [2, 183]. (B) smFRET histograms obtained in the presence of folding stabilizers: 0.4 M Na$_2$SO$_4$ (grey bars) and 20% (w/w) glycerol (red line).

In Tris buffer (pH 7.5), the sample shows three distinct populations with average FRET efficiencies of 1%, 55% and 99%, respectively (Figure 2-2 A, Table 2-1). We assign the 2nd and the 3rd FRET peak around to the co-existing $U_{exch}$ and the $F_{exch}$ states of drkN SH3, respectively [184]. Low or nearly-zero FRET peaks are often observed in single-molecule studies and are mostly attributed to donor-only or acceptor inactive or photobleached proteins. Alternatively, it would correspond to highly elongated conformations, in this case with end-to-end separations exceeding 90 Å. If the entire 60-residue drkN SH3 sequence was a single $\alpha$-helix with no loops, the end-to-end distance would be 60 res. $\times$ 1.5 Å/res. = 90 Å. Since this situation is unrealistic, the zero-FRET peak will be discarded in the further analysis. For the folded state, high FRET value yields an upper limit for $REE \approx 20$ Å, in agreement with the 3D structure, in which the N- and the C- termini are separated by $\sim$10 Å [2]. Based on the area under the peaks, the fraction populations of the folded and the unfolded states were estimated to be 0.53 and 0.47, respectively (Table 2-1). The FRET fractional population fraction of 0.53 for $F_{exch}$ at 20 °C is in reasonable agreement with previous values measured by NMR, i.e., 0.50 at 30 °C [24, 196] and 0.66 at 20 °C [182].
Table 2-1. Fitting parameters obtained for the smFRET histogram measured in Tris buffer (pH 7.5) (Figure 2-2) and in 6 M GdmCl (Figure 2-4) a.

<table>
<thead>
<tr>
<th></th>
<th><strong>Tris buffer (pH 7.5)</strong></th>
<th></th>
<th><strong>6 M GdmCl</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>E (%)</em></td>
<td><em>fraction</em></td>
<td><em>E (%)</em></td>
</tr>
<tr>
<td>Cluster 1</td>
<td>55.0 ± 19.6</td>
<td>0.53 ± 0.05</td>
<td>23.0 ± 16.4</td>
</tr>
<tr>
<td>Cluster 2'</td>
<td>-</td>
<td>-</td>
<td>91.3 ± 7.8</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>99.1 ± 2.5</td>
<td>0.47 ± 0.05</td>
<td>99.6 ± 3.6</td>
</tr>
</tbody>
</table>

a Data format: average efficiency ± HWHM, average fraction ± S.E.

To confirm the assignments of the FRET peaks, we performed experiments in solvents known to perturb the equilibrium between the folded and the unfolded state of drkN SH3. For instance, in solutions containing sodium sulfate or glycerol, the FRET population around $E = 55\%$ is absent and the smFRET histogram contains only the 0\% and the 100\% peaks (Figure 2-2 B). Sodium sulfate is the strongest salt in the Hofmeister series of anions and it is known to stabilize the folded state of the drkN SH3 domain and folded states in general by promoting hydrophobic interactions [24]. Glycerol is also known to stabilize proteins and prevent aggregation by shifting the folded protein ensemble towards more compact states and by destabilizing aggregation-prone partially unfolded intermediates [197].

2.4.1 Chemical denaturants lead to a heterogeneous set of conformations that include a high FRET cluster

Chemical denaturants, such as GdmCl and urea, cause proteins to lose most of their secondary and tertiary structure and to approach a random-coil state [12, 198]. smFRET studies of globular proteins such as Rnase H [199], protein L [187] and the cold shock protein CspTm [187, 200], and of IDPs such as the human prothymosin α (ProTα) [200], α-Synuclein [123, 201], Tau [201], Troponin I (TnIc) [202] and the yeast Sic1 [192] show the appearance of extended chain conformations in GdmCl, which become increasingly broader, more populated
and more extended as the denaturant concentration increases. In view of these precedents, for the drkN SH3 domain, the population of the high-FRET peak was expected to gradually diminish in GdmCl, accompanied by an increase of the intermediate-FRET peak. This peak was also expected to gradually shift towards lower $E$ values and become broader as [GdmCl] increases and the protein unfolds.

The smFRET data obtained for drkN SH3 under denaturing conditions are shown in Figure 2-3. Quite surprisingly, for a marginally stable protein, the high-FRET (~100%) population, although diminished, did not completely disappear even in the harshest denaturant conditions (8 M GdmCl and in 7 M urea). The peaks of the smFRET histograms are typically approximated by Gaussian distributions and sometimes, near the limits of the FRET range, by beta distributions. However, the choices of fitting functions were found not to be critical for data interpretation [194]. The smFRET histograms in this study were all decomposed into a minimum number of Gaussians (see later).

![smFRET histograms measured at various concentrations of GdmCl and urea.](image)

**Figure 2-3.** smFRET histograms measured at various concentrations of GdmCl and urea. Each histogram was constructed from a dataset consisting of ca. 10,000 intensity bursts and it was fitted to a
sum of four Gaussians (three at no denaturant). The non-zero FRET peaks are denoted 1, 2 and 2’, respectively, e.g., in the 8 M GdmCl histogram. The total area under the FRET peaks, excluding the zero-FRET peak (donor-only, inactive and/or photobleached acceptor) was normalized to unity.

The denaturant dependence of the smFRET histograms points to the presence of heterogeneous drkN SH3 conformations, with at least two distinct non-zero FRET clusters present at all denaturant concentrations. We define a FRET cluster as a population of protein conformations of similar end-to-end distance, which gives rise to a distinct peak in the smFRET histogram data. The results of decomposing the two denaturation series into Gaussian peaks, including average FRET efficiency, width and population fraction for each of them, are given in Table 2-2.

Table 2-2. FRET efficiencies and fractional populations obtained by Gaussian decomposition of drkN SH3 smFRET histograms in denaturant (Figure 2-3). *  

<table>
<thead>
<tr>
<th>[GdmCl] (M)</th>
<th>cluster 1</th>
<th>cluster 2’</th>
<th>cluster 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>52.6 ± 21.5 (0.59)</td>
<td>-</td>
<td>98.5 ± 2.5 (0.41)</td>
</tr>
<tr>
<td>1</td>
<td>50.5 ± 28.6 (0.61)</td>
<td>94.1 ± 5.2 (0.22)</td>
<td>98.9 ± 2.3 (0.17)</td>
</tr>
<tr>
<td>2</td>
<td>40.4 ± 25.3 (0.58)</td>
<td>89.1 ± 7.9 (0.22)</td>
<td>97.5 ± 3.6 (0.21)</td>
</tr>
<tr>
<td>4</td>
<td>31.7 ± 18.8 (0.53)</td>
<td>90.5 ± 6.7 (0.21)</td>
<td>97.8 ± 2.9 (0.25)</td>
</tr>
<tr>
<td>6</td>
<td>17.8 ± 16.2 (0.62)</td>
<td>89.6 ± 6.8 (0.19)</td>
<td>96.5 ± 4.0 (0.19)</td>
</tr>
<tr>
<td>8</td>
<td>20.8 ± 17.9 (0.69)</td>
<td>88.4 ± 7.3 (0.15)</td>
<td>96.6 ± 3.8 (0.16)</td>
</tr>
<tr>
<td>[Urea] (M)</td>
<td>cluster 1</td>
<td>cluster 2’</td>
<td>cluster 2</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
<td>29.7 ± 19.5 (0.72)</td>
<td>92.8 ± 6.0 (0.12)</td>
<td>99.1 ± 2.4 (0.16)</td>
</tr>
<tr>
<td>3</td>
<td>19.8 ± 17.7 (0.80)</td>
<td>91.7 ± 7.1 (0.10)</td>
<td>99.1 ± 3.0 (0.10)</td>
</tr>
<tr>
<td>7</td>
<td>12.3 ± 11.9 (0.69)</td>
<td>91.7 ± 6.6 (0.18)</td>
<td>98.2 ± 3.1 (0.13)</td>
</tr>
</tbody>
</table>
As discussed above, in the absence of denaturant, two Gaussians are sufficient to describe the data, corresponding to the unfolded (cluster 1, $E = 50-55\%$) and folded (cluster 2, $E = 90-100\%$) sub-populations of this SH3 domain. When the denaturant concentration was increased, cluster 1 shifted to lower FRET values, with the center peak reaching 21\% in GdmCl and 12\% in urea, and it increased as a fraction of the total histogram. This dominant cluster was satisfactorily fitted by a single Gaussian for both denaturation agents at all measured concentrations. The broad distribution and the significant shift to lower FRET is consistent with the overall expansion of proteins upon denaturation, typically measured in smFRET studies [187, 192, 200]. The remarkable difference compared to other proteins is that the unfolded fraction is already populated at zero denaturant due to unusual energetics, $\Delta G \approx 1$ kcal/mol, [2]. GdmCl and urea will diminish the transient secondary and tertiary structure in this cluster to a point where it may approach a random-coil state (see also below).

In the absence of denaturation agents, the high FRET region of the histograms ($E > 80\%$) is attributed to the folded state, $F_{exch}$, which is consistent with the close proximity of the N- and C- termini in the drkN SH3 solution structure (PDB 2A37). When denaturant was added, the average FRET efficiency of this population decreased overall and the peak broadened asymmetrically towards lower FRET (longer distance) values. To account for the asymmetric broadening, another Gaussian component (cluster 2’) was used for fitting. Cluster 2’ shifts from $E = 94\%$ in 1M GdmCl to $E = 88\%$ in 8 M GdmCl and accounts for approximately half of the high-FRET population. A similar trend was observed in urea, although the peak shift was somewhat smaller than in GdmCl (Table 2-2).

Although fitting the high-FRET population at high denaturant by two Gaussians does not necessarily imply the existence of two sets of protein conformations that are significantly different, the procedure can nevertheless serves to quantify the asymmetric broadening of this cluster towards lower efficiency values. The total fraction of the high-FRET population decreases from 0.41 at zero denaturant to 0.31 at the highest denaturant concentrations, with a minimum of 0.2 in 3 M urea.

---

Data format: average FRET efficiency ± HWHM (half-width at half-maximum), in %. For each FRET peak (cluster), the fractional population is given between brackets.
Overall, urea seems to be more effective at denaturing drkN SH3 than GdmCl. This is not surprising, since urea is thought to be better than GdmCl at destabilizing beta sheets [203] and more efficiently solvate the protein backbone and the side chains with little or no restrictions due to excluded volume effects [204]. While urea accumulates in the first solvation shell, guanidinium displays a longer-range electrostatic effect that does not perturb the structure of the solvent close to the protein. On the other hand, the ionic nature of GdmCl leads to screening of all electrostatic interactions in a protein. Surprisingly, the high-FRET drkN SH3 population does not disappear completely in either GdmCl or urea, as typically seen in smFRET denaturation studies on other proteins [187, 200, 205].

Apart from the two major FRET clusters described above, the histograms show some population density at intermediate FRET values ($E = 50–80\%$), particularly at GdmCl and urea concentrations above 3 M (Figure 2-3). There is no distinguishable peak in this range and this density may be caused by inter-conversion between the two major clusters [194, 206] or by photobleaching-induced “bridging” effects [206].

**Figure 2-4.** (A) smFRET histogram of drkN SH3 in 6 M GdmCl, constructed from ca. $10^5$ intensity bursts. The histogram was fitted to four Gaussians, and normalized to the area excluding the zero-FRET peak. The parametric fitting results are listed in Table 2. The insert shows a comparison between the high-FRET peaks measured in Tris buffer (blue bars) and 6 M GdmCl (red bars). (B) FRET histograms of drkN SH3 in 20% (v/v) formamide (black), 4 M GdmCl (red), and 50% (v/v) DMSO (blue). DMSO data was acquired with a different donor dye, i.e., A555 instead of BFL.
The different FRET clusters, corresponding to different sub-ensembles of the drkN SH3 conformations, are better resolved in the smFRET histogram shown in Figure 2-4 A. This data was acquired at 6M GdmCl and it was constructed from ~100,000 intensity bursts. Utilizing a much larger number of single-molecule events allowed for finer binning along the FRET efficiency axis and provided a higher signal-to-noise ratio than the denaturation series data (Figure 2-3). The histogram was fitted to a sum of three Gaussians, excluding the donor-only zero-FRET peak (Table 2-1).

In the low-FRET region, a broad peak (cluster 1) covers nearly the entire range between $E = 0\%$ and $E = 50\%$. With an average FRET efficiency of 23% and a fractional population of 0.68; this cluster is the predominant conformational sub-ensemble at 6M GdmCl. Cluster 1 exhibits a similar denaturant dependence, i.e. peak broadening and shift to lower FRET (overall expansion), as the chemically-denatured states of other proteins, which are typically described by a random-coil polymer model.

About one third of the total population shows an anomalously high FRET efficiency at this high denaturant concentration. This corresponds to a mixed population of conformations, all of which have the two ends of the chain in close proximity to each other, and it is described by a sum of two Gaussians of nearly equal area, which are centered at $E = 91.3\%$ (cluster 2') and $E = 99.6\%$ (cluster 2). The high-FRET peak is broader and more asymmetric than the $F_{exch}$ peak observed in Tris buffer, as its variance increased from $\Delta E \approx 5\%$ in Tris to $\Delta E \approx 12\%$ in 6 M GdmCl (insert of Figure 2-4 A). Importantly, NMR and bulk fluorescence studies indicated that the drkN SH3 domain is nearly completely unfolded in 2 M GdmCl [25], thus ruling out the presence of folded state conformers at high denaturant concentrations.

Organic solvents are commonly used to change the native structure of macromolecules by disrupting the hydrophobic interactions between hydrophobic side chains of amino acids. In order to verify whether non-polar hydrophobic interactions could cause the opposite ends of the protein to “stick” to each other in high denaturant concentrations, smFRET measurements of drkN SH3 in the presence of DMSO and formamide were performed.
Formamide and DMSO are widely used polar solvents that precipitate, crystallize and denature proteins [207]. In Figure 2-4 B, the smFRET histograms measured in solutions containing either 4 M GdmCl, 20% formamide, or 50% DMSO are shown. Interestingly, the formamide diminished the high-FRET population nearly five times more efficiently than GdmCl. An even more pronounced change was recorded in DMSO, for which the high-FRET peak is absent. This points to groups at opposite ends of the chain being involved in interactions that are facilitated by guanidinium and urea, thus causing the persistence of a dynamic, high-FRET population of end-labelled drkN SH3 under denaturing conditions (see Discussion).

2.4.2 The effect of different dye pairs

The significant fraction of drkN SH3 molecules with high FRET under denaturing conditions is highly unusual and it could be due to “sticky” fluorescent labels promoting attractive interactions between the two ends of the protein. As a control, we replaced the donor dye, the neutral BFL, with the negatively charged dye A555 and used the same acceptor dye, A647 (Figure 2-5). According to the manufacturer, A555 and A647 are both negatively charged, with net charges of -4 and -3, respectively, and are therefore unlikely to cause attractive interactions between the two ends. On the contrary, this FRET pair exhibits repulsive electrostatics and it could destabilize possible attractive intra-protein interactions.

Figure 2-5 shows the smFRET histogram data obtained for drkN SH3 labelled with A555 and A647 at different GdmCl and urea concentrations (Figure 2-5). The new dye pair has a larger Förster radius, $R_0 = 5.9 \pm 0.2$ nm, and consequently the histogram peaks are shifted to higher FRET efficiencies compared to the data measured with the BFL-A647 pair. For instance, the peak centered at $E = 23\%$ in 6 M GdmCl shifts to a peak centered at $E = 51\%$ with the new donor-acceptor pair, and the two high-FRET sub-populations shift to the right and merge within the $E \approx 100\%$ peak.
Figure 2-5. smFRET denaturation series of drkN SH3 measured with the A555/A647 donor-acceptor pair in GdmCl and urea solutions; both dyes are negatively charged.

When adding denaturant, the mid-FRET peak corresponding to the unfolded $U_{exch}$ state shifts slightly to the right (smaller $R_{EE}$) for [GdmCl] < 1 M, and then to the left (longer $R_{EE}$) for [GdmCl] > 1 M. The initial phase is consistent with overall chain compaction caused by the screening of electrostatic repulsion between the two ends of the protein [192]. In the absence of denaturant, the fraction of the high-FRET folded state (~0.2) is considerably smaller than the value measured with the other FRET pair (0.41, Table 2-2). This is likely the effect of electrostatic repulsion between the donor and acceptor dyes destabilizing the folded $F_{exch}$ conformation, and it highlights the importance of selecting fluorescent dyes for smFRET experiments that do not perturb the conformational equilibrium of the labelled protein. As the GdmCl concentration is increased, the high-FRET fractional population decreases sharply to ~0.05 in 2 M GdmCl, followed by an increase to ~0.15 in 7 M GdmCl. The electrostatic screening is absent in urea and the repulsion between the two dyes drives the high-FRET
fractional population to a low value of ~0.05, which remains roughly constant for urea concentrations above 3 M.

Our smFRET data suggests that: (i) the observed residual high-FRET population of the drkN SH3 protein in high denaturant concentration is unlikely to be caused primarily by an attractive interaction between fluorescent labels, and (ii) GdmCl, and to a lesser extent urea, promote the formation of intra- or inter-molecular contacts between residues near the N- and the C- termini of drkN SH3, which are stable on the diffusion timescale, i.e. ~100 µs. Due to the shallow energy landscape of unfolded, denatured and other disordered states, their conformational ensembles can be highly sensitive to conditions and other factors. The smFRET data are thus consistent with interactions within the drkN SH3 domain synergizing with denaturants, and possibly dye interactions, to generate conformational preferences leading to high FRET.

2.4.3 The high FRET cluster in denaturants is not caused by aggregation

In principle, the formation of a significant population of “cross-braced” SH3 aggregates could explain the presence and the properties of the high-FRET sub-population observed in our denaturation studies. However, these smFRET experiments were performed with the protein diluted to very low (pM-nM) concentrations. In addition, denaturants are known to be quite effective in monomer recovery of pre-formed aggregates, especially at high denaturant concentrations and smaller initial aggregate size [208]. Disulfide-bonded aggregates are the exception [209], but the mutated drkN SH3 sequence has only two (inserted) terminal cysteines, which are coupled to the fluorescent dyes and therefore unavailable with disulfide bond formation.

The hydrodynamic radius ($R_H$) of the fully folded and the chemically-denatured drkN SH3 was measured by FCS (Figure 2-6) and co-diffusion fraction of singly labeled SH3-BFL
and SH3-A647 was measured in different solvent conditions by dcFCS (Figure 2-7). A single-point mutation of a threonine to a glycine (T22G) was shown to fully stabilize the folded state of the drkN SH3 domain [2]. The T22G mutant was labelled with an amine-reactive TMR dye and the $R_H$ was measured by FCS. A value of $R_{H,F} = 14.4$ Å was obtained (Table 2-3), which is similar to the value obtained by NMR, $R_{H,F} = 15.6$ Å [210] within the experimental error of both methods.

We further investigated the possibility of drkN SH3 aggregation using FCS. Since the average radius of folded proteins is roughly proportional to the cubic root of their molecular weight, dimerization will lead to an increase of $R_H$ by at least 25%. Larger aggregates will result in a more significant increase of the $R_H$. If GdmCl or urea facilitate the aggregation of the drkN SH3 protein, the hydrodynamic radius is expected to increase with the GdmCl (urea) concentration. However, the results show that the $R_H$ values obtained by fitting the FCS curves to eq. (2-1) are virtually invariant for concentrations of GdmCl ranging from 2 M to 7 M (see Table 2-4). This suggests that high GdmCl concentrations do not promote the aggregation of drkN SH3. Virtually the same $R_H$ values were obtained in solutions of 20% formamide and 50% DMSO, indicating the absence of SH3 aggregation under these conditions as well.

**Figure 2-6.** Autocorrelation (ACF) data (symbols) and fitting according to eq. (2-1) (solid curves) for: WT SH3 in Tris buffer (pH = 7.5) (black), T22G SH3 (fully folded) in Tris buffer (pH = 7.5) (blue), and WT SH3 in 6M GdmCl (red).
The fitting of the FCS data of the wild-type protein under denaturing conditions required only one diffusive species. The obtained hydrodynamic radius is thus an average value for the entire denatured ensemble, namely corresponding to all the smFRET peaks combined. The solvent viscosity was determined by FCS using a free dye solution, was used to correct the diffusion parameters obtained for the denatured protein. After correction, the estimated average hydrodynamic radius of the entire denatured ensemble was \( R_{H,D} = 17.8 \pm 0.4 \) Å (Table 2-4). The ratio between the hydrodynamic sizes of the denatured state and of the folded state of drkN SH3 is \( R_{H,D} / R_{H,F} = 1.24 \pm 0.03 \).

**Table 2-3.** Hydrodynamic radii \( (R_H) \), radii of gyration \( (R_G) \) and shape factors \( (\rho) \) for different conformation states of drkN SH3. \(^{a,b}\)

<table>
<thead>
<tr>
<th></th>
<th>Folded, ( F_{exch} ) (native)</th>
<th>Unfolded, ( U_{exch} ) (native)</th>
<th>Denatured, ( D ) (6 M GdmCl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_H(\text{Å}) )</td>
<td>14.4 ± 0.2</td>
<td>18.9 ± 0.3</td>
<td>17.8 ± 0.4</td>
</tr>
<tr>
<td>( R_G(\text{Å}) )</td>
<td>12.0 ± 0.9</td>
<td>18.8 ± 1.0</td>
<td>20.6 – 21.9 (^b)</td>
</tr>
<tr>
<td>( \rho = R_G / R_H )</td>
<td>0.83</td>
<td>~1.0</td>
<td>1.16 – 1.23</td>
</tr>
</tbody>
</table>

\(^a\) \( R_H \) values were estimate by fitting the autocorrelation curves (Fig. 2-6) to eq. (2-1).

\(^b\) \( R_G \) values of the denatured ensemble were estimated from smFRET data (Table 2-1) or were previously measured by SAXS [210].

**Table 2-4.** The hydrodynamic radius \( (R_H) \) of drkN SH3 in GdmCl, formamide, and DMSO solutions.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>( R_H(\text{Å}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 M GdmCl</td>
<td>17.8 ± 0.4</td>
</tr>
<tr>
<td>3 M GdmCl</td>
<td>17.9 ± 0.4</td>
</tr>
<tr>
<td>4 M GdmCl</td>
<td>17.8 ± 0.3</td>
</tr>
<tr>
<td>6 M GdmCl</td>
<td>17.8 ± 0.3</td>
</tr>
<tr>
<td>7 M GdmCl</td>
<td>17.7 ± 0.3</td>
</tr>
</tbody>
</table>
Dual-color FCS (dcFCS) is a powerful technique to analyze nucleic acid and ligand-receptor interactions [132]. Here, dcFCS experiments were employed to examine the possibility of dye-dye, dye-protein and protein-protein interactions that could lead to SH3 aggregation. A single-cysteine mutant, SH3-C2, was labeled with either the donor dye, BFL, or the acceptor dye, A647. The positive control sample is a 40-base-pair oligonucleotide sequence labeled with Cy5 at the 5’ end and 6-FAM at the 3’ end. The data shown in Figure 2-7 indicate that, in contrast to the 40 base pair DNA, the co-diffusion of SH3-BFL and SH3-A647 is virtually inexistent in all solvents: Tris buffer (50 mM Tris, 150 mM NaCl, pH 7.5), 6 M GdmCl, 6 M urea and 50% v/v DMSO. This dcFCS data shows that there were no interactions between dye-labeled drkN SH3 monomers which were exposed to the same sample conditions as those during the labelling samples for smFRET experiments and thus rules out the presence of aggregates of drkN SH3 in high denaturant conditions.
Figure 2-7. Auto-correlation curves for blue (blue circles) and red (red circles) detector signals from BFL/6-FAM and A647/Cy5, respectively, and cross-correlation curve (black circles). Autocorrelation (ACF) and cross-correlation (CCF) data obtained for an equimolar mixture of SH3-C2-BFL and SH3-C2-A647 in a 6 M GdmCl solution. The samples are: (A) 40 base pair DNA conjugated with Cy5 and 6-FAM at 5’ and 3’, respectively in TE buffer (10mM Tris, pH 8.0), (B) BFL and A647 dye in Tris buffer (50 mM Tris, 150 mM NaCl, pH 7.5); SH3-C2 protein singularly labeled with BFL and A647 in (C) Tris buffer, (D) 6M GdmCl, (E) 6M urea, and (F) 50% DMSO (v/v).

2.5 Discussion

The chemically denatured state of proteins is generally described by a random-coil model [198] and it is often used as a proxy for the unfolded state ensemble. Accordingly, the high FRET (small $R_{EE}$) population that corresponds to the folded state ($F_{exch}$) should vanish under strong denaturing conditions. However, a sizeable high-FRET fraction of drkN SH3 persisted in very high concentrations of denaturant, in contrast to the typical chemical denaturation behavior of proteins observed in smFRET studies [186-188, 211]. Of potential relevance, NMR data points to the existence of significant transient structure in chemically denatured proteins [174-176], while computational studies emphasized the importance of non-native and non-local interactions along the folding pathways [179, 212, 213]. Note that the high-FRET peak in the presence of high [GdmCl] or [urea] is significantly broader than in Tris buffer (insert, Fig. 3), thus suggesting that it was not caused by the persistence of the folded $F_{exch}$ state, which gave rise to a narrow FRET peak at $E = 98\%$ (Figure 2-2). Moreover, previous NMR studies showed that drkN SH3 is nearly completely denatured at 2 M GdmCl [184, 214]. On the other hand, control dcFCS measurements performed at high denaturant concentrations ruled out protein oligomerization as a possible cause for this effect (Figure 2-7). What is then the physical origin of the significantly populated (20-35%), unusually high FRET ($E > 90\%$) cluster observed for drkN SH3 in urea and GdmCl?

Molar-range concentrations of denaturants are typically “good solvents” for proteins [198], and denatured proteins are often described as freely-jointed polymers [215]. Thus,
smFRET data on unfolded proteins are typically interpreted using homopolymer models that are either the Gaussian chain [200] or variants of the Sanchez model [216]. Recently, our groups developed a new method to infer conformational properties of heterogeneous disordered proteins from smFRET data using a coarse-grained SAW polymer model with physical excluded volume [190]. This sub-ensemble SAW-based inference was applied to an IDP protein, Sic1 from yeast, which showed distinct coexisting smFRET sub-populations [192]. Derived Sic1 conformational parameters, such as the hydrodynamic radius and the radius of gyration were consistent with those obtained in NMR and SAXS experiments. Here, we applied the same SAW framework to infer most probable $R_G$ values for each FRET cluster observed in the presence or in the absence of GdmCl (Table 2-5). Population-weighted $R_G$ averages were calculated according to eq. (2-5) and compared to results of SAXS experiments. To that end, different assumptions regarding the conformations responsible for the high-FRET cluster in 6 M GdmCl were considered (Table 2-5).

### Table 2-5. Radii of gyration ($R_G$, in Å) of different conformations of drkN SH3 estimated from the smFRET data or from previous ensemble SAXS data.

<table>
<thead>
<tr>
<th></th>
<th>FOLDED, $F_{exch}$</th>
<th>UNFOLDED, $U_{exch}$</th>
<th>DENATURED RANDOM COIL</th>
<th>DENATURED LOOPED</th>
<th>$&lt;R_G&gt;_{FRET}$</th>
<th>$&lt;R_G&gt;_{SAXS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRIS BUFFER</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PH 7.5)</td>
<td>12.0 ± 0.9$^a$</td>
<td>18.8 ± 1.0</td>
<td>-</td>
<td>-</td>
<td>15.5 ± 0.8</td>
<td>14.9 ± 0.5$^b$</td>
</tr>
<tr>
<td><strong>6 M GDMCL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.0 ± 0.9$^a$</td>
<td>-</td>
<td>22.5 ± 1.0</td>
<td>19.7 ± 0.8</td>
<td>21.9 ± 0.5$^b$</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ $R_G$ of the folded state was measured by SAXS [210] and adjusted for the presence of dye linkers.

$^b$ Ensemble average $R_G$ was measured by SAXS in Tris buffer [183, 210] and in 2M GdmCl [210].

$^c$ Population-weighted average $R_G$ calculated using eq. (2-5) with the fractions obtained by fitting the smFRET histograms (Table 2-1).

Under non-denaturing conditions, drkN SH3 coexists in two different states, a folded ($F_{exch}$) and an unfolded ($U_{exch}$) state, as demonstrated here (Figure 2-2) and in previous NMR studies [196]. Choy et al used sodium sulfate to stabilize the folded state and measured $R_{GF} = \ldots$
11.9 ± 0.5 Å using SAXS [210]. The presence of dye linkers at the two ends of the chain were found to be equivalent to adding between 4 and 16 residues [217], which results in the $R_{G,F}$ of the protein-dye construct being adjusted to 12.0 ± 0.9 Å.

Previous experimental studies and ENSEMBLE calculations were performed for the $U_{\text{exch}}$ state [218]. This work primarily probes chemically denatured states. Chemical denaturants significantly affect the $U_{\text{exch}}$ state, based on previously published results [2], and it is also likely that chemical denaturants affect the folded state, but the current data do not offer strong additional insights into either of these states. Nevertheless, For the $U_{\text{exch}}$ state, the SAW model estimated a radius of gyration $R_{G,F} = 18.9 ± 0.5$ Å based on an average FRET efficiency of 55%, and an average end-to-end separation of 43.0 ± 0.9 Å, which is very similar to the value obtained from the distribution of unfolded ENSEMBLE conformations [218] (Figure 2-8). The FRET population-weighted average $R_G$ was estimated to be 15.6 ± 0.8 Å (Table 2-3), which is close to the SAXS value for the drkN SH3 ensemble, i.e., $R_G = 14.9 ± 0.5$ Å [183]. The small difference is mostly due to the presence of dyes and linkers in the fluorescence experiments, which are expected to lead to a small increase the overall size of the chain. Another contributing factor is the relatively compact nature of the $U_{\text{exch}}$ state, which contains significant fluctuating structure and thus deviates from a random chain with excluded volume (see below). However, given the inherent limitations of the polymer model used for $R_G$ inference, the reasonable agreement between the smFRET and SAXS values is encouraging.
Figure 2-8. The distribution of end-to-end distance ($R_{EE}$) in the native drkN SH3 unfolded state, $U_{exch}$, generated from 1700 unfolded conformers obtained from the protein ensemble database ([183] and http://pedb.vib.be). The average separation distance obtained by Gaussian fitting (blue line) was $<R_{EE}>_U = 44.2 \pm 0.6 \, \text{Å}$. The vertical red bar indicates the value estimated by smFRET data (Table 2) and SAW model, $<R_{EE}>_U = 43.0 \pm 0.9 \, \text{Å}$.

The shape factor, defined as the ratio between the radius of gyration and the hydrodynamic radius ($\rho = R_G/R_H$), is commonly used to define the average conformational shape of a polymer. The shape factor $\rho_F$ of the drkN SH3 folded state was estimated to be 0.83 using the FCS-measured $R_{H,F}$ (Table 4) and 0.77 using the NMR-measured $R_{H,F}$ [210]. Both values are similar to the theoretical shape factor predicted for globular proteins, i.e. $\rho = (3/5)^{0.5} \approx 0.775$ [216, 219], and to experimental results on other proteins, e.g., $\rho \approx 0.75$ for the HIV-1 capsid protein [220].

The shape factor of the native unfolded drkN SH3 state ($U_{exch}$) was estimated to be around 1.0 (Table 4). Note that the $R_{H,U}$ is ~30% larger than $R_{H,F}$, in excellent agreement with previous NMR data [210]. As mentioned above, the FRET-inferred radius of gyration for $U_{exch}$ is slightly overestimated because this state is not a random chain. This implies that the actual value of $\rho_U$ is lower, probably between 0.775 and 1.0, which is consistent with an expanded protein conformation with sizeable fluctuating structure.

The smFRET histograms at high denaturant concentrations do not show a single uniform peak (Figure 2-3), which indicates that the denatured state ensemble is heterogeneous. In order to understand the nature of this heterogeneity, we applied the sub-ensemble FRET inference based on the SAW polymer model to estimate the most probable $R_G$ for each FRET peak of end-labelled drkN SH3 at 6 M GdmCl (Figure 2-4). Different premises about the nature of the protein conformations in the high-FRET cluster were used. As validation, the population-weighted average $R_G$ values, which were calculated based on the smFRET data (Table 2-1) and using eq. (2-5), were compared to the SAXS-measured $R_G$ (Table 2-3). In addition, the sub-ensemble $R_G$ values and the FCS-measured $R_H$ values were used to derive
shape factors for the denatured ensemble and for the residual high-FRET state, and thus help distinguish between different hypotheses (Table 2-5).

The predominant FRET population, i.e., cluster 1 in Figure 2-4, was modelled as a random self-avoiding chain with 66 residues, i.e., 60 amino acid residues between the labelling positions and 3 equivalent residues at each end accounting for the dye linkers. By imposing that $E_{\text{sim}}$ matches $E_{\text{exp}} = 23\%$, the radius of gyration of this random-coil sub-ensemble was estimated to be $R_{G,RC} = 22.5 \pm 1.0 \, \text{Å}$ (Table 2-5). The corresponding value obtained with the classical Gaussian chain model, which does not account for excluded volume, is much larger, i.e., $30.6 \pm 1.0 \, \text{Å}$. The expected value for a 66-residue chain based on the scaling relation for fully unfolded proteins [221] is $23.6 \pm 4.3 \, \text{Å}$, which is similar to the SAW smFRET inference. Moreover, since the Cα–Cα virtual bond length is $\approx 0.38 \, \text{nm}$, the Gaussian-inferred $R_G$ value translates into an inferred Kuhn length of $\sim 6$ residues, which is clearly unrealistic.

Since (dc)FCS controls ruled out the presence of aggregates in our single-molecule experiments, intermolecular FRET cannot cause the high-FRET peak at high denaturant (GdmCl and urea) concentrations. It is therefore reasonable to assume that this sub-ensemble (cluster 2 and 2’ in Figure 2-4) could arise from either a residual population of the folded state, $F_{\text{exch}}$, or a disordered looped state, $L$. The folded state is rather compact, i.e., $R_{G,F} = 12.0 \, \text{Å}$, and using it in eq. (2-5) to calculate the population-average radius of gyration of the denatured state, $R_{G,D}$, yielded a value of $19.7 \, \text{Å}$, which is more than $2 \, \text{Å}$ smaller than the SAXS value, $21.9 \, \text{Å}$ [210].

We next considered the alternate possibility, which is a looped ($L$) state in which the N- and C- termini are in close proximity. The simulations of the $L$ state were performed with an $n = 66$ SAW polymer, modified so that an attractive potential enforces two selected residues to be on average $5.7 \, \text{Å}$ apart. The “sticky” residues were chosen to be $l$ residues away from each end of the chain. Simulations were performed for $l = 5, 10$ and $13$, resulting in three $P(R_{EE}|l)$, which are shown in Figure 2-9. In each case the radius of gyration is $R_{G,L} = 15.8 \pm 0.2 \, \text{Å}$. Using the simulated $R_{EE}$ distribution $P(R_{EE}|l)$, we estimated the average FRET efficiency $\langle E \rangle(l)$ and compared it to the experimental values for the peaks 2 and 2’ from Table
2-1. The $\langle E \rangle = 91.3\%$ peak can be assigned to looped conformations with $l \approx 10$, i.e., a total of ~20 residues, are left dangling at the two ends. The $\langle E \rangle = 99.6\%$ peak can be assigned to a loop with $l \leq 5$, i.e. a total of less than 10 residues are left dangling at the two ends. Here these simulated $L$ states are used as putative sub-ensembles that contribute to the compact portion of the drkN SH3 unfolded ensemble at 6 M GdmCl. Based on Figure 2-9 B, we can infer that if the high-FRET population of drkN SH3 observed in 6M GdmCl (Figure 2-4 A, Table 2-1) corresponds to looped conformations, then $l < 10$ residues.

![Figure 2-9](image_url)

**Figure 2-9.** (A) The distribution of end-to-end distance of looped SAW chains with different dangling segment lengths at each end: $l = 5$ (red), 10 (blue), and 13 (green). (B) Average FRET efficiency ($\langle E \rangle$) estimated using eq. (2-4) for SAW looped chains with dangling segment lengths at each end of $l = 5$, 10 and 13. The solid line is a cubic spline interpolation between the three data points.

Choy et al [210] used 0.4 M sodium sulfate to stabilize the folded state of the drkN SH3 domain and obtained a radius of gyration of 11.9 ± 0.5 Å using SAXS. Hofmann et al. [222] estimated the increase in $R_G$ due to the fluorophore linker by assuming that the volume of gyration of the protein-dye construct is the sum of the individual radii of chain and dye ($V_G = V_{G,\text{protein}} + V_{G,\text{linker}}$).

$$R_G = \left[ R_{G,\text{protein}}^3 + R_{G,\text{linker}}^3 \right]^{1/3}. \quad (2-6)$$

Considering the linker as being approximately equivalent to a random walk of 4 to 16 residues [217], the $R_{G,\text{linker}}$ is between 3.1 and 6.2 Å. This results in the radius of gyration of the protein-dye construct to be between 11.47 Å (for the minimum SAXS derived and linker estimated $R_G$) and 12.89 Å (for the maximum SAXS derived and linker estimated $R_G$).
Alternatively one could estimate the radius of gyration of the folded or most compact state using (2-7)

\[ R_G = \left( \frac{3(N+1)v}{4\pi} \right)^{1/3} . \] (2-7)

Using \( N = 59 \) as the number of peptide bonds in the (unlabeled) protein and the weighted mean volume \( v \) of an amino acid (0.13 nm\(^3\)). This results in an \( R_G \) of 10.8 Å, approximately 8% smaller than the \( R_G \) measured by SAXS. By assuming the effect of the linkers is to add an offset of between 4 and 16 residues [217] the resulting \( R_G \) calculated with \( N \) between 63 and 75 will be between 11.1 and 11.7 Å. Based on these considerations, we assigned a range 11.1 to 12.9 Å, or approximately 12.0 ± 0.9 Å for the dual-labelled \( F_{\text{exch}} \) state of drkN SH3.

The \( R_G \) values of both looped states were estimated using the SAW formalism and were found to be almost identical, i.e., \( R_{GL} = 15.8 \pm 0.2 \) Å. Using this value in eq. (2-5) we obtained a population-average \( R_{GD} \) value of 20.6 ± 0.8 Å for the denatured ensemble (Table 2-3). This \( R_G \) value is closer, although not identical to the SAXS value, which was in fact measured at a different GdmCl concentration (2 M instead of 6 M). This, corroborated with the observed peak broadening (insert Figure 2-4 A), suggests that a dynamic looped conformational sub-ensemble most likely causes the appearance of the high-FRET peak at high urea and GdmCl concentrations.

The shape factor of the denatured drkN SH3 ensemble, \( \rho_D \), was calculated using the FCS-measured \( R_H \). We obtained a \( \rho_D \) value between 1.16 and 1.23 (Table 2-3), depending on whether the FRET or the SAXS \( R_G \) value was used. For a linear Gaussian chain, i.e., a polymer in a theta solvent, a value \( \rho = 8/3\pi^{\frac{1}{2}} \approx 1.51 \) was predicted by Tanford following the Zimm model [223]. This is the theoretical limit for \( n \to \infty \), however for \( n = 66 \) our coarse-grained simulations gave \( \rho = 1.25 \) for a Gaussian chain and \( \rho = 1.2 \) for a SAW chain. Coincidently, for a Gaussian ring polymer in a theta solvent the shape factor is \( \rho = \left( \frac{\pi}{2} \right)^{1/2} \approx 1.25 \) [224].
A similar shape factor, i.e., $\rho \approx 1.2$, was estimated for the denatured IgG domain of protein L using SMF methods [225]. In addition, $\rho \approx 1.06$ was found for several denatured polypeptides based on NMR and SAXS measurements [219]. This was interpreted as the proteins retaining some fluctuating structure even at high denaturant and thus being more compact than a random chain, or as an increase in $R_H$ due to binding of GdmCl or a change in the bound water layer. NOE and NMR chemical shift experiments suggested that drkN SH3 is almost completely denatured at 2 M GdmCl [184, 214]. It is thus plausible that the high-FRET drkN SH3 population observed in urea and GdmCl consists of solvated, disordered conformations, but which have some denaturant-mediated interactions between the N- and C-terminal regions.

The underlying mechanisms of protein destabilization by denaturants such as guanidinium salts and urea are not completely understood [203, 226]. What type of interactions could possibly lead to two ends of the drkN SH3 domain coming close together in high denaturant GdmCl and urea solution? Thirumalai and co-workers [204] used molecular dynamics simulations to estimate the interactions between hydrophobic and ionic solutes in aqueous GdmCl and urea solutions. They proposed that both denaturant solvents only mildly alter hydrophobic association, while they dramatically change electrostatic interactions by solvating the charged residues or by forming hydrogen bonds with the peptide backbone. Ramsden and co-workers [227] proposed that GdmCl and urea can tune the water structure to some extent to either increase or decrease the strength of the hydrophobic interactions. Godawat et al. [228] found evidence that small hydrophobic pairs are stabilized while large hydrophobic pairs are destabilized by adding the denaturant, while similar findings were reported recently by Khajehpour [229] based on fluorescence quenching experiments.

The high-FRET peak is absent in a 50% DMSO solution, suggesting that hydrophobic interactions plays an important role in the formation of the denatured drkN SH3 “looped” state. Note that nearly half of the amino acids in drkN SH3 sequence are non-polar. Contacts involved in the “loop” formation were found to be located within the terminal 5-10 residues at each end and probably involve aromatic residues and residues with non-polar side chains. Similarly, Chan and co-workers [212] reported that hydrophobic interactions in denatured Fyn
SH3 lead to a short-lived transient looped folding intermediate. In the case of drkN SH3 the looped state seems to be more stable, since the high-FRET peak appears as a distinct feature in the smFRET histogram, implying that the loop opening-closing rate is slower than the average diffusion time of the protein through the detection volume (~100 μs).

In a recent smFRET study, Muñoz and coworkers resolved two unfolded conformations of the chicken α-spectrin SH3 domain upon chemical denaturation [205]. Their results showed that a significant fractional population remains in the high FRET region at molar-range concentrations of urea and GdmCl. They assigned the denatured high-FRET population to the native folded conformation which gradually expands and becomes increasingly disordered due to breaking of intramolecular hydrogen bonds. In other smFRET studies of protein folding, the folded-state high-FRET peak decreases gradually and disappears upon adding denaturation agents [187, 199, 200]. In this connection, it is noteworthy that fluorescence decay analysis by Ittah and Haas suggested a prominent looped state in the early stage of the folding kinetics of the 58-residue bovine pancreatic trypsin inhibitor when nonlocal interactions bring the two chain ends of the protein to close proximity [230]. To our knowledge, the current study is the first time that single molecule experimental evidence of protein loop formation under denaturing conditions has been reported.

2.6 Conclusions

SMF spectroscopy experiments were employed to study the conformations of the drkN SH3 domain associated with the unfolded and the chemically denatured states. Under non-denaturing condition, drkN SH3 coexists in both folded and unfolded states and our FRET population fraction agrees with previous values measured by NMR. Sodium sulfate and glycerol were found to stabilize the folded state of drkN SH3 via increased hydrophobic interactions. When exposed to urea and GdmCl denaturants, the protein still exhibits two distinct FRET populations. The dominant conformation is expanded and shows a low FRET efficiency as expected. Surprisingly, the high-FRET (~100%) population did not disappear
even in the harshest GdmCl and urea conditions. $R_H$ values and zero co-diffusion fraction of drkN SH3 monomers suggest that GdmCl/urea do not promote the aggregation of drkN SH3. Different dye pairs were used to label the protein to rule out the possibility that the high FRET population is caused by the dye artifacts.

The high-FRET population absent in DMSO indicates that interaction between both termini is facilitated by GdmCl and urea, thus causing the high-FRET population perseveres under denaturant conditions and forms loop-like state. Although the underlying mechanism of loop formation remains unclear, FRET experiments in formamide and DMSO suggest that interactions between hydrophobic groups in the distal regions could contribute to the formation of the looped state. The looped state in drkN SH3 appears to be stable under the molecule diffusion time through the focal volume.

The size and shape of the drkN SH3 was interpreted using a coarse-grained, sub-ensemble based SAW polymer model that accounts for excluded volume. The model was applied to the sub-ensemble FRET data to estimate the most probable $R_G$ for each FRET peak of drkN SH3. For the unfolded state, the SAW model estimated $R_G$ is very similar to the value obtained from the distribution of unfolded ENSEMBLE conformations. The FRET population-weighted average $R_G$ in the presence of GdmCl is also close to the value obtained by SAXS. The SAW model infers that looped conformation is caused by the residues that are less than 10 amino acids away from the termini.

The underlying mechanism of looping and the inter-conversion rates between closed and open states could be further studied by single-molecule spectroscopy on various mutants targeting the end regions of the drkN SH3 sequence. Other osmolytes could be used to perturb intramolecular interactions in order to investigate the contribution of hydrogen bonds, non-polar contacts and salt bridges in the association of specific protein segments. It is worth pointing out that contacts between the termini are absent in the unfolded ensemble under non-denaturing condition. The loops are formed in the presence of denaturants and stabilized by denaturants, thus warning of more caution when using chemical denaturation to access the unfolded ensemble. Our findings also point to a potential synergy between the denaturants, the
intrinsic intraprotein interactions and potential dye interactions. In addition, heterogeneous denatured states are consistent with the idea that the free energy landscape of disordered/unfolded states is very malleable and underlying conformational biases could be amplified by certain chemical denaturants.
Chapter 3
Single-Molecule Dissection of the Conformations, Dynamics and Binding of the Disordered 4E-BP2 Protein

3.1 Abstract

Intrinsically disordered proteins (IDPs) play critical roles in regulatory protein interactions. Cap-dependent translation initiation is regulated by the interaction of eukaryotic initiation factor 4E (eIF4E) with disordered eIF4E binding proteins (4E-BPs) in a phosphorylation-dependent manner. Single molecule fluorescence resonance energy transfer (smFRET), fluorescence correlation spectroscopy (FCS), and fluorescence anisotropy decay (FAD) were used to detect and assess the structural changes and sequence-specific local chain motions of 4E-BP2 upon phosphorylation and upon binding to eIF4E. smFRET analysis reveals changes in the conformational ensemble responding to phosphorylation, denaturation, salt and pH modulation. It is shown that both hydrophobic and electrostatic effects play essential roles for the conformational landscape of IDPs. Polymer theory was applied to quantitatively describe and infer the chain dimensions from the experimental smFRET data. Nanosecond-scale dynamics in 4E-BP2 were observed by site-specific FCS, and were tentatively assigned to the transient chain contact formation leading to quenching of the fluorophore. Our data suggests that multi-site phosphorylation of 4E-BP2 slows down the proximal chain motions and can, in some case, influence distal regions. Segmental rotational correlation times measured by FAD at different sites along the chain provide a rough rigidity map of 4E-BP2 and can be used to evaluate its binding mode to eIF4E.

3.2 Introduction

Intrinsically disordered proteins (IDPs) are a class of proteins that lack well-defined three dimensional structures while still carrying out biological functions. [26-28]. IDPs play a
crucial role in mediating interactions with multiple partners and often function as protein interaction hubs [46, 47]. The protein complexes can undergo disorder to order transitions or remain dynamic [29]. The lack of stable folded structures observed in IDPs leads to the false argument that IDPs are (quasi-)similar to random coils. However, the transient secondary and tertiary structures, electrostatic interactions, and backbone torsion angle propensities make IDPs exhibit a wide range of compactness based on their amino acid sequence [26, 29, 35, 36].

Post-translational modifications (PTM), such as phosphorylation, acetylation, ubiquitination, etc. have been shown to induce significant structural reorganization of the IDPs and could even cause disorder-to-order transitions [29, 50, 51]. The change in the structure preference also regulates protein flexibility and protein-protein interactions [51-53]. Moreover, PTM highly relates to numerous developmental disorders and diseases [231] and plays a crucial role in cellular signaling and regulatory processes including metabolism [232, 233], cellular maintenance and regulation of cellular pluripotency [231, 234].

In the nervous system, translational control is compulsory for neurodegeneration [64], metabolic disorders, cellular transformation and cancer [65]. Cap-dependent initiation of translation is regulated by the interaction of eukaryotic initiation factor 4E (eIF4E) with disordered eIF4E binding proteins (4E-BPs) in a phosphorylation-dependent manner [53, 235, 236]. The simplified translational initiation scheme is shown in Figure 3-1. The eIF4E is a 24 kDa mRNA 5’ cap binding protein, which has been shown to be an oncogene and be involved in the induction of cellular transformation [237, 238]. The eIF4G, a scaffolding protein, plays a crucial role in docking several units of the translation initiation machinery, which allows them to assemble at 5’ cap of mRNA and recruit the ribosome [239].

The interaction between eIF4E and eIF4G is the fundamental step which starts the translation process. This interaction involves the canonical binding region YxxxxLϕ of 4E-BP2, where x and ϕ represent any amino acid and a hydrophobic amino acid, respectively. It is noted that this canonical binding motif is shared with eIF4G and is common to all 4E-BPs. Thus, the 4E-BP proteins regulate the translation by preventing the assembly of the eIF4F complex (consisting of eIF4E, eIF4G, and eIF4A as shown in Fig. 3-1) and the subsequent
ribosome recruitment to the mRNA. When 4E-BP2 is in non-phosphorylated or hypo-phosphorylated states (only one phosphorylation site is phosphorylated), it competes with eIF4G to bind eIF4E, whereas in the hyper-phosphorylated state (two (Thr 37 and Thr 46) or more than two sites get phosphorylated) it loses its binding affinity to eIF4E and allows eIF4G to bind eIF4E and initiate the translation [235]. Thus, (de)phosphorylation serves as an on-and-off switch that modulates cap-dependent translation initiation in eukaryotic cells.

**Figure 3-1.** Simplified interaction and phosphorylation scheme of 4E-BP2 in eukaryotic cells (figure courtesy of Sabelo Lukhele from Dr. Forman-Kay’s group [240], reproduced with permission.). 4E-BP2 shares the canonical eIF4E binding motif (YxxxxLɸ) with eIF4G. Non-phosphorylated 4E-BP2 outcompetes eIF4G for binding to eIF4E and it inhibits the translation. The phosphorylation of 4E-BP2 is mediated by mTOR with scaffold protein Raptor. The binding affinity of the hyper-phosphorylated 4E-BP2 is very low, so that eIF4G binds to eIF4E allowing the translation to start.

The 4E-BP2 protein plays an important role in controlling cell growth and proliferation via regulating mRNA translation [241] and is highly involved in immunity to viral infections [242]. The neural 4E-BP2 also functions in regulating synaptic plasticity, essential for learning, memory and autism spectrum disorder [66, 243]. Binding of IDPs often leads to ordering, yet the 4E-BP2/eIF4E complex has been shown to be highly dynamic with an exchanging bipartite interface [67]. A recent nuclear magnetic resonance (NMR) study reported that phosphorylation of 4E-BP2 reduces eIF4E binding affinity by inducing a folded beta sheet structure for residues 18-62 with long-range stabilizing interactions to residues outside the folded region [53]. However, important details of the structure dynamics of 4E-BP2 remain unknown, which prevent a clear mechanistic picture of the regulatory function of the eIF4E:4E-BP2 complex.
Single-molecule fluorescence (SMF) techniques were applied previously to measure the distribution of conformations, the peptide chain dimensions and dynamics in IDPs [123, 200, 222, 244]. Here, we used a combination of fluorescence spectroscopy techniques (FCS and FAD) to characterize the conformational changes and sequence-specific local chain motions of 4E-BP2 upon phosphorylation, denaturation, and upon binding to eIF4E. smFRET was used to measure intra-molecular distances at single-molecule level and to study changes in the conformational ensemble of 4E-BP2 as a function of phosphorylation, denaturation, salt and pH. Based on the Sanchez polymer theory [245], the radius of gyration ($R_g$) and the average mean squared end-to-end distance $1 \langle R_{EE}^2 \rangle$ were calculated from the smFRET data.

Typically, IDPs have a wide range of interchanging conformations, so that the dynamics information is important for their biological functions. The time scale of conformational dynamics in burst-mode SMF has an upper limit determined by the passage time of a protein through the focal volume (typically ~1 ms). On the other hand, FCS and FAD are sensitive to chain motions on the nanosecond-to-microsecond time scale, which is highly relevant for protein folding and for IDPs [246]. FCS was used to investigate intra-molecular kinetics in 4E-BP2 under non-phosphorylated (NP), hyper-phosphorylated (HP) 2 and denaturing conditions. In addition, FCS provided estimates for the inter-molecular contact rate involved in binding to eIF4E. The local, segmental flexibility of the peptide chain was measured by FAD in various sites of the protein. A coarse rigidity map of the 4E-BP2 was obtained in the NP and HP states, which was used to assess the mechanism of the binding to eIF4E.

---

1 The end-to-end distance refers to distance between two FRET probes.

2 Here, HP refers to as fully phosphorylated 4E-BP2.
3.3 Materials and Methods

3.3.1 Materials

The fluorescent probes used for labelling the 4E-BP2 protein for SMF experiments were: Alexa 488 (A488) maleimide, Alexa 647 (A647) maleimide (ThermoFisher Scientific, Canada) and Atto 488 (At488) maleimide (ATTO-TEC GmbH, Germany). Guanidinium chloride (GdmCl) (G9284, Sigma Aldrich) and urea (EM-9510, EMD Millipore) were used for protein denaturation. Immediately before use, urea was purified from hydrolysis products and ionic impurities as described previously [191]. GdmCl and urea solutions were adjusted to pH 7.4 for all the experiments unless stated otherwise. Each GdmCl and urea concentration was also confirmed by measuring the solution viscosity by FCS. All samples were diluted in phosphate-buffered saline (PBS) (pH 7.4) unless stated otherwise.

3.3.2 Protein expression and purification

Plasmid encoding the 4E-BP2 protein (residues 1 through 120) and eIF4E protein (residues 1 through 217) with kanamycin and chloramphenicol resistance were transformed into BL21-codonplus (DE3)-RIL competent E. coli cells. Bacterial cultures were grown overnight on Lennox Lysogeny broth (LB) (L7658, Sigma-Aldrich) – kanamycin/chloramphenicol Agar plate in a 37°C incubator. A single colony was inoculated into LB medium with 100 mg/L final concentration of kanamycin (KB0286, Bio Basic Inc., Canada) and chloramphenicol (CB0118, Bio Basic Inc., Canada) and shaken in a 37°C incubator for ~5 hours until cell density reached an OD600 of 0.9. A 1 mM final concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG) (IB0168, Bio Basic Inc. Canada) was then added and the medium was transferred to a 16°C incubator for overnight growth up to a cell density having an OD600 of around 1.3. Cells were lysed by sonication in a buffer containing 50 mM Tris-HCl, 300 mM NaCl, 2 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (646547, Sigma-Aldrich), 1 mM Ethylenediaminetetraacetic acid (EDTA) (EDT111, BioShop,
Canada), deoxyribonuclease I (DNASE1) and protease Inhibitor cocktail cOmplete (Roche Diagnostics, Laval, Quebec, Canada), pH 7.5. The 4E-BP2 protein was purified on a Ni column (17-5248-01 HisTrap HP, GE Healthcare) and an ion-exchange column (17-1153-01 HiTrap Q HP, GE Healthcare) with a linear gradient of NaCl (0-1 M) followed by HiLoad Superdex 75 PG gel filtration column (28-9893-34, GE Healthcare) in 50 mM Tris, 150 mM NaCl, 2 mM TCEP, 1 mM EDTA, 2 mM Benzamidine-HCl, pH 7.5. All the cysteine mutants were expressed and purified in a similar manner. The identity and the purity of the all protein samples were assessed by mass spectrometry.

Plasmid encoding the Mitogen-activated protein kinases (MAPK) with chloramphenicol resistance was transformed into BL21-codonplus (DE3)-RIL competent E. coli cells. Bacterial cultures were grown overnight on LB–chloramphenicol Agar plate in a 37°C incubator. A single colony was inoculated into LB medium with 100 mg/L final concentration of chloramphenicol and shaken in a 37°C incubator for ~4 hours until cell density reached an OD$$_{600}$$ of 0.7. 0.4 mM final concentration of IPTG was then added and the medium was transferred to a 16°C incubator for ~3 hours. Cells were lysed by sonication in a buffer containing 50 mM Tris-HCl, 300mM NaCl, 2 mM TCEP, 1 mM EDTA, DNASE1, protease Inhibitor cocktail cOmplete, pH 7.5. The MAPK protein was purified on a Nickle column and an ion-exchange column with a linear gradient of NaCl (0-1 M) followed by HiLoad Superdex 75 PG gel filtration column in 100 mM HEPES, 200 mM NaCl, 2 mM Dithiothreitol (DTT) (DB0058 Bio Basic, Canada), 0.2 mM EDTA, pH 7.5. The buffer used for phosphorylation is 50mM Tris-HCl, 20 mM MgCl$_2$, 1 mM EGTA, 2 mM DTT, 10mM ATP, pH 7.5 and the 4E-BP2:MAPK concentration ratio is 4:1. Phosphorylation takes overnight and the phosphorylation level was checked by mass spectrometry. The storage buffer is 50 mM HEPES, 100 mM NaCl, 1 mM DTT, 50% glycerol, 0.1 mM EDTA, 0.01% Brij 35(20150, Thermo Fisher Scientific), pH 7.5.
3.3.3 Sample preparation

All single cysteine proteins (C0/C35S/C73S (cysteine insertion at 0 position), S14C/C35S/C73S, C35/C73S, C73/C35S, C35S/C73S/S91C, C35S/C73S/C121 (insertion at 121 position)) were labeled with At488 maleimide by adding the At488 fluorophores to a 50 μL solution of 100 μM protein at a dye:protein molar ratio of 3:1. The double-cysteine mutant (H32C/C35S/C73S/S91C) was labeled with A488 maleimide and A647 maleimide by adding A488 and A647 to a 50 μL solution of 100 μM protein at a A488:A647:protein molar ratio of 1.3:3:1. TCEP was added at a 10× molar excess to the protein in order to reduce the disulfide bonds. All the maleimide-cysteine coupling reactions were performed in the PBS buffer at pH 7.4. Oxygen was removed by flushing the sample with argon gas in a desiccator for 5 min. The vial was capped tightly and shaken gently for 3 hours at room temperature. The excess dye was removed by size-exclusion chromatography using Sephadex G-50 gels (G5080, Sigma Aldrich) in a BioLogic LP system (731-8300, Bio-Rad). For dually labeled samples, the unlabeled and labeled species were further separated by MonoQ 5/50 GL column (GE Healthcare) in the AKTA protein purification system (18190026, GE Healthcare).

In order to estimate the Förster radius and the intensity corrections factors for smFRET analysis, donor-only and acceptor-only proteins were prepared using a similar protocol in which the protein was incubated for 3 hours with only one species of thiol-reactive dye. Using these control samples, the fluorescence quantum yields of the dyes attached to 4E-BP2 were found to be 0.76 for A488, 0.32 for A647 in the NP condition and 0.83 for A488 and 0.35 for A647 in HP condition. These represent ensemble-averaged and labelling-site-averaged values.

All samples were diluted to concentrations of 1–10 nM and 20–50 pM, which are most suitable for FCS/FAD and smFRET burst experiments, respectively. For a typical experiment, a sample solution of 30 μL was dropped on the surface of plasma-cleaned coverslip. Non-specific protein adsorption to the coverslip was prevented by adding 0.005% (v/v) Tween-20 (P2287, Sigma-Aldrich) to the solution, and bovine serum albumin (BSA) (15260-037, ThermoFisher Scientific) was used to coat the clean coverslips. All experiments were performed at 20 °C.
3.3.4 Instrumentation

Single-molecule fluorescence measurements were performed on a custom-built multiparameter fluorescence microscope that was described in detail elsewhere [121, 163, 247]. Pulsed 480 nm excitation is provided by frequency doubling the output of a femtosecond laser (Tsunami HP, Spectra Physics, Santa Clara, CA, USA) and continuous-wave (CW) 473 nm excitation comes from a solid state laser (Cobolt, Sweden).

In FCS measurements, the laser excites the sample at intensities of \( \sim 5 \text{ kW/cm}^2 \). The fluorescence from the sample was focused through a 50-\( \mu \text{m} \) pinhole and then divided into two channels by a non-polarizing cube beam-splitter. Each beam was focused onto a separate single-photon avalanche diode (SPAD) (PD5CTC, MPD, Italy). A pseudo-autocorrelation curve with logarithmic time binning and 24 bins per temporal decade was calculated from these two signals using a photon-by-photon algorithm [163]. In FAD measurements, the non-polarizing cube was replaced by a polarized cube.

In smFRET measurements, the laser excites the sample at intensities of \( \sim 70 \text{ kW/cm}^2 \). A 75-\( \mu \text{m} \) pinhole was placed in the emission path. A dichroic mirror (FF640-Di01, Semrock, Rochester, NY, USA) was used to split the fluorescence signals into donor and acceptor dye contributions. Red-sensitive SPAD detectors (COUNT-100, Laser Components, USA) were used in the acceptor channels. All the detectors are read and digitized by a multichannel time-correlated counting module (PicoHarp300, PicoQuant, Germany). The donor and acceptor signals were spectrally filtered using band-pass filters BP520/66m and HQ685/80 (Chroma, Bellows Falls, VT, USA). A custom-written Matlab code was used to identify and analyze dual-color single-molecule fluorescence bursts, from which a smFRET histogram was constructed.
3.3.5 FCS analysis

Assuming a 3D Gaussian shape of the detection volume, the fluorescence correlation function for free Brownian diffusion of a single molecular species with multiple relaxation components is given by [248]:

\[ G(\tau) = \frac{1}{N_{eff}} \left( \frac{1 + \frac{\tau}{\tau_d}}{1 + \frac{\tau}{s^2\tau_d}} \right)^{-0.5} \sum_i \left( 1 + a_i e^{-\frac{\tau}{t_i}} \right). \]  

(3-1)

In equation (3-1), \( N_{eff} \) is the average number of molecules in the focal volume and it is given by the inverse of the diffusion-related amplitude \( G_0 = 1/N_{eff} \cdot V_{eff} = \pi^{3/2} z_0 w_0^2 \). In addition, \( s \) is 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 (\( w_0^2 = 4D\tau_d \)) and hydrodynamic radius \( R_H \) of the molecule via the Stokes-Einstein equation [163]. \( a_i \) and \( t_i \) are the amplitude and the observed kinetic time constant of the \( i \)th relaxation. Different numbers of relaxation components in the fitting models were tested in the data fitting process. Fitting residual distributions were compared to obtain appropriate fitting results using minimal fitting parameters. Prior to each set of measurements, Rhodamine 110 was used to characterize the detection volume [163]. Contact on \( k_c \) and off \( k_o \) rate constants were calculated from \( a_i \) and \( t_i \) assuming a two state equilibrium between on and off fluorophore-quencher conformation [249].

\[ a_i = \frac{k_c}{k_o}, \]  

(3-2)

\[ t_i = 1/(k_c + k_o). \]  

(3-3)

3.3.6 FAD analysis

FAD monitors the rotation of the emission dipole during the lifetime of the excited state. We used “wobble-in-a-cone” model to describe the probe and segment motions [250]:
\[ r(t) = r_0 \left[ (1 - A) e^{-t/\rho_1} + A \right] e^{-t/\rho_2} + r_{\text{inf}}. \]  \hspace{1cm} (3-4)

\( \rho_1, \rho_2 \) are two anisotropy decay components and they were assigned to rotational correlation time of the dye and the local segment of the molecule. We can use this two-exponential decay equation to fit our fluorescence anisotropy data and extract rotational correlation times. \( A \) is a parameter describing the degree of motional restriction and is related to the half-cone angle \( \theta \).

\[ A = \left[ \frac{1}{2} (1 + \cos \theta) \cos \theta \right]^2. \hspace{1cm} (3-5) \]

Here, we neglect the protein global motion because the time scale of the overall protein motion is much longer than the fluorophore lifetime due to high backbone flexibility of the IDPs [251]. We cannot resolve the global motion correctly in the anisotropy decay curves with the additional global exponential term and the fitting results for the overall global tumbling times did not scale with the protein size (data not shown here). Similar phenomenon was found in Ref. [252]. Baseline offset \( r_{\text{inf}} \) was introduced to infer the presence of the slower global motion of the protein.

\subsection*{3.3.7 smFRET analysis}

The FRET efficiency \( E \) was calculated based on the number of detected photons in both donor (\( I_D \)) and acceptor (\( I_A \)) channels in each single-molecule burst [194]:

\[ E = \frac{I_A}{I_A + \gamma I_D}, \hspace{1cm} (3-6) \]

where \( \gamma \) is the correction factor for the differences in detection efficiencies of donor and acceptor channels and quantum yields of the dyes. In addition, corrections were applied on both \( I_D \) and \( I_A \) to subtract the background and the spectral cross talk. An \( E \) value was calculated for each single-molecule intensity burst and a FRET histogram was produced. The distance between the donor and acceptor probes \( R_{\text{EE}} \) determines the energy transfer efficiency according to the Förster equation:
\begin{equation}
E(R_{EE}) = \left[ 1 + \left( \frac{R_{EE}}{R_0} \right)^6 \right]^{-1},
\end{equation}

where \( R_0^6 = 8.79 \times 10^5 \frac{k^2 \Phi_D}{n^4} J(\lambda) \). The absorption and emission spectra used to estimate the spectral overlap integral \( J(\lambda) \) and the quantum yield \( \Phi_D \) were measured for dye-labelled 4E-BP2. At each solvent condition, the spectra, quantum yields of the donor and the acceptor and refractive indexes were measured and used in the FRET calculations. The \( k^2 = 2/3 \) assumption was confirmed by FAD measurements based on the rotational half-cone angles of the fluorophore. Figure 3-2 is a bar chart of the dye half-cone angles at 6 different labelling positions along the sequence. In each case, the rotational half-cone angle is \( 70^\circ \)– \( 80^\circ \), validating the general consumption that \( k^2 \) equals to \( 2/3 \) [253]. Using these data, the Förster radius, \( R_0 \), was estimated to be \( 5.1 \pm 0.2 \) nm in the NP state and \( 5.2 \pm 0.2 \) nm in the HP state. At 6 M GdmCl, the estimated \( R_0 \) values were slightly different, i.e., \( 5.0 \pm 0.2 \) nm and \( 5.1 \pm 0.2 \) nm in the NP and HP states, respectively.

**Figure 3-2.** Bar chart of half-cone angles of dye rotation in 0, 14, 35, 73, 91 and 121 positions of 4E-BP2 in different states. (blue) NP 4E-BP2 in PBS, (red) HP 4E-BP2 in PBS, (green) NP 4E-BP2 at 6 M GdmCl, (magenta) HP 4E-BP2 at 6 M GdmCl. Error bars are standard errors from data fits.
3.3.8 Polymer model for $R_G$ and $\langle R_{EE}^2 \rangle$ inference from smFRET data

In order to extract the information about chain dimensions, such as the radius of gyration, $R_G$, and the average mean squared end-to-end distance, $\langle R_{EE}^2 \rangle$, the mean FRET efficiency measured by smFRET can be expressed as [225, 254]:

$$\langle E \rangle = \int_0^\infty E(R_{EE})P(R_{EE})dR,$$  \hspace{1cm} (3-8)

where $P(R_{EE}) = \int p(R_{EE}|R_G)P_{\text{model}}(R_G)dR_G$ is the end-to-end distance distribution for a certain polymer model, which contains the radius of gyration, $R_G$, as a parameter. For a particular form of $P(R_{EE})$, a theoretical value of the mean FRET efficiency, $\langle E \rangle_{\text{th}}$, can be calculated using eq. (3-8). A search is performed until this value matches the experimentally measured smFRET efficiency, $\langle E \rangle_{\text{exp}}$; this procedure will thus provide the most likely $R_G$ and $\langle R_{EE}^2 \rangle$ values.

Here the Sanchez model was used, which uses a Flory-Fisk distribution [255] and it takes into account excluded volume and inter-residue interactions [216, 225, 245]. The end-to-end distribution of the Sanchez model involves two parts. The first part is the Boltzmann-weighted Flory-Fisk distribution [216, 225, 256, 257]

$$P_{\text{Sanchez}}(R_G) = Q^{-1}R_G^6 \exp \left( -\frac{\gamma R_G^2}{2 \langle R_{G,\theta}^2 \rangle} + n \left( \frac{1}{2} \phi \varepsilon - \frac{1-\phi}{\phi} \ln(1 - \phi) \right) \right).$$  \hspace{1cm} (3-9)

Here $Q$ is the normalization constant, $n$ is the number of monomers in the polypeptide chain, $\varepsilon$ is the mean-field interaction energy between two monomers, $\phi = 0.031n/R_G^3$ is the volume fraction of the chain, and the radius of gyration at the $\theta$ state $R_{G,\theta} = 0.225(n + 8)^{1/2}$ nm, which takes the length of the linkers of donor and acceptor fluorophores into account [190, 222, 257]. The $\theta$ state is the state that polymer-polymer and polymer–solvent interactions are balanced [258]. The second part is the probability function of end-to-end distance conditioning
upon the radius of gyration \( P(R_{EE}|R_G) \), which is the distance distribution of two random points inside a sphere [190, 216, 257]

\[
P(R_{EE}|R_G) = \frac{1}{\sqrt[5]{R_G}} \left[ 3 \left( \frac{R_{EE}}{\sqrt[5]{R_G}} \right)^2 - \frac{9}{4} \left( \frac{R_{EE}}{\sqrt[5]{R_G}} \right)^3 + \frac{3}{16} \left( \frac{R_{EE}}{\sqrt[5]{R_G}} \right)^5 \right].
\] (3-10)

### 3.4 Results and Discussion

#### 3.4.1 Towards a rigidity map of 4E-BP2: probing local chain flexibility

The structural flexibility of IDPs is essential for regulating their interactions with other proteins and their role in signaling processes [259]. We used FAD spectroscopy to measure the local, i.e., segmental, flexibility of the 4E-BP2 protein. The fluorescence anisotropy decay monitors the rotation of the emission dipole during the lifetime of the excited state of the fluorophore. As this dipole moment vector has a fixed orientation within the molecular frame, its rotation essentially tracks the rotational motion of the fluorophore. The data analysis quantifies the rotational freedom and speed of the fluorophore attached to the protein of interest. Inferred parameters could relate to the local rigidity of the protein, the friction within the immediate proximity of the dye, and the overall hydrodynamic radius of the protein [260].

Cysteine residues were inserted at 6 different positions, i.e., C0, C14, C35, C73, C91, and C121, to allow fluorescent labeling at those sites (see Figure 3-3 and Material and Methods). These single-Cys mutants provided a detailed picture of the peptide chain flexibility at specific points along the 4E-BP2 sequence; the approach reveals a local rather than a global average behavior of the whole protein. The local probing is particularly important for IDPs, for which the dynamics of one segment may be largely uncorrelated to that of other segments and could also be independent of the global protein dynamics.
The sequence of wild-type 4E-BP2. The red characters indicate the positions of the cysteine residue after single cysteine mutations (14, 35, 73, and 91). The cysteine insertion position 0 and 121 are not shown here.

The width of the instrument response function (IRF) is around 50 ps, which allows reliable measurements of anisotropy decay lifetimes as short as ~100 ps without using deconvolution. Figure 3-4 (A) shows a family of FAD curves in the apo (unbound) 4E-BP2 in the NP state. The decay curve measured for the free At488 is shown for comparison; the other curves were measured with the dye attached to different sites along the chain. The data clearly shows that the chain flexibility is site/region-dependent. For FAD curve fitting, we applied eq. (3-4) and used a common approach, which only considers the data points at $t > 2 \times \text{IRF}$ [261]. The anisotropy decay parameters for each single-Cys construct of both NP and HP 4E-BP2 were obtained by fitting the FAD data (Table 3-1). The rotational correlation times corresponding to segmental motion, $\rho_{\text{seg}}$, are shown in Figure 3-5 (A) and they range from 1.6 ns to 3.3 ns, depending on the labelling site.

Such diverse, non-uniform flexibility along the chain is most likely related to the nature of the amino acids around each site (Fig. 3-3). For examples, glycine and serine residues are the most flexible natural amino acids and proline, isoleucine, and valine are highly rigid residues [262]. As expected, at the terminal positions 0 and 121 increased flexibilities, i.e. faster rotational times, were observed (see Table 3-1). The N-terminal region is mainly composed of glycine and serine residues so that this region is the most flexible segment of the chain. In terms of C terminal, there is an isoleucine residue at the end of the chain so that the C terminal region is somewhat less flexible compared with the N terminal. Within a 10-amino-acid observation window [263, 264], the positions 14, 73 and 91 regions contain about 50% highly rigid amino acids, such as proline, isoleucine, and valine [262, 265], so that the chain appears to be more rigid in these regions. The slowest rotational correlation time was measured
at position 14 ($\rho_{seg} = 3.27$ ns). In addition to containing a significant amount of rigid amino acids, this segment also contains two positively-charged arginines, which stretch the segment further via electrostatic repulsion forces. On the other hand, three negatively-charged residues (2 glutamic acids and 1 aspartic acid) and one positively-charged lysine residue are located around position 91 within the observation window; that region ($\rho_{seg} = 2.59$ ns) appears to be more flexible compared to positions 14 ($\rho_{seg} = 3.27$ ns) and 73 ($\rho_{seg} = 2.86$ ns) due to the electrostatic attraction between the polar residues. Around position 35 ($\rho_{seg} = 2.48$ ns), there is a significant number of glycine residues compared with the number of prolines, so that the local chain flexibility is somewhat higher than that in other central regions.

**Figure 3-4.** Family of FAD curves of 4E-BP2 at different locations (0, 14, 35, 73, 91, and 121) in the apo state (A) and bound state (B). Magenta (position 0) and green (position 91) scattered dots are the raw data. Red and olive dash lines are the corresponding fits. They highlight the significant changes of FAD curve at the position 91, where secondary binding site is located, compared to flanking N terminal position 0. Bottom panel plots the corresponding residuals of the fits.
3.4.2 Local chain flexibility highlights the difference between intrinsically and denatured disordered states

4E-BP2 in NP state is largely disordered. Forman-Kay and co-workers [53] have recently shown that part of chain (18-62) in 4E-BP2 undergoes a disorder-to-order transition upon phosphorylation while rest of the chain still remains disordered and the binding affinity between 4E-BP2 and eIF4E is considerably diminished. We used FAD to probe the changes in local chain flexibility induced by phosphorylation. The segmental rotational correlation times are listed in Table 3-1 and the changes after HP are illustrated in Figure 3-5 (A). The main feature is that the local flexibility decreases only in the folded region (near position 35), whereas everywhere else on the chain becomes slightly more flexible. This agrees with the notion that well-defined tertiary structures are more stable and less flexible than the transient secondary and/or tertiary structures. The increased flexibility in the rest of the domain suggests that there is no significant phosphorylation induced structural changes in these regions [266]. It is likely that the increased segmental flexibility near the secondary binding site on 4E-BP2 prevents eIF4E binding [53]. Similar effects on chain flexibility related binding studies were also reported in [267-269].

![Figure 3-5](image)

**Figure 3-5.** (A) Segmental rotational correlation times of the NP (blue) and the HP (red) 4E-BP2 in PBS buffer. (B) Segmental rotational correlation times of the NP 4E-BP2 in PBS (blue), 6 M GdmCl (green) and the HP 4E-BP2 at 6 M GdmCl condition (magenta). Error bars come from the fitting of multiple data sets.
There is growing evidence that many IDPs under native or mildly denaturing conditions are more compact and behave quite differently than random coils [26, 35]. Here we apply FAD spectroscopy to examine to what extent intrinsically disordered states in 4E-BP2 differ from random coil states. As seen in Figure 3-5 (B) and Table 3-1, the segmental rotational dynamics is considerably slower for all the labelling sites in the NP state of 4E-BP2 in PBS buffer than that in 6 M GdmCl. Indeed, this result reinforces the idea that the disordered state of 4E-BP2 has distinct characteristics compared to random coils, most likely due to transient secondary structures. Table 3-1 and Figure 3-5 (B) also show that highly denatured states of both NP and HP 4E-BP2 have nearly identical chain flexibility. However, at the residues that are close to the folded region the results are still slightly different, suggesting that even though hydrophobic interactions and hydrogen bonds are largely disrupted at 6 M GdmCl, some residual secondary structure may persist in the phosphorylated 4E-BP2.

**Table 3-1.** Segmental rotational correlation times $\rho_{seg}$ of NP and HP 4E-BP2 in PBS buffer and 6 M GdmCl solvent in six different regions. Errors come from the fitting of multiple data sets.

<table>
<thead>
<tr>
<th>Residue #</th>
<th>NP PBS</th>
<th>HP PBS</th>
<th>NP 6 M GdmCl</th>
<th>HP 6 M GdmCl</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.61 ± 0.10</td>
<td>1.75 ± 0.13</td>
<td>0.89 ± 0.10</td>
<td>0.82 ± 0.08</td>
</tr>
<tr>
<td>14</td>
<td>3.27 ± 0.26</td>
<td>2.56 ± 0.15</td>
<td>1.54 ± 0.10</td>
<td>1.27 ± 0.04</td>
</tr>
<tr>
<td>35</td>
<td>2.48 ± 0.33</td>
<td>3.13 ± 0.07</td>
<td>1.31 ± 0.04</td>
<td>1.57 ± 0.05</td>
</tr>
<tr>
<td>73</td>
<td>2.86 ± 0.14</td>
<td>2.35 ± 0.14</td>
<td>1.43 ± 0.08</td>
<td>1.43 ± 0.19</td>
</tr>
<tr>
<td>91</td>
<td>2.59 ± 0.15</td>
<td>2.01 ± 0.05</td>
<td>1.42 ± 0.07</td>
<td>1.49 ± 0.09</td>
</tr>
<tr>
<td>121</td>
<td>1.90 ± 0.18</td>
<td>1.43 ± 0.08</td>
<td>0.95 ± 0.06</td>
<td>0.96 ± 0.09</td>
</tr>
</tbody>
</table>
3.4.3 Segmental motion analysis reveals the protein-protein interaction interface

As discussed above, FAD is an extremely useful technique to probe the local chain flexibility that is hidden in ensemble behavior. Employing the same principle, we studied local segmental flexibility in the presence of eIF4E to gain insight into the binding interface of 4E-BP2.

To rule out the possibility that the fluorophore At488 used to label the protein has non-specific interactions with eIF4E, FCS measurements were performed to examine the diffusion time of the At488 in the absence and presence of eIF4E (5 nM At488 and 1 µM eIF4E). Figure 3-6 (A) and the fitting results demonstrate that At488 indeed does not bind to eIF4E. After the clearance of the non-specific interaction between At488 and eIF4E, FCS measurements of 4E-BP2 and 4E-BP2/eIF4E complex were performed to obtain the $R_H$ of the apo and the bound states. The concentrations used for the experiments are 5 nM 4E-BP2 with/without 500 nM eIF4E. The high concentration of eIF4E (unlabeled) ensured that virtually all of 4E-BP2 (labelled) was in the bound form. Figure 3-6 (B) shows that the FCS curve shifts to the right, i.e., longer diffusion times, upon addition of eIF4E. This corresponds to an increase of $R_H$, which is consistent with the formation of the 4E-BP2/eIF4E complex. The $R_H$ values estimated by fitting the FCS curves to eq. (3-1) are 29.0 ± 0.5 Å for the NP 4E-BP2 and 36.6 ± 0.8 Å for the 4E-BP2/eIF4E complex. Figure 3-6 (C-D) are bar charts of the $R_H$ values measured in the apo (C) and bound states (D), where the fluorescent probe was located at six different Cys sites. For both cases, it was found that the $R_H$ estimation does not depend on the labelling site.
Figure 3-6. (A) FCS curves of At488 dye alone (red) and mixture of At488 and eIF4E (blue). (B) FCS curves of 4E-BP2 labeled with At488 (red) and mixture of At488 labeled 4E-BP2 and eIF4E (blue). Hydrodynamic radii of 4E-BP2 alone (C) and eIF4E bound 4E-BP2 (in eIF4E saturation condition) (D). Each bar in (C-D) indicates different At488 probe locations. Error bars are standard errors from data fits.

Figure 3-4 (B) shows the anisotropy decay curves for the 4E-BP2/eIF4E complex for the six different labeling sites. In contrast with the apo state (see Fig. 3-4 (A)), these curves decay to non-zero asymptotic values, which indicate that the rotational motion of the 4E-BP2 chain around each labeling site was constrained upon binding to eIF4E. The FAD curves visually convey the idea that the positions 73 and 91 are close to the binding interface, as they change most upon binding. Two dashed lines (red and olive) in Figure 3-4 (A-B) underscore the magnitude of the changes of FAD curves in the position 91 compared to the N terminus.
Figure 3-7. (A) Segmental rotational correlation times of the NP 4E-BP2 (blue) and the eIF4E-bound 4E-BP2 (red) in six different positions (0, 14, 35, 73, 91, and 121). (B) Half-cone angles of the At488 fluorophore labeled on 4E-BP2 in six different locations (0, 14, 35, 73, 91, and 121) in the apo state (blue) and in the bound state (red). All data were measured in PBS buffer, pH 7.4. Error bars come from the fitting of multiple data sets.

The segmental rotational correlation times $\rho_{\text{seg}}$ and the half-cone angles $\theta_{\text{dye}}$ at each location along the chain were estimated using eqs. (3-4) and (3-5) (Table 3-2) and they are plotted in Figure 3-7. The disordered extended C-terminal region is significantly affected, i.e., $\rho_{\text{seg}}$ nearly doubles, upon binding to eIF4E whereas there is little change in the N-terminal region. These results demonstrate that segmental motion parameters measured by FAD can be used to extract information about the binding contributions of different regions in a protein as it interacts with another protein. The changes in the $\theta_{\text{dye}}$ match the changes of $\rho_{\text{seg}}$, which is consistent with the fact that the rotational freedom of the dye label on 4E-BP2 is significantly hindered upon binding to eIF4E. The results also urge caution when considering the $\kappa^2$ factor in smFRET studies of 4E-BP2 in the presence of eIF4E.

Table 3-2. Segmental rotational correlation times $\rho_{\text{seg}}$ and dye half-cone angles $\theta_{\text{dye}}$ in the apo and bound states in six different regions. Errors come from the fitting of multiple data sets.

<table>
<thead>
<tr>
<th>Position</th>
<th>Apo ($\rho_{\text{seg}}$) (ns)</th>
<th>Apo ($\theta_{\text{dye}}$) (°)</th>
<th>Bound ($\rho_{\text{seg}}$) (ns)</th>
<th>Bound ($\theta_{\text{dye}}$) (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.64 ± 0.10</td>
<td>77.43 ± 0.66</td>
<td>1.93 ± 0.13</td>
<td>69.86 ± 0.28</td>
</tr>
<tr>
<td>14</td>
<td>3.27 ± 0.26</td>
<td>70.88 ± 1.97</td>
<td>3.36 ± 0.11</td>
<td>61.28 ± 0.41</td>
</tr>
</tbody>
</table>
3.4.4 Intra-chain kinetics for 4E-BP2

The 4E-BP2 is disordered, however the transient secondary structure observed by the nuclear Overhauser effect in NMR measurements implies that the protein exhibits a dynamic local structure [53]. Measuring the local chain kinetics is thus essential to understand these dynamics and serves as complimentary information to smFRET and NMR conformation data due to the fast interconversion between different conformers in IDPs. To gain insight into the dynamics of the 4E-BP2 in NP, HP and denatured conditions, FCS analysis was used to measure the chain kinetics at various sites within 4E-BP2.

Figure 3-8 shows FCS data, fitting curves and fitting residuals of 4E-BP2 in the NP state (Fig. 3-8 (A-B)) and the HP state (Fig. 3-8 (C-D)). Four different fitting models were tested for each sample, ranging from one kinetic component (1F) to four kinetic components (4F) in addition to one diffusion component (1D) (see eq. (3-1)). The 1D+3F, 1D+2F and 1D+2F models return reasonable diffusion times and flat residuals in NP, HP (except for N terminal) and denatured situations, respectively. Each fitted kinetic component has about an order of magnitude difference in time scale. However, we cannot rule out the possibility of other decay components that were not resolved because they lie too close to the detected lifetimes or because they lie outside the temporal window of the FCS measurements (10^{-8} – 10^{-2} s).
Figure 3-8. (A) FCS data (grey line) of 4E-BP2 labeled with At488 at residue 73 in NP condition (A) and HP condition (C) showing conformational fluctuation and diffusion components. Colored lines are the fitting curves based on four different fitting models. Red, green, blue and magenta colors represent one diffusion component with one, two, three or four kinetic components, respectively. Insets in (A) and (C) are zoom in of the $10^{-8}$ to $10^{-6}$ second. (B, D) Weighted residuals for each fitting model tested.

In the NP state, the chain kinetics varies along the sequence, which was also found in other IDP, e.g., the N-terminal domain p53-TAD [249]. Although local kinetics are heterogeneous, 3 kinetic components were required for all labelling sites and they fall into three time scales: 100 ns –1 µs, 10 –30 µs, and ~100 µs. We label them as $\tau_1$, $\tau_2$ and $\tau_3$ components. Our results are in agreement with previous study of CspTm protein, for which dynamics in the range from 300 ns to 200 µs was found [254].

The control FCS experiment on the dye alone (At488) did not show a photophysical process in the sub-µs time scale. We can thus assign the 100 ns –1 µs component as a unique feature of 4E-BP2 in the absence of phosphorylation. This kinetic component is most likely caused by the modulation of the dye’s fluorescence by the fast motion of amino acid quenchers in and out of the dye’s proximity (typically ≤ 10 Å). Tryptophan (Trp) and Tyrosine (Tyr) are
the most efficient fluorescence quenchers [244, 270]. Although methionine (Met) and histidine (His) are less efficient compared with Trp and Tyr, five His residues and three Met residues could provide additional quenching. The sub-μs chain dynamics corresponds to the interconversion between different conformers within the disordered conformational ensemble. This assignment agrees with the values extracted from model polypeptides where probe and quencher are separated by the entire length the polypeptide chain. The contact times ranges from 300 ns to 4 μs based on the amino acid sequence of the polypeptide [271, 272].

The intermediate component, having a lifetime of 10-30 μs, is absent in both the HP and the denaturation states (see below). These kinetics may correspond to the contact between distal parts of the chain or to interconversion between transient secondary structures and disordered structures. Dynamics on this time scale could also be caused by the formation of non-random chain configurations [249]. The slowest, 100 μs lifetime is viscosity dependent (see below), and thus can be assigned to the concerted motions of the protein chain as suggested in [270]. The concerted motions could result from domain movement of the chain [273] or motion of transient or residual secondary structure elements with flexible linkers [274]. The contact rates corresponding to τ₁ component were calculated using eqs. (3-2) and (3-3). The contact rates in six different locations along the chain were plotted in Figure 3-9; for NP 4E-BP2 these rates are in the range of $10^4 - 10^5$ s⁻¹.

3.4.4.1 The effect of phosphorylation

For the phosphorylated state, only two kinetic components were needed at each labelling position (except for the N terminus), which fell into two time regions: 1–10 μs, and ~100 μs. We label them as τ₁ and τ₃ components. The 1–10 μs component can be assigned to quenching by aromatic residues similar to the sub-μs lifetime observed in the NP state. The quenching dynamics slow down upon phosphorylation presumably because the average quenching efficiency drops as a result of protein expansion due to electrostatic repulsion introduced by the phosphate groups. Similar to the NP case, the ~100 μs time constant can be assigned to the concerted motions of the protein chain.
The contact rates corresponding to \( \tau_1 \) component were calculated using eqs. (3-2) and (3-3). As shown in Figure 3-9, upon phosphorylation, chain motions slow down not only in segments near the phosphorylation sites, but also in the distal regions. The FCS results are in agreement with previous NMR data and indicate that phosphorylation stabilizes long-range interactions to residues outside the folded region \([53]\). Similarly, long-range modulation of chain motions in the N terminal domain p53-TAD were also found to be induced by phosphorylation \([249]\).

One possible factor to stabilize 4E-BP2 in the HP state could be the steric effect introduced by several phosphate groups \([275, 276]\). It was suggested by Fersht and co-workers that transient interaction network within IDPs triggered by phosphorylation spreads local structural changes to global changes \([249]\). The folded region contains two Tyr residues in close proximity to the fluorophore at position 35. It is also likely that fluorophore contacts made with these quenchers track the fluctuation between disordered and folded structures.

Figure 3-9. At488-amino acid quenching rates of 4E-BP2 in NP (blue) and HP (red) conditions. The arrows indicate the changes in amplitude and rate constant upon HP. Error bars are standard errors from data fits.
3.4.4.2 The effect of denaturation

Segmental flexibility of 4E-BP2 in strong denatured solvent was different than under native condition (Fig. 3-5 (B)), pointing to significant transient secondary and tertiary structures in the 4E-BP2 native disordered state. Here we use the FCS technique to probe the chain kinetics of 4E-BP2 at 6 M GdmCl.

As with what we found upon phosphorylation, 1D+2F model can fit all the FCS curves reasonably well. The two kinetic components are located in the 1–10 µs and ~100 µs regimes and we label them as τ₁ and τ₃ components. We assign 1–10 µs component as the time scale for amino-acid quenching to the fluorophore and ~100 µs as the time scale of the concerted motion of the denatured protein chain. The slowing down in the quenching dynamics is because end-to-end distance increases in the presence of 6 M GdmCl. The reason for ~100 µs assignment in all the cases is that this component correlates with the diffusion time of the molecule. As the diffusion time increases because of the viscosity increase in 6 M GdmCl solution, the ~100 µs component also increases for all six labelling sites. For both NP and HP 4E-BP2, the ratio between the diffusion time and the relaxation time of the concerted motion is 3.2 ± 0.2 under denaturing conditions, and 2.5 ± 0.3 in PBS buffer. As noted above, the faster 1–10 µs lifetime is similar to the time scale of fluctuations measured by FCS for the chemically denatured cytochrome c [277].

Our FCS results in physiological buffer and 6 M GdmCl solvent further prove that IDPs exhibit significant deviations from random coils in native state. In Figure 3-10, it is shown that quenching rates are significantly reduced in all six locations along the protein chain upon 6 M GdmCl denaturing agents. It may be due to the end-to-end distance increases as examined in the smFRET studies or solvent viscosity increases in 6 M GdmCl or guanidinium ion binding to the protein chain [271]. In addition, because of the increase in fluctuations as the protein denatures, the amplitude of the contact rates strongly increase. A similar trend was also observed in a FCS study of apomyoglobin [270].
Figure 3-10. At488-amino acid quenching rates of NP 4E-BP2 in PBS buffer (blue), at 6 M GdmCl solution (green) and HP 4E-BP2 at 6 M GdmCl solution (red). The arrows indicate the changes in amplitude and rate constant upon denaturation. Error bars are standard errors from data fits.

It is noted that the contact rates in all six regions probed have very similar values at 6 M GdmCl. This suggests that the denatured protein most likely behaves like a random coil. Hagen et al. [278, 279] studied contact formation in unfolded cytochrome c and found that the contact rate for 50 residue polypeptide chain is approximately $2.5 \times 10^4$ s$^{-1}$. Based on the $k_c \sim n^{1.5}$ scaling law expected for random coils [272], the contact rate for the 4E-BP2 (120 a.a.) is calculated to be around $1 \times 10^4$ s$^{-1}$. This empirical estimation closely agrees with our experimental data shown in the Figure 3-10 (green).

The FCS data indicate that denaturants slow down local chain contact rates (see Fig. 3-10), whereas FAD data suggest the local segments become more flexible along the chain (see Fig. 3-5). These seemingly contradictory results can be explained by first understanding the underlying difference between the FCS-measured local chain contact rates and the FAD-measured local chain flexibility. The two techniques probe distinct states of the local chain and their time scales are also different. FCS provides information on sub-microsecond-to-millisecond dynamics, while FAD provides access to the (sub)nanosecond time scale.
Secondly, the problem is solved by considering the effect of the denaturant solvent on the structural properties of protein; Krieger et al. [271] pointed out that high concentration denaturants, as good solvents, reduce the strength of intra-molecular interactions like hydrophobic interaction, hydrogen bonding and van der Waals interactions relative to protein-solvent interactions. This makes unstructured protein chains more flexible and leads to a dynamic behavior expected for an unperturbed chain [271]. Figure 3-10 also compares the quenching rates and amplitudes in the denatured NP state to those in the denatured HP state. The data shows that the local chain dynamics are mostly similar in both cases except at the residues that are close to the folded region (position 35). This observation, as with our anisotropy decay results, suggests that some residual secondary structure could persevere in the phosphorylated 4E-BP2 at 6 M GdmCl.

3.4.5 Inter-chain kinetics for 4E-BP2 bound to eIF4E

The amino-acid quenching effect can also be used to investigate the inter-molecular on and off rates between 4E-BP2 and eIF4E. Since the eIF4E sequence (PDB: 2GPQ) has 8 Trp and 6 Tyr residues (at least 3 Trp and 2 Tyr are exposed on the surface from 3D structure visualization and solvent accessible surface areas calculations using POPS algorithm [280]) the amino-acid quenching effect is expected to be much stronger in the 4E-BP2/eIF4E complex than that for 4E-BP2 alone.

To ensure that nearly all 4E-BP2 molecules are in the bound state, the FCS experiments were done under eIF4E saturation condition (5 nM 4E-BP2 and 500 nM eIF4E). In the 4E-BP2/eIF4E complex, three kinetic, non-diffusion decay components were needed to fit the data. The three dynamic time scales are the same as measured for NP 4E-BP2. The on and off rates could be calculated using eqs. (3-2) and (3-3). It is shown that the on and off rates for regions around sites 73 and 91, which are near the secondary binding site (78-82) [67, 281], increase significantly in the bound state. This is possibly due to a Trp 73 residue found at the binding interface of eIF4E (Fig. 3-11 (D, E) & Fig. 3-12 (D, E)). The on rate between 4E-BP2 and eIF4E in the secondary binding sites ranges from $1 \times 10^6$ s$^{-1}$ to $3 \times 10^6$ s$^{-1}$ and the off rate
is $\sim 1 \times 10^7$ s$^{-1}$. It was found that the on and off rates for regions around sites 35 and 121 also experience a change upon complex formation (Fig. 3-11 (C, F) & Fig. 3-12 (C, F)).

The data also indicate that both on and off rates remain the same in position 0 and 14 because these two regions are quite remote from both the canonical (54-60) and the secondary (78-82) binding sites [67, 281]. The changes in on and off rates highlight the highly dynamic structure of 4E-BP2/eIF4E complex and also validate our interpretation of the sub-microsecond dynamics in the NP 4E-BP2 being caused by quenching by aromatic amino acids.

**Figure 3-11.** Inter-molecular interaction on rates between 4E-BP2 and eIF4E, where fluorescent probe located in six different locations (0, 14, 35, 73, 91, and 121) on the 4E-BP2 sequence. The arrows indicate the changes in amplitude and rate constant upon eIF4E binding. Error bars are standard errors from data fits.
Figure 3-12. Inter-molecular interaction off rates between 4E-BP2 and eIF4E, where fluorescent probe located in six different locations (0, 14, 35, 73, 91, and 121) on the 4E-BP2 sequence. The arrows indicate the changes in amplitude and rate constant upon eIF4E binding. Error bars are standard errors from data fits.

3.4.6 The linear size(s) of 4E-BP2

smFRET measurements were performed to examine the conformations of 4E-BP2 in its native state and the conformational changes upon phosphorylation, denaturation, salt titration and pH modulation. A double-cysteine mutant (C32C91) labeled with A488 and A647 was used in the smFRET experiments. Each smFRET histogram was fitted to a Gaussian distribution. The center FRET efficiency $\langle E \rangle$ and the half-width at half-maximum (HWHM) $\sigma$ for each sample and solvent conditions are reported in Table 3-3.

3.4.6.1 The effect of phosphorylation

Phosphorylation is thought to reduce tertiary contacts between IDP and its binding partner [282] and enhance binding by induced conformational change [283]. A recent NMR study reported that phosphorylation of 4E-BP2 at positions 37 and 46 induces residues 18-62
to form a 4-stranded β-sheet structure and diminish binding affinity to its partner [53]. Characterizing the conformational changes of IDPs upon phosphorylation is therefore essentially important to understand their biological function. The smFRET histogram for the HP 4E-BP2 is shown in Figure 3-13 (F). Surprisingly, \( \langle E \rangle \) did not increase upon HP, as it would be expected on the basis that 4E-BP2 undergoes a disorder-to-order transition and becomes more compact. Instead, the center peak value \( \langle E \rangle \) shifts from 0.57 to 0.30 upon HP, corresponding to an increase of the average donor-acceptor distance.

The reason for this expansion could be that the phosphorylation leads to a change in the electrostatic charge from 0 to -10 within the 2-Cys region (C32C91) probed by smFRET. Thus, in the HP state the electrostatic repulsions leads to an increase in the end-to-end distance. Electrostatic interactions are known to play a crucial role in determining the dimensions and the polymer behavior of IDPs. Our results are in agreement with both theoretical [36] and experimental [192, 200] studies on IDPs. It should be noted that the folded region only occupies half of the region between two cysteines, the topology of the 4E-BP2 upon HP could also make this protein more extended.
**Figure 3-13.** smFRET efficiency histograms of NP (A-E) and HP (F-J) 4E-BP2 at different concentrations of GdmCl and urea. Solid lines are Gaussian fits to the data; red solid lines are the overall fitting profiles and blue lines are individual Gaussians peaks. The minor peak near $E \sim 0$ corresponds to molecules without (active) acceptors. The red dash lines indicate the locations of the peak FRET efficiency for the non-zero Gaussian; the fitting parameters are given in Table 3-3.
### 3.4.6.2 The effect of denaturant

In order to assess whether 4E-BP2 behaves like a random-coil or like a premolten globule, smFRET experiments were performed in denaturation solvents, such as GdmCl and urea. Random-coil-like IDPs have similar dimensions as those of denatured proteins, whereas premolten globule IDPs are much more compact [244]. The data and the analysis results are shown in Figure 3-13 and Table 3-3. The \( \langle E \rangle \) value shifts monotonously to smaller values with the increase of [GdmCl] for the NP 4E-BP2. The \( \langle E \rangle \) value changes from ~0.6 at PBS to ~0.2 at both 6 M GdmCl and 6 M urea, indicating that the 32-91 region in the natively disordered protein is significantly more compact than that in the chemically-denatured state. This is in line with the conclusions regarding the 4E-BP2 segmental flexibility and chain kinetics in the previous sections, and suggests that 4E-BP2 should be classified as a premolten-globule type of IDP.

An interesting effect occurs in the HP 4E-BP2: with the increase of the GdmCl concentration, the FRET peak exhibits a rollover behavior, which is absent in NP 4E-BP2. The \( \langle E \rangle \) value first increases slightly in 1 M GdmCl, indicating a collapse of the disordered state (Fig. 3-13 (G)). Then, the \( \langle E \rangle \) value decreases with the increase of [GdmCl] as a result of the denaturant-induced chain expansion, similar to the NP 4E-BP2 and to typical smFRET unfolding behavior [225, 244]. The rollover behavior is caused by the electrostatic screening of the phosphate charge by the ionic denaturant GdmCl, which initially makes the protein become more compact, as previously observed in IDPs [200].

<table>
<thead>
<tr>
<th>[GdmCl] (M)</th>
<th>NP ( \langle E \rangle )</th>
<th>( \sigma )</th>
<th>( \sigma_{sn} )</th>
<th>HP ( \langle E \rangle )</th>
<th>( \sigma )</th>
<th>( \sigma_{sn} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.57</td>
<td>0.128</td>
<td>0.072</td>
<td>0.30</td>
<td>0.130</td>
<td>0.069</td>
</tr>
<tr>
<td>1</td>
<td>0.40</td>
<td>0.134</td>
<td>0.068</td>
<td>0.40</td>
<td>0.142</td>
<td>0.069</td>
</tr>
</tbody>
</table>

**Table 3-3.** Non-zero FRET peak efficiencies \( \langle E \rangle \) and widths \( \sigma \) are extracted from the Gaussian fits of the smFRET histograms at different GdmCl and urea concentrations. FRET peak width under shot noise limit \( \sigma_{sn} \) is calculated based on eq. (3-11).
<table>
<thead>
<tr>
<th>3</th>
<th>0.25</th>
<th>0.158</th>
<th>0.060</th>
<th>0.28</th>
<th>0.169</th>
<th>0.062</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.20</td>
<td>0.148</td>
<td>0.054</td>
<td>0.18</td>
<td>0.164</td>
<td>0.052</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[urea] (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

\(^a\) The errors for the mean FRET efficiency are ± 0.01.

\(^b\) The errors for the fitted FRET peak width are ± 0.010.

\(^c\) The errors for the FRET peak width at signal to noise limit are ± 0.006 calculated based on eq. (3-11).

### 3.4.6.3 Conformation exchange dynamics: the width of smFRET peaks

The shot-noise limit of the width of a FRET efficiency peak, \(\sigma_{sn}\), is given by [284]:

\[
\sigma_{sn} = \sqrt{\langle E \rangle (1 - \langle E \rangle) \langle N^{-1} \rangle},
\]

where \(\langle N^{-1} \rangle\) is the mean of the inverse number of photons in a burst.

The width \(\sigma\) obtained from Gaussian fitting and the shot-noise width \(\sigma_{sn}\) calculated from eq. (3-11) are listed in Table 3-3. It is noted that \(\sigma\) is 2-3 times larger than \(\sigma_{sn}\). The donor photophysics such as triplet-state formation, photobleaching and intensity variation across the excitation volume could broaden the width of the FRET distribution. We recall that the time scale of intra-chain protein dynamics measured in FCS was in the (sub)microsecond range. Therefore, it appears reasonable that the FRET broadening could be caused by interconversion among various conformations while the protein diffuses through the detection volume (~ 1ms time scale). This is also expected from the perspective of the rugged and shallow energy landscape of IDPs, which permits transitions between states in different energy wells on a wide range of time scales. Each molecule samples several conformations during the time that it takes to diffuse through the focal volume resulting in the peak broadening beyond the shot noise.
limit. The peak center corresponds to the weighted average FRET efficiency of the individual conformational states.

3.4.6.4 Non-cooperative conformational transition

smFRET denaturation studies on globular proteins, such as the cold shock protein CspTm [187, 200, 254], the B domain of Protein A (BDPA) [285] and the chymotrypsin inhibitor 2 (CI2) [286] showed a cooperative transition between two FRET efficiency peaks. In those experiments a narrow high-FRET peak is observed in the absence of denaturant, corresponding to the folded. When the denaturant is added, a lower FRET peak appears, corresponding to the denatured, unfolded state. With increasing denaturant concentration, the population of the new peak increases at the expense of the initial high-FRET peak. The population fraction of the unfolded peak follows a sigmoidal curve as a function of denaturant concentration, which is typical to a 2-state cooperative transition.

Such a typical cooperative unfolding behavior has not been observed in our smFRET measurements on 4E-BP2. As seen from Figure 3-13, only one FRET peak was observed for both NP and HP 4E-BP2, which gradually shifts to lower efficiency values as the molecule expands in denaturant. The data points to a significant amount of rapid interconverting conformations rather than a small group of distinct structural conformers [244]. This non-cooperative transition from native to denatured states was also observed in other IDPs, such as N terminal domain of HIV-1 integrase (IN) [200], human prothymosin α (Pro α) [200], prion-determining domain of Sup35 [244] and BBL domain [287]. However, it is unusual for a stable folded protein. In HP 4E-BP2, despite the folded structure, non-cooperative unfolding also occurs in our case. The reason is that FRET probes (position 32 and 91) flank both folded (18-62) and disordered structures (62-120) to cover all the phosphorylation sites (37, 46, 65, 70, 83). It is possible that the conformation or topology change in disordered part overwhelms the change in the disorder-to-order region upon phosphorylation.
3.4.6.5 The effect of salt

Intra-chain electrostatic interactions are essential factors determining physical properties of IDPs, in particular their compactness. Charge effects on the 4E-BP2 conformations were studied via smFRET measurements at different salt (KCl) concentrations. The FRET efficiency histograms and their corresponding Gaussian fitting parameters are shown in Figure 3-14 and Table 3-4, respectively. In the NP protein, the net charge is zero in the region between the two Cys residues, so that $\langle E \rangle$ was expected to be independent of the salt concentration; the experiment confirmed this trend (Fig. 3-14 (A-C)). However, $\langle E \rangle$ takes a slightly higher value in no salt condition, as shown in Fig. 3-14 (A). This is in agreement with the polyampholyte theory [200, 288], according to which polypeptide chains with balanced positive and negative charges exhibit a collapse at low ionic strength because of the attractive net electrostatic forces between opposite charges in the polymer chain. In the presence of salt, it screens the attractive interactions so that the chain somehow gets expanded.
**Figure 3-14.** smFRET efficiency histograms of NP (A-C) and HP (D-F) 4E-BP2 at different concentrations of KCl. Solid lines are Gaussian fits to the data; red solid lines are the overall fitting profiles and blue lines are individual Gaussians peaks. The minor peak near $E \sim 0$ peaks corresponds to molecules without (active) acceptors. The red dash lines indicate the locations of the peak FRET efficiency for the non-zero Gaussian; the fitting parameters are given in Table 3-4.

In the HP protein, the chain gains a charge of -10 because of the five added phosphate groups, which are all located between the FRET sites. As shown in Figure 3-14 (D-F), the histogram continuously shifts to higher FRET values with increasing [KCl]. This is in agreement with salt-induced charge screening [192, 200]. An interesting question arises when comparing the $\langle E \rangle$ value for NP 4E-BP2 (0.59) to the value for HP 4E-BP2 (0.41) at 3 M KCl (Table 3-4). If the chain expansion in HP condition was solely due to electrostatic repulsion, one would expect that at 3 M salt all the phosphate groups in the chain would be screened. Then, why are the two $\langle E \rangle$ values so different?

**Table 3-4.** Non-zero FRET efficiencies extracted from the Gaussian fits of the FRET efficiency histograms at different KCl concentrations.

<table>
<thead>
<tr>
<th>KCl (M)</th>
<th>NP $\langle E \rangle$</th>
<th>HP $\langle E \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.64 ± 0.12</td>
<td>0.23 ± 0.13</td>
</tr>
<tr>
<td>0.15</td>
<td>0.59 ± 0.15</td>
<td>0.32 ± 0.16</td>
</tr>
<tr>
<td>3</td>
<td>0.59 ± 0.13</td>
<td>0.41 ± 0.14</td>
</tr>
</tbody>
</table>

*a The error values are the HWHM of the FRET efficiency histogram.

To answer this question, we first need to calculate the Debye screening length $\lambda$, which is expressed as [289]

$$
\lambda = \sqrt{\frac{\varepsilon_0 \varepsilon_\text{f} T}{2e^2 N_A I_s}},
$$

(3-12)
where $\varepsilon_0$ is the vacuum permittivity, $\varepsilon$ is the relative permittivity, $k_B$ is the Boltzmann constant, $T$ is the temperature, $e$ is the electron charge, $N_A$ is Avogadro’s number, and $I_s$ is the ionic strength. The Debye screening lengths were calculated to be 7.87 Å and 1.76 Å for 150 mM KCl and 3 M KCl solutions under $\varepsilon = 80$ [290] and $T= 293$ K.

The electric field $E(Q_1, r)$, at a distance $r$ from a spherical charge $Q_1$ with radius $R$ is [290]

$$E(Q_1, r) = \left(\frac{Q_1}{4\pi\varepsilon_0}\right) \left(\frac{1}{r^2}\right) \left(\frac{r+\lambda}{R+\lambda}\right) e^{-(r-R)/\lambda},$$

(3-13)

and the force associated with this electric field and charge $Q_2$ is $F(r) = Q_2 E(Q_1, r)$. With this we can estimate the force ratio between the low salt (L) and the high salt (H) conditions is

$$F_{ratio} = \frac{F_L(r)}{F_H(r)},$$

(3-14)

which provides a measure for the magnitude of the effective charge screening between two charged groups in high salt conditions. Figure 3-15 is the plot of the force ratio of the electrostatic repulsion between two phosphate groups at 150 mM KCl vs. 3 M KCl, assuming that the phosphate group is a sphere of radius $R = 2.5$ Å [290].

![Figure 3-15](image)

**Figure 3-15.** The force ratio for two phosphate groups screened by 150 mM KCl and 3 M KCl as a function of separation distance $r$. The inset shows a zoom in for distances less than 10 Å.
The curves in Figure 3-15 show that $F_{ratio}$ increases steeply with distance, with a value exceeding 15 for a separation between the phosphate groups of 10 Å. The shortest distance between two phosphates located at Ser 65 and Thr 70 in the 4E-BP2 is ∼12 Å based on the distance measure of the peptide structure generated from its amino acid sequence under PDB Utility Server, Dr. Adriaan Bax group, NIH, USA. This suggests that a 3 M KCl solution provides overwhelming screening of the electrostatic repulsion between the phosphates within the 4E-BP2 chain. However, the longer end-to-end distance was observed for HP 4E-BP2 at 3 M KCl compared to NP 4E-BP2 at 150 mM KCl (Fig. 3-14 (B) and (F) and Table 3-4). This implies that additional intra-molecular interactions other than electrostatic repulsion are introduced upon phosphorylation and the overall chain topology changes after the protein gets HP. This suggestion is in agreement with recent NMR results showing that helix propensities are reduced significantly throughout the sequence and folded beta sheet structure is formed upon phosphorylation [53].

3.4.6.6 The effect of pH

Amino acids have different pKa values so that changing the pH value of the solvent will modulate the total charge and the charge distribution along the chain. We performed smFRET measurements on both phosphorylation states of 4E-BP2 at pH 3, 7.4 and 10 to examine how the pH modulates the dimensions of 4E-BP2. The FRET efficiency histograms are shown in Figure 3-16; the net charge between residue 32 and 91 and FRET values of the center of the Gaussian non-zero peak are listed in Table 3-5. In each phosphorylation state, $\langle E \rangle$ values at pH 7.4 are the same as the value at pH 10. Although the net charge increases by 2 units from pH 7.4 to pH 10, only the two Cys amino acids used for labelling shift from neutral to -1 charge whereas the charges of all other amino acids remain the same (see Fig. 3-17 (A)). The results indicate that the charge changes at the labeling sites do not influence the $\langle E \rangle$ values.

At pH 3, the histograms shift to the high-FRET range, with the $\langle E \rangle$ values virtually the same for the two phosphorylation states of the protein (Fig. 3-16 (C) and (F)). The transition
to high FRET efficiency for the HP 4E-BP2 can be explained from the perspective of the charge balance, since the net protein charge between residues 32 and 91 reduces from -10 to 1 at pH 3. However, the FRET histogram of the HP 4E-BP2 at pH 3 is very different compared to that of the NP 4E-BP2 at pH 7.4 (Table 3-5), which had a similarly low net electrostatic charge. Furthermore, the net charge of the NP 4E-BP2 increases dramatically when pH drops from 7.4 to 3. However, this does not lead to a smaller \( \langle E \rangle \) value, rather, a higher \( \langle E \rangle \), which is also very close to the HP situation.

It is obvious that these smFRET data do not completely obey the electrostatic repulsion trend, i.e., that \( \langle E \rangle \) values should shift to lower values and result in chain expansion when the net charge of the polypeptide chain increases and transit to higher values and result in chain compaction when the net charge decreases.

**Figure 3-16.** pH dependent smFRET efficiency histograms of NP (A-C) and HP (D-F) 4E-BP2. The solid lines represent Gaussian fitting, with fitting parameters given in Table 3-5. Red lines are the
overall fitting curves and blue lines are individual Gaussian peaks. The zero-FRET peak corresponds to molecules without (active) acceptors.

Why does 4E-BP2 collapse at low pH? To answer this, we first examined the quantum yields of the protein-conjugated fluorophores. They were found to be similar to the values measured at pH 7.4. This implies that the change of the protonation state of 4E-BP2 has little effect on the photophysical properties of the dyes. Protein aggregation could be another cause for high FRET. Since it was reported that low pH accelerates the aggregation of α-synuclein [291], we used FCS to measure the $R_H$ of 4E-BP2 at pH 3. In both NP and HP cases, $R_H = 27.6 \pm 0.6 \, \text{Å}$, which is slightly smaller than $R_H$ values in physiological buffer ($R_{H,NP} = 29.0 \pm 0.5 \, \text{Å}, R_{H,HP} = 31.1 \pm 0.6 \, \text{Å}$). In the absence of an increase in $R_H$, we conclude that the collapse of the 4E-BP2 is not due to protein aggregation at low pH.

**Table 3-5.** The net charge of the 4E-BP2 protein between residues 32 and 91 at different pH values and the non-zero FRET efficiencies estimated by Gaussian fitting of the smFRET histograms.

<table>
<thead>
<tr>
<th></th>
<th>NP Charge</th>
<th>NP $&lt;E&gt;$ a</th>
<th>HP Charge</th>
<th>HP $&lt;E&gt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH 7.4</td>
<td>0</td>
<td>0.57 ± 0.13</td>
<td>−10</td>
<td>0.30 ± 0.13</td>
</tr>
<tr>
<td>pH 10</td>
<td>−2</td>
<td>0.54 ± 0.15</td>
<td>−12</td>
<td>0.29 ± 0.16</td>
</tr>
<tr>
<td>pH 3</td>
<td>+6</td>
<td>0.71 ± 0.11</td>
<td>+1</td>
<td>0.74 ± 0.11</td>
</tr>
</tbody>
</table>

a The error values are the HWHM of the FRET efficiency histogram.

The fundamental difference between pH 7.4 and pH 3 is the protonation of 6 amino acids (4 aspartic acids, 1 glutamic acid, and 1 histidine) within the 32-91 region. It has been suggested that the protonated amino acids promote the formation of intra-chain hydrogen bonds and increase the hydrophobic interactions at low pH conditions [257]. These intramolecular interactions could compensate or even overwhelm the electrostatic repulsion forces between the protonated residues at low pH. We also calculated the mean hydrophobicity value, which increases from 11.65 at pH 7 to 13.64 at pH 2 on a 100-point scale for the entire chain...
The hydrophobicity distribution was plotted in Figure 3-17 (B). It is shown that there are more positive peaks (more hydrophobic at low pH) than negative peaks for both the entire chain and the region flanked by the FRET probes. In addition, there are five outstanding positive peaks within the FRET probes and two of them are located near the probe positions. The increased hydrophobicity could possibly collapse the region within the FRET probes. Or it is likely that the increased hydrophobicity at the probes area bring the two probes closer. It is also possible that the increased hydrophobic interaction between N- and C- termini bring the regions where FRET probes are located close to each other. It should be noted that chain compaction at low pH does not significantly change the polymer behavior of an IDP [17, 294] thus these collapsed states at low pH are not folded states.

**Figure 3-17.** (A) Charge distribution plot of NP 4E-BP2 at pH 10, 7.4 and 3. (B) (top) Hydrophobicity distribution plot of 4E-BP2 at pH 7 (green) and pH 2 (orange); (bottom) Hydrophobicity difference between pH 2 and pH 7 (pH 2 – pH 7). The hydrophobicity values at pH 7 are from [292] and pH 2 values are from [293] and they were scaled to a maximal value 100. Blue stars indicate phosphorylation sites and magenta circles are fluorophore locations. Grey shaded areas are outside the FRET probes region.

### 3.4.6.7 $R_G$ and shape factor inference from smFRET data

The dimensions of IDPs depend critically on the environment. A quantitative approach for estimating the dimensions of IDPs makes use of the mean FRET efficiency $\langle E \rangle$ measured
in smFRET experiments [200, 222]. The globule-to-coil transition theory of Sanchez [245] can be used to extract the chain dimensions information, such as $R_G$ and $\sqrt{\langle R_{EE}^2 \rangle}$, from $\langle E \rangle$. The Sanchez polymer model describes the chain dimensions by taking into account the excluded volume effect and intra-chain interactions [225, 245].

Table 3-6 shows the $R_G$ and $\sqrt{\langle R_{EE}^2 \rangle}$ values of NP 4E-BP2 obtained by using the Sanchez model and the smFRET $\langle E \rangle$ values measured in various denaturant and salt concentrations. The ratio between $R_G$ and $R_H$ is known as the shape factor and is expected to be around 1.5 for a random coil polymer [295]. From FCS measurements at 6 M GdmCl, we found that $R_H$ is $30.4 \pm 0.8$ Å for NP 4E-BP2. Combining these values with $R_G$ values from Table 3-6, a shape factor of $\sim 1.2$ for NP 4E-BP2 was obtained. Our results are in agreement with the results of Sherman and Haran [225], which showed that $R_G/R_H \sim 1.2$ for the IgG binding domain of protein L at 7 M GdmCl.

**Table 3-6.** $R_G$ and $\sqrt{\langle R_{EE}^2 \rangle}$ values of NP 4E-BP2 under different solvent environment inferred from smFRET data using Sanchez model.

<table>
<thead>
<tr>
<th>NP</th>
<th>$R_G$ (Å)</th>
<th>$\sqrt{\langle R_{EE}^2 \rangle}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS</td>
<td>21.7 ± 1.2</td>
<td>52.5 ± 0.5</td>
</tr>
<tr>
<td>1M GdmCl</td>
<td>27.3 ± 1.6</td>
<td>65.3 ± 0.9</td>
</tr>
<tr>
<td>3M GdmCl</td>
<td>33.9 ± 3.8</td>
<td>81.8 ± 1.4</td>
</tr>
<tr>
<td>6M GdmCl</td>
<td>37.0 ± 5.5</td>
<td>89.3 ± 1.9</td>
</tr>
<tr>
<td>6M urea</td>
<td>36.8 ± 5.1</td>
<td>88.9 ± 1.8</td>
</tr>
<tr>
<td>0M KCl</td>
<td>19.9 ± 1.2</td>
<td>48.3 ± 0.8</td>
</tr>
<tr>
<td>3M KCl</td>
<td>21.3 ± 1.2</td>
<td>51.4 ± 0.6</td>
</tr>
</tbody>
</table>
Corroborated with the segmental rotational data at 6 M GdmCl (Fig. 3-5 (B)), these estimates of the $R_g/R_H$ ratio suggest that 4E-BP2 probably does not expand sufficiently to reach complete random coil dimensions at 6 M GdmCl. Sherman and Haran [225] also suggested that hydration and the binding of the guanidinium ion to denatured proteins would lead to an increased $R_H$ and a reduced shape factor.

3.5 Conclusions

Single molecule fluorescence techniques, such as smFRET, FAD and FCS, were used to study the conformations, dynamics and binding of 4E-BP2. FAD method was employed to obtain the local chain flexibility at various points within the 4E-BP2 protein. A heterogeneous local flexibility behavior was observed throughout the protein sequence and regional amino acid composition was found to play an essential role in determining the rigidity map of the chain. It was observed that the segmental flexibility is hindered only in the region of the protein that folds upon phosphorylation, whereas the rest of the chain surprisingly becomes more flexible. We hypothesize that the increased segmental flexibility in the binding sites might result in the low binding affinity to eIF4E. In chemical denaturants, the segmental rotational correlation times at different positions along the chain are significantly faster, implying that the local segments of 4E-BP2 become more flexible. This also indicates that the intrinsically disordered 4E-BP2 has distinct structural properties compared to random coils. Anisotropy curves decay to asymptotic levels in the presence of eIF4E, which clearly shows how local tumbling motions in 4E-BP2 become constrained upon binding to eIF4E. The segmental chain flexibility close to the binding sites, even extending to C terminal, are considerably affected by binding, whereas there is little effect in the N-terminal region. These results also illustrate how the FAD method can be applied to probe the relative binding strength locally, at the level of individual protein segments.

Intra-molecular dynamics of 4E-BP2 was accessed by the FCS technique. Heterogeneous local quenching kinetics were observed. In native 4E-BP2, three kinetic
components were required in all the probe positions. In HP and chemically denatured states, only two kinetic components were found for all labelling sites except at the N-terminus (which needed an extra sub-μs component). The time scales of these kinetic components differ in an order of magnitude. The fastest time component in all cases was assigned to be introduced by proximal amino acid quenching. The quenching dynamics slows down upon phosphorylation and denaturation probably is because the quenching efficiency is reduced due to the expansion of the chain observed in smFRET measurements. Long-range interactions in 4E-BP2 were inferred because local dynamics slows down upon phosphorylation not only in the regions that are close to the phosphorylation sites, but also in the distal areas.

We also examined the inter-molecular kinetics in the 4E-BP2/eIF4E complex system using the FCS approach. As with 4E-BP2 in native condition, three dynamic components were needed to fit the data. The quenching rates near the binding sites increase significantly because multiple tryptophan and tyrosine residues are involved in the binding interface of eIF4E. The on and off rates in the secondary binding sites are on the order of $10^6$ s$^{-1}$ and $10^7$ s$^{-1}$, respectively, highlighting the dynamic nature of the binding complex.

smFRET measurements were performed to study the 4E-BP2 chain dimensions in different solvent conditions. Remarkably, the mean FRET efficiency $\langle E \rangle$ decreases from $\sim$0.6 to $\sim$0.3 after HP instead of the increasing expected for a disorder-to-order transition. The chain expansion is caused by the increase of the net protein charge after phosphorylation. In addition, the $\langle E \rangle$ value of the HP 4E-BP2 experienced a rollover phenomenon with the increase of [GdmCl]. The initial collapse of the disordered state is due to charge screening by ions in solution. Similarly, $\langle E \rangle$ transits to higher values with increasing [KCl]. These results demonstrate that electrostatics plays a critical role in modulating the dimensions and compactness of IDPs. Based on the Debye length estimates, we calculated the charge-charge interaction force ratio between low and high salt conditions. We inferred that additional intramolecular interactions, such as hydrophobic interaction and hydrogen bonding, must take place upon protein phosphorylation and lead to more compact protein conformations. The smFRET data at pH 3 suggest that these interactions could compensate or even overcome the electrostatic repulsion between phosphate groups. A non-cooperative transition was found in a
denaturation smFRET series on the 4E-BP2 in both phosphorylation states, which points to rapid interconverting species in a dynamic ensemble. $R_g$ and $\sqrt{\langle R_{EE}^2 \rangle}$ values were inferred from a Sanchez polymer model and the smFRET data. Estimates of the shape factor suggest that 4E-BP2 is not fully expanded to the state of a random-coil polymer at 6 M GdmCl.

Our work presented here is the initial step of a broader collaborative effort with the Prof. Julie Forman-Kay and Hue Sun Chan groups to characterize the conformation and dynamics of 4E-BP2. This study will bring together single molecule fluorescence, NMR and computational methods to paint a coherent picture of this IDP that is highly involved in with neurodegenerative disorders, such as autism spectrum diseases.
Chapter 4
Lipids and Fluorophores --- Friends or Foes?

4.1 Abstract

Non-specific interactions between lipids and fluorophores can cause potential artifacts in single-molecule studies of membrane proteins in cells or of cytosolic proteins encapsulated in liposomes. To gain insight into these effects, interactions between 9 commonly used fluorophores that cover the full visible spectrum range and 6 commonly used lipids that include cationic, zwitterionic and anionic types were examined. Fluorescence correlation spectroscopy probed the diffusion of dyes in the presence and in the absence of set amounts of lipid vesicles in order to quantify these interactions. Partition coefficients and free energies of partitioning for different fluorophore-lipid pairs were obtained. This data will provide the single-molecule and the lipid research community new insights into how to perform fluorescence spectroscopy in a lipid environment while minimizing the non-specific interaction between fluorescent probes and the lipids. Lipids with different head groups and with different degrees of chain saturation were investigated. Hydrophobic and electrostatic interactions, steric effects and polarity-induced shifts in the fluorescent emission spectrum were discussed. Fluorescence images of the dye either encapsulated in giant unilamellar vesicles or upon incubation on supported lipid bilayers helped visualize the colocalization between the dye and the lipid and thus the interaction between the two molecules.

4.2 Introduction

Protein-lipid interaction has been one of the major areas of research in biophysics and biochemistry. However, the mechanisms by which the lipid membrane modulates the protein structure, dynamics and function to sustain a wide range of cellular processes are still not fully understood [77]. With the development of various fluorescence-based techniques and the
advent of brighter and more photostable organic fluorophores, investigations of protein-lipid interactions using fluorescently labeled proteins have become common [123, 296-301]. Meanwhile, studies of conformations of membrane proteins in lipid vesicles, nanodiscs or live cells via fluorescence techniques are emerging [302-304].

Haran and co-workers [305] pioneered the encapsulation of cytosolic proteins into 100 nm lipid vesicles for single-molecule fluorescence (SMF) studies. The vesicles were then immobilized on supported lipid bilayers (SLB) via biotin-avidin interaction; this scheme helped prevent non-specific protein-surface interactions. Following that, trapping protein molecules into lipid vesicles to study protein conformation, dynamics and interactions became a rather common tool [117, 118, 120, 122, 306, 307]. Recently, the encapsulation protocol was also implemented in studying the conformations of intrinsically disordered proteins [157, 308].

The most remarkable feature of single-molecule spectroscopy is to identify heterogeneous, rare and transient states of the biological systems. However, the underlying questions facing researchers are whether conjugated fluorophores exhibit non-specific interactions with the protein or with the lipid membranes in the studies of membrane proteins or of cytosolic proteins encapsulated in lipid vesicles. The non-specific interactions refer to any interaction that introduces artifacts in the proposed experiments. Here, they could be π-π interaction introduced by aromatic rings in the fluorophore and amino acids in the protein, or hydrophobic interaction caused by the fluorophore and hydrophobic residues in the protein, or hydrophobic and electrostatic interactions between fluorophores and lipids. These effects must be measured and understood using proper controls, and a cautionary approach must be taken when analyzing and interpreting the SMF experimental data. Molecular dynamics simulation [309] has been performed for two commonly used fluorophores (Alexa 488 and Alexa 594) to address the fluorophore-protein interaction. Vesicle-fluorophore dialysis method [310] has been used to characterize the fluorophore-lipid interaction; however, a systematic quantification of the interaction between popular SMF fluorophores and various type of lipids has not been performed using single-molecule techniques.
In this chapter, we focus on finding the degree to which commonly used fluorophores interact with lipid membranes in SMF studies. This will provide new insights into selecting fluorophore-lipid pairs that are suitable for studying the conformations of proteins trapped in lipid vesicles and the interactions between proteins and lipids at single-molecule resolution. We chose nine commonly used fluorophores in SMF studies and six different lipids that are widely used for making lipid vesicles. The fluorophores have different structure, charge and hydrophobicity; the lipids selected also have different charge, head groups and chain saturation. Fluorescence correlation spectroscopy (FCS) was corroborated with standard fluorescence imaging techniques to visualize and quantify the partition of each fluorophore partition into different lipid membranes. FCS is a technique that is based on the analysis of fluorescence intensity fluctuations of molecules diffusing in and out of the femto-liter focal volume and has single-molecule sensitivity [193]. It is a powerful tool to measure concentration, translational diffusion and chemical reactions [311-313]. The fact that large unilamellar vesicles (LUVs) are much bigger than the average size of a fluorophore makes FCS an ideal tool for resolving the diffusion times between free and vesicle-bound fluorophores and to estimate the bound fraction even when the interaction is rather weak. It has been applied to study the interaction between lipid vesicles and different peptides and proteins [299, 314-317]. Wide-field and total-internal reflection fluorescence (TIRF) imaging of giant unilamellar vesicle (GUV) and SLBs were also performed to provide a visualization of the interaction between fluorophores and lipid bilayers.

4.3 Materials and Methods

4.3.1 Materials

Alexa 488 and Alexa 647 were purchased from Molecular Probes (Thermo Fisher Scientific, Canada). Atto 488, Atto 532, Atto 565, Atto594 and Atto 647N were obtained from ATTO-TEC (ATTO-TEC GmbH, Siegen, Germany). Fluorescein (FL) was purchased from Sigma-Aldrich (cat. # F2456, Sigma, USA). Cyanine 5 (Cy5) was purchased from Lumiprobe
(Lumiprobe Corporation, Florida, USA). All the fluorophores except FL have a maleimide conjugation group, which is used for thiol chemistry to label cysteine residues in the proteins.

Bodipy-FL C5 lipid (cat. # D3834) was purchased from Thermo Fisher Scientific. 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) (cat. # 850375), 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG) (cat. # 840475), 1,2-dioleoyl-sn-glycero-3-phospho-L-serine (DOPS) (cat. # 840035), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (cat. # 850457), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) (cat. # 850355), 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) (cat. # 890890) and Cholesterol (Chol) (cat. # 700156) were purchased from Avanti Polar Lipids (Alabama, USA). All the lipids were dissolved in 95% chloroform and 5% methanol, unless stated otherwise. All the phospholipid concentrations were accessed by phosphorus assay [318, 319]. All the lipid solutions were stored in parafilm wrapped glass vials in -20 °C.

4.3.2 Sample preparation

4.3.2.1 Small and large unilamellar vesicles preparation

A desired lipid quantity from the lipid stock was transferred to a new vial. The lipid solution was dried under nitrogen to form lipid films and was left in vacuum desiccator overnight to remove all the chloroform/methanol residues. The lipid films were hydrated for 2 hours in a phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂SO₄, 1.8 mM KH₂PO₄, pH 7.4). Small unilamellar vesicles (SUVs) with a diameter of 50 nm and large unilamellar vesicles (LUVs) with a diameter of 100 nm were prepared using the extrusion method [320]. In short, the hydrated vesicle solution was pushed 35 times through a porous polycarbonate membrane filter (50 nm or 100 nm) back and forth using a Mini-Extruder apparatus (cat. # 610000, Avanti Polar Lipids, USA) to produce SUVs or LUVs. DPPC vesicles were extruded at temperature of around 55 °C. The other vesicles were prepared at room temperature. Fluorescent LUVs were prepared by mixing 0.01 mol% Bodipy-FL lipid. The sizes of vesicles were assessed by dynamic light scattering (DLS). After using the 50 nm
membrane filter, the average diameter for DOPC vesicles is 60.0 ± 4.4 nm. The average vesicle diameters for DOTAP, POPC, DOPC, DPPC, DOPS and DOPG LUVs after the 100 nm membrane filter were 104.3 ± 10.9 nm, 118.9 ± 7.0 nm, 115.6 ± 11.1 nm, 114.8 ± 5.8 nm, 110.0 ± 14.8 nm, 108.3 ± 19.7 nm, respectively. Here, the errors were derived from the standard deviation of several mean values from multiple measurements. The highest lipid stock concentration prepared in PBS buffer was 80 mg/mL. All the lipid vesicles were stored in 4 °C and were used within 3 days after preparation.

4.3.2.2 Giant unilamellar vesicle preparation

Giant unilamellar vesicles (GUVs) were prepared according to the method outlined in [86]. In short, two Indium-Tin-Oxide (ITO) coverslips were rinsed in Milli-Q water and then 95% ethanol two times, followed by drying with argon gas. The ITO coverslips were then sonicated in 50% methanol/50% chloroform for 10 minutes and then dried again with argon. A piece of taped aluminum foil was attached to each ITO coverslip for the purpose of connecting voltage across two ITO coverslips. Lipids were diluted to 5 mg/mL in the 95% chloroform/5% acetonitrile. A 50-µL lipid solution was dropped on the coverslip and spin coated at 600 rpm to form a thin lipid film. A 1-mm poly(dimethylsiloxane) spacer was sandwiched between the two ITO coverslips and the chamber was filled with a 100-µL solution of 100 mM sucrose. In the fluorophore encapsulation case, the final fluorophore concentration of 50 nM was used. A sine wave (1 V_pp, 10 Hz) was applied via a function generator for 2 hours and a 3 V_pp square wave at 5 Hz was applied to detach the vesicles from the surface. The remaining solution from the chamber was pipetted using syringe and transferred to a microscope coverslip for wide field imaging.

4.3.2.3 Supported lipid bilayer preparation

A solution of 10 mg/mL SUVs in a PBS buffer containing 4 mM CaCl_2 was sandwiched between clean coverslips. After incubation for about 2 hours at room temperature, the
coverslips were separated and rinsed with PBS buffer thoroughly [90]. Supported lipid bilayer (SLB) formation was verified by fluorescence recovery after photobleaching (FRAP) experiment using 90 mol% DOPC and 10 mol% Bodipy-FL lipids (data not shown).

4.3.2.4 Organic fluorophore preparation

All fluorophore powders were dissolved in anhydrous dimethyl sulfoxide (DMSO) to be 20 mM stock. Each stock solution was diluted to nM concentrations into the PBS buffer for FCS measurements. The concentrations of all fluorophores were examined by FCS before the lipid-fluorophore interaction studies to guarantee that the fluorophore concentration used in all FCS experiments were the same unless stated otherwise. All FCS experiments were performed under fluorophore-hydrolysis conditions, in which the fluorophores were kept at 4 °C for 3 days after dilution from the stock solutions to PBS buffer, in order to mimic the circumstances of real single-molecule experiments.

4.3.2.5 Polyethylene glycol surface

The preparation of polyethylene glycol (PEG) surface was followed by the protocol written by T. Ha group [307]. Briefly, the coverslips were rinsed and sonicated for 30 minutes in the following sequence: Milli-Q water --- MeOH (cat. # 34860-1L-R, Sigma Aldrich, USA) --- acetone (cat. # 270725-1L, Sigma Aldrich) --- fresh 1 M KOH solution (cat.# 6160-1-70, Caledon Lab. Canada). The coverslips were then raised with Milli-Q water and dried with argon gas. Fresh 230 g/L m-PEG-silane (cat. # mPEG-silane, MW 5,000, Lysan Bio. USA) in 10 mM sodium bicarbonate buffer (pH 8.5) was used to coat the coverslips. Incubation was performed in the dark overnight in a humidified PEG incubation container at room temperature. After incubation, each coverslip was washed by 10 mM sodium bicarbonate buffer, followed by Milli-Q water and then dry with argon gas. The PEG coated coverslips were stored in the vacuum desiccator in the dark before use. The mixture of fluorophore and lipid solution was
first incubated for 5 min and then dropped onto the PEG coated coverslip to prevent non-specific adsorption to the coverslip.

4.3.3 Instrumentation

4.3.3.1 Dynamic light scattering

Dynamic light scattering (DLS) measurements were performed on a Malvern Zetasizer Nano-ZS instrument (Malvern Instruments Ltd., Worcestershire, UK). Standard calibration beads samples were used to confirm the accuracy of the instrument. The data recorded provided the mean diameter and the full distribution of vesicle sizes. All DLS measurements were carried out at 20 °C.

4.3.3.2 Fluorescence emission spectrum

Fluorophore fluorescence emission measurements were acquired on QuantaMaster PTI spectrofluorimeter (Photon Technology International, Canada) equipped with a red-sensitive photomultiplier tube (R928P, Hamamatsu). 480 nm, 530 nm and 630 nm wavelengths were used to excite the blue-excitation fluorophores (Alexa 488, Atto 488 and FL), green-excitation fluorophores (Atto 532, Atto 565 and Atto594) and red-excitation fluorophores (Alexa 647, Atto647N and Cy5), respectively. The starting points for collecting all the emission spectra were 10 nm red shifted from the excitation wavelengths. All emission spectra were collected at 20 °C.

4.3.3.3 FCS experiments

FCS measurements were performed on a custom-built confocal microscope. The samples were excited with continuous-wave (CW) lasers through an oil-immersion objective
(1.40NA/100X UPlanSApo, Olympus, USA). 488 nm (TECBL-488, World Star Tech, Canada), 532 nm (TECGL-532, World Star Tech, Canada) or 633 nm (TECRL-633, World Star Tech, Canada) lasers were employed for excitation depending on the absorption spectrum of the fluorophore. The fluorescence beams were routed through a 30-µm diameter pinhole and a combination of long-pass and band-pass filters. The fluorescence signal was divided into two quasi-equal channels using a non-polarizing cube beamsplitter and detected by two single photon-counting avalanche photodiodes (SPCM-CD 3017, Perkin Elmer Optoelectronics, Canada). The following combination of emission filters were used: (1) BLP01-488 (Semrock, USA) and HQ 520/66 (Chroma, USA) for Alexa 488/Atto 488/FL, (2) BLP01-532 (Semrock, USA) and HQ 575/80 (Chroma, USA) for Atto 532, (3) BLP01-532 and HQ 600/80 (Chroma, USA) for Atto 565/Atto 594, and (4) BLP01-647 (Semrock, USA) and HQ 685/80 (Chroma, USA) for Alexa 647/Atto 647N/Cy5. TTL signal output from the detectors was fed into a 4-channel hardware correlator (Flex02-01D, Correlator.com), which provided autocorrelation and cross-correlation curves.

4.3.3.4 Fluorescence lifetime experiments

Fluorescence lifetime measurements were performed on a custom-built multiparameter SMF microscope that was described in detail elsewhere [121, 163, 247]. Pulsed laser excitation at 480 nm or at 527 nm was provided by frequency-doubling the output of a tunable femtosecond oscillator (Tsunami HP, Spectra Physics, Santa Clara, CA, USA). The fluorescence from the sample was focused through a 75-µm diameter pinhole and then divided into two detection channels by a polarizing cube beamsplitter. Each beam was focused onto a single-photon avalanche diode (SPAD) (PD5CTC, MPD, Italy). The microscope objective and the emission filters were the same as described in the FCS section above. Fluorescence intensity and anisotropy decay curves on the nanosecond scale were acquired using a routed multichannel time-correlated single-photon counting system (PicoHarp300, PicoQuant GmbH, Germany).
4.3.3.5 Fluorescence imaging

The imaging was performed on a custom-built total internal reflection fluorescence microscope (TIRFM) described in detail in Chapter 1. The fluorophores were excited either at 473 nm, 532 nm or 633 nm depending on their absorption spectrum. All CW lasers were controlled by an acousto-optic tunable filter (Gooch & Housego, USA). The excitation light was reflected by a quad-edge dichroic mirror (Di01-R405/488/532/635, Semrock, USA) and illuminated the sample by the collimated beam after passing through an oil-immersion microscope objective (1.45NA/60X Plan-Apochromat, Olympus, USA). The fluorescence emission signals were collected using the same objective and transmitted through the quad dichroic mirror. In order to remove scattering and cross-talk signals, the fluorescence images were collected using a combination of a long-pass and a band-pass filters. The filters used in the imaging were the same as the ones used in the FCS measurements. Images were collected using a cooled, ultrasensitive electron-multiplied charge-coupled device (EMCCD, DU-897BV, Andor Technology, USA). The wide-field mode was used for imaging the GUVs and the TIRF mode was applied to image the SLBs.

4.3.4 LogD calculations

LogD is the log of the distribution coefficient $K_D$, which is the ratio of the sum of the concentrations of all species in non-polar solvent to the sum of the concentrations of all species in polar solvent. A molecule with positive logD value is hydrophobic and a molecule with negative logD value is hydrophilic. The more positive the logD is, the more hydrophobic the molecule is. LogD values at pH 7.4 were calculated using Marvin Sketch software (ChemAxon, Cambridge, MA) based on the structures of lipids and fluorophores and the PBS solvent condition, shown in Appendix A and B. Consensus method (based on model proposed by [321] and PHYSPROP© Database) was used in logD calculation and tautomerization (isomerization reaction where tautomers are interconverted) and resonance were considered in the logD calculations in the software [310].
4.3.5 FCS analysis

The fluorescence intensity correlation function for multiple molecular species with different diffusion and triplet components is given by [193]:

\[ G(\tau) = \frac{1}{\Sigma_{l} q_{l}^{2} N_{l}} \sum_{l} q_{l}^{2} N_{l} \left( 1 + \left( \frac{\tau}{\tau_{d,l}} \right) \right)^{-1} \left( 1 + \frac{1}{s^2} \left( \frac{\tau}{\tau_{d,l}} \right) \right)^{-0.5} \left( 1 + K_{l} e^{-t/\tau_{k,l}} \right). \]  (4-1)

In eq. (4-1), \( s \) is the ratio between the axial and the lateral radii of the ellipsoidal detection volume, \( q_{l} \) is the molecular brightness, \( N_{l} \) is the average number of molecules in the detection volume, \( \tau_{d,l} \), \( K_{l} \) and \( \tau_{k,l} \) are the diffusion time, triplet fraction, and triplet lifetime, respectively, for the \( l^{th} \) molecular species.

In our lipid-fluorophore interaction experiments, only free fluorophores and fluorophore-bound vesicles will give off fluorescence, so that eq. (4-1) can be reduced to the following two species form:

\[ G(\tau) = \frac{1}{N_{\text{tot}}^{2}} \left[ f \left( 1 + \left( \frac{\tau}{\tau_{d1}} \right) \right)^{-1} \left( 1 + \frac{1}{s^2} \left( \frac{\tau}{\tau_{d1}} \right) \right)^{-0.5} \left( 1 + K_{1} e^{-t/\tau_{k1}} \right) + (1 - f) \left( 1 + \left( \frac{\tau}{\tau_{d2}} \right) \right)^{-1} \left( 1 + \frac{1}{s^2} \left( \frac{\tau}{\tau_{d2}} \right) \right)^{-0.5} \left( 1 + K_{2} e^{-t/\tau_{k2}} \right) \right]. \]  (4-2)

Prior to each set of measurements, a dilute solution of a dye with known diffusion coefficient (Rhodamine 110 for blue excitation, Rhodamine 6G for green excitation and Alexa 647 for red excitation) was used to characterize the detection volume [163]. Control FCS experiments in PBS buffer were performed for each fluorophore used in this study to obtain the concentration, diffusion coefficient and triplet information. The estimated diffusion times for the free fluorophores were set to be the initial guess value of \( \tau_{d1} \) for the unbound fluorophore species in eq. (4-2). 95% and 105% of the fluorophore diffusion time were set to be the lower and upper boundaries of \( \tau_{d1} \) and 90% and 110% of the fluorophore triplet fraction and time were set to be the lower and upper boundaries of \( K_{1} \) and \( \tau_{k,1} \). These boundary settings were based
on multiple measurements of the fluorophore only sample under the same laser excitation condition.

Converting from hydrodynamic radius of the LUVs measured in DLS to diffusion time via the Stokes-Einstein equation [163], an average diffusion time was chosen as the initial guess for the fluorophore-bound LUV species \( \tau_{d2} \). The DLS size distribution was used to set lower and upper boundaries of the \( \tau_{d2} \) in eq. (4-2). It is noted that in all FCS experiments \( \tau_{d2} \) should be in fact a distribution of diffusion times, corresponding to the vesicle size distribution measured by DLS. However, it was found that using an average \( \tau_{d2} \) value to represent the distribution of \( \tau_{d2} \) has negligible effect on the general FCS curve shape and potential fitting results (see Figure 4-1). In addition, the vesicle size obtained using the one-component FCS fitting function for 0.01 mol% Boidpy-FL DOPC LUVs (102.22 ± 3.2 nm in diameter) was similar to the average vesicle size measured by DLS. Thus, a single component was used in the fitting routine to describe the diffusion of the vesicles. The parameters \( K_2 \) and \( \tau_{k,2} \) include the fluorophore triplet formation [128], photobleaching and blinking [322-324] upon binding to a lipid vesicle.

**Figure 4-1.** Simulations of a group of FCS curves with diffusion times using average values (dash lines) and normal distributions (solid lines) of each component in different bound fraction percentages (red: 0%, green 20%, blue: 50%, magenta: 80%). 50 ± 3 µs and 5000 ± 1500 µs used in the simulation were similar to the diffusion time of free fluorophores and LUVs in our experiments.
By fitting the measured correlation curve to the model described by eq. (4-2), one can obtain the fraction of unbound fluorophore, $f$. It is noted that the $f$ in eq. (4-2) is the pseudo-free fluorophore fraction. Only when the brightness of the free fluorophore and the fluorophore bound complex are identical, the pseudo-free fluorophore fraction becomes the real free fluorophore fraction. Otherwise, the brightness ratio $\beta = q_2 / q_1$ needs to be taken in account to obtain the real free fluorophore fraction, $f' = \frac{\beta^2 f}{1-f+\beta^2 f}$. Global analysis on sets of independent FCS data, consisting of one type fluorophore and one type of lipid at various lipid concentrations, were performed to reduce the statistical correlations among different fitting parameters.

4.3.6 Theory of partition coefficient and free energy of partitioning

Different from the distribution coefficient, $K_D$, which is the ratio of the sum of the concentrations of all species of the compound in a mixture of two immiscible phases, the partition coefficient ($K_p$) is the concentration ratio of a specific compound in a mixture of two immiscible phases in a state of equilibrium. The interaction between fluorophores and lipids can be described as two immiscible phases, i.e. aqueous buffer and hydrophobic lipid bilayer, whereby $K_p$ was used here to quantify the degree of fluorophore-lipid interaction. The partition coefficient is defined as [325]:

$$K_p = \frac{[\text{Dye}]_{\text{lipid}}/[\text{Lip}]+[\text{Dye}]_{\text{lipid}}}{[\text{Dye}]_{\text{water}}/([\text{Water}]+[\text{Dye}]_{\text{water}})}.$$

(4-3)

In all our experimental conditions, the water concentration $[\text{Water}] = 55.5 \, M$ which is much higher than the dye concentration in water, and the accessible lipid concentration $[\text{Lip}]$ is also much greater than the lipid bound dye concentration, so that the eq. (4-3) can be simplified to

$$K_p = \frac{[\text{Dye}]_{\text{lipid}}/[\text{Lip}]}{[\text{Dye}]_{\text{water}}/[\text{Water}]}.$$

(4-4)
The free energy of partitioning $\Delta G$ associated with the partition coefficient $K_p$ is

$$
\Delta G = -RT\ln K_p,
$$

where $R$ is the ideal gas constant and $T$ is the absolute temperature. An order of magnitude change in $K_p$ at 20 °C will result in 5.61 kJ/mol change in $\Delta G$.

The lipid-bound fluorophore fraction $\xi$ can be related to the partition coefficient $K_p$ after simple algebraic manipulation,

$$
\xi = \frac{K_p[lip]}{[Water]+K_p[lip]}.
$$

Thus, from the vesicle-bound fluorophore fraction at various lipid concentrations, we can extract the partition coefficient, $K_p$, of a specific fluorophore-lipid pair and also the standard free energy for their interaction.

### 4.4 Simulations

Multiple fluorophores can bind to the same vesicle if fluorophore concentration is high. Multi-binding phenomenon could introduce quenching of fluorescence, the existence of different brightness species, and anti-cooperative phenomenon (due to charge-charge repulsion). To avoid these artefacts and to ensure ideal experimental conditions, we estimated the average number of fluorophores per vesicle in various fluorophore concentrations and the corresponding partition coefficients. Assuming that the LUVs are monodispersed and that their morphology is not influenced by the number of fluorophores bound to the LUVs, the average number of lipid bound fluorophores per LUV can be described as [301]

$$
\langle k \rangle = \frac{m[dye]_{tot}}{[lip]_{tot}},
$$
where $m$ is the average number of lipid molecules in a vesicle, $[Dye]_{tot}$ is the total fluorophore concentration and $[Lip]_{tot}$ is the total lipid concentration. A 100-nm diameter vesicle with a lipid head group size of 0.63 nm$^2$ [326] was used in the simulation. Simulations were performed with $K_p = 10^1 - 10^6$ and a family of curves with $K_p = 10^3 - 10^6$ are shown in Figure 4-2. The data show that 1 nM is an appropriate fluorophore concentration in the FCS experiments to ensure that there is at most a single fluorophore bound to a lipid vesicle even at very high $K_p$ values. Thus, the concentration of fluorophore used in all the FCS measurements was 1 nM, unless stated otherwise.

**Figure 4-2.** Simulated average number of lipid-bound fluorophores per vesicle in various fluorophore concentrations 0.1 nM to 4.6 nM. (A) $K_p = 10^3$, (B) $K_p = 10^4$, (C) $K_p = 10^5$, (D) $K_p = 10^6$. Grey dashed line indicates an average of one fluorophore bound per vesicle. Inset in (D) is the zoom in from 0 to 2 mM lipid concentration.

Using eq. (4-6), a family of bound fraction curves with various partition coefficient $K_p$ values were also simulated (Figure 4-3) with a 1 nM fluorophore concentration in each case. The family of simulated curves shows, as expected, that at a given lipid concentration, the greater the partition coefficient is, the larger the bound fraction will be. More importantly,
Figure 4-3 provides us an estimation of the degree to which fluorophores interact with lipid membranes when only one FCS measurement for a given lipid concentration is performed (especially other fluorophores and lipids that are not studied in this chapter). The maximal lipid concentration used in the simulation was set to be the highest concentration in forming stable vesicles (private communication with Avanti Polar Lipids, Inc.). The maximal $K_p$ value was set to be the highest value found in the literature related to ligand-membrane interaction [327].

![Graph](image)

**Figure 4-3.** Simulated vesicle-bound fraction of fluorophores (1 nM solution) interacting with 100-nm lipid vesicles in diameter at the following $K_p$ values: $10^1$ (red), $10^2$ (orange), $10^3$ (yellow), $10^4$ (light green), $10^5$ (green), $10^6$ (dark green), $10^7$ (cyan).

### 4.5 Results

#### 4.5.1 The brightness control

Eq. (4-1) indicates that brightness of the fluorescence species matters in the FCS analysis. The brightness of a given fluorophore is proportional to the product of the quantum yield and molar extinction coefficient. Because molar extinction coefficient is an intrinsic property of a fluorophore [99], it does not change upon lipid interaction. Therefore, brightness is determined by the quantum yield of the fluorophore. Here, we performed fluorescence
lifetime experiments to determine how the quantum yield and thus the relative brightness of the fluorophore changes upon interaction with the lipids.

Figure 4-4. Fluorescence intensity decays curves of dyes in the presence of varying amounts of lipids in LUV morphology: (A) Alexa 488 with DOTAP; (B) Atto 565 with DOPG; (C) Atto 594 with DOPC. For all samples, the final fluorophore concentration was 1 nM. The fitting errors of the lifetimes are 0.01 ns.

Figure 4-4 shows examples of fluorescence lifetime data and fitting for three different fluorophores in the presence of different amounts of lipids. The fluorescence intensity decay curves were fit to an exponential decay function starting from \( t = 2 \times \text{IRF} \) (IRF = 50 ps) to extract the lifetime parameters [261]. The exemplary lifetime values are presented in the legend of Figure 4-4. It is shown that the fluorophore-lipid mixtures share very similar lifetime values as the free fluorophores. The other blue and green excitation fluorophores behave similarly. Thus, we conclude that the fluorescence quantum yields and thus the molecular brightness of these fluorophores are not affected by their interaction with lipid vesicles. Due to the instrument limitations, the lifetime experiments on red excitation fluorophores could not be performed, but their basic chemical structures do not differ a lot from the blue or the green fluorophores, so we expect that their brightness will be largely invariant when bound to lipid vesicles.
4.5.2 Partition coefficient and free energy of partitioning

Figure 4-5 shows three examples of FCS titration curves and the corresponding increase of the lipid-bound fractions estimated for each of them. All the FCS curves were normalized to 1 at the time point of $10^{-5}$ sec. It is clear that the FCS curves shift to the right with increasing lipid concentration and the degree of the shift depends on the the lipid-fluorophore pairs being studied (Figure 4-5 (A-C)). Using eq. (4-2), the fractions $\xi$ of lipid-bound fluorophores at each lipid concentration were estimated for different fluorophore-lipid pairs. The larger shift indicates a higher bound fraction as demonstrated in Figure 4-5 (D-F).

![Figure 4-5](image)

Figure 4-5. Families of normalized FCS curves of (A) Atto 488, (B) Atto 594, (C) Atto 647N with different DOPC lipid concentrations. Scattered dots are the raw data and the solid lines are the fitted curves using eq. (4-2). Lipid-bound fractions $\xi$ of (D) Atto 488, (E) Atto 594, (F) Atto 647N are shown as a function of DOPC concentration. Scattered dots are the values extracted from the FCS fitting. Error bars come from the fitting of three independent data sets (same as below). Red solid curves represent fittings using eq. (4-6).

The partition coefficient, $K_p$, for each fluorophore-lipid pair can be obtained by fitting the bound fraction $\xi$ of the fluorophore as a function of lipid concentration using eq. (4-6) and the free energies of partitioning $\Delta G$ can be calculated using eq. (4-5). Table 4-1 and Appendix
C contain the $K_p$ and $\Delta G$ values, respectively, which describe the interaction between 9 commonly used single-molecule fluorophores and cationic (DOTAP), zwitterionic (DOPC) and anionic (DOPG) lipids.

**Table 4-1.** Partition coefficients $K_p$ of different fluorophores interacting with DOTAP, DOPC, and DOPG lipids. Error bars are based on fitting three independent data sets.

<table>
<thead>
<tr>
<th></th>
<th>DOTAP (+)</th>
<th>DOPC (+-)</th>
<th>DOPG (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexa 488</td>
<td>242 ± 17</td>
<td>149 ± 6</td>
<td>41.7 ± 4.7</td>
</tr>
<tr>
<td>Atto 488</td>
<td>298 ± 14</td>
<td>253 ± 15</td>
<td>73.7 ± 4.7</td>
</tr>
<tr>
<td>FL</td>
<td>197 ± 11</td>
<td>245 ± 26</td>
<td>61.6 ± 3.2</td>
</tr>
<tr>
<td>Atto 532</td>
<td>270 ± 18</td>
<td>189 ± 5</td>
<td>166 ± 18</td>
</tr>
<tr>
<td>Atto 565</td>
<td>529 ± 66</td>
<td>(1.7 ± 0.1)×10³</td>
<td>(2.5 ± 0.2)×10³</td>
</tr>
<tr>
<td>Atto 594</td>
<td>(1.4 ± 0.9)×10⁴</td>
<td>(3.5 ± 0.8)×10³</td>
<td>(2.6 ± 0.7)×10³</td>
</tr>
<tr>
<td>Alexa 647</td>
<td>(2.3 ± 0.6)×10⁴</td>
<td>15.3 ± 1.7</td>
<td>13.4 ± 2.0</td>
</tr>
<tr>
<td>Atto 647N</td>
<td>(3.3 ± 0.4)×10³</td>
<td>(1.1 ± 0.2)×10⁴</td>
<td>(1.5 ± 0.2)×10⁶</td>
</tr>
<tr>
<td>Cy5</td>
<td>(1.3 ± 0.4)×10⁵</td>
<td>(3.0 ± 0.7)×10³</td>
<td>(2.3 ± 0.4)×10⁶</td>
</tr>
</tbody>
</table>

Larger $K_p$ value corresponds to stronger partitioning of fluorophore into the lipid membrane. Comparing the $K_p$ values in Table 4-1, we found that the blue and green excited fluorophores exhibit smaller $K_p$ than red fluorophores in general. The partition coefficients of Atto 647N and Cy5 are on the same order of magnitude of the partition coefficients of fluorescent membrane probes, such as 1,6-diphenyl-1,3,4-hexatriene (DPH) and its derivatives, whose partition coefficients into DPPC membrane were determined to be $0.6 \times 10^6$ – $3 \times 10^6$ [328]. These fluorophores cannot be used to label protein in the study of protein interaction with lipids because some proteins themselves do not interact such strong with lipids.
For example, Middleton and Rhoades showed that the $K_p$ of Alexa 488 labeled $\alpha$-synuclein with POPC lipid is $\sim 800$ using FCS [300] and Melo, et al., used same method and stated that the $K_p$ of Alexa 488 labeled lysozyme with POPC lipid is $\sim 10^4$ [301]. Our results show that DOPC and POPC have similar partition strength with Alexa 488 (see section 5.5 and Figure 4-8 for details) whose $K_p$ value is $\sim 150$ (Table 4-1). This $K_p$ value is 6 times smaller than the value in the case of $\alpha$-synuclein with POPC lipid. However, it was found that the partition coefficient of a small molecule, anti-inflammatory drug, Ibuprofen with DOPC lipid is only 65 using the UV-Vis sum frequency generation technique [329]. This result suggests that of the fluorophores studied in this work, only Alexa 647 should be used to label a molecule of interest when studying protein-lipid interactions. Therefore Table 4-1 and Appendix C provide us references for the study of (macro)molecule-lipid interactions. When the $K_p$ and $\Delta G$ values obtained using fluorescence methods are similar to the ones listed in the aforementioned tables, attention needs to be placed on the re-examination of the interaction using different techniques. These $K_p$ and $\Delta G$ values might actually be a result of the fluorophore interacting with the lipid membrane rather than the molecule being studied. In the following sections, several potential factors that proliferate the interaction between lipids and fluorophores are elaborated.

### 4.5.3 The electrostatic interaction

To examine the role of electrostatic interaction in lipid-fluorophore interaction, FCS experiments in varying salt (NaCl) concentrations were conducted. Figure 4-6 shows the FCS data obtained for Atto 647N in the presence of 4 mM DOTAP in 0 M, 14 mM and 140 mM NaCl. The data clearly shows that the correlation decay curve shifts to faster decay times for decreasing [NaCl]; the corresponding lipid-bound fractions $\xi$ of Atto 647N changes from $\sim 10$ % to zero. This trend suggests that electrostatics contributes to the overall interaction between charged fluorophores and charged lipids. Here, the strength of interaction between the positively-charged Atto 647N and the positively-charged DOTAP is diminished at low salt concentrations due to lack of sufficient charge screening by counterions in solution. Higher salt concentration leads to screening of charge repulsion and favors binding between...
fluorophores and lipids of the same charge. It should be noted that no change in vesicle size was found at both low and high ionic strengths.

Figure 4-6. Normalized FCS curves of 1 nM Atto 647N in the presence of 4 mM DOTAP at different NaCl concentrations: Grey: Atto 647N alone; Red: 0 M NaCl; Green: 14 mM NaCl; Blue: 140 mM NaCl. Scatter dots are the data points and solid lines are fitting curves using eq. (4-2).

4.5.4 The effect of lipid head groups

A typical lipid molecule is composed of a hydrophilic head group and one or two hydrophobic acyl chains. The hydrophilic head groups will spontaneously expose themselves to the aqueous buffer environment whereas the fatty acyl chains will be buried together away from water due to hydrophobic effects. In single-molecule experiments of water-soluble proteins trapped in liposomes [117, 118, 121, 151, 305], the proteins and the fluorophore labels will primarily be exposed to the lipid head groups of the inner layer of the vesicle. The scheme of the encapsulation of a protein in a vesicle is shown in Figure 1-15 in Chapter 1. Investigating the effect of different types of head groups on the strength of the fluorophore-lipid interaction will provide us insights into what type of lipids to select in order to make “inert” lipid vesicles for SMF studies of cytosolic proteins immobilized via the liposome encapsulation assay.
Figure 4-7. (A) Normalized FCS curves illustrate the interaction of Atto 565 (1nM, black) with lipids (50 mM) of the same charge but different head groups: DOPS (red) and DOPG (blue). The scatter dots are data points and solid lines are fitting curves. (B) Bar chart of the lipid-bound fractions, $\xi$, of different fluorophores in the presence of DOPS (red) or DOPG (blue) lipids. All fluorophore concentrations were 1 nM. The lipid concentrations were 50 mM for Alexa 488, FL, Atto 532, Atto 565 and 15 mM for Atto 647N, Cy5 fluorophores.

PG, PS, and PA are three of the most common and most important anionic phospholipids in the cell membrane and they play key roles in cell viability [330], and cycle signaling [331]. Since the PA lipids alone are not stable and will form a hexagonal phase [332], PG and PS lipids were the focus of this chapter. The structures of DOPG and DOPS used here can be found in Appendix A. The structure difference between DOPG and DOPS is their head group. PG has two hydroxyl groups and PS has a carboxylic acid group and a protonated amine group. The sizes of the two head groups are similar to PC lipids [333], but their transition temperatures and lipid membrane packing densities are different (See Discussion).

Figure 4-7 (A) shows the FCS curves of Atto 565 interacting with DOPS and DOPG. It is clear that the FCS curve shifts to longer diffusion times when the lipid head group changes from PS to PG. Figure 4-7 (B) compares the lipid-bound fractions $\xi$ of different blue, green and red fluorophores in the presence of DOPS and DOPG lipids. Whereas the blue and green fluorophores bind less to these lipids than their red counterparts, it is clear that all of them exhibit a stronger interaction with DOPG than DOPS.
4.5.5  The effect of acyl chain saturation level

The degree of saturation of the acyl chain plays an essential role in membrane fluidity and elasticity and thus influence the intercalation and the interaction of molecules with the lipid membrane. Here, fully unsaturated DOPC, half saturated-half unsaturated POPC and fully saturated DPPC were used to investigate the extent to which the saturation of the acyl chain affects the lipid-fluorophore interaction.

**Figure 4-8.** (A) Normalized FCS curves illustrate the interaction of 1 nM Atto 647N fluorophores with 5 mM DOPC (red), POPC (blue) and DPPC (green) lipids. The scatter dots are data points and the solid lines are fitting curves. (B) Bar chart of the lipid-bound fractions, $\xi$, of different fluorophores interacting with DOPC (red) and POPC (blue) lipids. All fluorophore concentrations were 1 nM. The lipid concentrations were 30 mM for Alexa 488, FL, Atto 532, Atto 565 and 5 mM for Atto 647N, Cy5 fluorophores.

Figure 4-8 (A) shows the FCS curves of Atto 647N interacting with DOPC, POPC and DPPC lipids. It is evident that the interaction becomes weaker as the degree of chain saturation increases, as illustrated by the shift to faster decay times when non-saturated lipids are replaced by partially- and fully-saturated lipids. Figure 4-8 (B) summarizes the lipid-bound fractions, $\xi$, of six different fluorophores in DOPC and POPC, respectively. Our data indicates that the saturation level of the fatty acyl chain is an important factor for the dye-lipid interaction strength. In general, unsaturated lipids are more susceptible to exhibit favorable interactions with SMF fluorophores.
4.5.6 The effect of cholesterol

It is well known that in the cell membrane, cholesterol and sterol restrict the lipid chain motions and increase the rigidity of the neighboring lipid molecules in the bilayer [68]. In addition, cholesterol has been shown to affect drug encapsulation and release efficiency and drug loading capacity [334, 335]. Thus, by incorporating cholesterol into the lipid bilayer, we attempted to study the effect of chain rigidity on the interaction between fluorophores and lipids. This could also have potential applications for single-molecule measurements of custom-made drug delivery systems. At high concentrations, the cholesterol can precipitate out of lipid bilayers (~ 50 mol% [336]), can change the morphology of lipid vesicles [337] and it affects the phase transition and packing of lipids [338, 339]. To avoid these unwanted effects, our experiments were conducted in 15 mol% and 30 mol% cholesterol.

![Figure 4-9](image)

**Figure 4-9.** Normalized FCS curves of 1 nM Atto488 (A), Atto565 (B) with 20 mM DOPC-Chol lipids and 1 nM Atto 647N (C) with 5 mM DOPC-Chol lipids. Red: 0 mol% Chol; Green: 15 mol% Chol; Blue: 30 mol% Chol. The scatter dots are data points and the solid lines are fitting curves. Insets in (A) and (B) show a zoom in $10^{-4} \sim 10^{-2}$ sec range.

From Figure 4-9, it can be seen that the correlation decay curves shift only slightly to the left (faster diffusion times) upon incorporating cholesterol into the lipid bilayer. Based on the fitting results, the lipid-bound fraction, $\xi$, of the fluorophore changes from 2 % to 1 % for Atto 488, from 24 % to 20 % for Atto 565, and from 27 % to 24 % for Atto 647N, as when the cholesterol percentage is increased from 0 mol% to 30 mol%. These reductions are small and
they are close to the error margins of the estimates, suggesting that integrating up to 30% cholesterol into lipid vesicles does not affect the fluorophore-lipid interaction. Adding cholesterol may be beneficial for studying specific lipid-protein interactions, but the typical fluorophores used for labelling will be quasi-insensitive to the presence of cholesterol.

4.5.7 Fluorescence imaging on GUV

The typical diameter of the GUV (> 1 µm) is comparable to the size of a cell (prokaryotic cells: 0.1−5 µm, eukaryotic cells: 10−100 µm). As such GUVs are widely used in membrane biophysics because they can be easily observed under an optical microscope while enabling to differentiate biological phenomena that occur in the lipid membrane vs. the inner aqueous environment. Here, DOPC GUVs were prepared in order to visualize the localization of the fluorophores encapsulated within them.

Figure 4-10. Wide-field fluorescence images of (A) DOPC-Bodipy-FL GUV, (B) DOPC GUV encapsulating Atto 488, (C) DOPC GUV encapsulating Atto 647N, showing localization of the fluorophores in the lipid vesicles. Scale bar is 5 µm.

Wide-field fluorescence images of Atto 488 and Atto 647N fluorophores encapsulated into GUVs are shown in Figure 4-10. A GUV containing 0.05 mol% Bodipy-FL (fluorescent) lipid is shown as a reference. It thus appears that Atto 488 fluorophores diffuse freely inside the GUV, whereas Atto 647N is mostly anchored onto the lipid membrane. The reason why
Atto 647N molecules did not fully cover the lipid membrane is that the fluorophore concentration used in encapsulation process was limited to 50 nM, in order to prevent excess fluorophores from diffusing inside the vesicle and affect the imaging contrast. An identical fluorophore concentration was used for Atto488. This fluorophore concentration is calculated based on the bound fraction $\xi$ in FCS measurements (> 60 %) and size of the vesicles.

4.5.8 Single-molecule imaging of fluorophores on supported lipid bilayers

50 µL solutions containing 0.5-30 µM fluorophore were loaded on SLBs and incubated for 10 minutes at room temperature. Then the coverslip was rinsed with PBS buffer to remove the non-attached dye molecules. The SLB was then imaged under a TIRFM and the number of fluorophore molecules that remained on the SLB was counted in the region of $36 \times 36 \mu m^2$. At least five different regions of each sample were recorded and analyzed. Note that the dye molecules that transiently interact with the bilayer were not observed as they were washed away during the rinsing process.

![Figure 4-11](image)

**Figure 4-11.** TIRF images of DOPC SLB with 5 µM Atto 488 (A), Atto 565 (B), and Atto 647N (C). Red circled dots represent fluorophores attached to the SLB after a 10-min incubation followed by rinsing (mostly each dot is one single molecule, some brighter ones could result from more than one fluorophores occupy the same diffraction-limited spot). Scale bar is 5 µm.
Figure 4-11 shows the effect of Atto 488, Atto 565 and Atto 647N fluorophores interacting with DOPC SLB. Comparing with the partition coefficients, $K_p$, shown in Table 4-1, it is evident that fluorophores with higher $K_p$ display more molecules in the same region of interest. Three fluorophore concentrations, 0.5 µM, 5 µM, and 30 µM, were tested and the numbers of fluorescent molecules observed on the lipid bilayer in the region of $36 \times 36 \, \mu m^2$ for each case were converted to the density of the absorbed fluorophore. The density values are summarized in Table 4-2. Whereas the low binding-affinity fluorophore Atto488 exhibits a similar density on SLB with the increase of the fluorophore concentration, the density of high binding-affinity Atto 647N fluorophore increases sharply with the increase of the fluorophore concentration in the initial incubation solution.

Table 4-2. Density of fluorescent molecules observed on DOPC SLB. Unit: #/100 µm²

<table>
<thead>
<tr>
<th>Fluorophore (µM)</th>
<th>Atto 488</th>
<th>Atto 565</th>
<th>Atto 647N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>27.8 ± 11.6</td>
<td>41.7 ± 6.9</td>
<td>209.9 ± 38.6</td>
</tr>
<tr>
<td>5</td>
<td>38.6 ± 18.5</td>
<td>163.6 ± 27.8</td>
<td>921.3 ± 333.3</td>
</tr>
<tr>
<td>30</td>
<td>41.7 ± 8.5</td>
<td>327.2 ± 71.8</td>
<td>3314.8 ± 593.4</td>
</tr>
</tbody>
</table>

4.5.9 The shift of emission spectra

Besides the photostability and brightness, the environmental sensitivity of the fluorophore in a lipid environment is a major consideration in SMF experiments. The shift of fluorescence emission spectrum can potentially inform one on how strong the fluorescence properties of the fluorophore change upon interactions with lipids. The emission spectra of all the fluorophores studied here were systematically examined in the presence of DOTAP (+), DOPC (+-) and DOPG (-) lipids. For these experiments, the concentration of fluorophores was set to 0.2 µM and that of the lipids at 5 mM. For most of the dyes studied here, the emission spectra are not affected by the presence of the lipids. The results of affected fluorophore-lipid
pairs were listed in Table 4-3 (data with zero shift in the emission spectrum are not shown). Please note that the spectral shift also depends on the concentration of the lipids. For example, FL shows +6 nm shift and +9 nm shift with 1 mM and 2.5 mM DOTAP, respectively.

In addition, we observed that the emission intensity of fluorophores increase upon interaction with the lipids and the magnitude of the increase depends on the combination of fluorophore and lipid. Because the detection efficiency of the detector in the spectrofluorimeter is relatively low compared to the single photon counting detector used in the confocal microscope, and the sample volume is much higher, relatively large amounts of fluorophore are required to perform this study. In this case, multiple fluorophores may interact with the lipid vesicle at the same time causing quenching, anti-cooperative binding or other complex photophysics, which are beyond the scope of this chapter. However, the spectrum shift warns that measures need to be taken in pairing proper fluorophore and lipid for single-molecule fluorescence measurements.

Table 4-3. Shifts of the fluorescence emission spectra for different fluorophores interacting with DOTAP, DOPC and DOPG lipids. “+” sign before the numbers indicates red shift of the spectrum. Fluorophore-lipid pairs with zero spectrum shift are not listed.

<table>
<thead>
<tr>
<th>Emission Spectra Shift (nm)</th>
<th>DOTAP (+)</th>
<th>DOPC (+-)</th>
<th>DOPG (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>+10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Alexa647</td>
<td>+3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Atto647N</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
</tr>
<tr>
<td>Cy5</td>
<td>+4</td>
<td>+8</td>
<td>+15</td>
</tr>
</tbody>
</table>
4.6 Discussion

The interactions between common SMF fluorophores and different type of lipids were quantified using LUVs and a FCS assay. The size of liposomes (100 nm in diameter) is more than two orders of magnitude larger than the size of organic dyes. Since FCS measures diffusion, it is an excellent tool to study the binding of fluorophores to lipid vesicles, ensuring that even small bound fractions of the dye can be resolved. Using a 2-component diffusion model to fit the FCS data (eq. (4-2)) [299, 301], the lipid-bound fractions $\xi$ of each fluorophore at different lipid concentrations were obtained. Based on the bound fractions, $\xi$, the partition coefficient $K_p$ was obtained for each fluorophore-lipid pair (eq. 4-6, and Table 4-1). All the $K_p$ values were larger than one, indicating that the fluorophore-lipid interaction was spontaneous. The larger the partition coefficient, the stronger the fluorophore-lipid interaction and the lower the free energy of association.

As expected, our results show that the hydrophobic effect affects the interaction between fluorophores and lipids. The logD values of the fluorophores studied here are listed in Appendix B. The more positive the logD value is, the more hydrophobic the fluorophore is. For example, the logD values of Atto 647N and Cy5 are 3.09 and 3.84, which are the most positive logD values in the list. In Table 4-1, it is clear that the partition coefficients of Atto 647N and Cy5 are extremely large compared to other fluorophores and their values are similar to fluorescent membrane probes [328]. Conversely, the dyes with more negative logD values, such as Alexa 488 and Alexa 647, whose logD values are -10.49 and -8.10, weakly interact with the lipids studied here.

In the case of Atto 565, which is moderately hydrophilic based on its logD value, the $K_p$'s in all three types of lipids have intermediate values between low (Alexa 488) and high (Atto647N) values. Interestingly, the hydrophobic fluorophore FL (logD = 3.86) behaves similarly to the hydrophilic fluorophores Alexa 488 and Atto 488. From a structural point of view, the FL molecule is small (332 Da) and does not have extended carbon chains beyond the aromatic rings (see Appendix B). It is possible that these carbon chains may be attracted by and even intercalate into the lipid bilayer, thus favoring fluorophore-lipid binding.
Our data also suggest that electrostatic attraction between fluorophores and lipids plays a crucial role in determining the interaction strength. Table 4-1 shows that negatively charged fluorophores interact stronger with cationic DOTAP than with anionic DOPG, whereas positively charged fluorophores prefer to interact with the anionic lipid. For instance, the negatively charged Alexa 647 fluorophore (charge -4), hardly binds to anionic DOPG or with zwitterionic DOPC. However, it binds quite strongly to cationic DOTAP lipid. It suggests that the electrostatic repulsion between same charge pairs disfavor binding. It is also possible that the increase in the diffusion time of Alexa 647 in DOPG lipid solution is primarily due to an increase in viscosity. Using single diffusion component model, by assuming no interaction between Alexa 647 and DOPG, the diffusion time of Alexa 647 increases 1.2 times with 70 mM DOPG lipids. Accessing viscosity values in the lipid solution is usually performed using a viscometer, which require milli-liter to liter volume of sample. Even if the increase in the diffusion time purely results from viscosity, this small viscosity increase is still within the errors margin of the $K_p$ reported here. Furthermore, experiments performed at different salt concentrations show that charge screening is modulated by salt ions and it influences the fluorophore-lipid interaction. Thus, the thermodynamic parameters reported here (partition coefficients, free energies of partitioning) result from both hydrophobic and electrostatic contributions. Depending on the fluorophore-lipid pair, sometimes the hydrophobic factor dominates, and sometimes the electrostatic interaction overwhelms the hydrophobic interaction, such as Alexa 647 with DOTAP.

Partition coefficients $K_p$ and free energies of partitioning $\Delta G$ provide insights into three different directions. Firstly, these values can be directly used to select which fluorophores should be used to label proteins for SMF studies in a lipid environment in order to avoid unwanted artefacts due to interactions between fluorophores and lipids. For studies of protein-lipid interaction and for encapsulating proteins into liposomes for protein-protein or protein ligand interactions, only one fluorophore is needed to label the protein in most cases. It is recommended to label the protein of interest with a blue-excitation fluorophore such as Alexa 488 or Atto 488, because theses dyes show little or no interactions with either cationic, zwitterionic or anionic lipids. This provides researchers the opportunity to choose a proper
lipid composition in case the protein itself may exhibit non-specific interactions with lipids. In addition, the blue dyes are much brighter and photostable than FL and Alexa 647 under physiological pH conditions. In the case of studying the conformation of proteins by encapsulating proteins into lipid vesicle, two different fluorophores are required to label the protein. Depending on the requirement of the Förster radius and feasibility of the instruments, blue-red or green-red fluorophore pairs are commonly used by many research groups. To meet the requirement for a short Förster radius, blue-red fluorophore pair is suggested because of the small overlap between donor emission spectrum and acceptor absorption spectrum. The donor choice should again be Alexa 488 (or Atto 488) and the acceptor should probably be Alexa 647. If a longer Förster radius is required, Atto 532 could be used as the donor and Alexa 647 as the acceptor since the spectral overlap is larger.

Secondly, the partition/free energy values inform us about how to pair fluorophores and lipids when one partner is fixed. The partition coefficient $K_p$ and free energy of partitioning $\Delta G$ in Tables 4-1 and Appendix C offer direct comparison in terms of interaction strength among commonly used fluorophores and lipids. If the protein or the peptide needs to be labeled with a certain fluorophore for specific purposes, the choice of lipid could be made based on the data from Tables 4-1 and 4-2.

Thirdly, our results provide correction factors in the study of protein/peptide-lipid interaction using fluorescence techniques. There are two contributions to the partition coefficient obtained in the protein/peptide-lipid interaction. One contribution is from the specific interaction between protein/peptide and lipid. The other contribution arises from the interaction between the fluorophore label and the lipid. In order to obtain the real partition coefficient between the protein/peptide and the lipid, corrections need to be made to offset the dye-lipid interaction.

In terms of choosing lipid head groups with different electrostatic charges, our results demonstrate that overall the anionic lipid DOPS has lower binding affinity to the fluorophores than the anionic lipid DOPG. Two possible reasons could explain this observation. Firstly, the logD values of DOPS and DOPG are 7.56 and 9.62, respectively. DOPG is thus more
hydrophobic than DOPS, meaning the hydrophobic fluorophores are more likely to interact with DOPG lipid rather than DOPS lipid. Secondly, from the lipid bilayer packing perspective, it was shown that chain packing density of DOPS is larger than that of DOPG [340]. Therefore, the higher packing density provides fluorophore less opportunity to interact with the DOPS lipid bilayer.

In terms of the impact of the saturation of the acyl chain, our results show that fully unsaturated DOPC interacts stronger with the fluorophores than half-saturated POPC and fully-saturated DPPC. This effect is especially obvious when the hydrophobicity of the fluorophores is high. In addition, the logD values of DOPC, POPC and DPPC are 11.19, 10.66, and 10.14, respectively. From the hydrophobicity perspective, similar to the case of DOPG and DOPS, fluorophores exhibit more favorable interactions with DOPC due to stronger hydrophobic interaction.

The phase transition temperatures of DOPC, POPC and DPPC lipids are -17 °C, -2 °C, and 41 °C, respectively. All our measurements were performed in 20 °C, so that the DPPC lipids were in the gel phase while DOPC and POPC were in the fluidic phase. The hydrophobic acyl chains are extended and more ordered in the gel phase than in the fluidic phase, so that the chain packing density of DPPC would be the largest among these three lipids. Because of this highly impermeable structure, fully-saturated DPPC lipids interact the least with the fluorophores. Forming DPPC vesicles requires a relatively high temperature due to its high melting temperature (41 °C). Encapsulating fluorescently labeled proteins into DPPC lipid vesicle is not feasible all the time because the high temperature could potentially denature the protein [341]. Thus, in this chapter, we were more focused on the comparison between DOPC and POPC. For DOPC, there are two unsaturated double bonds on each acyl chain. Each double bond produces a kink in the chain, thus disrupting the ordered structure of the lipid bilayer more heavily than POPC which has only one mono-unsaturated alkyl group. The disordered structure offers more flexibility of the lipid molecule in the lipid bilayer and in turn provides a higher chance to interact with fluorophores.
GUV and SLB imaging provided means to better visualize the interaction between fluorophores and lipids as well as to obtain an independent measure of interaction strength, i.e., the density of adsorbed fluorophores on SLBs. In contrast to the observation that Atto 488 freely diffuses inside the GUV, the anchoring of Atto 647N on the inner lipid wall suggests that the hydrophobic fluorophores are able to intercalate into the lipid bilayer due to the hydrophobic interaction.

Changes in emission spectrum and intensity could result from various factors specific to the lipid membrane, such as the polarity of the local environment, formation of excimers [342] and restricted rotational motion of the fluorophore [343]. Therefore, special attention needs to be paid when designing, performing and analyzing SMF experiments and proper controls are required to account for unwanted fluorophore-lipid interactions.

**4.7 Conclusions**

In this chapter, interactions between different types of lipids and fluorophores were systematically investigated using FCS. 9 fluorophores that are widely used in SMF studies and 6 commonly used lipids were considered. Lipid-bound fluorophore fractions were extracted from FCS measurements. Partition coefficients $K_p$ and free energies of partitioning $\Delta G$ of the lipid-fluorophore pairs were obtained to quantify the interaction strength and gain insight into its physical mechanism. A large $K_p$ (or a very negative $\Delta G$) corresponds to a strong interaction between lipid and fluorophore. Our results show that hydrophobic fluorophores, such as Atto 647N and Cy5, interact strongly with each type of lipids considered in this study. Their partition coefficients are several orders of magnitude larger than those hydrophilic fluorophores, such as Alexa 488 and Atto 488. This suggests that the hydrophobic interaction is a major component of the interaction between lipids and fluorophores. Our results also show that negatively charged fluorophores tend to be attracted to the cationic lipids, such as DOTAP, while positively charged fluorophores bind more to anionic lipids, such as DOPG. FCS data in
different salt concentrations further confirms that attractive electrostatic interaction also contributes to the process of lipid-fluorophore interaction.

To examine how lipids with same charge but different head group structures influence lipid-fluorophore interaction, the interactions between DOPG and DOPS with the same fluorophore were compared. The data showed that DOPS interacts less with the fluorophore than DOPG because of its lower logD value and higher lipid membrane packing density. Based on these results, DOPS rather than DOPG should be considered to form a lipid bilayer system for SMF studies to prevent the binding fluorophores to the lipid surface.

Next, we studied the effect of the lipid chain saturation levels on fluorophore-lipid interaction. Our results show that fully-saturated DPPC interacts less to its fluorophore partner compared to half-saturated POPC and fully-unsaturated DOPC. However, DPPC is not commonly used in single-molecule encapsulation, conformation or interaction studies because of its high transition temperature. Therefore, POPC is preferred rather than DOPC in selecting neutral zwitterionic lipids for SMF studies.

Moreover, our results show that incorporating cholesterol into the lipid bilayer does not have a significant influence on the absorbance of fluorophores onto the lipid membrane. This result provides good news for researchers in the field of drug encapsulation and delivery because cholesterol has often reported to have a positive impact on drug loading and release.

In addition to the spectroscopic studies, microscopic images were obtained to visualize the interaction between lipids and fluorophores. GUV and SLB imaging upon incubation with solutions of fluorophores show that hydrophobic fluorophores have the ability to intercalate in the lipid bilayer, while hydrophilic fluorophores do not anchor on the lipid membrane.

In conclusion, our study shows that blue-excitation fluorophores, such as Alexa 488 and Atto 488, are better options to label proteins for fluorescence studies performed in a lipid environment. More specifically, Blue-red pair, such as Alexa 488-Alexa 647, and green-red pair, such as Atto 532-Alexa647, are better choices for smFRET studies of the conformations of proteins using the lipid vesicle encapsulation. Admittedly, only a small portion of the
fluorophores in current market were studied in this chapter. However, the main goals of this research were reached. Firstly, we implemented an FCS assay to systematically and quantitatively study the lipid-fluorophore interaction. FCS is a common fluorescence technique to study protein-lipid interactions. Therefore, a direct comparison between the partition coefficients obtained in the protein-lipid interaction studies and the current study can be made. The comparison of $K_p$ or $\Delta G$ offers the baseline correction to the specific interaction between protein and lipid, to account for the presence of the dyes. Secondly, it is the first time, to our knowledge, that the study of fluorophore-lipid interaction considers the effects of different head group structures, different degrees of chain saturation and different molar percentage mixing of cholesterol. The results provide a reference for future studies and a systematic method about how to find compatible pairs of lipids and fluorophores in fluorescence studies.
Concluding Remarks and Future Directions

The main objective of this thesis is to determine the conformations and dynamics of disordered protein ensembles by developing and applying single-molecule fluorescence techniques. Initially, I studied a protein folding model system, the N-terminal SH3 domain of the Drosophila drK protein, which is known to exist in an equilibrium between the folded state and a partially unfolded state under normal conditions. smFRET measurements on drkN SH3 in physiological buffer were performed and two distinct FRET populations were observed: a high FRET population corresponding to the folded state and a medium-FRET population corresponding to unfolded conformers. To better understand unfolded states, we exposed the drkN SH3 domain to chemical denaturants. Quite surprisingly, the protein was found again in two distinct FRET states. The dominant conformation is expanded and shows a low FRET efficiency as expected. However, a high-FRET 20-30% population fraction was also present. We ruled out the aggregation of drkN SH3 in GdmCl and urea.

A coarse-grained self-avoiding polymer model was applied to the smFRET data to estimate the most probable radius of gyration for each FRET sub-ensemble. Using the model we inferred that the high-FRET population in high denaturant conditions is a most likely a looped state caused by interactions between residues that are less than 10 amino acids away from the two termini. Similar experiments in other osmolytes such as DMSO and formamide suggest that interactions between small hydrophobic groups in the distal regions contribute to the formation of the looped conformation. These results challenge the idea that proteins are completely solvated at molar concentrations of denaturants and caution against the conventional use of chemical denaturants to study unfolded/disordered states of proteins. Remarkably, this study succeeds in bringing together FRET, SAXS, NMR and computational methods to paint a coherent picture of the physical properties of a “simple” model system. Future work could involve single-molecule studies on various mutants and in the presence of different osmolytes to better understand the mechanism of loop formation and to measure interconversion rates between open and closed conformations.
Later, I examined the conformational and dynamical properties of the disordered neuronal protein 4E-BP2 in the non-phosphorylated and the fully phosphorylated states. Our single-molecule data provide insight into the binding and the energetics this IDP that is highly involved in regulating translational initiation and is associated with neurodegenerative disorders and cancers, such as autism spectrum diseases. The results presented here are the initial steps of a broader collaborative effort with the Forman-Kay and Chan groups to characterize the conformation and dynamics of 4E-BP2. Mutagenesis studies [53] show that phosphomimetic mutants T37D, T37E, T46D or T46E do not mimic phosphorylation states to induce folding of the 4E-BP2. G39V and/or G48V mutations unfold the fully phosphorylated 4E-BP2 and resume the binding affinity for eIF4E. Going forward, SMF studies of the conformations and the dynamics of these mutants could help understand the phosphorylation-induced 4E-BP2 folding mechanism.

Our FCS and FA results show that chain dynamics of C-terminal intrinsically disordered region (C-IDR) are affected by phosphorylation and by binding to eIF4E. The biological significance of the C-IDR in 4E-BP2 remains an open question. It was found that the additional binding attenuation occurs via a mechanism that depends on the presence of both the folded domain and the C-IDR phosphorylation sites (pS65, pT70, and pS83) [53]. How does C-IDR phosphorylation sites regulate binding is still an open question. In the future, investigating the conformations of 4E-BP2 in different phosphorylation states and with C-terminal truncation mutants could help figure out the structural-functional role of the C-IDR in 4E-BP2.

In non-phosphorylated and hypo-phosphorylated states, 4E-BP2 binds to eIF4E with strong binding affinity ($K_d \sim 3$ nM). This binding process induces a disorder-to-order transition [53]. Phosphorylation of T37 and T46 can induce regions of the protein become folded β-sheet structure [53]. However, this β-sheet structure is not stable enough to completely reduce the binding ability to eIF4E. As suggested by NMR, there is an order-to-disorder transition upon binding [53]. smFRET measurements could be performed to probe the order to disorder transitions of 4E-BP2 upon binding to eIF4E under these specific phosphorylation conditions.
Additional smFRET experiments on other double-Cys constructs could help paint a more complete picture of the structural ensemble of 4E-BP2 in the apo, bound and phosphorylated states. Novel polymer models considering the effect of folded regions in IDPs are critically needed to extract the most probable $R_G$ and $\sqrt{\langle R_{EE}^2 \rangle}$ values. The obtained distance information could serve as experimental restraints for the ENSEMBLE program [218] to obtain more accurate 4E-BP2 conformational ensembles. The 4E-BP2 ensemble data in different phosphorylated conditions could also help understand exactly how phosphorylation of 4E-BP2 regulates the eIF4E binding. Furthermore, by encapsulating a donor-labelled 4E-BP2 and an acceptor-labelled eIF4E inside unilamellar lipid vesicle with proper size, the interaction kinetics and timescales of this interaction can be probed.

A key driving force behind the interest in IDPs is their strong association with various diseases, such as cardiovascular disease and neurodegenerative disease, including Parkinson’s, and Prion diseases [54, 56]. Completing the research program discussed above will move us toward a much more complete understanding of 4E-BP2’s function in cells, and will help setup a fluorescence-based high-throughput assay to screen small molecules that stabilize or destabilize the folded conformation to identify lead compounds for therapeutics to treat cancer and autism spectrum diseases.
References


Appendix A:
Lipid structures, logD values, charges and phase transition temperatures

<table>
<thead>
<tr>
<th>Lipids</th>
<th>Structures</th>
<th>Hydrophobicity (logD)</th>
<th>Charge</th>
<th>T&lt;sub&gt;m&lt;/sub&gt; (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOPC</td>
<td><img src="image" alt="DOPC Structure" /></td>
<td>11.19</td>
<td>0</td>
<td>-17</td>
</tr>
<tr>
<td>POPC</td>
<td><img src="image" alt="POPC Structure" /></td>
<td>10.66</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>DOPG</td>
<td><img src="image" alt="DOPG Structure" /></td>
<td>9.62</td>
<td>-1</td>
<td>-18</td>
</tr>
<tr>
<td>DOPS</td>
<td><img src="image" alt="DOPS Structure" /></td>
<td>7.56</td>
<td>-1</td>
<td>-11</td>
</tr>
<tr>
<td>DOTAP</td>
<td><img src="image" alt="DOTAP Structure" /></td>
<td>9.60</td>
<td>+1</td>
<td>&lt;5</td>
</tr>
<tr>
<td>DPPC</td>
<td><img src="image" alt="DPPC Structure" /></td>
<td>10.14</td>
<td>0</td>
<td>41</td>
</tr>
</tbody>
</table>
The structure and \( T_m \) data were obtained from Avanti Polar Lipids, Inc. The hydrophobicity values were calculated using Marvin Sketch software (Chem Axon, Cambridge, MA). The charge values were from [1].

### Appendix B:

**Fluorophore structures and their charge and logD values**

<table>
<thead>
<tr>
<th>Fluorophore&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Structure&lt;sup&gt;b,c&lt;/sup&gt;</th>
<th>Charge&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Hydrophobicity (logD)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexa488-M&lt;sup&gt;c&lt;/sup&gt;</td>
<td><img src="image" alt="Alexa488-M Structure" /></td>
<td>-1</td>
<td>-6.75</td>
</tr>
<tr>
<td>Alexa488-MH&lt;sup&gt;c&lt;/sup&gt;</td>
<td><img src="image" alt="Alexa488-MH Structure" /></td>
<td>-2</td>
<td>-10.49</td>
</tr>
<tr>
<td>Atto488-M</td>
<td><img src="image" alt="Atto488-M Structure" /></td>
<td>-1</td>
<td>-4.25</td>
</tr>
<tr>
<td>Compound</td>
<td>Structure</td>
<td>Charge</td>
<td>pKa</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Atto488-MH</td>
<td><img src="image1" alt="Atto488-MH Structure" /></td>
<td>-2</td>
<td>-7.99</td>
</tr>
<tr>
<td>Fluorescein</td>
<td><img src="image2" alt="Fluorescein Structure" /></td>
<td>0</td>
<td>3.86</td>
</tr>
<tr>
<td>Bodipy-FL-M</td>
<td><img src="image3" alt="Bodipy-FL-M Structure" /></td>
<td>0</td>
<td>-3.96</td>
</tr>
<tr>
<td>Compound</td>
<td>Structure</td>
<td>Charge</td>
<td>pK_a</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Bodipy-FL-MH</td>
<td><img src="image" alt="Bodipy-FL-MH" /></td>
<td>-1</td>
<td>-3.5</td>
</tr>
<tr>
<td>Atto532-M</td>
<td><img src="image" alt="Atto532-M" /></td>
<td>-1</td>
<td>-2.93</td>
</tr>
<tr>
<td>Atto532-MH</td>
<td><img src="image" alt="Atto532-MH" /></td>
<td>-2</td>
<td>-6.68</td>
</tr>
<tr>
<td>Atto565-M</td>
<td><img src="image" alt="Atto565-M" /></td>
<td>+1</td>
<td>1.60</td>
</tr>
<tr>
<td>Compound</td>
<td>Charge</td>
<td>Stability</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Atto565-MH</td>
<td>0</td>
<td>-2.14</td>
<td></td>
</tr>
<tr>
<td>Atto647N-M</td>
<td>+1</td>
<td>2.53</td>
<td></td>
</tr>
<tr>
<td>Atto647N-MH</td>
<td>0</td>
<td>3.09</td>
<td></td>
</tr>
<tr>
<td>Cy5-M</td>
<td>+1</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>Material</td>
<td>Structure</td>
<td>Time (h)</td>
<td>Stability (h)</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------</td>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>Cy5-MH</td>
<td><img src="image" alt="Cy5-MH Structure" /></td>
<td>0</td>
<td>3.84</td>
</tr>
<tr>
<td>Alexa647-M</td>
<td><img src="image" alt="Alexa647-M Structure" /></td>
<td>-3</td>
<td>-4.26</td>
</tr>
<tr>
<td>Alexa647-MH</td>
<td><img src="image" alt="Alexa647-MH Structure" /></td>
<td>-4</td>
<td>-8.1</td>
</tr>
</tbody>
</table>

* M represents maleimide linker before hydrolysis, MH represents maleimide linker after hydrolysis. SE represents succinimidyl ester before hydrolysis, SEM represents succinimidyl ester after hydrolysis.
b The structures of the fluorophores used in Chapter 2, 3, and 4. The structure data were obtained from Thermo Fisher Scientific and ATTO-TEC.

c The hydrophobicity values were calculated using Marvin Sketch software (Chem Axon, Cambridge, MA) at physiological condition.
Appendix C:
Free energy of partitioning

The free energy of partitioning $\Delta G$ between different fluorophores and the cationic DOTAP, zwitterionic DOPC, and anionic DOPG lipids, respectively. Error bars are based on the results from three independent datasets.

<table>
<thead>
<tr>
<th>$\Delta G$ (kJ/mol)</th>
<th>DOTAP (+)</th>
<th>DOPC (+/-)</th>
<th>DOPG (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexa 488</td>
<td>-13.36 ± 0.18</td>
<td>-12.19 ± 0.09</td>
<td>-9.07 ± 0.27</td>
</tr>
<tr>
<td>Atto 488</td>
<td>-13.87 ± 0.12</td>
<td>-13.47 ± 0.14</td>
<td>-10.47 ± 0.15</td>
</tr>
<tr>
<td>FL</td>
<td>-12.86 ± 0.13</td>
<td>-13.38 ± 0.26</td>
<td>-10.03 ± 0.13</td>
</tr>
<tr>
<td>Atto 532</td>
<td>-13.62 ± 0.17</td>
<td>-12.76 ± 0.07</td>
<td>-12.43 ± 0.26</td>
</tr>
<tr>
<td>Atto 565</td>
<td>-15.25 ± 0.31</td>
<td>-18.07 ± 0.13</td>
<td>-19.07 ± 0.24</td>
</tr>
<tr>
<td>Atto 594</td>
<td>-22.61 ± 1.89</td>
<td>-19.78 ± 0.58</td>
<td>-19.03 ± 0.68</td>
</tr>
<tr>
<td>Alexa 647</td>
<td>-24.32 ± 0.71</td>
<td>-6.63 ± 0.27</td>
<td>-6.28 ± 0.37</td>
</tr>
<tr>
<td>Atto 647N</td>
<td>-19.71 ± 0.27</td>
<td>-22.56 ± 0.44</td>
<td>-34.53 ± 0.31</td>
</tr>
<tr>
<td>Cy5</td>
<td>-28.47 ± 0.72</td>
<td>-30.64 ± 0.61</td>
<td>-28.47 ± 0.72</td>
</tr>
</tbody>
</table>
Appendix D:
Total internal reflection fluorescence microscope (TIRFM) user operation and diagnosis manual

1. Hardware startup. Turn on lasers, AOTF, MadCity Z stage, ASI X-Y stage and Thorlabs motorized stage. Wait for 30 minutes until lasers get warmed up. Check the laser power and record it in the log book. If the user finds a decreasing power trend, please report to corresponding lab member or PI.

2. Software startup. Use your UTORid to login into the TIRF computer. Open the C:\ drive. The TIRF LabVIEW data acquisition program shortcut has been created in the C:\ drive. Open the LabVIEW program and run it by clicking the arrow on the top left of the graphical user interface (GUI).

3. Set up the initial values for the hardware control modules. Set the z-stage position to 20 μm and the lateral position for the motorized stage to 22.8 mm for wide-field imaging or 18.6 mm for TIRF imaging. The value of 22.8 mm was determined experimentally by observing and noting where the excitation beam hits the center of dichroic mirror and reflects back to the center of the objective mount. The value of 18.6 mm comes from experimental results obtained from images of 20 nm bead in order to find the position with the maximum intensity. Sometimes, the excitation alignment will be off due to possible external reasons, such as subtle movements through touch by the user or improper adjustment of optical elements, meaning that 22.8 mm and 18.6 mm are no longer optimal. How to quickly check the excitation alignment? The quickest way is to place a bare coverslip on the sample holder (see detailed procedure in step # 5) and move the motorized stage to 18.6 mm. At 18.6 mm, the excitation mode is total internal reflection so that you should be able to see two beam spots on the dichroic mirror (one from excitation beam and one from scattering) by adjusting the objective lens ring mount. If users cannot see two beam spots with roughly equal size at the same horizontal level, users need to re-align the excitation path. If users can observe that, you can skip step # 4.

4. Re-align the TIRF excitation path. 1) Check the laser beam position on iris I1 (see Figure 1-8) to make sure that the beams hit the center of the iris and exhibit Gaussian profile
on the iris. Sometimes adjusting the width of the iris is necessary to have a better assessment. If the status of the laser beams are not described as in 1), the corresponding mirrors (M1, SP for 633 nm laser line; M2, DM2 for 532 nm laser line; M3, DM1 for 473 nm laser line; M4, M8 for 405 nm laser line) need to be tuned to obtain that correct alignment. SP, DM2, DM1 and M8 are recommended to be adjusted first because the positions of M1-M4 influence the laser lines that enter into the confocal setup. Try to avoid adjusting M9 and pinhole mount since they are shared by all four laser lines. 2) Flip off FL1 and the beam should hit exactly on the drawn circle on the right of TIRF box panel. If not, the periscope needs to be adjusted. Please note that adjusting the periscope mirrors are rarely necessary. Please check all the above steps and perform the action. 3) Flip back FL1 and check the beam position on iris I2. Same as the I1 case, the beams should be placed at the center of the iris. If not, FL1 needs to be adjusted. 4) At this moment, you should be able to see two well-shaped beam spots as described in step # 3. If yes, you can skip the following re-alignment steps. If no, this means the 4 band DM does not send beam straight up and/or the objective holder is not placed horizontally, so proceed to following steps. 5) Take the sample holder and objective out. Set lateral position to be 22.8 mm. The beam should be Gaussian profile and hit on the drawn circle on the top of the TIRF box. Put the sample holder back and place a test mirror on the top of the sample holder and look at the reflected beam on the back of I1. Adjust the position of 4 band DM to overlap the incoming beam and the reflecting beam. Then take the sample holder out and place the test mirror on the top of gimbal mount. Look at the beam spot on the back of I1 again and adjust the gimbal mount to overlap with the incoming and reflected beams. After these are completed, put the objective and the sample holder back. Now you should be able to observe the two beam spots by adjusting the objective lens ring mount.

5. Check ASI X-Y stage and make sure the objective is located at the center of the sample holder. If not, X and Y offsets correction are needed in the LabVIEW program.

6. Emission alignment check. Clean the objective using the Thorlabs lens-paper with a drop of EtOH. After EtOH evaporates, drop a small amount of oil on the top of the objective. Place 200 nm Tetraspeck beads on the holder. Select the excitation wavelength. In the LabVIEW program, laser power percentages was used to avoid the inconsistency due to laser
power dropping with time. Percentage vs. power chart is stored in the C:\ drive for reference. It is recommended that users measure the power every time one performs experiments and keeps a record of it. For 200 nm beads, 1-3 % laser power of any excitation source is recommended and no EM gain is needed to image 200 nm beads. Always start with low intensity to avoid saturating the EMCCD. The user can select “Run till abort” mode in the camera setting to check the alignment. The default exposure time can be used for alignment purpose. Set the temperature of the EMCCD to -80 ºC. When the temperature reaches this value, the green LED light on the LabVIEW interface will light-up. Now, you can press the shutter open button and start the acquisition. We first will focus on the procedure for one channel alignment. If everything is aligned, the user should be able to see 200 nm beads spread out over the imaging field of view within the LabVIEW program, each being about 5 by 5 pixels. Focusing of the bead image can be performed by adjusting the objective lens ring mount to find optimal z-position for the objective. The beads should spread throughout the region of image interest so that users move to several different imaging areas, there should be a good coverage of bright beads. If clear regions of even spread beads is not observed then it can be concluded that the emission path is not well aligned. For example, bead molecules should never be seen near the bottom part of imaging area, rather must be centered within the viewing area. To align the bead images to the center of the view, one can start by adjusting M14 as the first action. Tune one of the knobs of the M14 kinetic mount so that the bead image will be centered in the region of interest. In most of the situations, it will work very well. If not, both M10 and M14 need to be adjusted. For dual channel, the above check is also necessary prior to moving to two color alignment. In addition, several other steps mentioned below need to be examined. By default, donor/parallel polarized signal go to the left half of the view and acceptor/perpendicular polarized signal go to the right half of the view. Select 4 beads that are located at the top/bottom/left/right on the left side view and record the pixel position. Adjust M12 (for anisotropy) or M14 (for FRET) to shift the bead images on the right side view to the corresponding position. For example, if x dimension range is 512 pixel, and pixel position on the left side view is (x1, y1), then same bead on the right side view will be (x1+512/2, y1). Click on the molecule and the x and y positions will be shown on the right column of the LabVIEW program. Sometimes, 4 sampling beads cannot meet this criterion at the same time.
What needs to be done is to adjust the one channel alignment in order to make sure that horizontal and vertical pixel positions corresponding to the same bead on both sides of the field of view, and are not more than 2 pixels apart in either direction. After this is done, save the aligned image in TIFF format for future data analysis, such as for image transformation and G factor calibration. It is recommended to save the alignment image every time before and after each experiment. It should also be noted that the dimension of the alignment file should be the same as all other measurement files for analysis.

7. Place samples. Drop a small amount of oil on the objective and place the coverslip/coverslide/cell dish on the sample holder. For coverslips, the screw-on cap can be used to fix the position of the coverslip. For coverslide, the slide holder with a metallic clamp is necessary to anchor the slide. Do your best to maintain an even field of illumination by making sure that the slide or slip is not tilted in any way.

8. Set the Acquisition parameters. Set the laser intensity value (scale in 100%), ALEX-related experiments in the laser control panel and camera data acquisition parameters in the camera control panel. Now you are ready to acquire data by clicking the “Start Acquisition” button.

9. Real sample measurement. For immobilized molecule measurements, make sure the density of the molecule is not very high so that each molecule can be well identified. Most of the time, oxygen scavenger, such as PCA/PCD, glucose oxidase and anti-blinking chemicals, such as Trolox, β-mercaptoethanol are needed to produce photostable and non-blinking intensity time trajectories. Power, EM gain, exposure time are parameters frequently adjusted to optimize data collection. Save image stacks as TIFF format for later data analysis.

10. Background measurement. Apply the same laser intensity and camera parameters such as exposure time, EM gain, vertical speed, etc., used in the real samples for the background measurement. Use buffer only samples as background measurements. Collect a single image and save as TIFF format. It should be noted that the dimension of the alignment file should be the same as all other measurement files.
11. Flatness / pixel correction / leakage measurements. Apply the same laser intensity and camera settings such as exposure time, EM gain, vertical speed, etc used in the real samples for the background measurement. Use dilute dye solution (~10 nM concentration) as sample. For flatness/pixel correction, only one sample is needed, while for leakage measurement involved in smFRET measurements, both donor and acceptor dyes are needed for calibration. Collect a single image frame and save it in TIFF format. It should be noted that the dimension of the alignment file should be the same as all other measurement files.

12. Final stage. After the user finishes the experiments, the user needs to close the camera shutter and raise the camera temperature to room temperature. Then the user needs to set values of MadCity Z stage, ASI X-Y stage and Thorlabs motorized stage to zero. Next, stop the LabVIEW program and turn off all the hardware. Take the sample out of the sample holder, clean the objective with lens-paper with a drop of EtOH and slide up the TIRF microscope door.
Statement of Contributions

Chapter 1:

I developed a new version of the total internal reflection microscope (TIRFM) in the Gradinaru lab from a previous setup developed by Dr. Amir Mazouchi and Dr. Baoxu Liu. I wrote a multi-featured and user-friendly data acquisition program in LabVIEW to control and integrate all hardware in the microscope, based on an old version written by Dr. Baoxu Liu and by an undergraduate intern, Daniel Jacobs. I implemented the alternating laser excitation (ALEX) modality on TIRFM. I developed the data analysis program in Matlab to analyze both spectroscopic data and microscopic images. The program has been used to analyze the distribution of DNA oligos conjugated to quantum dot biosensors (Bioconjugate Chemistry 2014), supramolecular organization of the M2 muscarinic receptor and the Ga1 protein (Journal of American Chemical Society 2016), size of DNA ladder constructs for diagnostic assay and adenosine A2A receptor conformations and ligand interactions. I also developed time-lapse fluorescence recovery after photobleaching (FRAP) method to minimize unwanted photobleaching during the fluorescence recovery process. This method was applied to study lateral diffusion coefficient of the lipids coated on polystyrene bead (PS). The lipid coated PS samples were prepared by Dr. Qasim Saleem and Amy Petretic. The results have been published in Biomacromolecules 2015, in which I am a co-first author. The multi-parameter fluorescence confocal microscope (MPFCM) was developed by Dr. Baoxu Liu and the ALEX modality on MPFCM was developed by another graduate student, Gregory Gomes. The dual-
color fluorescence correlation microscope (dcFCM) was built by a graduate student, Yuchong Li. I also developed several sample protocols based on the existing literature, and finalized them with the assistance of other lab members. These protocols include protein labelling and purification, component cells, plasmid proliferation, spin coating, lipid vesicle and lipid bilayer preparation, coverslip surface modification, etc.

Chapter 2:

I investigated conformational heterogeneity of N-terminal Src homology 3 domain of Drosophila downstream of receptor kinase (drkN SH3) domain using single-molecule fluorescence spectroscopy and an excluded-volume polymer model. Dr. Amir Mazouchi designed and performed the experiments in the early stage. The SH3 protein was initially provided by a technician, Hong Lin, working with Prof. Julie-Forman-Kay (The Hospital for Sick Children). Later on, I expressed and purified the SH3 protein with the assistance of Hong Lin. I labeled SH3 with different fluorophores and performed protein purification and mass spectroscopy. I designed and performed single-molecule Förster resonance energy transfer (smFRET) measurements to characterize the heterogeneous conformations of SH3. I employed fluorescence correlation spectroscopy (FCS) and fluorescence cross correlation spectroscopy (FCCS) to obtain the hydrodynamic radius ($R_H$) in different solvents and rule out the possibility of aggregation. I measured the tryptophan fluorescence of SH3 protein to obtain the denaturation curve. I proposed possible mechanisms to explain the persistence of the high-FRET peak in high denaturant conditions. The LabVIEW code for the smFRET burst analysis was developed by Dr. Baoxu Liu. Self-avoiding walk (SAW) polymer model was developed
by Prof. Hue Sun Chan and Dr. Jianhui Song. Gregory Gomes analyzed the simulated SAW data. A technician in our lab, Abdullah Bahram, did some of the FCS experiments. The Amberlite MB-1 resin and the basic protocol for urea purification were provided kindly by Prof. Voula Kanelis (Chemistry, UofT). Those results have been published in *Biophysical Journal 2016*, in which I am a co-first author.

**Chapter 3:**

I used single molecule techniques to investigate conformations, dynamics and binding of the disordered eukaryotic initiation factor 4E (eIF4E) binding proteins 2 (4E-BP2). I made the double cysteine mutant plasmids under the guidance of Dr. Alaji Bah from the group of Prof. Julie Forman-Kay, and he expressed and purified the eIF4E and 4E-BP2 protein samples. I did the fluorophore labelling of 4E-BP2 and subsequent purification by size-exclusion chromatography and ion exchange chromatography. I also did mass spectroscopy on 4E-BP2 and labelled 4E-BP2 samples in order to assess the quality of the sample. I designed and performed smFRET measurements to reveal changes in the conformational ensemble corresponding to phosphorylation, denaturation, salt and pH modulation. I designed and performed time-resolved fluorescence anisotropy decay (FAD) measurements to obtain the local chain flexibility at various points along the 4E-BP2 sequence and measure the response to eIF4E binding, phosphorylation and denaturation. I designed and performed FCS measurement to study the size ($R_H$) and the intra-molecular chain dynamics of 4E-BP2 and inter-molecular kinetics between 4E-BP2 and eIF4E. I applied the Sanchez polymer model to extract the end-to-end distance ($R_{EE}$) and radius of gyration ($R_G$) with the help of Gregory
Gomes. Gregory Gomes wrote the Matlab program to analyze single-molecule fluorescence bursts detected in the ALEX mode. This work is currently being prepared for publication in collaboration with Prof. Chan and Forman-Kay.

Chapter 4:

I investigated the interaction between fluorophores and lipids using single molecule fluorescence methods. I prepared the vesicles and the supported lipid bilayers. I performed dynamic light scattering (DLS), fluorescence lifetime, fluorescence emission spectroscopy and fluorescence imaging measurements. I performed simulations of the fluorophore-lipid interaction to determine proper experimental conditions. I employed the FCS technique to quantify the interactions between commonly used fluorophores and different types of lipids. I studied lipids with different head groups and with different degrees of chain saturation. I also investigated cholesterol and salt effect on the interaction. An undergraduate student, Dan Yomo, performed some of the measurements at the initial phase of the project. This work is currently being prepared as a manuscript for publication.
List of Publications

1. Single-Molecule Dissection of the Conformations, Dynamics and Binding of the Disordered 4E-BP2 Protein
   **Z. Zhang**, J. Dawson, A. Bah, G. Gomes, J. Song, H. S. Chan, J.D. Forman-Kay, C. C. Gradinaru
   In preparation

2. Choosing the right fluorophore for single-molecule fluorescence studies in lipid environment
   **Z. Zhang**, D. Yomo, C.C. Gradinaru
   In preparation

3. Conformations of a Metastable SH3 Domain Characterized by smFRET and an Excluded-Volume Polymer Model
   A. Mazouchi#,** Z. Zhang**#, A. Bahram, G.-N. Gomes, H. Lin, J. Song, H. S. Chan, J. D. Forman-Kay, C. C. Gradinaru
   Biophysical Journal, 2016, 110 (7), pp 1510–1522
   #: These authors contributed equally to this work

4. Single Lipid Bilayer Deposition on Polymer Surfaces Using Bicelles
   Q. Saleem#, **Z. Zhang**#, A. Petretic, C. C. Gradinaru, and P. M. Macdonald
   Biomacromolecules, 2015, 16 (3), pp 1032–1039
   #: These authors contributed equally to this work

5. Isolation of Monovalent Quantum Dot–Nucleic Acid Conjugates Using Magnetic Beads
6. Liposome-Coated Hydrogel Spheres: Delivery Vehicles with Tandem Release from Distinct Compartments
   Q. Saleem, Z. Zhang, C. C. Gradinaru, and P. M. Macdonald
   Langmuir, 2013, 29 (47), pp 14603–14612

7. Single-Molecule Analysis of the Supramolecular Organization of the M₂ Muscarinic Receptor and the Gα₁ Protein
   Journal of American Chemical Society, 2016, 138 (36), pp 11583–11598

8. Synthesis of Stable Multifunctional Perfluorocarbon Nanoemulsions for Cancer Therapy and Imaging
   D. A. Fernandes, D. D. Fernandes, Y. Li, Y. Wang, Z. Zhang, D. Rousseau, C. C Gradinaru, and M. C. Kolios
   Langmuir, 2016, 32 (42), pp 10870–10880