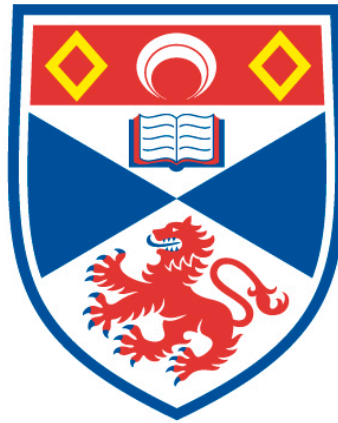


SINGLE-MOLECULE STUDIES OF SURFACE-IMMOBILISED AND  
FREELY DIFFUSING RNA STRUCTURES

Euan Stuart Shaw

A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews



2016

Full metadata for this item is available in  
St Andrews Research Repository  
at:  
<http://research-repository.st-andrews.ac.uk/>

Identifiers to use to cite or link to this thesis:  
DOI: <https://doi.org/10.17630/10023-12974>  
<http://hdl.handle.net/10023/12974>

This item is protected by original copyright

# Single-molecule studies of surface-immobilised and freely diffusing RNA structures

Euan Stuart Shaw



University of  
St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD  
at the University of St Andrews

February 2016



## Abstract

During the process of transcription, a messenger RNA (mRNA) copy of the genetic information encoded in the DNA is built. As mRNA is constructed, secondary and tertiary motifs may form, which combine into intricate structures through a process called RNA folding, enabling RNA to perform biological functions beyond transporting genetic information, including gene regulation and catalytic self-cleaving processes. Facilitating RNA folding are divalent ions, located site-specifically within the structure, and monovalent ions which bind non-specifically to the phosphate backbone, shielding the negative charge and allowing the motifs to move into close proximity and interact. Three-way RNA junctions are among the smallest biologically-active RNA structures and are known to mediate both gene regulatory and catalytic processes. In the first part of this thesis, I use single-molecule total-internal reflection fluorescence microscopy with Förster resonance energy transfer (TIRFM-FRET) to characterise the folding and function of two of these structures: the adenine riboswitch and the hammerhead ribozyme. Using single-molecule Förster resonance energy transfer (sm-FRET), I extract information on both the prevalent conformations of these molecules at specific chemical conditions and kinetic information on structural rearrangements which occur on both the molecular and global levels. Building on this knowledge built up of the folding pathway of the adenine riboswitch induced by monovalent ions, I moved to develop a method in which the competing interplay between monovalent ions and urea, an unfolding reagent, is exploited to isolate and overpopulate a transient intermediate state identified on the folding pathway. Although chemical



denaturants are commonly used to investigate the structures of proteins, their application to RNA folding is still in its infancy. For the first time, I demonstrate that this approach allows the manipulation of the folding dynamics of RNA, forcing the structure into a state which is ordinarily poorly-populated. I speculate that this could enable a detailed characterisation of these states by NMR and other high resolution ensemble techniques. Finally, I move on to expanding the range of single-molecule techniques available in St Andrews. Despite the power of single-molecule TIRFM-FRET, it requires surface immobilisation, which can compromise biological function through further modifications to the natural form of the sample under investigation. To overcome this problem, I implement single-molecule fluorescence correlation spectroscopy (sm-FCS), which probes freely diffusing samples in solution. After testing this sm-FCS system various test structures, I upgrade it for dual-colour fluorescence cross-correlation spectroscopy (sm-FCCS), and finally to multi-parameter fluorescence detection (sm-MFD), where the fluorescence lifetime of the sample is also returned. The capabilities of these three techniques are tested by examining protein-DNA interactions, RNA structure and vesicle morphology.

# Declarations

## 1. Candidate's declarations

I, Euan Stuart Shaw, hereby certify that this thesis, which is approximately 56,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2011 and as a candidate for the degree of Doctor of Philosophy in September 2011; the higher study for which this is a record was carried out in the University of St Andrews between 2011 and 2015.

---

Euan S. Shaw      February 2016

## **2. Supervisor's declaration**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

---

J. Carlos Penedo    February 2016

### 3. Permission for publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

No embargo on printed copy but embargo on all of electronic publication of thesis for a period of two years on the following ground: publication would preclude future publication.

---

Euan S. Shaw

February 2016

---

J. Carlos Penedo

February 2016

## Acknowledgements

These four years have been simultaneously one of the toughest but also one of the most rewarding experiences of my life and I owe a great deal of thanks to many people who have helped me practically and personally through the rollercoaster of pitfalls and successes that is PhD life. It is quite easy to divide all of these names into two groups, professional and personal, but the fact of the matter is that the lines aren't distinct at all. From practical advice on tackling problems from family to nights out with work colleagues, my support network has been one huge and collective entity.

To my girlfriend, Aly, you have been with me through the ups and downs and patiently supported me. Thank you so much for all of the gestures big and small that have carried me through this process, I love you so much.

I want to say a big thank you to my family, my mum and dad and my brother, you suffered the whole journey with me and helped me out no end in all sorts of ways. Some debts will never be repaid, and even if I could I know you wouldn't allow it.

To my colleagues past and present, Paul, Jose, Steve, Mike, Kaley, Cibran and Francisco, thank you all for your discussions, help and the laughs provided throughout my time. Paul, thank you in particular for all of the compliments. Thank you also to Scott in stores for your help throughout the

PhD. In biology thank you to everyone in Malcolm White's group, to Clare, Rich and Shirley in particular, and the Level 3 biology lab in general for keeping the physicist from blowing the place up.

Thank you to all of my friends for offering distractions from work when needed and trying to keep me sane – in particular to Tom, Sam and Clare from home, and Phil, Kieran, James, Simran, Anja, Lotty and, of course, Aly from St Andrews. Also a special mention must go to all of the footballers, from Last Minute to the Whip Inn Boys (Lee, Steve, Grandad, Emily, Al, Andy, Guil, Neil, Iosu, Cedric, Callum, Grant, John, Phil, Dan, Meg, Francesco). The injuries were (are) all worth it.

The final and most important person I want to thank is my supervisor, Dr Carlos Penedo. Working under his supervision during my undergraduate project made the decision when he encouraged me to apply for a PhD a no-brainer. His constant support and guidance throughout my PhD cannot be put into words, right from teaching me the basics of MATLAB coding in week 1 to reading and offering chapter corrections late into the night and early into the morning in week 204. Thank you so much for everything.

## List of abbreviations

<b>Abbreviation</b>	<b>Meaning</b>
2-AP	2-aminopurine
ADC	Analogue to digital converter
AFM	Atomic force microscopy
APD	Avalanche photodiode
BSA-bi	Biotinylated bovine serum albumin
CRISPR	Clustered regularly interspaced short palindromic repeats
Cy3	Cyanine 3 dye
Cy5	Cyanine 5 dye
DNA	Deoxyribonucleic acid
$E_{app}$	Apparent FRET
EM-CCD	Electron multiplying charge coupled device
EPR	Electron paramagnetic resonance
FRET	Förster resonance energy transfer
FWHM	Full width half maximum
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (buffer)
HMM	Hidden Markov modelling
IRF	Instrument response function
MCP-PMT	Multi-channel plate photomultiplier tube
miRNA	Micro RNA
mRNA	Messenger RNA
NA	Numerical aperture
ND	Neutral density
NMR	Nuclear magnetic resonance
NSOM/SNOM	Near-field scanning optical microscopy/scanning near-field optical microscopy
o-TIRFM	Objective-type total internal reflection microscopy
PMT	Photomultiplier tube
PSF	Point spread function

p-TIRFM	Prism-type total internal reflection microscopy
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
siRNA	small interfering RNA
sm-FCCS	Single-molecule fluorescence correlation spectroscopy
sm-FCS	Single-molecule fluorescence cross-correlation spectroscopy
sm-FRET	Single-molecule Förster resonance energy transfer
sm-MFD	Single molecule multi-parameter fluorescence detection
snRNA	Small nuclear RNA
SPAD	Single-photon avalanche diode
T50	50 mM Trizma hydrochloride (2-Amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride)
TAC	Time-to-amplitude converter
TCSPC	Time-correlated single photon counting
TEMED	Tetramethylethylenediamine
TIRFM	Total internal reflection fluorescence microscopy
tRNA	Transfer RNA



# Table of contents

<b>Abstract</b>	<b>i</b>
<b>Declarations</b>	<b>iii</b>
1. Candidate's declarations.....	iii
2. Supervisor's declaration.....	iv
3. Permission for publication.....	v
<b>Acknowledgements</b>	<b>vi</b>
<b>List of abbreviations</b>	<b>viii</b>
<b>Table of contents</b>	<b>x</b>
<b>Chapter 1: Introduction to fluorescence and single-molecule methods</b>	<b>1</b>
1.1. Origins of the theory of fluorescence.....	1
1.2. Principles of photoluminescence.....	2
1.3. Fundamental properties of fluorescent media.....	4
Absorption spectroscopy and the Beer-Lambert Law.....	5
Excitation and emission spectra.....	6
1.3.1. Stokes Shift.....	7
1.3.2. Kasha's Rule.....	8
1.3.3. Franck-Condon Principle and the Mirror Image Rule.....	9
1.3.4. Fluorescence lifetime.....	10
1.3.5. Fluorescence quantum yield.....	12
1.4. Nanometre-sized distance measurements using fluorescence: Förster Resonance Energy Transfer (FRET).....	12
1.4.1. Basic concepts of FRET.....	12
1.4.2. Common FRET pairs in fluorescence spectroscopy.....	19
1.4.3. Bulk methods to measure FRET.....	21
Steady-state FRET measurements.....	21
Fluorescence lifetime-based measurements.....	23
1.5. Single-molecule spectroscopy.....	29
1.5.1. Advantages of single-molecule spectroscopy and the origins of the field.....	29
1.5.2. Single-molecule techniques.....	31
Resolution limits from the diffraction limit of light.....	33
Principles of confocal microscopy.....	34
Principles of TIRFM.....	36
Prism-type TIRFM (p-TIRFM).....	39
Objective-type TIRFM (o-TIRFM).....	40
1.6. Measuring FRET at the single-molecule level (sm-FRET).....	43
1.6.1. Surface-immobilised sm-FRET techniques.....	43
1.6.2. Limitations of surface-immobilised single-molecule fluorescence techniques.....	47
1.6.3. Fluorescence correlation spectroscopy: biomolecular dynamics from microseconds to seconds.....	49
1.6.4. Single-molecule multiparameter fluorescence detection (sm-MFD).....	51

<b>Chapter 2: Structure and function of RNA</b>	<b>54</b>
2.1. Deoxyribonucleic acid (DNA)	57
2.2. Ribonucleic acid (RNA)	60
2.3. RNA folding	63
2.4. The functional roles of RNA	65
2.5. RNA as enzymes catalysing chemical reactions	68
2.6. Riboswitches	74
2.6.1. Riboswitch architecture	74
2.6.2. Mechanisms of gene regulation by riboswitches	77
2.6.3. Gene regulation by riboswitches takes place co-transcriptionally	79
2.6.4. Ligand-binding mechanisms: induced fit vs conformational selection models	83
2.6.5. Riboswitch classes and structural diversity	85
2.6.6. Riboswitches as novel antibiotic targets	86
2.7. Single-molecule FRET studies of riboswitches	87
The first sm-FRET studies on a riboswitch: structure, function and folding pathway of the adenine riboswitch	88
<b>Chapter 3: Materials and Methods</b>	<b>94</b>
3.1. Preparation of the RNA Structures	94
3.1.1. Adenine Riboswitch	94
RNA aptamer purification	97
3.1.2. The hammerhead ribozyme	97
Synthesis of the RNA structures	97
Purification of the RNA structures	100
3.2. Channel preparation of single-molecule spectroscopy	101
3.2.1. Cleaning protocol prior to construction	101
3.2.2. Constructing imaging chambers for single-molecule spectroscopy	102
3.3. Surface immobilisation for single-molecule spectroscopy	103
3.4. Single-molecule additives to improve dye photostability	104
Preparation of single-molecule additives	105
3.5. Bulk spectroscopic methods for RNA studies	107
3.5.1. UV-Vis spectroscopy to determine RNA concentration	107
3.5.2. Lifetime measurements using a pulsed excitation fluorimeter	108
3.6. Analysis of bulk data for dye-labelled RNA structures	109
3.6.1. Calculating fluorescence lifetimes from TCSPC data	109
3.6.2. Inter-dye distance measurements from TCSPC data	110
3.7. Single-molecule mapping and recording using prism-type total internal reflection fluorescence microscopy	111
3.7.1. Technical details of the p-TIRFM setup	111
3.7.2. Calibration of the p-TIRFM system and data acquisition	114
3.8. Analysis of single-molecule FRET data	116
3.8.1. Steady-state FRET histograms	116
3.8.2. Studying dynamics using Hidden Markov Modelling (HMM)	117
Fitting the FRET trajectory using a Hidden Markov Modelling Program	122
Extracting dynamic information from HMM-fitted FRET data using MATLAB	123
3.9. Fitting data and error analysis	126
3.9.1. Fitting procedures for single-molecule data	126
Steady-state FRET histograms	128
Dwell time histograms	129

3.9.2. Error analysis and the propagation of errors.....	129
Steady-state histograms.....	129
Dwell time histograms.....	130
<b>Chapter 4: Influence of monovalent ions on riboswitch folding</b>	<b>131</b>
4.1. Introduction.....	131
4.2. Steady-state monovalent ion-induced folding of the adenine riboswitch.....	132
4.3. Investigating the sodium-induced folding dynamics using Hidden Markov Modelling..	140
Post-HaMMY filtering of “fast transitions”.....	142
4.4. Na <sup>+</sup> -induced folding dynamics of the adenine riboswitch aptamer.....	146
4.5. Ligand-induced stabilisation of the folded adenine riboswitch complex using Na <sup>+</sup> ions as tertiary folding agents.....	156
4.6. Conclusions.....	161
<b>Chapter 5: Manipulating RNA dynamics to isolate transient folding states</b>	<b>164</b>
5.1. Introduction.....	164
5.1.1. Chemical denaturation as a method to manipulate RNA dynamics.....	164
5.1.2. Previous single-molecule FRET studies on the adenine aptamer using single- molecule chemical denaturation.....	165
5.2. Comparative analysis of the relative unfolding efficiency of different chemical denaturants at the single molecule level.....	170
5.3. Influence of urea in the monovalent ion-induced folding pathway of the adenine riboswitch.....	176
5.4. Single-molecule dynamics of the adenine riboswitch under chemical denaturation.....	184
5.5. Investigating the ligand-induced stability of a chemically denatured adenine riboswitch.....	191
5.6. Conclusions.....	197
<b>Chapter 6: From folding to catalysis: Studies of small RNA enzymes</b>	<b>200</b>
6.1. Introduction to catalytic RNA structures.....	200
6.2. The hammerhead ribozyme: a three-way RNA junction with catalytic function.....	202
6.2.1. Identification and action of the hammerhead ribozyme.....	202
6.2.2. Structural studies of the hammerhead ribozyme.....	203
6.2.3. Folding pathway of the hammerhead ribozyme at the ensemble-FRET level.....	206
6.2.4. Single-molecule FRET studies of the hammerhead ribozyme.....	208
6.3. Results and discussion.....	209
6.3.1. Single-molecule folding dynamics of the minimal hammerhead ribozyme.....	209
6.3.2. Single-molecule analysis of conformational changes involving P1 and P2 stems in the minimal hammerhead ribozyme.....	210
6.3.3. Single-molecule analysis of conformational changes involving P1 and P3 stems in the minimal hammerhead ribozyme.....	215
6.4. Using sm-FRET to investigate the folding dynamics of the natural hammerhead ribozyme.....	220
6.5. Conclusions.....	223
<b>Chapter 7: Implementation of single-molecule techniques with freely- diffusing molecules</b>	<b>227</b>
7.1. Introduction.....	227
7.1.1. Fluorescence imaging without surface immobilisation.....	227

7.1.2.	The property of diffusion .....	228
7.1.3.	Single-molecule fluorescence techniques to measure diffusion .....	230
7.1.4.	Detection of single-photon emission .....	231
7.1.5.	Fluorescence correlation spectroscopy and cross-correlation spectroscopy .....	232
	Single detector: Autocorrelation .....	232
	Two detectors: Cross-correlation .....	235
7.2.	Experimental realisation of single-molecule FCS .....	237
7.2.1.	Excitation: designing a single-molecule confocal microscope .....	237
7.2.2.	Resolving the emission from a confocal spot onto a point detector .....	239
7.2.3.	Electrical diagram to collect emission signals and allow FCS resolution .....	241
7.3.	System verification using different fluorescently-labelled species .....	244
7.3.1.	Calculating the confocal volume using a sample with known diffusive properties .....	244
7.3.2.	Protein-DNA interactions observed using FCS .....	246
7.3.3.	Monitoring the swelling and solubilisation of lipid vesicles with Triton X-100 using FCS .....	248
7.4.	Moving from single- to two-colour FCCS .....	253
7.5.	System modifications to optimise for two-colour detection .....	255
7.6.	Extracting FRET information from individual intensity bursts .....	256
7.6.1.	Extracting the background level from each channel .....	257
7.6.2.	Extracting and analysing the individual fluorescence bursts .....	259
7.7.	Experimental test of non-immobilised sm-FRET for a doubly-labelled DNA structure .....	261
7.8.	Towards single-molecule multi-parameter fluorescence spectroscopy (sm-MFD) .....	264
7.8.1.	System modifications to measure fluorescence lifetimes .....	264
7.8.2.	Extracting polarisation information from single-molecule emission: fluorescence anisotropy .....	267
	Fluorescence anisotropy .....	267
	Experimentally determining fluorescence anisotropy .....	268
	Correction of emission polarisation on detection efficiencies .....	269
7.8.3.	Electrical circuitry for sm-MFD .....	270
7.9.	Experimental verification of the sm-MFD system: Examining a previously characterised structure .....	272
7.9.1.	SAM-I riboswitch studies using sm-FRET shows three conformational states .....	272
7.9.2.	Investigating SAM-I riboswitch population changes using sm-MFD .....	275
	Examining the unfolded SAM-I aptamer using sm-MFD .....	276
	Examining the magnesium-induced intermediate state using sm-MFD .....	280
	Using sm-MFD to study the ligand-bound state of the SAM-I riboswitch .....	284
7.10.	Conclusions .....	289
<b>Chapter 8: Conclusions and future directions</b> .....		<b>292</b>
	Conclusions from experimental work .....	292
	Future directions .....	301
<b>Appendices</b> .....		<b>303</b>
I.	Deriving the Förster radius ( $R_0$ ) from first principles .....	303
II.	Uncertainty calculations and combinations .....	307
III.	Deriving the fluorescence correlation function $G(\tau)$ .....	308
IV.	List of publications .....	313
<b>References</b> .....		<b>315</b>



# Chapter 1: Introduction to fluorescence and single-molecule methods

## 1.1. Origins of the theory of fluorescence

Throughout the various fields of science, the observation and detection of photoluminescence is now one of the most fundamental of measurements [1]. Whether present through natural or manipulated means, the recording of fluorescent signals offers a wealth of structural data at the macroscopic, cellular and even sub-cellular levels.

Nicolás Monardes, a Spanish botanist, was the first to observe and use the phenomenon in a scientific context [1]. In 1565, he used the emissive colour resulting from a chemical reaction between the wood of *Lignum nephriticum* and water in order to identify counterfeit samples [1]. Further important milestones followed in the nineteenth century: Sir David Brewster observed the transmission of red light through an alcohol solution of chlorophyll under illumination from sunlight in 1833, and Sir John Herschel observed blue emission from a solution of quinine in 1845 [2, 3]. However, both of these papers believed the luminescence to have originated as a consequence of scattering, and it wasn't until the work by Sir George Gabriel Stokes on quinine that the emission after absorption model was proposed [4, 5]. Stokes was the first to use the term “fluorescence”, as a means of distancing himself

from the original terms “dispersive reflection” and “true internal dispersion”.

## 1.2. Principles of photoluminescence: From absorption to emission

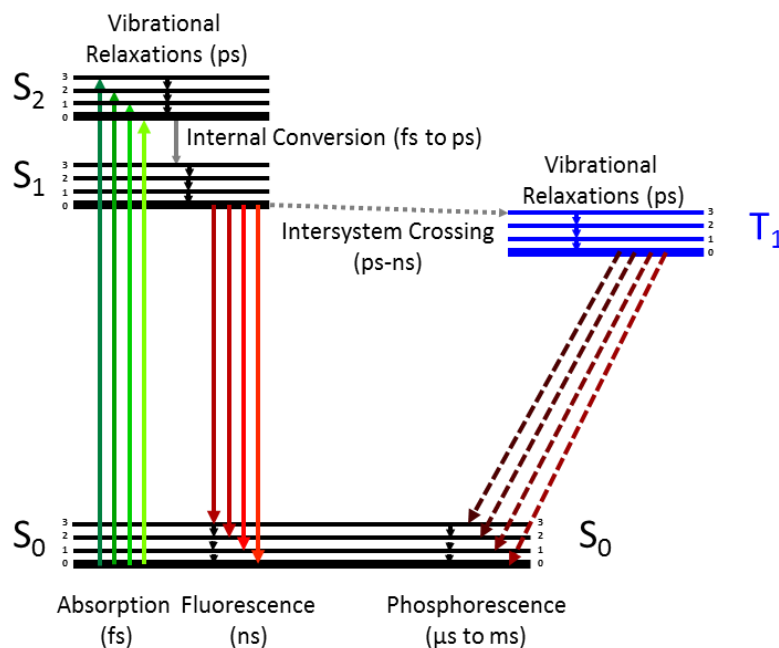
One of the first quantum mechanical milestones was the discovery of the quantisation of electron energy levels radiation. In Bohr’s model of the atom, he postulated that electrons may move up or down through these energy levels with the absorption or emission of electromagnetic radiation of a suitable energy [6]. If an electron absorbs a photon whose energy is equal to that of the band gap between its current energy level ( $S_0$ ) and a higher, unoccupied level ( $S_1$ ) then it will move to this state. When it has been promoted, it must return to its ground state ( $S_0$ ) and will do so spontaneously after a brief period of time. This relaxation occurs first by vibrational relaxations within the same electronic state to reach the lowest point within the energy band and then by the emission of a photon to move between energy levels [7]. This second stage, the emission of a photon, is known as photoluminescence.

One may divide photoluminescence into two mechanisms: fluorescence and phosphorescence. Both are emissive processes, the steps of which are most clearly illustrated by a Jablonski diagram, as shown below in Figure 1.1, although the lifetimes associated to each vary enormously.

The cycle of fluorescence occurs in four stages: the first stage is the molecule reaching an excited state, usually due to the promotion of an electron to a higher electronic state ( $S_1$ ), for instance by providing a photon of appropriate energy for the electron to cross the energetic band gap [8]. The molecule then loses some energy due to vibrational relaxations, before reaching the lowest vibrational level of the excited electronic state. At this point, the molecule may return to the ground state by emitting a photon or by a non-radiative process [9].

In the case of phosphorescence, however, there is one significant difference: as the molecule relaxes through the excited energy band ( $S_1$ ), it undergoes a forbidden transition from the singlet  $S_1$  level to the triplet  $T_1$  level [10]. This process is known as inter-system crossing. In order to return to the singlet ground state, it must undergo a second forbidden transition. Although forbidden, quantum mechanically it is possible, though as a result of the kinetics of the transition it is a much slower process than fluorescence emission, usually by three to six orders of magnitude [9].

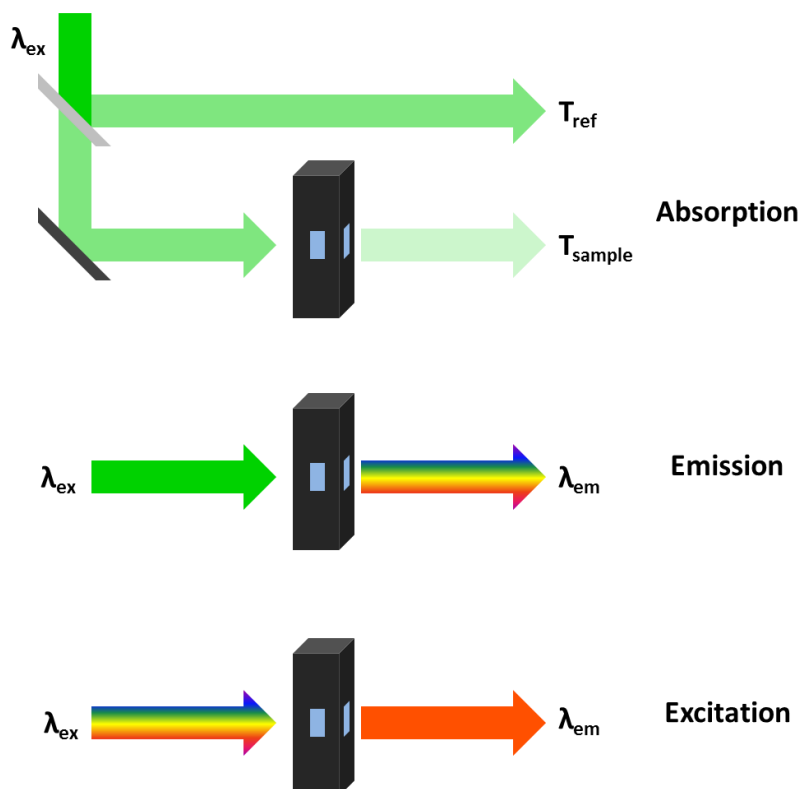




**Figure 1.1:** Illustration of a Jablonski Diagram. The paths for each process which can occur are shown, along with their associated timescales. The solid lines are indicative of a permitted transfer where the Pauli exclusion principle is not violated, these being singlet-singlet transitions. The dotted lines indicate a forbidden transfer, where the multiplicities of the initial and final states differ. One must note that the permitted transitions occur on significantly shorter timescales than their forbidden analogues (e.g. fluorescence and phosphorescence)

### 1.3. Fundamental properties of fluorescent media

One typically characterises the fluorescence properties for a given sample in one of three ways, through recording the absorption, excitation or emission spectra. The three mechanisms are summarised in Figure 1.2 below.



**Figure 1.2:** Summary of the three experimental designs used to probe the fluorescence for a sample. The colours of the block arrows are indicative of the wavelengths examined at each end: the excitation wavelengths  $\lambda_{\text{ex}}$  and the emission wavelengths  $\lambda_{\text{em}}$ . The solid arrows denote a fixed wavelength, whereas the multi-coloured arrows denote a swept wavelength. The intensity of the beams, in the case of the absorption scheme, is denoted by the colour intensity of the reference ( $I_{\text{ref}}$ ) and sample beams ( $I_{\text{sample}}$ ).

### Absorption spectroscopy and the Beer-Lambert law

The absorption of the sample is measured by the scheme illustrated in Figure 1.2. The sample is illuminated by a single wavelength, the peak absorption wavelength of the sample, and the transmission of this light through the sample recorded by a detector. This intensity is compared to a reference beam, of the same initial intensity, through Equation 1.1 shown below:

$$T = \frac{I_{sample}}{I_{ref}} \quad (1.1)$$

$$A = -\log_{10} T \quad (1.2)$$

The transmission allows the absorbance ( $A$ ) to be calculated through Equation 1.2. At the same time, the transmission may also be defined in terms of the molar absorption coefficient ( $\epsilon$ ), the concentration ( $C$ ) and the length of the sample propagated by the excitation beam ( $\ell$ ) (Equation 1.3).

$$T = 10^{-\epsilon\ell C} \quad (1.3)$$

Through recording the transmission through the sample, and using a known cuvette size and a dye of known spectral properties, the concentration of the dye can be determined, which may allow properties like the labelling efficiency to be determined for a specific sample, for example in nucleic acids which themselves absorb in the ultraviolet spectral range.

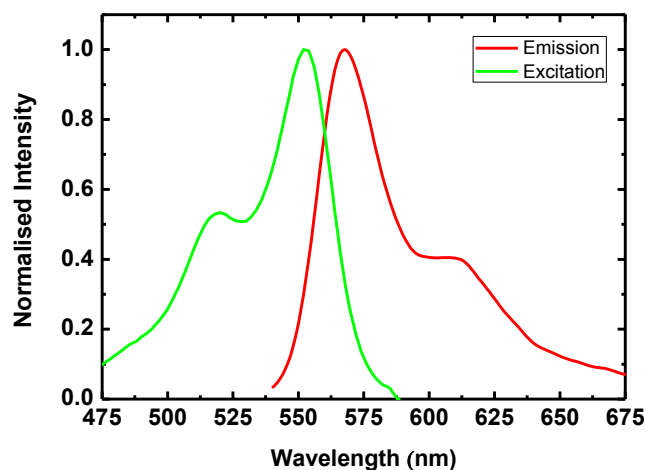
### **Excitation and emission spectra**

The full spectral properties of the sample and its dye may be tracked by sweeping the wavelengths of both excitation and emission light. The light swept determines which spectrum is resolved. By sweeping the fluorescence wavelength detected whilst holding excitation constant, one resolves the emission spectrum for the sample, which gives the intensity of light of all wavelengths emitted.

If, however, the wavelength detected is held constant and the excitation wavelength is scanned with the emission recorded at a constant wavelength, one obtains a plot of the wavelength dependence of emission at the peak wavelength, and hence the excitation spectrum.

### 1.3.1 Stokes Shift

In his 1852 paper, Stokes proposed that for a fluorescent species the wavelength of the light absorbed must always be less than that of its fluorescence emission [4]. This rule is what is known as the Stokes shift, yet the first statement of this property of light emission was proposed ten years earlier by Edmond Becquerel, a French physicist working on the UV-stimulated light emission of calcium sulphide. The key difference between these is the emissive nature of each sample; where quinine is a fluorescent material, calcium sulphide is phosphorescent [1]. One can see the Stokes shift by simply plotting the emission and excitation spectra for a fluorescent dye together, as shown in Figure 1.3 for the cyanine dye Cy3.



**Figure 1.3:** The excitation and emission spectra of a fluorescent molecular label, Cy3, normalised and plotted against wavelength in nanometres. The excitation spectrum is shown in green and the emission in red. The spectra were normalised to the peak values of each trace for ease of comparison.

### 1.3.2 Kasha's Rule

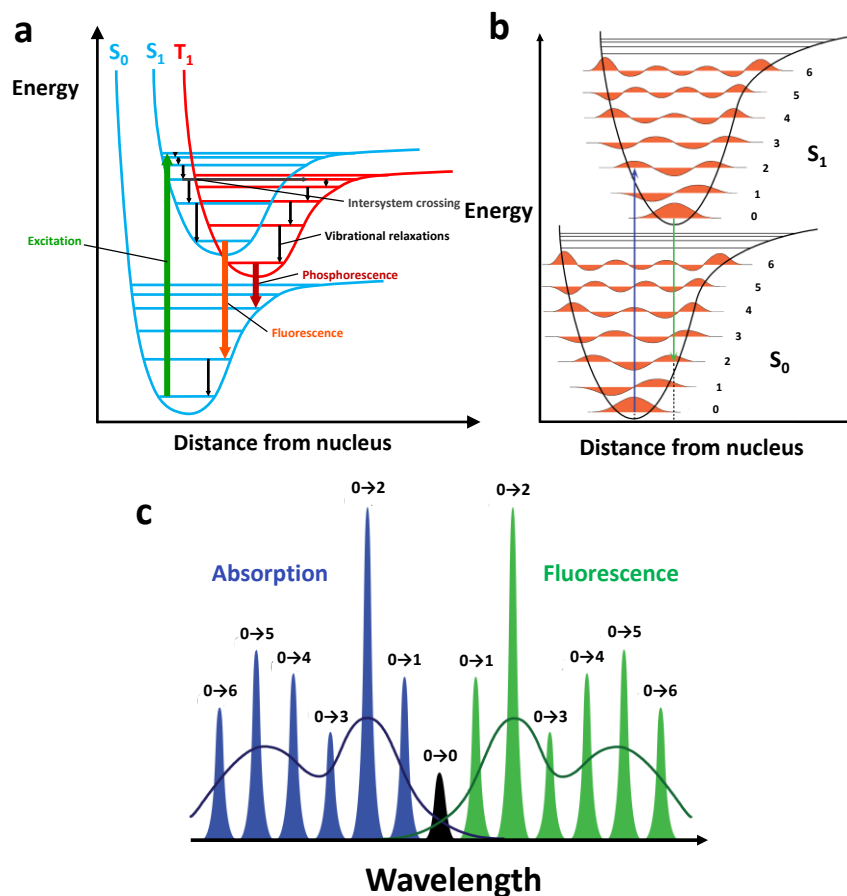
One observes that regardless of the choice of excitation wavelength, providing that it falls within the suitable range, the emission spectrum will always have the same composition i.e. the emission curve will have the same shape [9]. This phenomenon is Kasha's rule, named after the American chemist Michael Kasha, who identified it in 1950 [11]. The explanation of Kasha's rule can be interpreted from a Jablonski diagram. The simplified example in Figure 1.1 shows that the molecule decays non-radiatively through the energy levels until it reaches the lowest excited state, provided that they are sufficiently close in energy for this to be possible. Quite simply, the choice of excitation wavelength determines the starting point of this internal conversion. This decay is very fast, typically of

the order of picoseconds, six orders of magnitude faster than typical fluorescence lifetimes [9].

### 1.3.3 Franck-Condon Principle and the Mirror Image Rule

The Jablonski diagram shown in Figure 1.1 may be plotted with an additional dimension, this being the distance of the electronic states from the nucleus, exhibiting the distance-dependent shape of the Morse potential wells (Figure 1.4a). As one may also observe from the excitation and emission spectra in the example above in Figure 1.3, the curves seem to be almost mirror images of each other - this is known as the mirror image rule, which, although not universally true, holds for many common fluorescent molecules [9].

The mirror image rule is a consequence of the Franck-Condon principle, which states that the process of absorption, the emission of a photon and the associated rearrangement of the electrons within the electronic levels, happens on a much shorter timescale than any nuclear motion. By working under the assumption that the energy levels have the same form with regard to the band structure, and considering these plotted against distance from the nucleus (Figure 1.4), the relative probabilities for transitions between energy levels within the bands can be derived by drawing a vertical line from the ground state up through the excited state and calculating the probability that a transition between each level will occur. This is illustrated in Figures 1.4b and 1.4c.



**Figure 1.4:** a) Jablonski diagram with the nuclear coordinate factored in, showing the positional variation of the energy levels. The diagram is the same as Figure 1.1 with the only difference being the omission of the second excited state  $S_2$ . b) The wave functions of the electronic states associated with each vibrational state within an electronic state. The blue line indicated absorption and the green line indicated fluorescence. c) The contributions from each vibrational transition to the absorption and emission spectra. The magnitude of each is governed by the overlap of the wave functions for the two states in question, centred upon the initial state. The envelopes shown for both the absorption and emission spectra come from inhomogeneous broadening effects. Adapted from [12].

### 1.3.4 Fluorescence lifetime

When a molecule is promoted to an excited state, it may exist there for a period of time before relaxing back to the ground state. The relaxation process to the ground state takes place via a combination of radiative and

non-radiative processes. The latter include thermal relaxations via energy exchange with the surrounding solvent, collisional quenching with solvent molecules and many other processes [6]. The change in intensity of emission ( $I$ ) as a function of time from a fluorescent sample is expressed by Equation 1.4, where  $k_{total}$  is defined as the total decay rate for the sample.

$$\frac{dI}{dt} = -k_{total}I \quad (1.4)$$

The total rate of decay ( $k_{total}$ ) comprises of radiative ( $k_r$ ) and non-radiative components ( $k_{nr}$ ) as expressed in Equation 1.5. One defines the fluorescence lifetime  $\tau$  of a sample as the inverse of the total rate of decay for a fluorescent species [9], as stated in Equation 1.6.

$$k_{total} = k_r + k_{nr} \quad (1.5)$$

$$\tau = \frac{1}{k_r + k_{nr}} \quad (1.6)$$

$$\tau_{natural} = \frac{1}{k_r} \quad (1.7)$$

The natural fluorescence lifetime ( $\tau_{natural}$ ) is defined as the rate of decay due exclusively to photon emission and therefore is defined as in Equation 1.7.



### 1.3.5. Fluorescence quantum yield

The fluorescence quantum yield ( $\Phi$ ) for a fluorescent species is a measure of its quantum efficiency, namely of the efficiency of conversion from an absorbed photon to an emitted photon. Specifically it is calculated through Equation 1.8 as the ratio between absorbed and emitted photons:

$$\Phi = \frac{\text{Number Emitted Photons}}{\text{Number Absorbed Photons}} = \frac{k_r}{k_r + k_{nr}} \quad (1.8)$$

The ideal fluorophore would clearly have a quantum yield as close to 1 as possible, which would signify an ideal label with 100% emission per absorbed photon.

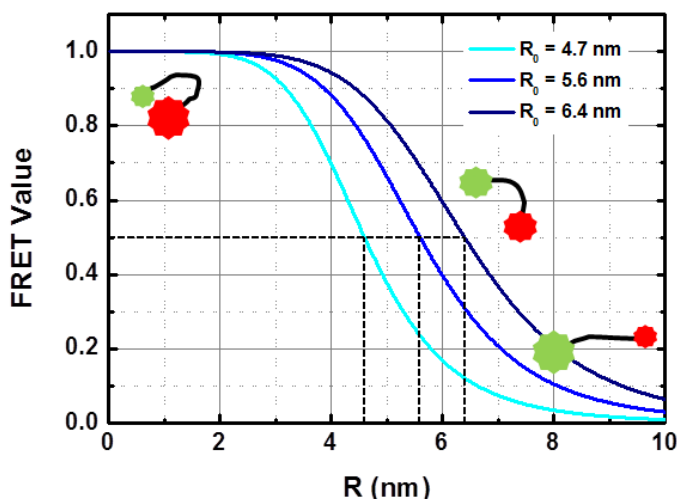
## 1.4. Nanometre-sized distance measurements using fluorescence: Förster Resonance Energy Transfer (FRET)

### 1.4.1. Basic concepts of FRET

In 1948, Theodor Förster theorised that a non-radiative transfer of energy could occur between two molecules in close proximity, specifically within the range that their molecular dipole moments can interact [13]. In addition to this distance requirement, the absorption spectrum of one of the molecules, known as the acceptor, must overlap with the emission spectrum of the other, called the donor [14]. A pair of dye molecules whose spectra fit this criterion is called a FRET pair.

The transfer rate ( $k_T$ ) for this dipole-dipole interaction is given by Equation 1.9 below. The strong dependence of the rate of energy transfer with the inverse of the sixth power results in a very sensitive intermolecular ruler, though on a distance scale limited to a range of 2 to 8 nanometres [9], [15]. A plot illustrating this effective range is shown in Figure 1.5 below.

$$k_T = k_D^0 \left(\frac{R_0}{R}\right)^6 = \frac{1}{\tau_D^0} \left(\frac{R_0}{R}\right)^6 \quad (1.9)$$



**Figure 1.5:** Plot showing the variation in the FRET value as a function of dye separation for a specific FRET pair with a known Förster radius. Simplified diagrams showing the proximity of the dyes to each other are placed at the areas along the curve to which they apply; the high FRET portion of the curve has the dyes very close together, whereas the low FRET region has them at their greatest separation.

The rate of transfer also depends upon the emission rate for the donor ( $k_D^0$ ), which may also be expressed in terms of the acceptor-free lifetime of the donor ( $\tau_D^0$ ), and on the Förster radius of the FRET pair ( $R_0$ ) which is defined by Equation 1.10. This radius is the dye separation required for the

rates of donor emission and energy transfer to be equal, and depends on the orientation factor ( $\kappa$ ) between the dyes, the quantum yield of the donor ( $\Phi_D$ ), Avogadro's number ( $N_A$ ), the refractive index of the medium ( $n$ ) and the spectral overlap integral ( $J(\lambda)$ ).

The spectral overlap integral ( $J(\lambda)$ ) is also defined in Equation 1.10, and is a function of acceptor extinction coefficient ( $\epsilon_A$ ) and the normalised intensity spectrum of the donor ( $I_\lambda^D$ ) and the wavelength ( $\lambda$ ).

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

$$\text{where } J(\lambda) = \int_\lambda I_\lambda^D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda \quad (1.10)$$

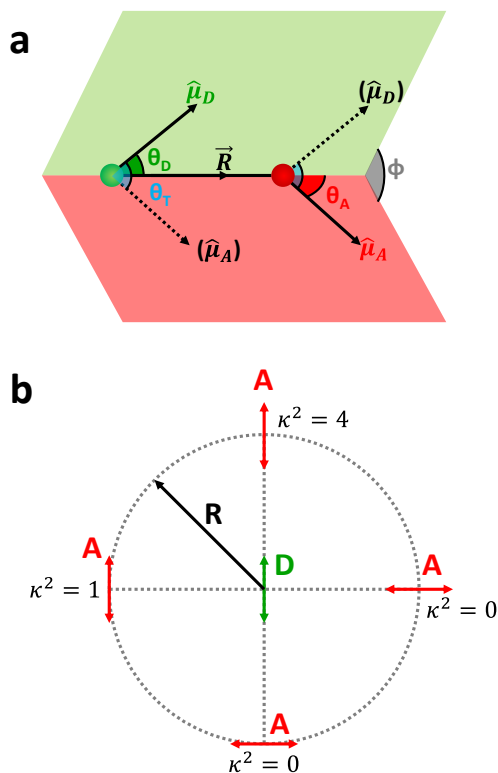
$$\text{and } \kappa^2 = (\cos\theta_T - 3\cos\theta_D \cos\theta_A)^2$$

$$\text{or } \kappa^2 = (\sin\theta_D \sin\theta_A \cos\phi - 2\cos\theta_D \cos\theta_A)^2$$

The orientation factor ( $\kappa$ ), along with the inter-dye separation, further influences the recorded FRET value. The value is related to the angles formed when considering the interaction between the dipole moments of the donor and acceptor dyes ( $\hat{\mu}_D$  and  $\hat{\mu}_A$ ) and their separation vector ( $\vec{R}$ ). A simplified image showing the angles involved between the dipoles and the separation vector ( $\vec{R}$ ) is shown below in Figure 1.6a.

Although it is possible to calculate the exact experimental range of orientation factors for a specific dye labelling scheme using fluorescence polarisation spectroscopy (see chapter 7, section 8 for more detail on this topic), it is generally assumed that the motion of the two dyes is fast and

isotropic and, as a result, a value of  $\kappa^2 = 2/3$  is used [16]. Should one or both of the dyes be rigidly confined, however, an orientation factor of  $\kappa^2=0.475$  is used [16]. The orientation factor can be checked by assessing the polarisation of emission light, by using a polarizing beamsplitter to separate s- and p-components for a point detection scheme, or in the case of a widefield study by using a Wollaston prism to create a polarisation-based offset between each component and resolving each image to separate halves of an EM-CCD chip [17].



**Figure 1.6:** Diagrams summarizing the orientation factor involved in defining the Förster radius. a) Interaction of the donor dipole-separation vector ( $\hat{\mu}_D$  and  $\vec{R}$ ) and acceptor dipole-separation vector planes ( $\hat{\mu}_A$  and  $\vec{R}$ ), highlighting the key angles between each component used to calculate the orientation factor. b) Extreme cases for the interacting dipoles, with the orientation factors associated with each scenario shown alongside the figure. These cover the range of possible values that the orientation factor can take, i.e.  $0 \leq \kappa^2 \leq 4$ .

The relationship between the efficiency of energy transfer and the distance is given by Equation 1.11.

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \quad (1.11)$$

This equation requires prior knowledge of the Förster radius, which can be calculated from Equation 1.10 if the spectral overlap and the quantum yield of the donor in the absence of acceptor are known. For fluorescence labels covalently attached to the biomolecule using a long carbon linker, the fluorophore is assumed to be freely rotating around this linker and thus the orientation factor can be assumed to adopt a value of  $2/3$ . Controls to ensure this assumption is correct are required for each FRET pair and labelling position and can be performed by checking the fluorescence anisotropy of the donor and acceptor. The efficiency of energy transfer ( $E_{FRET}$ ) is the magnitude that is experimentally determined, and from this, the D-A distance can be extracted using Equation 1.11. At ensemble level, one may calculate the FRET efficiency ( $E_{FRET}$ ) in different ways. The first method relies on monitoring the intensity of the donor in the absence and presence of acceptor and calculating  $E_{FRET}$  as stated in Equation 1.12

$$E_{FRET} = 1 - \left(\frac{I_{DA}}{I_D}\right) \quad (1.12)$$

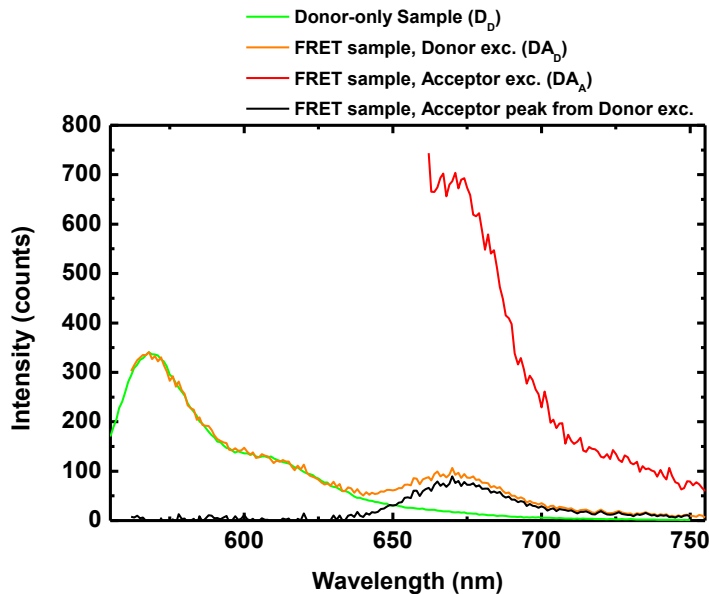
where  $I_D$  is the emission intensity where only the donor is present, and  $I_{DA}$  is the emission intensity of the donor dye where the donor and acceptor are both present on the sample

The second method is based on monitoring the lifetime of the donor in the presence and absence of the acceptor and calculating the FRET efficiency from Equation 1.13.

$$E = 1 - \left( \frac{\tau_{DA}}{\tau_D} \right) \quad (1.13)$$

Here,  $\tau_D$  is the lifetime where only the donor is present, and  $\tau_{DA}$  is the lifetime of the dye where the donor and acceptor are both present on the sample.

The last method involves monitoring the sensitised emission of the acceptor following the RatioA method. For this, three scans must be taken: donor excitation of a structure with only the donor present ( $D_D$ ), donor excitation for the structure with both FRET dyes ( $DA_D$ ) and direct acceptor excitation of the dual-labelled construct ( $DA_A$ ). One first normalises the  $D_D$  and  $DA_D$  spectra and subtract  $DA_D - D_D$ , as shown in the upper panel of Figure 1.6. This isolates the acceptor-only component that is excited through energy transfer.



**Figure 1.6:** Overlay of the three scans required to isolate the RatioA peak in order to determine the extent of energy transfer. The traces are coded as  $D_D$  (green),  $DA_D$  (yellow) and  $DA_A$  (red), with the acceptor peak due to donor excitation in black.

The  $DA_A$  scan gives a quantitative indication of the amount of acceptor present in the sample. This would correspond to a 100% energy transfer scenario. As a result, one may quantify the extent of energy transfer by taking the ratio of the intensities of the peak wavelengths, as shown in Equation 1.14; this value is known as the RatioA, which indicates the ratio of the areas of each spectrum.

$$RatioA = \left( \frac{I_A^D}{I_A^A} \right) \quad (1.14)$$

At the single-molecule level, the FRET efficiency is calculated using the intensities of the donor ( $I_D$ ) and acceptor dyes ( $I_A$ ) separately, and is stated in Equation 1.15.

$$E = \frac{I_A}{I_D + \alpha I_A} \quad (1.15)$$

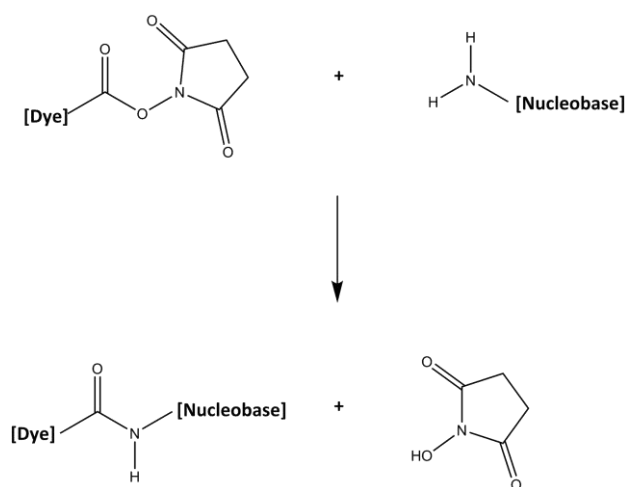
In addition to the intensities, a correction factor ( $\alpha$ ) is introduced, which accounts for the leakage from the spectral overlap between the emission spectra and direct excitation of the acceptor from the source [18]. When this correction factor is not considered, the FRET ratio calculated is known as the apparent FRET ( $E_{app}$ ).

#### 1.4.2. Common FRET pairs in fluorescence spectroscopy

Some common FRET pairs are listed in Table 1.1 below. In order for two fluorescent dyes to be an effective FRET pair, their spectral overlap must be sufficient to allow an efficient energy transfer rate when they are in close contact, yet not too great so as to prevent an adequate separation of the donor and acceptor emission intensities. Among other important requisites, the donor species must have strong absorption and its absorption spectra should allow the excitation of the donor with minimal direct excitation of the acceptor. As previously mentioned, the donor should have high quantum yield and its rotational freedom should be high enough to ensure that the  $2/3$  approximation of the orientation factor is preserved. Moreover, the fluorescent dyes (donor and acceptor) should be commercially available with chemical groups that allow site-specific attachment to the nucleic acid sequence. Covalent attachment to the nucleic acid structure at a particular



base, sugar or phosphate group is normally performed either during solid-phase chemical synthesis using phosphoroamidite chemistry or post-synthetically. For the latter, the most common method is the incorporation of an amino-linker at a specific location, followed by chemical coupling with an N-hydroxyl-succinimide (NHS) ester derivative of the dye (see Figure 1.7 below).



**Figure 1.7:** Scheme for labelling a nucleic acid with a site-specific modification inserted during synthesis. The reactive groups involved in the reaction are shown, along with the structure of the links and the by-product of the reaction. Typically the amino-linker is attached to the nucleobase, though this is not exclusively the case.

In the table below, along with the specific dye molecules constituting each FRET pair, the peak excitation wavelengths for the donor dyes, the Förster radii for each FRET pair and the peak emission wavelengths for each member of the pair are also listed.

**Table 1.1:** A list of common FRET pairs and their fluorescence properties [14, 18–22].

<b>Donor</b>	$\lambda_{\text{exc}}(\text{max})$ (nm)	$\lambda_{\text{em}}(\text{max})$ (nm)	<b>Acceptor</b>	$\lambda_{\text{em}}(\text{max})$ (nm)	$R_0$ (Å)
Fluorescein	494	521	Cy3	560	56
Fluorescein	494	521	Cy5	667	44
Alexa488	495	519	Alexa647	668	39
Alexa488	495	519	Cy5	667	52
Cy3	547	560	Cy5	667	53

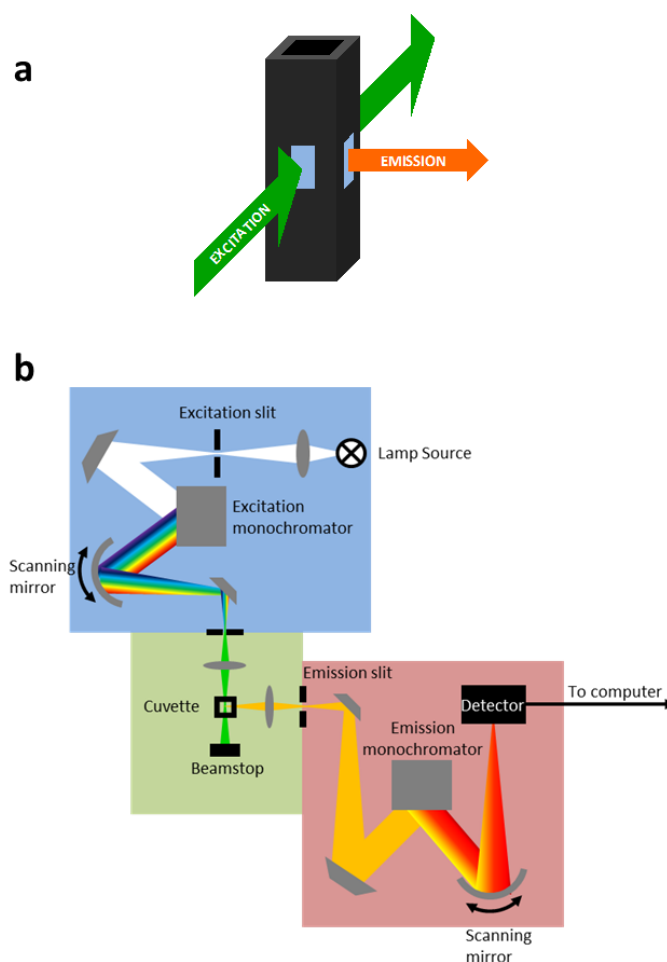
### 1.4.3. Bulk methods to measure FRET

Bulk methods in fluorescence spectroscopy are all those experimental procedures performed in a conventional fluorimeter using a microlitre-sized cuvette, or simply a microcuvette. Here, millions of biomolecules labelled with the FRET pair are simultaneously targeted and an average  $E_{\text{FRET}}$  magnitude is extracted. The advantage of using a bulk measurement system is the speed of data acquisition; one may acquire all of the information required in a single scan and extract the relevant parameters using the methods previously described in order to calculate the FRET efficiency. The experimental configurations used to calculate bulk-FRET parameters through steady-state or time-resolved approaches are briefly described below.

#### Steady-state FRET measurements

In the case of a doubly labelled sample, the simplest mechanism for extracting FRET information is by performing a steady-state scan on the system using a conventional fluorimeter and a microcuvette. One excites the entire volume of the sample at the peak excitation wavelength of the donor

and collects the donor and acceptor emission emitted at 90 degrees, as shown in Figure 1.8a. This helps to minimise the collection of scattered light in the fluorescence emission, which can cause damage to the sensitive detectors used in such devices. FRET was observed for the first time in a fluorescence experiment using a steady state setup investigating the emission of fluorescent tryptophyl monomers in peptide chains [23].



**Figure 1.8:** a) Diagram showing the excitation and fluorescence collection vectors. The microcuvette has windows on three sides, which allows the fluorescence to be recorded from a perpendicular vector to the emission pathway. b) Simplified schematic of a conventional steady-state fluorimeter, showing all key components involved in its operation. The operational details are elaborated in the main text below.

The optical schematic for a fluorimeter is laid out in Figure 1.8b. The light from the source is collimated and then directed into an excitation monochromator where the excitation light is split by a diffraction grating to separate its wavelength components. A combination of a steering mirror and a slit is used to isolate the excitation wavelength required, which is then focussed into the microcuvette. The emission is, as shown in Figure 1.8a, collected at  $90^\circ$  to the excitation and its intensity modulated by an excitation slit. The fluorescence emission from the sample is split by a second diffraction grating and scanned across the detector using a second steering mirror in order to resolve its emission spectrum.

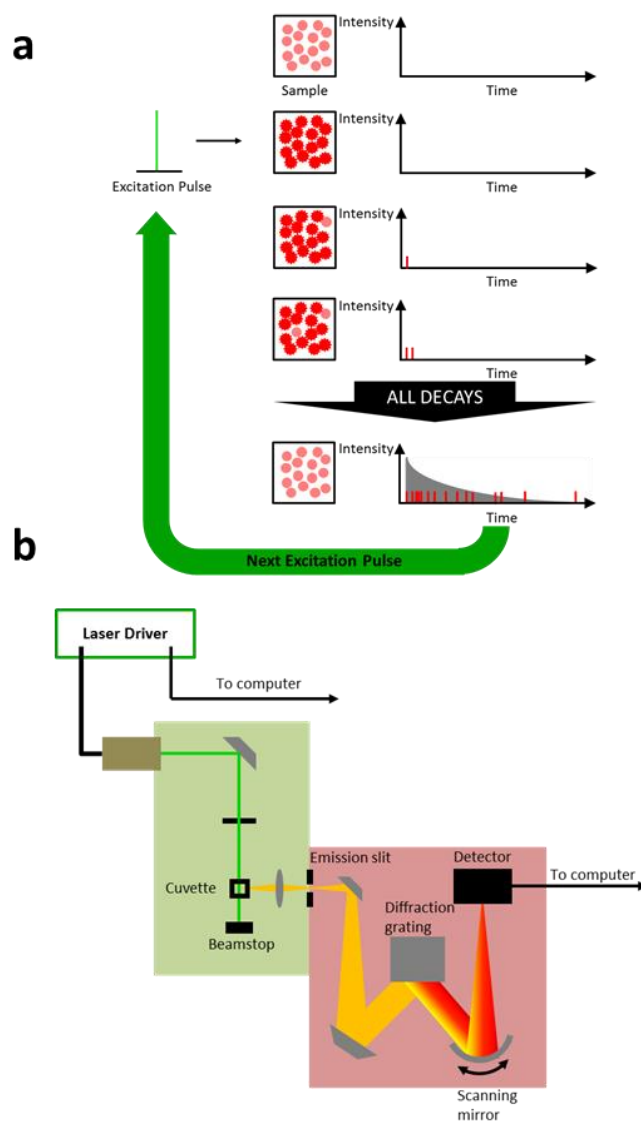
In order to estimate the FRET of a suitable sample, the donor is excited at its peak wavelength and fluorescence collected from a wavelength of at least 10 nm greater in order to prevent the direct measurement of scattered light. The light is polarised at an angle of  $54.7^\circ$ , commonly known as the magic angle, this eliminates the influences of polarised emission fluctuations on the spectra due to changes in molecular orientations, and allows the true FRET value to be obtained following the procedures described in section 1.4.1.

### **Fluorescence lifetime-based FRET measurements**

An alternative means of measuring the FRET between two fluorescent labels on a sample is to use time-correlated single photon counting (TCSPC). As in the case for a steady-state bulk measurement, one requires two

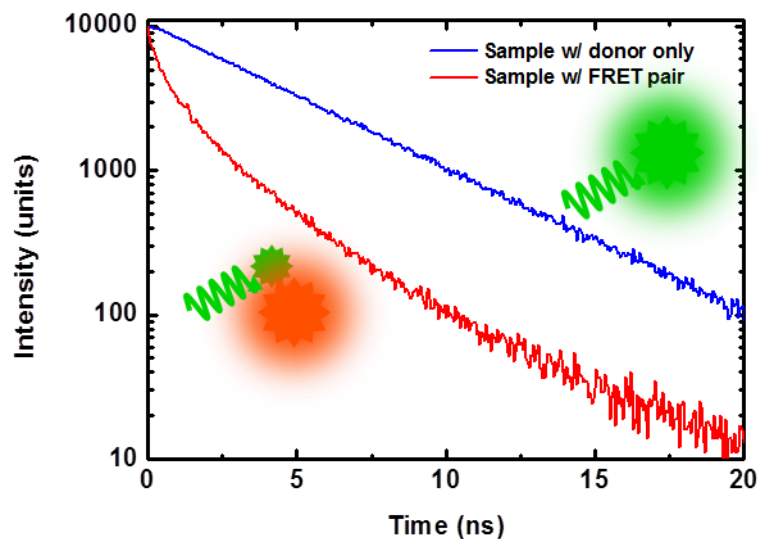
differently labelled constructs; one with a donor dye only and one with both the donor and acceptor present.

Consider the scenario outlined below in Figure 1.9a: a picosecond laser pulse excites the FRET donor, and the intensity of the light emitted by the donor is collected as a function of time by a detector. The emission intensity profile follows that of an exponential decay, as described in Equation 1.4, and the acquisition of this exponential decay of fluorescence intensity is described in Figure 1.9a below.



**Figure 1.9:** a) Simplified description of how the time-dependent decay of the donor following fast excitation by a pulsed laser is obtained from a TCSPC experiment. A sample, initially completely in the ground state, is excited by a laser pulse with a short pulse width, typically  $\sim 50$  ps. Each molecule relaxes in time through the time window of acquisition, before all are re-excited by a new laser pulse. b) Optical schematic of a TCSPC system for measuring bulk samples. The pulsed excitation source is directed through the cuvette holding the sample. As for the steady-state schematic, emission is collected perpendicular to the emission and is then split into its wavelength components before being directed into the detector. To increase time resolution, photomultipliers used in steady-state measurements with temporal responses in the 0.5-1 ns regime are replaced by microchannel plates (time response in the ps regime). See main text for details of TCSPC operation.

This process of photon counting lies at the heart of TCSPC. The optical layout for such a system is shown in Figure 1.9b. A sequence of short pulses ( $\sim 50$  ps FWHM) is generated by a laser driver at high frequency ( $\sim 40$ - $80$  MHz), and these are passed through the sample. As for the steady-state system, in order to minimise scattered light the fluorescence is collected at  $90^\circ$  to the excitation. A diffraction grating and a scanning mirror combine to allow wavelength-specific detection, which is set to the donor emission peak. The fluorescent decay for both the donor-only and donor and acceptor-labelled samples are obtained using the same parameters, and only the donor emission region is used for detection. As a result of the energy transfer process being a competitive mechanism to deactivate the donor from the excited state, the fluorescence lifetime of the donor decreases (Figure 1.10). The magnitude of this decrease correlates with the D-A distance, so that the closer they are, more efficient energy transfer and faster decay of the donor following the excitation pulse (Figure 1.10). By collecting the fluorescence decay of the donor in the absence and presence of acceptor it is possible to determine the energy transfer rate and efficiency and thus to quantify the D-A separation following Equations 1.13 and 1.11, respectively. As in the steady-state mode, a magic angle configuration is used to avoid polarisation artefacts.

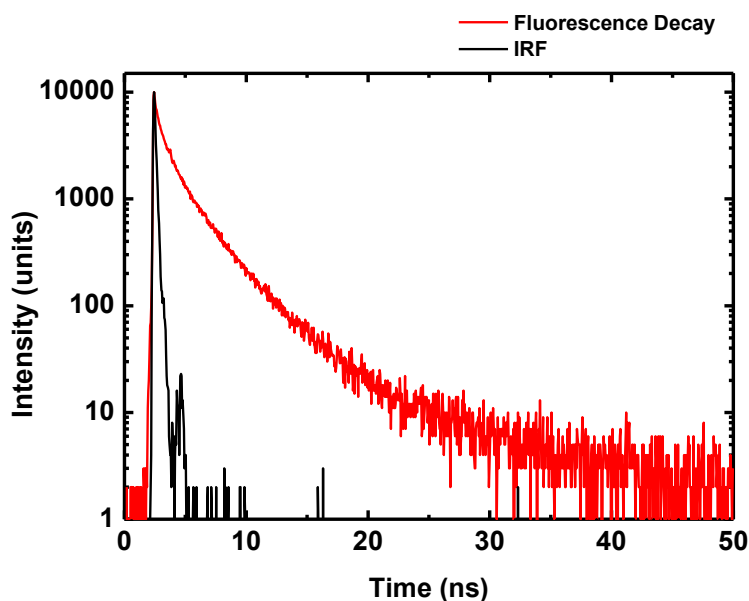


**Figure 1.10:** Effect of the presence of an acceptor dye on the fluorescence lifetime of the donor dye (fluorescein) attached to an RNA structure (adenine riboswitch). The straight line on the logarithmic scale demonstrated by the blue curve is indicative of a single exponential decay, where only one lifetime component is present. The red curve, however, shows a two-component decay process for the excited state of the donor dye; direct emission from the donor itself and energy transfer to the acceptor dye (Cy3). These two processes act together to deplete the excited population more quickly, and as a result the decay is much faster.

The intensity decay curve recorded is not, however, the decay from the dyes alone. The instrument response time plays a part in the measured decay because laser pulses are not infinitely short ( $\delta$ -pulse) and have a finite width (Figure 1.11). As a result, this profile must be taken into account and extracted from the curve in order to isolate the true fluorescence decay, as stated in Equation 1.16. The temporal profile of the laser pulse is obtained using a scattering medium and recorded at the wavelength of excitation. With this scan recorded, a deconvolution of the instrument response signal and the measured intensity must be performed, finally giving the intensity of fluorescence alone as shown in the figure below.



$$I_{measured}(t) = IRF(t) \otimes I_{fluorescence}(t) \quad (1.16)$$



**Figure 1.11:** Illustration of the instrument response function (IRF, in black) and the decay obtained from excitation with this pulse profile (red) for the adenine riboswitch labelled with fluorescein and Cy3. It can clearly be seen that the pulse has a finite width, and thus its contribution to the decay curve obtained must be removed. The decay curve has been fitted using a bi-exponential decay curve. The extraction of an average lifetime from this two component fitting process is covered in more detail in Chapter 3.

Having isolated the fluorescence component from the recorded decay, this is then used to extract the average lifetime of the donor in the presence and absence of acceptor. Depending on the specific location of the donor and the nature of the donor itself, the decay in the absence of acceptor may be fitted to a mono- or multi-exponential decay, with each component representing a relatively different positioning of the donor. Each of these components of the decay may exhibit slightly different rates of energy transfer. The average lifetimes with and without acceptor for an exponential fit with  $n$  components are calculated using Equation 1.17.

$$\tau_{average} = \frac{\sum_{i=1}^n A_i \tau_i}{\sum_{i=1}^n A_i} \quad (1.17)$$

This average is then implemented in Equation 1.13 to extract the average FRET efficiency from which to calculate the average FRET distance. The calculation performed for extracting the average lifetime from a bi- or multi-exponential decay curve is discussed in more detail in Chapter 3, Section 6.

## 1.5. Single-molecule spectroscopy

### 1.5.1. Advantages of single-molecule spectroscopy and the origins of the field

When one chooses to measure samples at the bulk level, a general idea of how all of the population is behaving is gathered, but the finer details of how the population is distributed are lost in the averaging process. Consider the example of the heights of pupils in a classroom: when one looks at the average, a single number will be returned. Though useful if compared to the average height of a different class, when this sample set is considered on its own this value doesn't provide the entire set of information that could be extracted if no averaging was performed.

Now consider examining all of the heights individually. This would not necessarily show you an average value immediately, not without calculation as for the bulk case, but what it would show would be how the distribution of heights is spread out for the class; statistics like the median and mode would become apparent, offering further insight into the statistics of this

sample population. This simple example offers the basis for the motivation for pursuing spectroscopy at the single-molecule level, namely this additional detail that may be obtained. So far, I have only mentioned the ability of single-molecule techniques to resolve steady-state heterogeneity (distribution of species in equilibrium). However, another fundamental advantage of single-molecule fluorescence compared to conventional fluorescence is the possibility of extracting the dynamic behaviour of individual molecules. Because single molecules are not synchronised in their fluctuations between different states, a bulk or ensemble characterisation can only retrieve partial information. Even if it was possible to synchronise all molecules to an identical initial state, as commonly done using T-jump or stopped-flow techniques, after a certain time they will become non-synchronised. The ability to follow the time-dependent changes in the conformation of molecules one by one bypasses this limitation. In single-molecule techniques because each molecule is monitored individually, there is no need for synchronisation and short-lived and low populated conformational states that will be hidden in conventional many-molecules FRET can be uncovered.

From an experimental perspective, single-molecule experimentation was first realised in 1961 by Boris Rotman, where the activity of individual enzymes was observed [24], before in 1976 Thomas Hirschfeld observed dye-labelled globulin proteins drifting through an excitation volume generated by a total reflection microscope [25]. The significant advances in the field

came first in 1989 by Moerner et al, where the absorption spectrum of a small hydrocarbon (pentacene) was resolved under near absolute zero conditions [26]. A year later, Orrit et al performed the same experiment but instead of detecting the absorption spectra for single pentacene molecules they recorded fluorescence, a study that constituted the first single-molecule fluorescence measurement [27]. In recognition of his contribution to super-resolution microscopy in performing this first experiment at the single-molecule level, William Moerner, along with Stefan Hell and Eric Betzig, was awarded the 2014 Nobel Prize in Chemistry [28].

The realisation of near-field scanning optical microscopy (NSOM or SNOM) in 1993 heralded a significant advancement in the field, allowing single-molecule spectroscopy to be performed at room temperatures [29], and further advances in technology and innovation in combining techniques have resulted in a huge variety of experimental designs being possible [30].

### 1.5.2. Single-molecule techniques

The principal issue with early single-molecule experiments was the signal-to-noise ratio. The low signals retrieved meant that the samples had to be cooled to temperatures close to absolute zero in order to increase the quantum yield of individual fluorophores and thus be able to isolate the sample data with certainty from the background. The error for a photon-counting experiment is described by Equation 1.18 below [31]. The paper by Thompson et al discusses the origins of the equation in great detail.

$$\sigma_{\mu_i} = \sqrt{\frac{s_i^2}{N} + \frac{a^2}{12N} + \frac{8\pi s_i^4 b^2}{a^2 N^2}} \quad (1.18)$$

The accuracy of locating a single molecule from the background noise is given by the term  $\sigma_{\mu_i}$  and the equation may be divided into three components, each one characterising a separate contribution that hinders the resolution of sample from noise. The first term ( $\frac{s_i^2}{N}$ ) governs the photon noise and is the detection area of the molecule (width of the point spread function in the x-y plane) divided by the number of detected photons. The second term ( $\frac{a^2}{12N}$ ) arises from the quality of the detector. This term is the area of one pixel ( $a^2$ ) divided by the number of photons detected, and gives an idea of the detection efficiency of the detector, in terms of how high its resolution is i.e. its pixel count. The final term is a combination of both sample and detector properties. This term is a measure of how the detector noise and background fluorescence from a sample affect the accuracy of the localisation of each molecule.

In order to reduce this background signal to a level where the fluorescence is very clear and easily resolvable, one must have a very low excitation volume that prevents significant background light from being excited. Two of the most common diffraction-limited fluorescence imaging techniques used in single-molecule spectroscopy are confocal microscopy and total internal reflection fluorescence microscopy (TIRFM), which form the basis of

many of the super-resolution or sub-diffraction-limited techniques available [30].

### Resolution limits from the diffraction limit of light

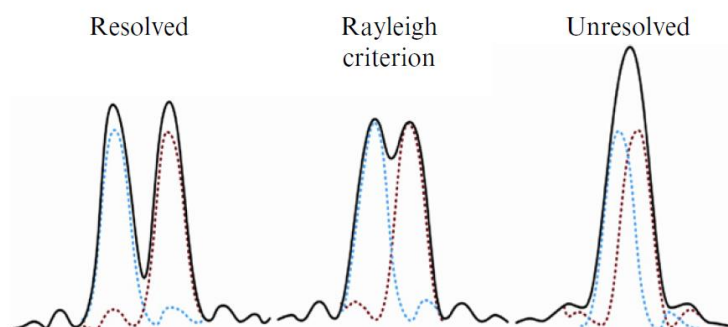
In 1873, it was proposed by Ernst Abbe that the primary factor in limiting the optical resolution of a microscope is the diffraction of the light passing through it. The intensity of emitted light from a molecule is not a point source but is spread out across a small ellipsoidal volume, defined by a point spread function (PSF), shown in Figure 1.12. Abbe proposed that the minimum spot dimensions  $w_{xy}$  and  $w_z$  for a system using a wavelength of light  $\lambda$  and focussing through a medium of refractive index  $n$  at an angle to the normal of  $\theta$  are defined by Equations 1.19a and 1.19b.

$$w_{xy} = \frac{\lambda}{2n \sin(\theta)} = \frac{\lambda}{2 NA} \quad (1.19a)$$

$$w_z = \frac{2\lambda n}{(n \sin(\theta))^2} = \frac{2\lambda n}{(NA)^2} \quad (1.19b)$$

The term  $NA$  is known as the numerical aperture and applies specifically to the objective of the microscope. High NA objectives typically offer values of up to 1.6, leading to a diffraction limit when considering mid-visible light of 550 nm of  $w_{xy} = 180$  nm and  $w_z = 500$  nm [32]. The differences arise from the origin of the expressions; whereas the lateral (x-y) resolution may be described by ray optics, the axial (z) resolution is derived from ray optics and, as such, depends on the periodic contributions in both the plane of

focus and the planes along the axial axis, hence is a function of the square of the numerical aperture [33].

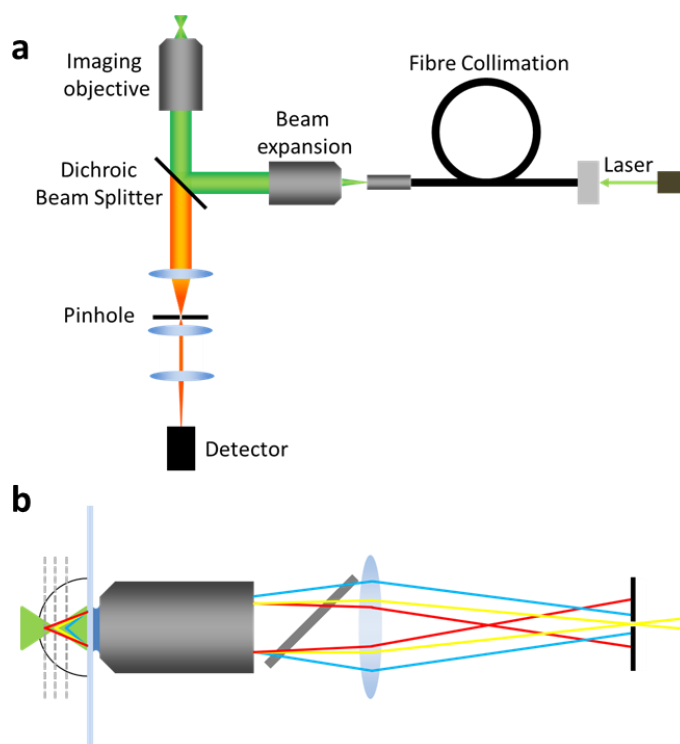


**Figure 1.12:** Illustration of the three stages of single-molecule resolution. In the first case, the two PSFs are clearly resolvable and the two molecules may be examined individually. The second situation is the Rayleigh criterion, which represents the limit at which two molecules may be separated. The final scenario is where the molecules are in too close proximity for their individual PSFs to be untangled. This corresponds to an unresolved system. Figure is adapted from [34].

### Principles of confocal microscopy

The idea of confocal microscopy was proposed by Marvin Minsky in 1955 [35], though the combination of this setup with laser excitation, performed in 1978 by Thomas and Christoph Cremer, provided the real advance that popularised the technique to biophysicists [36]. The mechanism in schematic form is summarised in Figure 1.13a below. A collimated, expanded laser beam is directed into the back aperture of a high numerical aperture microscope objective. This objective focusses the laser light to a diffraction limited spot whose dimensions are given by Equations 1.19a and 1.19b above. The same objective collects the fluorescence emitted by the sample back down the excitation path.

The most critical component of all in the confocal microscopy setup is the pinhole; this restricts the detection volume to one plane within the sample volume. Non-optimal positioning of this component will result in a poorer signal to noise ratio, reducing the effectiveness of the system. The importance of the pinhole position is shown in Figure 1.13b.



**Figure 1.13:** a) Schematic of a basic confocal microscopy system with one detector. In this system, beam expansion is provided by a microscope objective after fibre collimation has been performed to form a Gaussian beam profile. b) Diagram illustrating how the position of the pinhole can influence the imaging plane (grey dashed lines) within the sample (outlined in black). The figure shows the correct positioning of a pinhole to image fluorescence from the centre of the excitation focal spot (yellow lines). The pinhole prevents off-plane emission (blue and red lines) from passing through to the detector.

This technique, with the point detector systems available, such as photomultiplier tubes and avalanche photodiode detectors, allows a time resolution of data acquisition down to the nanosecond time regime.



However, as a point detection technique, in order to build up an image one must scan either the excitation beam or the sample itself. This necessary raster scanning makes the technique less attractive for use in wide field application for studying real-time dynamic processes.

### Principles of TIRFM

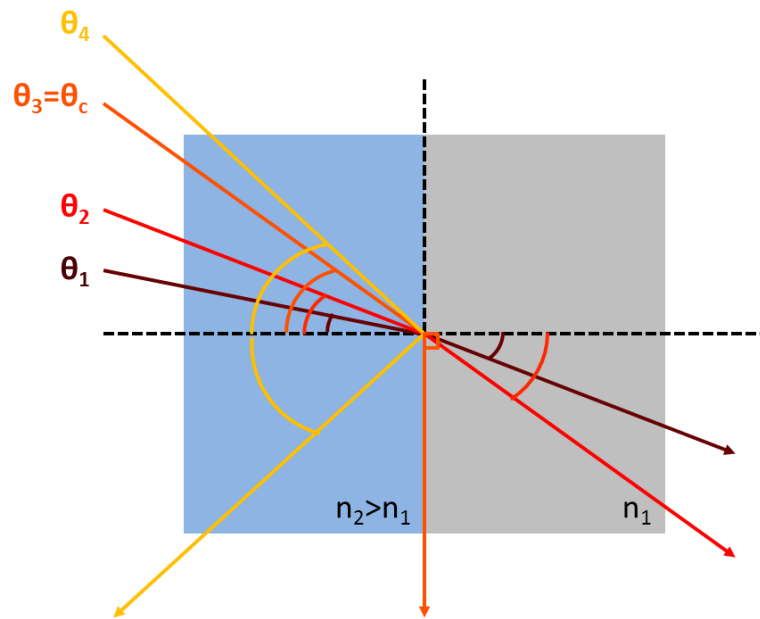
The technique was first realised in 1981, which Daniel Axelrod constructed the first TIRF microscope and investigated the interaction between cells and a solid substrate [37]. His initial design was based around the notion that an illumination source, when reaching a boundary between two different refractive indices at an angle of incidence  $\theta_1$ , will be subject to Snell's law (Equation 1.20):

$$\frac{n_2}{n_1} = \frac{\sin(\theta_1)}{\sin(\theta_2)} \quad (1.20)$$

If one considers a situation where  $n_1 > n_2$ , then at a certain angle of incidence the angle of refraction  $\theta_2$  will be  $90^\circ$ . This angle is known as the critical angle of incidence ( $\theta_c$ ) and is calculated for refractive indices  $n_1$  and  $n_2$  through Equation 1.21 below.

$$\frac{n_2}{n_1} = \sin(\theta_c) \quad (1.21)$$

For the situation where  $\theta_1 > \theta_c$  the incident beam will be reflected at an angle equal to that of the incident beam. This phenomenon is known as total internal reflection. These three scenarios are illustrated for four incident beams in Figure 1.14 below.



**Figure 1.14:** An illustration of Snell's Law acting upon four beams upon a refractive index boundary between two media of  $n_1$  and  $n_2$ .  $\theta_1$  is incident at a sharp angle, and passes through into the lower index medium exiting at an increased angle.  $\theta_2$  enters at a higher angle, and as a result exits having experienced a greater degree of refraction.  $\theta_3$  is incident at the critical angle  $\theta_c$  and as a result the beam leaves the boundary at  $90^\circ$  to the normal.  $\theta_4$  is an angle greater than the critical angle and as such the beam sees total internal reflection, leaving the boundary at the same angle that it is incident with.

By using an incident excitation beam angle in excess of the critical angle, this forces total internal reflection to occur. However, the light is not all reflected, as a fraction propagates through the boundary in the form of an evanescent wave. The intensity profile of the wave ( $I(t)$ ) is given by Equations 1.22a to 1.22c [38].

$$I(t) = I_0 e^{-\frac{z}{d}} \quad (1.22a)$$

$$I_0^{\parallel} = I_1^{\parallel} \frac{4\cos^2\theta_1(2\sin^2\theta_1 - n^2)}{n^4\cos^2\theta_1 + \sin^2\theta_1 - n^4} \quad (1.22b)$$

$$I_0^{\perp} = I_1^{\perp} \frac{4\cos^2\theta_1}{1 - n^2} \quad (1.22c)$$

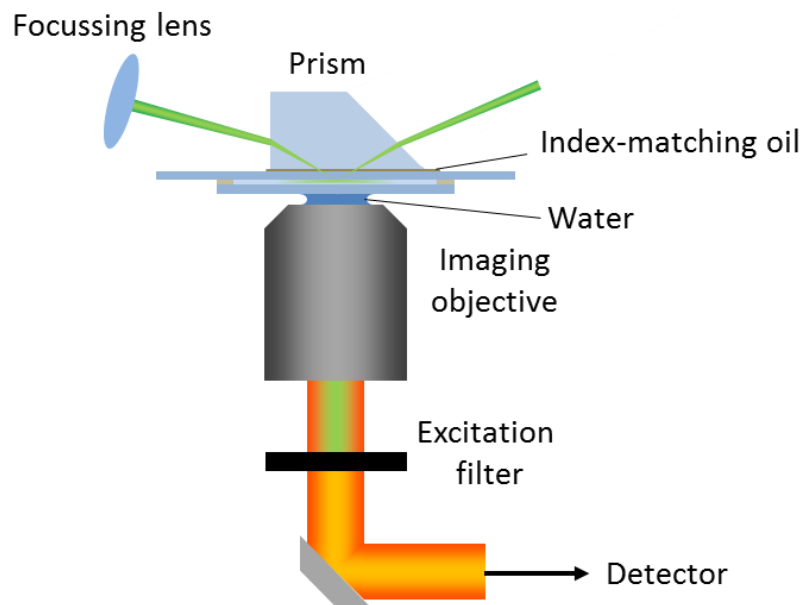
This wave is short-range in its penetration depth  $d$ ; typically one will find the penetration depth of the field to be in the range of 100-200 nm, though this value itself depends upon the excitation wavelength ( $\lambda$ ), refractive indices of the two media at which the excitation beam is reflected ( $n_1$  and  $n_2$ ) and the angle of incidence ( $\theta_1$ ) [14]. Specifically, this rate of decay through the medium is given by:

$$d = \frac{\lambda}{4\pi\sqrt{n_1^2\sin^2(\theta_1) - n_2^2}} \quad (1.23)$$

This sharply decaying intensity is the key to limiting the excitation volume, and combining the selective excitation of the dyes closest to the slide surface whilst the background signal is minimised by the low volume maintains a good signal to noise ratio. The size and intensity of the evanescent wave are tuneable by adjusting the angle of incidence of the laser. The strong signal to noise ratio and the ability to acquire large amounts of data in a single experiment over long timescales using this technique are two of the principal reasons why TIRFM is used most specifically to examine surface-immobilised samples.

### **Prism-type TIRFM (p-TIRFM)**

The original conformation of TIRFM used a quartz prism placed upon the sample holder to generate the evanescent wave [37]. A laser is focussed onto a quartz prism ( $n_{\text{quartz}}=1.55$ ) that is in contact with a quartz slide. Refractive index-matching oil is used to ensure that the laser beam passes through the interface between the prism and the quartz slide unperturbed. With the sample being in a water-based solvent, this results in the first refractive index boundary ( $n_{\text{water}}=1.33$ ). The excitation beam experiences total internal reflection and is directed away from the boundary at the same angle of incidence, generating an evanescent field at the same time that penetrates into the water media. In order to eliminate the excitation light from the emission collected, a long-pass filter is used, blocking the transmission of the lower wavelength laser light. The schematic used in modern TIRFM setups is shown below in Figure 1.15.

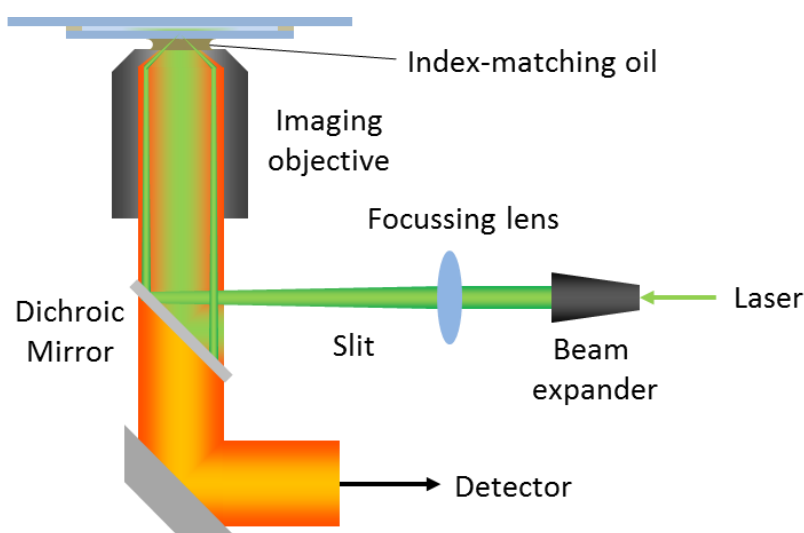


**Figure 1.15:** Schematic of a p-TIRFM setup to extract fluorescence from a slide-based sample. The laser is directed into the prism at an angle in excess of the critical angle. The index matching oil helps the beam to pass through into the slide unperturbed, creating the evanescent wave at the boundary between the quartz and sample. The objective collects the emission from the sample and collimates it, before passing through a filter to eliminate any scattered light from the excitation source, leaving the emission light only which is processed by optics tailored to the purpose of the experiment.

### Objective-type TIRFM (o-TIRFM)

Eight years later, Axelrod proposed a second configuration for TIRF, where the excitation was passed through the objective used to collect the emission [39]. The advances in technology had led to the production of high NA objectives which allowed a laser to be passed through the objective and create an angle of incidence with the sample holder sufficient to generate an evanescent wave at the glass-water boundary. In order to achieve this evanescent wave, the laser must first be expanded in size and focussed

using a lens that will give the tightest focus at the position at which the evanescent field is generated. The lens that is used to achieve this is typically mounted on a fully-adjustable stage in order to fine-tune the angle of incidence and calibrate the system with ease. The laser light is sent into the objective using a 45° dichroic mirror, which also prevents any scattered excitation light from reaching the detector. The excitation system used in modern o-TIRFM experiments is shown in Figure 1.16 below.

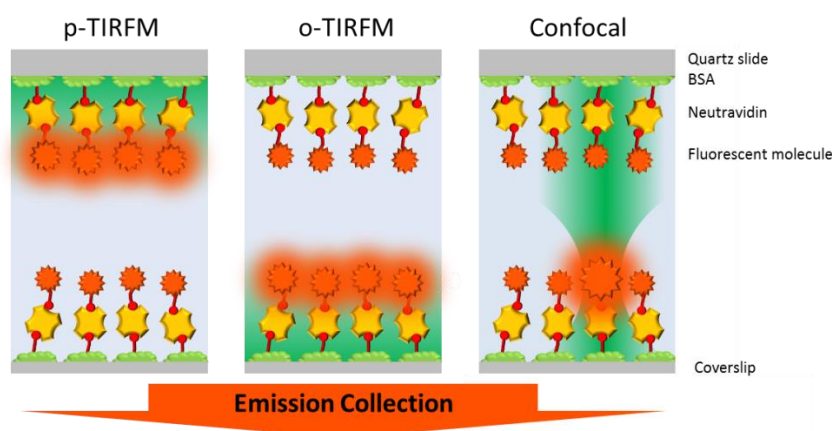


**Figure 1.16:** Optical schematic of the excitation mechanism for an o-TIRFM experiment using a slide chamber-based sample. The excitation beam is passed through the back aperture of the oil-immersion objective at a position which allows it to follow the path shown, generating an evanescent wave at the lower surface of the sample chamber, the opposite face to the p-TIRFM system. The high NA of the objective allows angles of incidence in excess of the critical value to be achieved. The same dichroic which directs the excitation into the objective prevents scattered light from reaching the detector.

It is most commonly the particular experiment rather than the resolution properties which dictates the choice between p-TIRFM and o-TIRFM. The presence of the prism above the quartz slide can be very restrictive in terms

of sample access. As a result, for experiments requiring constant sample access to replenish buffers or change the chemical conditions one typically finds that an o-TIRFM setup is used. However, since the excitation and collection of the emission light are performed within the same objective, the background noise from an o-TIRFM setup will be higher than for a p-TIRFM setup, a fact which can be crucial particularly for low intensity emission.

The three excitation schemes for the single-molecule discussed above are summarised in Figure 1.17 below. One typically uses linear single-polarisation excitation beams as this offers the ability to monitor polarisation changes (see Chapter 7 for more detail).



**Figure 1.17:** The selective excitation mechanisms for each of the single-molecule imaging techniques discussed in this section, where the sample is considered to be surface immobilised in a slide-coverslip chamber (see Chapter 3, section 2). p-TIRFM sees excitation of only the fluorescent molecules closest to the evanescent wave, which is, in this configuration, generated at the boundary between the quartz slide and the sample buffer (water-based). o-TIRFM also sees excitation only in the evanescent field, but in this case the evanescent wave is generated at the boundary between the sample buffer and the coverslip. Confocal microscopy uses the objective to focus a collimated beam to a diffraction-limited spot and selects only a point in the sample. The emission is collected through the coverslip in all of the configurations.

## 1.6. Measuring FRET at the single-molecule level (sm-FRET)

### 1.6.1. Surface-immobilised sm-FRET techniques

In the previous sections, I covered the use of FRET-based experiments to provide information about biomolecular structures and have established the advantages of applying single-molecule techniques to provide details about static and dynamic heterogeneity that would otherwise remain hidden using ensemble averaging methods. The next logical step is to explore how the combination of both methods in the field of single-molecule FRET is revolutionising our understanding of biomolecular structure and function [40].

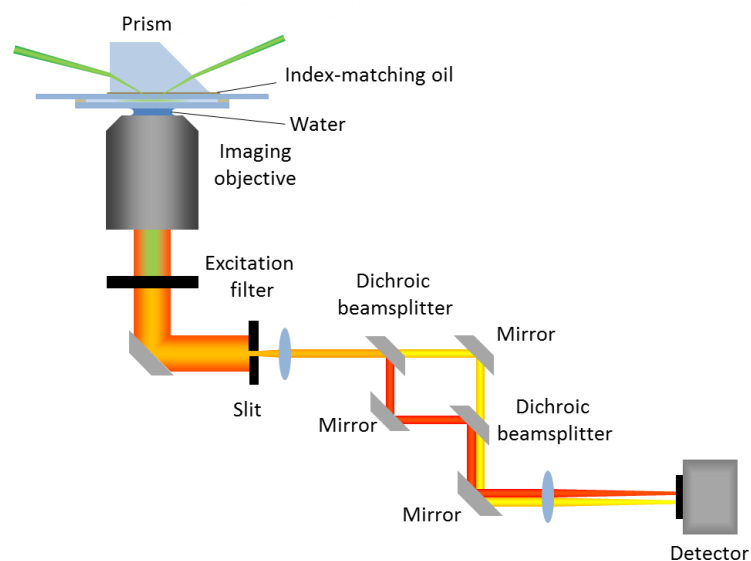
Betzig's first fluorescence experiment performed at room temperature was reported 1994 [29], but it took a further two years for the first experiment combining FRET with single-molecule spectroscopy to be published [41]. Although a hugely significant milestone, this paper reported work performed in non-aqueous conditions; the DNA imaged was dried onto slides and then imaged using NSOM.

The means of realising aqueous sm-FRET was provided by the work of Funatsu et al in 1995, where single-molecule imaging of myosin molecules reacting with adenosine triphosphate (ATP) was performed using epifluorescence and p-TIRFM [42]. In 1999, the same group combined TIRFM with a doubly-labelled coil protein (tropomyosin) and realised aqueous sm-FRET for the first time [43]. Around the same time, work by Schültz et al



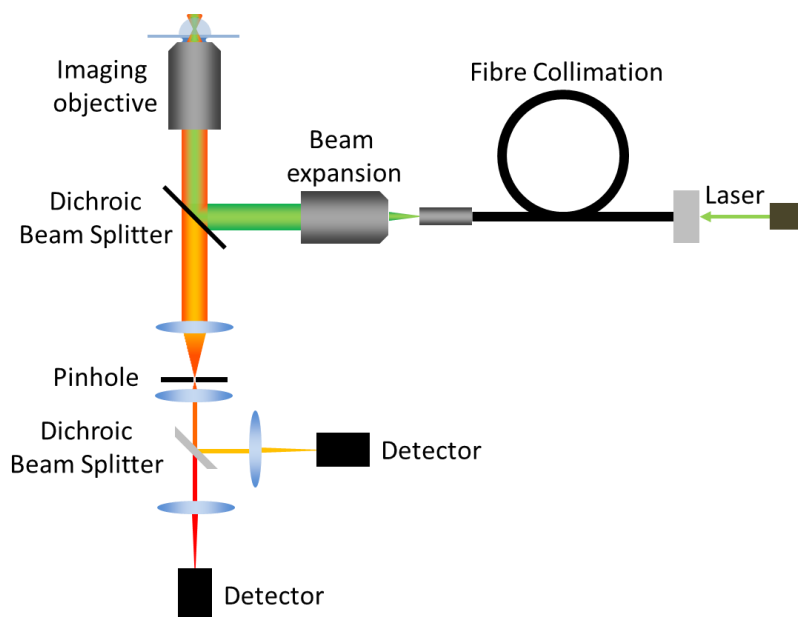
measured sm-FRET for biotinylated fluorophores binding to a streptavidin protein using an alternating excitation scheme to verify the presence of each dye species [44], before Ha et al reported using a similar technique but in confocal microscopy mode later in the same year [45]. A more detailed discussion of the specific means of surface immobilisation is presented in the Materials and Methods section (Chapter 3, section 3).

The detection schemes for sm-FRET experiments depend upon the excitation mechanism. For a TIRFM-based experiment, one typically uses an optical layout as presented in Figure 1.18. The light is collected by the imaging objective and shaped using slits to narrow the imaging region to a rectangular shape. This narrowed emission light is then collimated and passed through an optical relay system consisting of two regular and two dichroic mirrors, which offsets the emission from each of the dye species present on the sample (Figure 1.12). Finally, these two separate components are focussed onto different areas of the detector, typically an EM-CCD camera that allows a millisecond time resolution. This setup allows us to record movies of many thousands of frames in length, allowing long-term molecular dynamic behaviours to be tracked [14].



**Figure 1.18:** Optical setup to resolve the emission for a two-dye system using a p-TIRFM excitation configuration. The detector used is an EM-CCD camera, and the emission from each dye is diverted to a separate half of the detector chip.

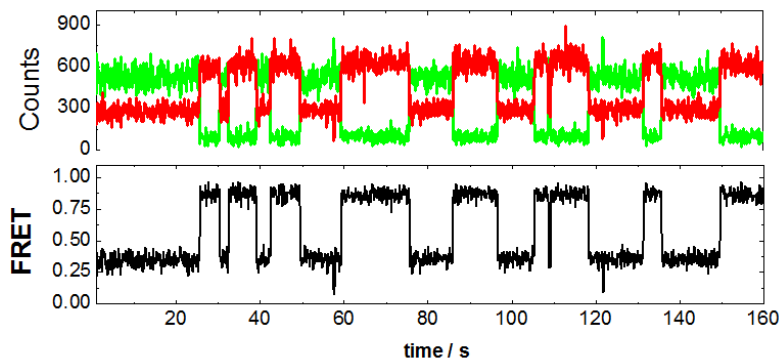
If one wishes to use a confocal excitation scheme because of its intrinsically higher time resolution, then the detection optics must be altered significantly. Due to the point excitation mechanism used in confocal microscopy, detectors such as avalanche photodiodes are the best option because of their small active area, of the order of micrometres in size, and their high temporal resolution. APDs usually have a sub-nanosecond temporal resolution, six orders of magnitude faster than an EM-CCD camera. A standard experimental design for two-colour confocal microscopy is shown in Figure 1.19 below.



**Figure 1.19:** The setup used to realise sm-FRET for a confocal excitation scheme using two detectors to collect the emission from each of the dyes. The fibre and objective collimation creates a perfectly Gaussian collimated beam, which helps to form the best possible excitation spot in conjunction with the pinhole.

The beam expansion and collimation are provided by the combination of an objective and optical fibre, and this light is directed into the back aperture of a high NA oil-immersion objective. Emission from the sample is collected by the same objective and focussed through a pinhole to eliminate off-plane light, as described in Figure 1.13b. The light is then re-collimated, before being split by a dichroic mirror and focussed onto two separate APDs.

In both setups, TIR and confocal, the experimental outcome of the single-molecule FRET experiment consists of photon trajectory of donor and acceptor intensities as depicted in Figure 1.20. From here, the apparent FRET efficiency can be extracted using Equation 1.15.



**Figure 1.20:** Sample traces from a single-molecule FRET experiment (adenine riboswitch), which indicate the intramolecular dynamics of a single molecule. The upper panel shows the changes in intensity of the donor (green) and acceptor (red) dyes. The lower panel shows the associated apparent FRET changes, derived from Equation 1.15.

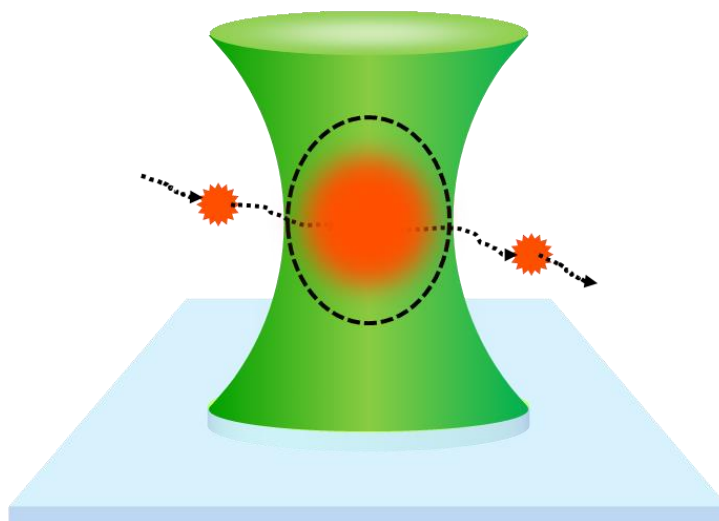
### 1.6.2. Limitations of surface-immobilised single-molecule fluorescence techniques

Tethering the sample to the surface allows long-term imaging of a single molecule, allowing its dynamic behaviour to be analysed continuously for as long as the FRET pair remains active before one of the two fluorophores photobleaches irreversibly.

However, in order to immobilise a sample to a slide one must introduce an additional structural modification, further perturbing the sample from its natural form and increasing the chance of compromising its functionality. Thus, controls are always needed to ensure the surface-immobilised biomolecule remains active and its structure is not disrupted by the presence of the quartz slide. A further complication arises from the slide itself; biomolecules can interact non-specifically with quartz slides. This is particularly relevant for some proteins that have strong tendency to stick to the quartz surface. Although passivation schemes such as the one described

in Chapter 3, section 4 can significantly reduce this risk, the coverage offered is not perfect and passivation methods often need to be additionally fine-tuned for specific biomolecules.

Confocal microscopy may be used as an excitation base for imaging single molecules freely diffusing above the quartz slide with no need for surface immobilisation (Figure 1.21). The planar selection that the pinhole position offers, combined with focussing the laser within the buffer by moving the objective lens position, allows this to be achieved easily using a correctly aligned microscope-based system. By combining this with using very low sample concentrations, one can ensure that only one sample molecule is in the excitation volume at any given time.



**Figure 1.21:** Single-molecule detection for freely diffusing molecules using a laser focussed to a confocal spot (indicated by the dotted circle). The dye molecule is only excited, and therefore only emits, whilst it is within the confocal volume.

Once the emission is collected by the objective and the planar selection has been performed by the pinhole, one may analyse the fluorescence in a number of ways, and the flexibility that such an excitation scheme offers

has led to the recent development of single-molecule multi-parameter fluorescence detection (sm-MFD) where several fluorescence observables may be probed simultaneously [46–48].

A further limitation in surface-immobilised single-molecule techniques comes from the interpretation of zero-FRET peaks. In a surface-immobilised sm-FRET experiment, a zero-peak ( $E_{\text{app}} < 0.12$ ) arises due to the unavoidable leakage of the donor emission into the acceptor channel, which is normally  $\sim 10\%$ . Commonly used dichroic mirrors have a steep transition curve from full reflectance to full transmittance, but even for the best mirrors available, these transitions take place within a few nanometres (10–20 nm), thus leading to the leakage effect. As a result, labelled samples immobilised on the surface that lack the acceptor due to incomplete labelling or acceptor photobleaching will give an apparent FRET efficiency of  $\sim 0.1$ – $0.15$  [11].

### **1.6.3. Fluorescence correlation spectroscopy: biomolecular dynamics from microseconds to seconds**

A molecule within a system in the liquid or gaseous state is not confined in space and time. Instead, thermal fluctuations and concentration gradients within the medium result in a constant molecular rearrangement. This rearrangement is characterised by a random walk, or Brownian motion, which gives a measure of how far the molecule is expected to travel in a period of time ( $t$ ) [9].

Due to the non-linear motion exhibited by a molecule whilst moving in solution, it is more prudent to calculate the properties of motion in terms of an area covered. One does so by looking at the mean squared displacement of the molecule, as shown by Equation 1.24 shown below:

$$\langle r(t)^2 \rangle = 6Dt \quad (1.24)$$

In this equation, ( $D$ ) is defined as the diffusion coefficient, which is an estimate of how far the molecule will travel in one second. This property will clearly depend on the size and mass of the molecule. The early work on diffusion by Fick explained the macroscopic property of diffusion, yet the microscopic behaviour wasn't fully characterised until Einstein's work in 1905 [49]. He derived the expression for the diffusion coefficient in terms of the properties of both the solvent and sample, shown in Equation 1.25, where  $k_B$  is Boltzmann's constant ( $1.38 \times 10^{-23} \text{ JK}^{-1}$ ),  $T$  is the temperature of the system in kelvin,  $\eta$  is the viscosity of the solution and  $R_H$  is the hydrodynamic radius of the sample. Knowing that the mass of a sample is directly proportional to its volume (linked through the density), then one may infer that the diffusion coefficient is inversely proportional to the mass of the sample ( $m$ ) cubed (Equation 1.26).

$$D = \frac{k_B T}{6\pi\eta R_H} \quad (1.25)$$

$$D \propto \frac{1}{\sqrt[3]{m}} \quad (1.26)$$

The combination of Equations 1.25 and 1.26 allows the mass of the sample to be directly linked to its mobility within a buffer of known viscosity, subject to the correct calculation or knowledge of the diffusion coefficient, information not readily available from surface-immobilised schemes [9].

The measurement of this diffusion coefficient is the basis of single-molecule fluorescence correlation spectroscopy (sm-FCS). Here, the molecular motion through a confocal excitation spot (with a cross-sectional area  $w_0^2$ ) is tracked by resolving the average of the temporal durations of fluorescence bursts and fluctuations ( $\tau_D$ ) caused by molecules drifting through the excitation volume. The diffusion coefficient is then calculated through Equation 1.27.

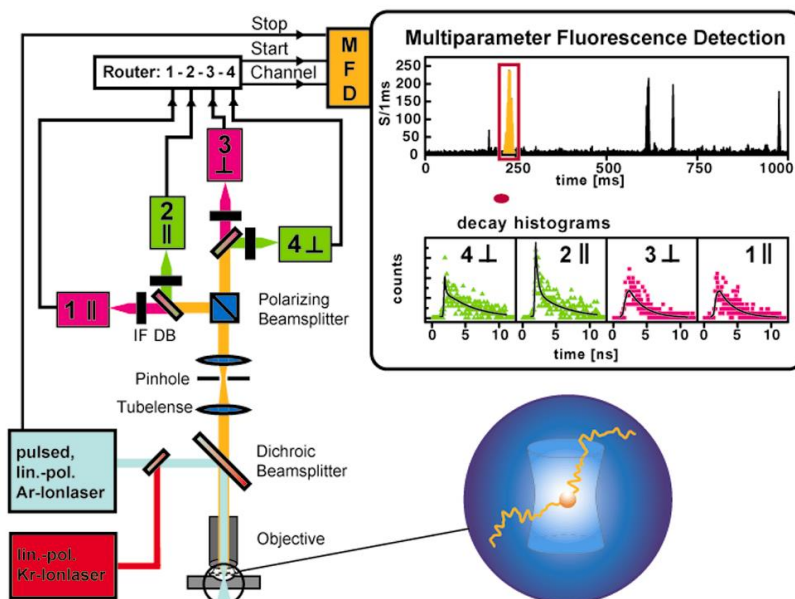
$$D = \frac{w_0^2}{4\tau_D} \quad (1.27)$$

The mathematical process of correlation itself is discussed in detail in Chapter 7, as this technical chapter focusses on experimental realisation of a sm-FCS system.

#### 1.6.4. Single-molecule multiparameter fluorescence detection (sm-MFD)

The technique was developed by Seidel et al and is based around a confocal excitation scheme using a pulsed laser source [46]. The technique is performed using the optical setup shown in Figure 1.22 below.





**Figure 1.22:** Experimental setup for single-molecule multi-parameter fluorescence spectroscopy. Transit of a freely diffusing molecule through the confocal detection volume results in bursts of fluorescence. The emission is separated by polarisation and spectral region before detection and storage. Taken from [46].

There are two key differences between the setups for confocal-based sm-FRET and sm-MFD and both are in place to boost the flexibility of the system. The first difference is in excitation; the use of a pulsed excitation source offers the capability to perform single-molecule lifetime measurements via time-correlated single-photon counting (sm-TCSPC). With the commonplace option to driver pulsed lasers in a continuous mode, then this also offers the option to perform conventional burst analysis.

In order to extract single-molecule lifetime measurements, a timing system must be established. One such schematic of a single-molecule MFD system is shown in Figure 1.22. The second difference is in the excitation path optics. In conventional confocal microscopy to image sm-FRET, whether this is a free diffusion or surface immobilised scheme, two detectors

are employed, as outlined in Figure 1.13. When one upgrades this setup to a sm-MFD system, however, one makes use of four detectors.

The emission light is divided first by polarisation, and then by wavelength, a system which allows imaging resolution of both sm-FRET and anisotropy, allowing the actual FRET value to be resolved, as opposed to the apparent FRET that is extracted from TIRFM-based measurements [46].

The specifics in the timing circuitry and optical arrangements are discussed in more detail in Chapter 7, where the technical aspects are explored in experimentally realising the method.

## Chapter 2: Structure and function of RNA

The biological functions performed within organisms of all classes result from key interactions between a huge variety of molecules and molecular complexes, including proteins, nucleic acids, lipids, sugars and small metabolites. The study of how these processes happen, the intricate rearrangements and reactions that occur both inter- and intra-molecularly, is the basis of the field of biochemistry [50]. The early work on structural determinations of these macromolecular structures heralded a new area within the field of biochemistry that has come to be known as molecular biology. This new field was driven in particular by the advances in the isolation of specific cellular constituents and technological developments, leading to the realisation of techniques such as X-ray crystallography, nuclear magnetic resonance, in addition to the fluorescence techniques discussed in Chapter 1. Many of these crucial functions are performed or mediated by large molecular structures, categorised as biological macromolecules. There are two main classes of these biological macromolecules found within a cell: proteins and nucleic acids. By working either individually or cooperatively, they are responsible for most of the processes that give rise to life.

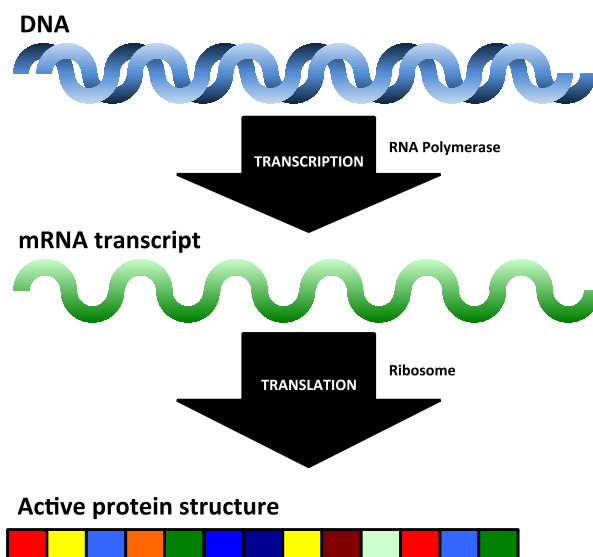
Proteins are the main workhorses within the cell and perform a huge array of functions. Such is their importance to a cell that, in the bacteria

*Escherichia coli*, proteins constitute over 15% of the mass of the bacterium, with approximately 3,000 different species present in every cell [50]. Some of the capacities that proteins are used for are in signalling, structural maintenance and the regulation of metabolite concentration in the cell. Furthermore, proteins form multimeric complexes that catalyse chemical reactions within the cell. These complexes are known as enzymes [51]. Structurally, proteins consist of a long chain of amino acids, covalently linked by peptide bonds. The secondary structure is a result of hydrogen bonding between the various residues, and the establishment of this compacted structure is what makes the proteins biologically active, condensing the molecule and forming binding sites for target molecules, whether these be small metabolites or other large macromolecules such as nucleic acids [50].

The second class of biological macromolecules is, as stated above, the nucleic acids. This classification has two members, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Though the differences between the constituents of the two are very subtle, the overall effect is significant.

The three classes of biological macromolecules are all vitally important in the flow of genetic information within a cell. The description of the transfer of genetic information from DNA to RNA and this into proteins is known as the central dogma of molecular biology (Figure 2.1) and was first stated by Francis Crick in 1958, then formalised by the same author twelve years later [52, 53]. This establishes that the genetic information encoded within

the DNA sequence is read by the RNA polymerase machinery and use to generate an RNA copy of the DNA sequence. This RNA copy is then read by the ribosome to make the proteins that perform most of the functions within the cell.



**Figure 2.1:** The central dogma of molecular biology establishes that DNA genetic code is read by RNA polymerase machinery and the nascent RNA is translated by the ribosome into the proteins that constitute the main functional structures within the cell

Because the majority of the biophysical techniques developed and the studies carried out in this thesis are related to the understanding of the structure and function of nucleic acids, in the next sections I will briefly introduce both DNA and RNA, with special emphasis on understanding the catalytic and regulatory functions of RNA that constitute the main goals of this work.

## 2.1. Deoxyribonucleic acid (DNA)

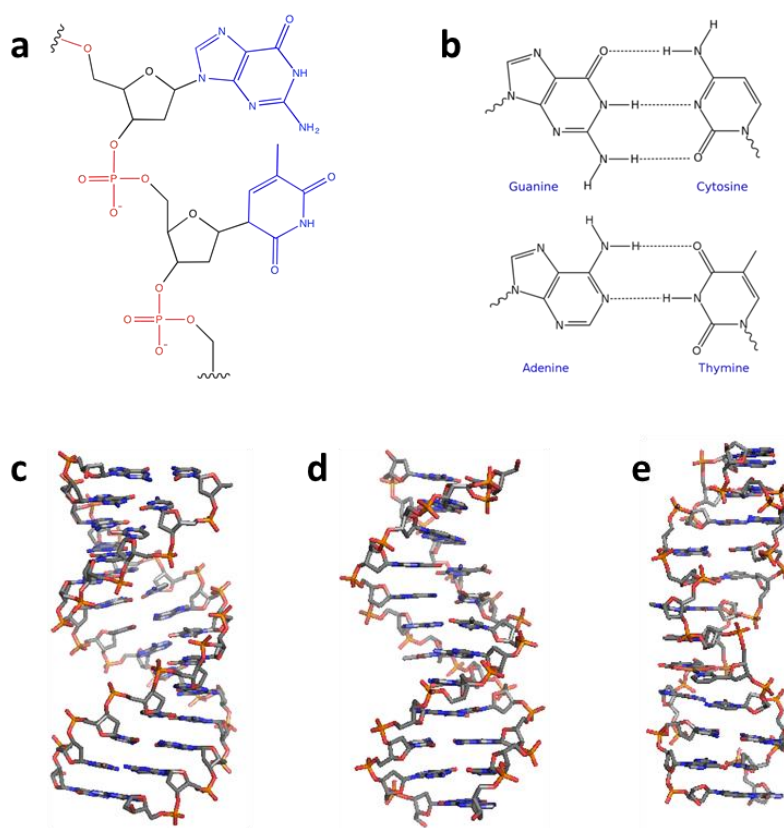
DNA is the more well-known of the two nucleic acids structures found in a cell, responsible for carrying the genetic code in the cell. It is a long polymer chain consisting of a backbone of deoxyribose sugar monomers joined by phosphate groups, as shown in Figure 2.2a. Bound to each sugar is one of four specific molecules, known as nucleobases. In the specific case of DNA, these nucleobases are adenine (A), thymine (T), cytosine (C) and guanine (G).

In contrast to RNA, DNA does not typically exist in a single-stranded form; instead hydrogen bonds form between A-T and C-G nucleobases, shown in Figure 2.1b. These complementary base pairs are known as Watson-Crick pairs and bind two complementary strands together firmly in a helical structure. The most common form is a B-DNA helix, but there are also biologically active A-DNA and Z-DNA helices that will form under certain conditions, as illustrated in Figure 2.1c i)-iii). B-DNA is the commonly found DNA structure within cells of all organisms. The structural parameters that characterise the different DNA and RNA helices are summarised in Table 2.1. The hydrogen bonds between A-T and G-C or A-U in the case of RNA are a fundamental feature of the double helix structure that contributes to its thermodynamic stability but they are not the only force involved in stabilising this structure. An additional important contribution is due to base stacking. Bases stack above each other roughly

perpendicular to the helical axis, favouring interactions between the electron clouds of the bases, a stabilising effect known as  $\pi$ - $\pi$  stacking.

The identification of the helical structure of DNA in 1953 won the Nobel Prize in medicine or physiology for Watson, Crick and Wilkins nine years later, having used the work of Franklin and Gosling to formulate their model [53–55]. These early X-ray studies were carried out using concentrated solutions of DNA that had been drawn into thin fibres. The B form was observed at high humidity and corresponds to the average structure commonly found in cells and contains 10 base pairs per turn and a wide major groove and a narrow minor groove (Table 2.1). The separation between adjacent base pairs is 0.34 nm, and thus a helical turn corresponds to a length of 3.4 nm (Figure 2.2c). The A-form is observed under conditions of low humidity and it is more compact with 11 base pairs per turn (Figure 2.2d). The bases exhibit a large tilt with respect to the helical axis and contain a central hole absent in B-DNA. A-DNA is adopted by RNA-RNA and RNA-DNA hybrid helices and by DNA-DNA helices in complexes with certain proteins. In contrast to B-DNA and A-DNA, which are both right-handed helices, Z-DNA is a left-handed helix and has a zig-zag appearance (Figure 2.2e). This zig-zag appearance arises from the rotation of the glycosidic bond that connects the base to the 1' position of the 2'deoxyribose. This bond can be in one of two configurations: *syn* and *anti*. In right-handed B-DNA, this bond is always in the *anti* conformation, whereas in Z-DNA, it alternates between *syn* and *anti* conformation depending on the nature of

the nucleobases. The fundamental repeating unit in Z-DNA is usually a purine-pyrimidine dinucleotide, where pyrimidine residues are in the anti conformation and purine bases in the syn conformation. This alternative configuration between syn and anti configurations gives Z-DNA its left-handed properties.



**Figure 2.2:** a) Structure of the backbone of deoxyribonucleic acid, showing the alternating covalently-linked phosphate (red) and deoxyribose sugars (grey), as well as the positions at which the nucleobase molecules are found (blue). b) Chemical structures of the four nucleobases of DNA, showing the hydrogen bond base pairing network for adenine – thymine and guanine – cytosine. c)-e) The three biologically active helical structures for DNA, c) A-DNA (PDB ID: 4OKL [56]), d) B-DNA (PDB ID: 4PWM [57]), e) Z-DNA (PDB ID: 1JES [58]).



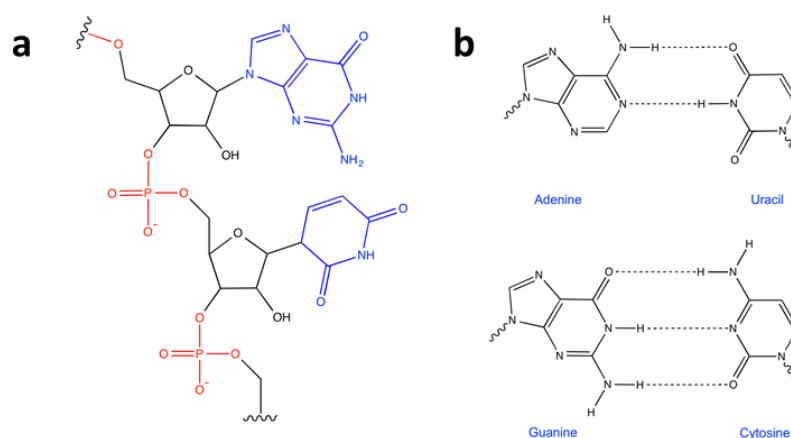
**Table 2.1:** A comparison of the structural properties of A, B and Z DNA derived from X-ray analysis. Adapted from [59].

Property	Helix Type		
	A	B	Z
Overall proportions	Short, broad	Long, thin	Elongated, slim
Rise per base pair	2.3 Å	3.32 Å	3.8 Å
Helix-packing diameter	25.5 Å	23.7 Å	18.4 Å
Helix rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per helix repeat	1	1	2
Base pairs per turn of helix	~11	~10	~12
Rotation per base pair	33.6°	35.9°	~60° per 2 bp
Pitch per turn of helix	24.6 Å	33.2 Å	45.6 Å
Tilt of base normal to helix axis	+19°	-1.2°	-9°
Base pair mean propeller twist	+18°	+16°	~0°
Helix axis location	Major groove	Through bps	Minor groove
Major-groove proportions	Extremely narrow but very deep	Wide, int. depth	Flattened out on helix surface
Minor-groove proportions	Very broad, but shallow	Narrow, intermediate depth	Extremely narrow, very deep
Glycosyl-bond conformation	anti	anti	anti at C, syn at G

## 2.2. Ribonucleic acid (RNA)

Just like DNA, ribonucleic acid (RNA) is built up from four nucleobase molecules covalently bonded to a backbone of alternating sugar and phosphate groups. The differences alluded to above refer to the sugar, which in the case of RNA is ribose, which contains a hydroxyl group at the 2' position, and to one nucleobase which is an unmethylated thymine molecule as shown in Figure 2.3a and 2.3b. Although these are small changes in the building blocks for RNA as opposed to those for DNA, the global effect is extremely important to determine the functional difference between DNA and RNA, and for instance, the presence of the 2' hydroxyl confers RNA its catalytic function. The other fundamental difference between both nucleic

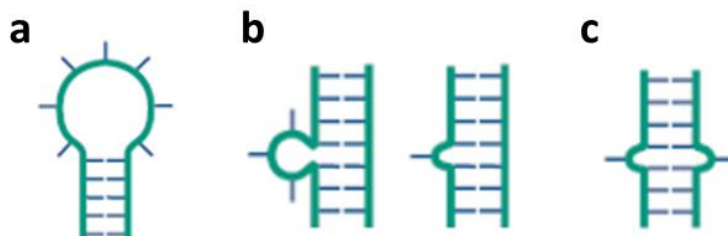
acids is that RNA exists primarily in a single-stranded form. Except for in some viruses, RNA is not the genetic material and therefore it does not need to be permanently in a double helix to protect the genetic information from damage and chemical attack. The means of protecting the genome in the single-stranded viruses are covered in detail in reference [60].



**Figure 2.3:** a) Chemical structure of the RNA backbone. The red subunits are the phosphate groups, the black subunits are the ribose sugars and the nucleobases are shown in blue. b) Chemical structures of the nucleobases of RNA. Guanine and cytosine are the same as in DNA, and base paired in the same way.

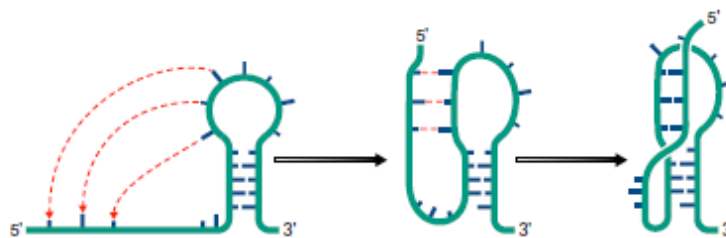
The formation of helical motifs through Watson-Crick base pairing of complementary nucleotide sequences is possible in RNA as it is in DNA. However, it is unusual for RNA to form long double helices in Nature. The most common situation found in cells is to have RNA structures showing a mixed single-stranded and double-stranded character. This is because stretches of single-stranded RNA sequences can fold back on themselves forming local segments of double helix structures similar to A-DNA. If the complementary sequences are next to each other, the formation of the

double RNA helix may lead to the formation of stable hairpins, bulges or RNA loop secondary structures as shown in Figure 2.4.



**Figure 2.4:** An RNA molecule containing alternating complementary sequences and single-stranded regions can form a) hairpin structures, b), bulges, and c) loop structures, provided that the complementary sequences are close to each other. Adapted from [59].

The flexibility of the single-stranded RNA backbone confers RNA the possibility of adopting a range of tertiary structures including pseudoknots (Figure 2.5), loop-loop interactions and many other tertiary RNA motifs. The formation of these structures usually involves the formation of non-canonical (non-Watson-Crick) base pairing interactions such as G-U pairs, base triples (U-A-U) and base-backbone interactions.



**Figure 2.5:** RNA folding steps involved in the formation of a tertiary pseudoknot interaction from a pre-formed secondary hairpin motif. The tertiary contacts, base pairing between the tail and loop, are shown as the red dotted lines. Adapted from [59].

The conformational changes taking place in the RNA that allow the single-stranded sequence to adopt this three-dimensional structure is commonly referred as RNA folding mechanism. The adoption of these intricate 3D-RNA structures is crucial for RNA to perform many of its functional roles and I will devote the next section to provide a summary of the factors modulating the RNA folding process.

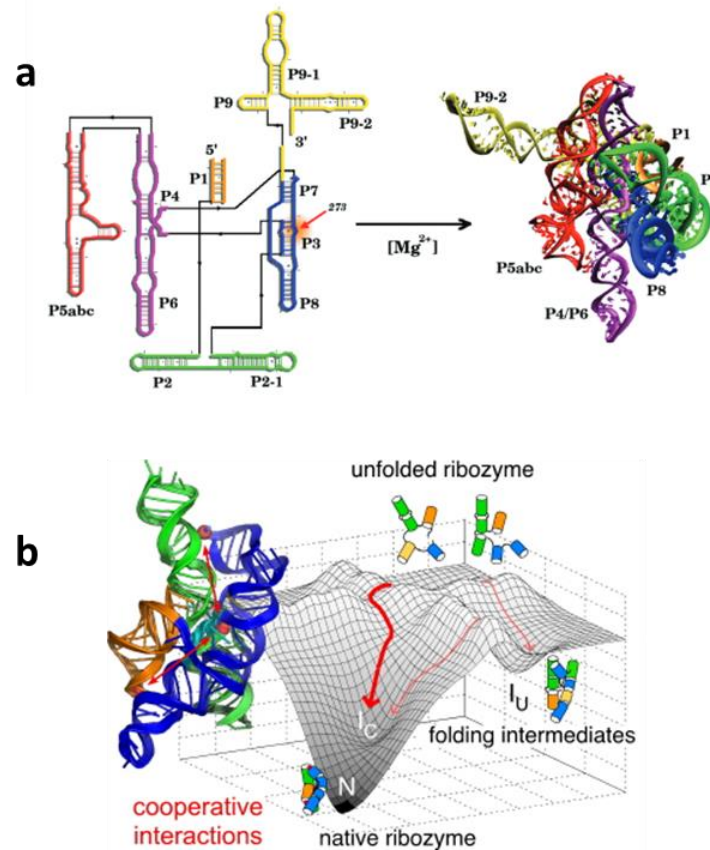
### **2.3. RNA folding**

RNA folding studies aim to identify the sequence of events and conformational changes from which a single-stranded RNA sequence evolves into a three-dimensional structure with a defined biological function [61]. The RNA folding field shares some common features with the protein folding field, as both are influenced by similar, although not identical driving forces. These include hydrophobic, electrostatic and solvent-dependent interactions. The formation of an RNA native structure in vitro is commonly divided into two levels of organisation. The first level involves the formation of the secondary structures and usually takes place on the sub-millisecond timescale. These early events are then followed by late and slower folding events ( $\sim$  seconds) where secondary structural motifs are brought together to form the compact, active structure. In both steps, the addition of mono- ( $\text{Na}^+$ ,  $\text{K}^+$ ) and divalent metal ions ( $\text{Mg}^{2+}$ ) is critical to neutralise the negative charge of the RNA backbone and allow this to

collapse and form secondary structures at the local scale that can progress further to form the native 3D-conformation [59, 61] (Figure 2.6a). A number of studies in recent years in the area of RNA folding have led us to conclude that RNA folding takes place in a rugged folding landscape [62] (Figure 2.6b). This rugged nature arises from the presence of several competing interactions including electrostatic repulsion between the charges of the phosphate backbone, base stacking and hydrogen bonding [62]. Although a complete theoretical framework for the influence of metal ions on RNA folding and structure is still lacking, the qualitative aspects leading to charge neutralisation and RNA folding induced by metal ions can be summarised if we consider two main metal-ion contributions observed from crystal structures of RNA molecules. Firstly, ions localised within the vicinity of the RNA phosphate backbone acting as diffusive charge screening agents; and secondly, discrete ions that interact specifically with certain sites in the RNA native structure [61]. A rugged landscape also explains several features commonly found in RNA folding and related to each other: the presence of misfolded intermediates [63], the slow rate of tertiary folding observed for certain RNA species [64, 65] and the folding heterogeneity observed from single-molecule FRET studies from molecule to molecule [66, 67].

In summary, a number of factors including the ability to exist as single- and double-stranded regions, the polyelectrolyte character of the RNA backbone and the role played by metal ions contribute to the complexity of

RNA folding and the astonishing variety of RNA structures able to execute a continuously growing range of cellular functions.



**Figure 2.6:** a) Schematic of the secondary structure of the most extensively studied group I intron from *Tetrahymena*. This secondary structure has a number of paired helices indicated by P1 through P9. The addition of excess  $Mg^{2+}$  stimulates a transition to a compact tertiary structure [68]. The structure on the right is a model proposed by Westhof and Michel [69]. b) The energy landscape for an RNA ribozyme folding, showing the evolution from unfolded states through structural intermediates  $I_C$  (compact) and  $I_U$  (non-standard) to the stable native structure  $N$  [64].

## 2.4. The functional roles of RNA

It was initially assumed that RNA simply carries transiently the genetic information from the DNA from which it was read off to the ribosome; this is

the origin of the term messenger RNA. In the process of transcription, the RNA polymerase enzyme locates a specific sequence along the DNA and binds to the duplex, either itself in eukaryotic cells or through a group of proteins called transcription factors. Once bound, the enzyme forms a transcription bubble by breaking the double bonds between a length of the DNA duplex and then moves along one of the strands, constructing the complementary messenger RNA (mRNA) sequence as it reads the DNA template. When the RNA polymerase reaches the termination point of the gene, the mRNA and DNA template are released from the complex, allowing the base pairs to reform in the duplex. Further processing occurs on the mRNA before reaching the ribosome, such as the removal of some non-coding regions that are interspersed within the gene sequence of eukaryotes.

However, this picture of RNA as a mere transient carrier of genetic information has changed since the discovery that RNA is also responsible for many other tasks in both normal and disease-related cellular states [50, 51, 65]. These RNA structures that do not take the form of mRNA are generally referred as non-coding RNAs because they do not encode proteins [66, 67]. The growing number of RNA structures actively involved in many cellular processes, the diversity of biological functions and their abundance has led to the hypothesis of an RNA world, preceding DNA and proteins, where all functions were carried out by RNA molecules [70]. Many types of RNA have been found to act catalysing chemical reactions [71], a feature that was previously thought to be exclusive of proteins. In eukaryotes,

transfer RNA (tRNA) and ribosomal RNA (rRNA) have been known to play crucial roles in the translation process of mRNA into proteins since almost half a century ago [72].

Small nuclear RNAs (snRNAs), microRNAs (miRNAs) and small interfering RNAs (siRNAs) are further important categories of non-coding functional RNAs. snRNAs play critical roles in the regulation of RNA splicing and are tightly associated with proteins in complexes called small nuclear ribonucleoproteins. U1 to U6 particles are involved in splicing pre-mRNA into mature mRNA. miRNAs are small elements comprising 22 to 26 nucleotides (nt) that function in gene regulation by binding to target mRNAs and repressing translation. Since their discovery in *C. elegans* [73], miRNAs have been found in flies, mice and humans and it has been shown that they play significant roles in cancer and other diseases [74].

Small interfering RNAs are sequences of 21 to 25 nt length that participate in gene regulation. One strand of a double-stranded siRNA is incorporated into the RNA-induced silencing complex (RISC) to inhibit the expression of mRNA with a sequence complementary to the siRNA. siRNAs have evolved as a defence mechanism against double-stranded RNA viruses [74].

In prokaryotes, gene regulation is also modulated by mRNA regulatory elements: so-called riboswitches. Riboswitches are RNA sensors that detect a specific metabolite and regulate the expression of proteins associated with the transport or biosynthesis of the metabolite they bind. Importantly, the



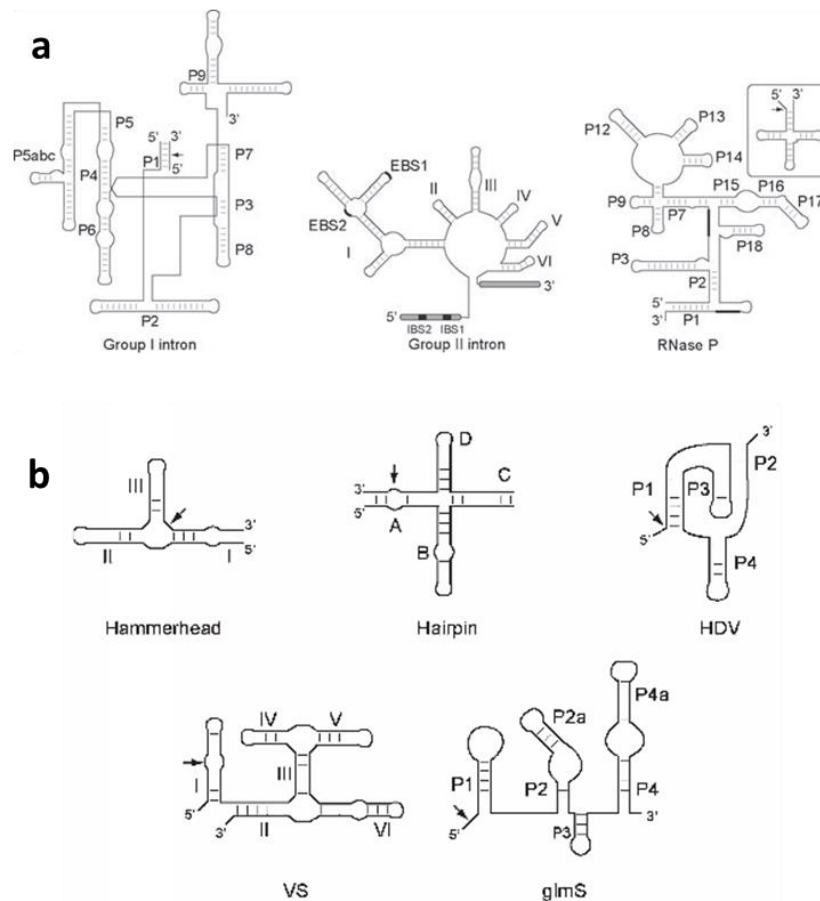
regulatory function of riboswitches is performed without the assistance of any protein. Since their discovery, riboswitches have been found that respond to a variety of metabolites including nucleotides, amino acids and sugars [71]. Because the understanding of riboswitch structural dynamics is one of the main aims of this work, the structure of these mRNA elements will be explored in more detail in the next sections.

Lastly, RNAs with enzymatic activity, commonly referred as ribozymes, constitute a specific class of RNAs discovered in the early 1980s by Thomas R. Cech, who was studying RNA splicing in the ciliated protozoan *Tetrahymena thermophile*, and by Sidney Altman and Norman Pace, who were studying the bacterial RNase P complex [75]. In 1989, Cech and Altman shared the Nobel Prize in Chemistry for their "discovery of catalytic properties of RNA". As for riboswitches, the structure and functional dynamics of ribozymes are explored in more detail in the next sections.

## **2.5. RNA as enzymes catalysing chemical reactions**

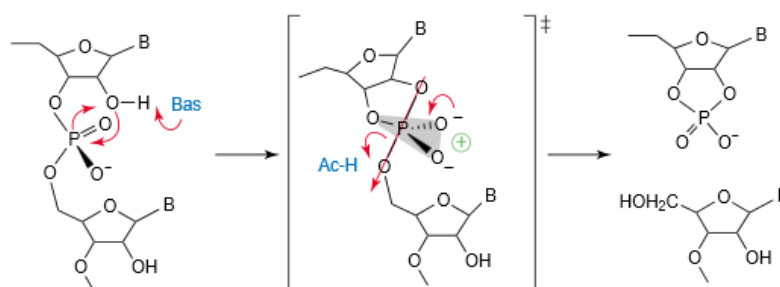
As mentioned above, certain RNA molecules have the ability to catalyse biochemical reactions within the cell. Before the discovery of ribozymes, proteins were thought to be the only biochemical entity within the cell capable of catalysing chemical reactions. The most common activity of ribozymes is the ligation (building) and cleavage (splitting) of both DNA and RNA [76–79] and in peptide bond formation. Within the ribosome, the RNA

component acts as a ribozyme to link amino acids during protein synthesis [80]. Natural ribozymes are usually classified according to their relative size. Large ribozymes (>300 nt) include the self-splicing group I and II introns and the RNase P (Figure 2.7a). Small nucleolytic ribozymes (Figure 2.7b) are a mechanistically distinct class that includes the hammerhead, hairpin, hepatitis delta virus (HDV), Varkud satellite (VS), glmS and the recently discovered twister ribozyme [81]. Several excellent reviews on ribozymes have been published [82, 83].



**Figure 2.7:** a) Secondary structure of naturally occurring large ribozymes. Cleavage sites are indicated by arrows. b) Secondary structure of small nucleolytic ribozymes. Cleavage sites are indicated by arrows. Taken from [84].

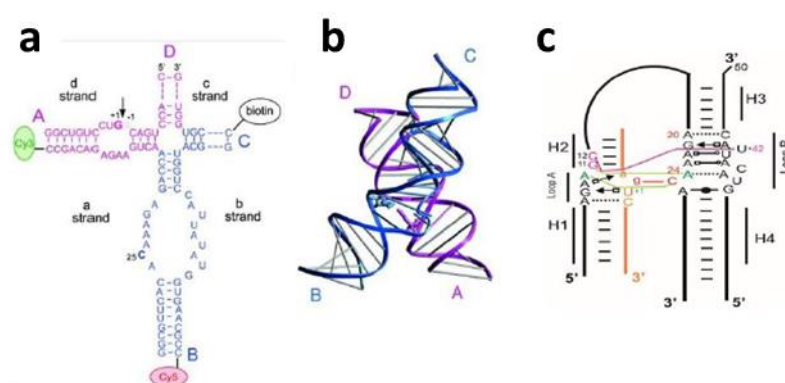
All of these small ribozymes perform a self-cleaving trans-esterification reaction initiated, as described in Figure 2.8, by nucleophilic attack of the 2'-hydroxyl group of the adjacent phosphorous, generating a hydroxyl and a cyclic phosphate terminus that can be used to catalyse the inverse (ligation) reaction. The reaction mechanism is an  $S_N2$  reaction and is accompanied by an inversion of configuration at the sugar phosphate backbone (Figure 2.8).



**Figure 2.8:** The chemical mechanism of the cleavage reaction of nucleolytic ribozymes. Cleavage follows an  $S_N2$  mechanism and it is initiated by attack of the 2'-hydroxyl group on the adjacent phosphorous to generate a cyclic phosphate. The transition state has a bi-pyramidal geometry. Taken from [85].

Crystal structures have obtained for most of the small ribozymes and models in solution for their folding processes have been proposed using biophysical techniques such as FRET, EPR, NMR and others. Of the small ribozymes, the hairpin ribozyme is the one most extensively studied using both ensemble and single-molecule FRET techniques [86–90]. In contrast, there have been very few studies at single-molecule level regarding the folding pathway of the VS [91] and the hammerhead ribozymes [92]. The latter is the subject of further analysis in this thesis.

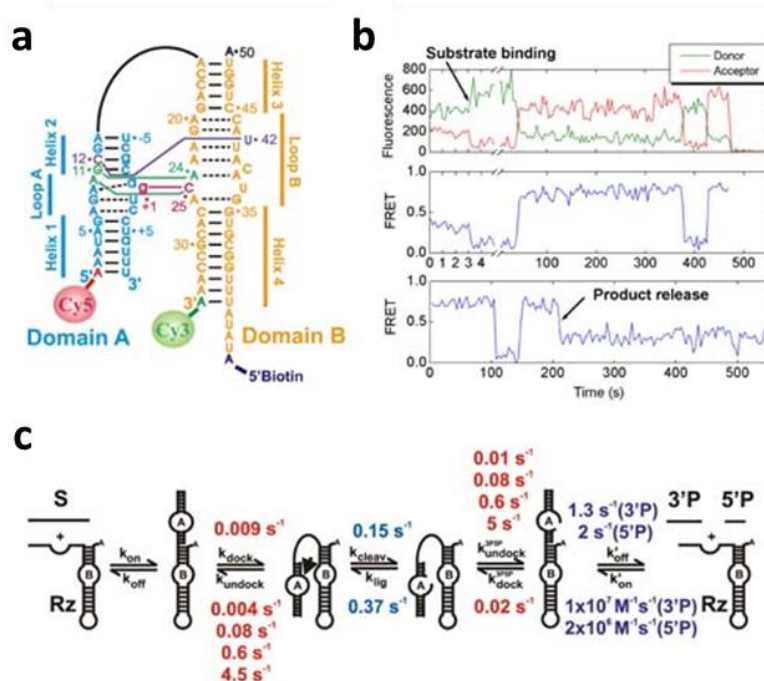
The folding pathway and the catalytic function of the hairpin ribozyme have been extensively studied at single-molecule level and it has become a model of how single-molecule FRET can help to elucidate the structure and dynamics of functional RNAs. Here, I briefly described the main findings from studies and I refer the reader to excellent reviews in the area for more details [84, 93–96].



**Figure 2.9:** The hairpin ribozyme. a) Secondary structure of the hairpin ribozyme containing arms A, B, C and D. The position of the FRET labels using for ensemble and single-molecule characterisation is shown. b) Crystal structure of the hairpin ribozyme. c) Secondary structure of the minimal form of the hairpin ribozyme showing the network of interactions between domains A and B. Taken from [90].

The hairpin ribozyme (Figure 2.9) is derived from the minus strand of the satellite RNA of the tobacco ringspot virus and is organised around a 4-way helical junction and contains two internal loops in stems A and B (Figure 2.9a). The crystal structure of the folded ribozyme showed that A stacks on D and B on C, forming an antiparallel structure. The ribozyme substrate interacts with helices 1 and 2 of domain A (Figure 2.9a). The nucleotides in the internal loops are highly conserved and any modification negatively influences folding and catalysis. The stability of the folding state is achieved

through several interactions between loops in domains A and B. Interestingly, most studies on the hairpin ribozyme have been performed on the minimal form, which lacks the helical junction and has stems A and B connected by a stretch of nucleotides. Although the minimal form is catalytically active and folds into the native state in the presence of mono- and divalent metal ions, it requires much higher, non-physiological concentrations of  $\text{Mg}^{2+}$  (>50 mM). Both the minimal and full forms of the hairpin ribozyme have been extensively characterised using FRET [86, 97–102]. From these studies it was shown that the minimal hairpin ribozyme exhibits a significant kinetic heterogeneity with a single loop-loop docking rate but four undocking rates, suggesting that the ribozyme undocks from 4 different loop-loop docked states (Figure 2.10) [103]. As a result of this folding heterogeneity, the ribozyme also exhibited heterogeneous cleavage kinetics. Similar folding heterogeneity was also observed in the 4-way junction, but the  $\text{Mg}^{2+}$  concentration required for folding was reduced to physiological concentrations. The folding and catalytic heterogeneity observed by single-molecule experiments is a feature that could not be retrieved from ensemble measurements. To further demonstrate that such heterogeneity is not an artefact of single molecule interactions with the surface, the Ha group performed sm-FRET experiments encapsulating the hairpin ribozyme within a vesicle [104].



**Figure 2.10:** Single-molecule analysis of the hairpin ribozyme. a) Secondary structure of the minimal form of the ribozyme (2-way junction) showing the position of the Cy3 donor and the Cy5 acceptor dyes. b) A fluorescence time trace of a single ribozyme that shows binding of a substrate, docking and undocking. Green and red lines are the fluorescence signals from the donor and acceptor, respectively. The blue line is the corresponding FRET value. The substrate-free ribozyme gives FRET ~ 0.4. Upon binding of a substrate, the FRET time-trace exhibits stochastic transitions between two levels (0.2 and 0.8) that correspond to the undocked and docked states, respectively. (Lower panel) FRET time-trace of a single complex of the ribozyme showing docking, undocking and cleavage. The release of products after cleavage recovers the ribozyme to a substrate-free state with FRET ~ 0.4.

In these conditions, the ribozyme remains freely diffusing in a solution environment within the vesicle. Even in such conditions, the ribozyme exhibits a similar degree of dynamic heterogeneity suggesting that this is an intrinsic feature of the RNA. Lastly, the hairpin ribozyme showed a strong memory effect, where each molecule rarely switches between the different docked state. These studies on the hairpin ribozyme illustrate the

complexity of RNA folding and function and emphasise the power of sm-FRET to reveal important information that otherwise will remain hidden using conventional approaches.

## 2.6. Riboswitches

### 2.6.1. Riboswitch architecture

Riboswitches regulate the expression of a gene in an mRNA transcript by sensing and binding a specific metabolite in a binding pocket formed in the native state of the mRNA. Ligand binding triggers a structural rearrangement of the mRNA sequence that regulates the expression of the downstream genes. Since their discovery in 2002 [105], there have been 34 riboswitch structures identified, which target ligands such as amino acids, nucleobases, ions and sugars as well as much larger coenzyme structures [106]. These structures are found predominantly in prokaryotic bacterial cells, though one class has been discovered also to be present in plant and fungal eukaryotic cells [107].

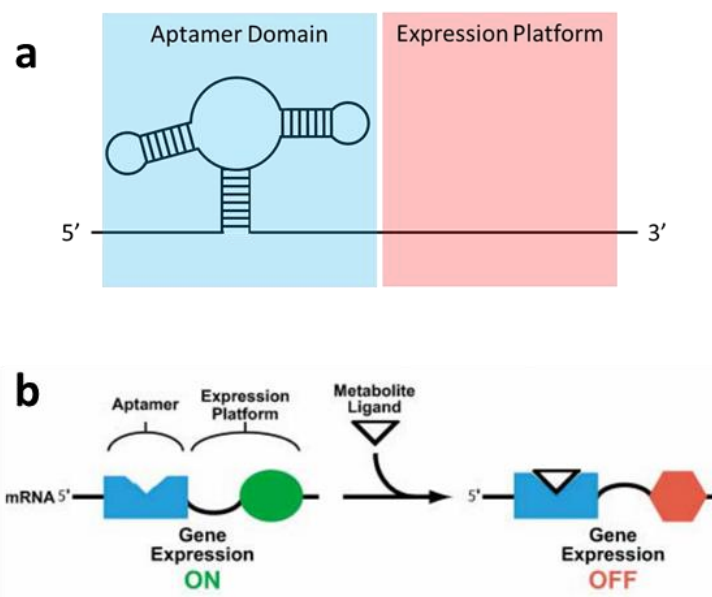
The structure of the riboswitch itself may be divided into two parts:

1. The aptamer domain is the region that senses the metabolite, and it is a network of secondary structural motifs held together by tertiary interactions, as shown in Figure 2.5 [108]. Though a riboswitch typically has one aptamer, the glycine riboswitch is unique in having two separate ligand-binding aptamer domains interacting with each

other and controlling a single expression platform [109]. The aptamer domain is highly conserved and has been evolutionarily optimised to respond to a specific ligand [15]. Aptamer domains usually show an intricate network of contacts with most of the atomic groups of the ligand to ensure a very high ligand-binding affinity and specificity. Usually, nucleotide sequence and structure of the aptamer domain for a given metabolite is conserved across organisms and this conserved sequence is normally used to define the identity of each riboswitch.

2. The expression platform is typically located immediately downstream of the aptamer domain (Figure 2.5) but in some cases the two domains overlap. The role of the expression platform is to signal the ligand-binding event and transduce this into changes in its structure to modulate gene expression [108]. The expression platform is mostly non-conserved and exhibits a high level of plasticity since it can adopt a broad range of secondary structures to regulate gene expression.





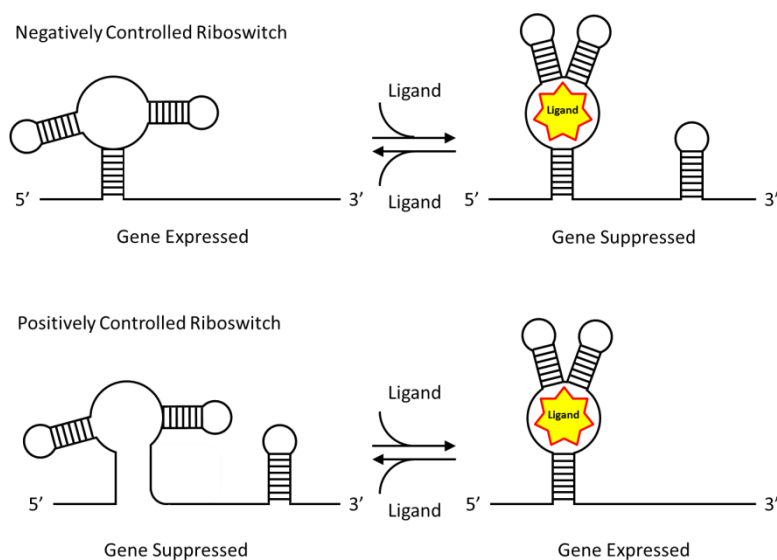
**Figure 2.5:** a) Simplified depiction of the two main parts of a riboswitch: the aptamer domain, responsible for sensing the appropriate ligand, is shaded blue and the expression platform, where the code for the gene is read from by the RNA polymerase, is shaded pink. b) General model of riboswitch function. Binding of the target ligand to the aptamer domain stabilises induced a change in the structure of the expression platform that results in a change in gene expression of the downstream genes.

The fact that ligand binding is exclusively achieved by the aptamer domain (Figure 2.5) has facilitated the study of the aptamer-ligand interaction by itself using chemical probing, NMR, X-ray and fluorescence techniques. The lengths of the aptamer domains range from ~50 to 200 nucleotides [110], whereas the expression platforms are far more diverse in structure, size and sequence. Depending on the particular riboswitch, affinities have been reported ranging from 5 nM for the SAM riboswitch [111, 112] to 30  $\mu$ M for the glycine riboswitch [109]. Thus, riboswitches have evolved to bind tightly to their metabolites with affinities fine-tuned to carry out their biological role. Importantly, each aptamer domain has been shown

to discriminate strongly against chemicals structurally similar to the target metabolite [105, 111, 113–116].

### 2.6.2. Mechanisms of gene regulation by riboswitches

The function of the riboswitch is, as stated above, to regulate gene expression. Despite their simplicity, riboswitches are able to modulate gene expression by a large array of mechanisms of which transcription attenuation and translation inhibition are the most commonly observed [108]. The guanine and adenine riboswitches constitute a good example of negatively and positively regulated gene expression using a transcription termination mechanism, respectively (Figure 2.6).



**Figure 2.6:** Representative examples of down- and up-regulation of gene expression by transcription termination in a) the guanine, and b) the adenine riboswitches.

In the case of the guanine riboswitch, a terminator hairpin is formed when the guanine ligand is bound resulting in gene repression (Figure 2.6a). In the adenine riboswitch, however, the ligand-bound state is associated with the disruption of the terminator structure leading to the activation of gene expression (Figure 2.6b).

Gene regulation by translation inhibition takes place in those riboswitches where the expression platform containing the ribosome binding site (RBS) is in a structure where the RBS is inaccessible, thus preventing the mRNA from being translated. As an example, translation inhibition has been characterised in detail in the miniROSE (repressor of heat shock gene expression) element. It has been shown the miniROSE thermo-sensing riboswitch controls translation by regulating the length of a simple helical domain. As the temperature increases, the riboswitch changes from a conformation where the Shine-Dalgarno (SD) sequence is base-paired within the helical structure to a conformation in which the SD sequence is exposed, thus leading to gene expression, as shown in Figure 2.7.



**Figure 2.7:** Representative example of gene expression control by translation inhibition in the miniROSE element. The miniROSE riboswitch positively regulates gene expression by changing at high temperatures to a conformation that exposes the Shine-Dalgarno sequence. Taken from [108].

In addition to transcription termination and translation inhibition, some riboswitches control gene expression by a self-cleaving mechanism such as in the case of the *glmS* riboswitch, which can be considered to act as a ribozyme [117–119].

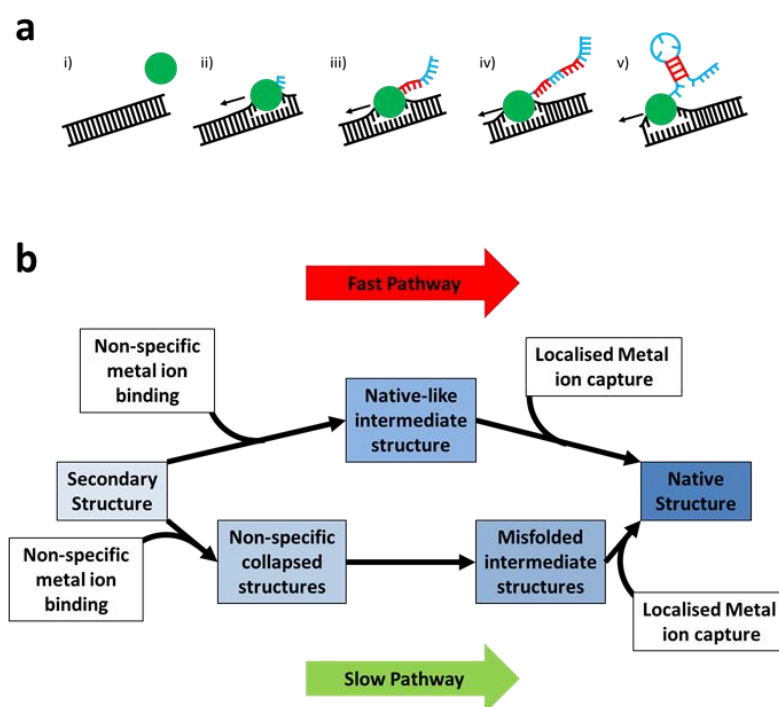
A riboswitch may also regulate genes that aren't located immediately downstream from the structure itself. In this case, rather than being considered as a *cis*-acting structure (acting upon an adjacent gene) it is said to be a *trans*-regulatory mechanism (acting upon a non-adjacent gene or a separate molecule) [120]. Two mechanisms by which this can occur are by physical interference, as postulated in the instance where the simultaneous transcription of two genes from the same DNA template forms tRNA- and SAM-binding riboswitches which prevent transcription from proceeding, and by base pairing with a non-adjacent coding region [70, 121].

### **2.6.3. Gene regulation by riboswitches takes place co-transcriptionally**

Riboswitches form as the mRNA transcript is read from the DNA template, the first part of the central dogma of molecular biology as discussed in Section 2.3 and summarised in Figure 2.3. This is summarised in Figure 2.6a for the generation of a general RNA hairpin. The figure shows an RNA polymerase (RNAP) moving to the DNA duplex containing the genetic information (i) before it binds to the duplex by creating a transcription bubble and starts to build the mRNA transcript (ii). As it proceeds, it generates complementary sequences (iii and iv, red sections)

which can pair up, leading to the formation of secondary structural motifs, in this case a simple hairpin (v). Tertiary structure formation is mediated by the presence of metal ions. An mRNA structure may collapse into its stable, native form by either binding metal ions at specific locations within the structure, locking motifs in position, or by a non-specific charge shielding of the negatively-charged phosphate spine of the mRNA [122].

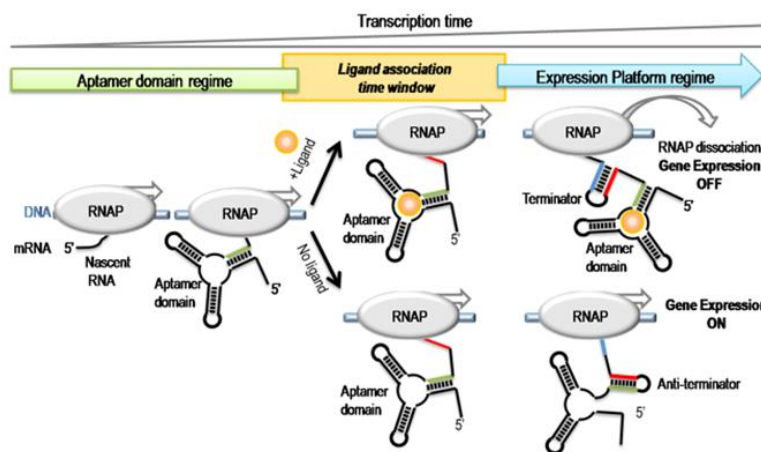
The process of condensing into the native conformation is encoded within the RNA sequence and it is summarised in Figure 2.6b. The process may proceed either directly to a structure resembling the final, stable native state, or depending on the balance between the transcription rate and the folding rate, it may ‘misfold’ leading to a number of possible non-native conformational arrangements [63, 122, 123]. These non-native structures need to be resolved before the RNA structure can further progress to the native state.



**Figure 2.6:** a) A step-wise description of the generation of a secondary structure being generated from an mRNA transcript. b) A summary of the pathways followed in RNA forming its tertiary structure from the secondary structures generated as described in a).

In the context of riboswitch function *in vivo*, in addition to the balance between the rates of transcription and the metal ion-induced folding of the secondary and tertiary folding, the outcome of the gene regulation process is modulated by the ligand-binding rate. This is schematically represented in Figure 2.7 below, where the nascent RNA transcript emerging from the RNAP starts folding as it is being transcribed. After enough of the mRNA sequence has been transcribed, the aptamer domain is ready to accept the incoming ligand. As transcription continues, there is a limited time window where the aptamer domain with the appropriate ligand-binding pocket is accessible. If the ligand-binding rate is fast enough, the aptamer-ligand complex is formed and this triggers a particular conformational change in

the expression platform, resulting in one of the two possible outcomes of the genetic decision. On the other hand, if ligand binding is too slow to compete with the transcription and folding rates then the aptamer-ligand complex is not formed and the expression platform adopts the competing structure that results in the opposite regulatory outcome.



**Figure 2.7:** Gene regulation takes place co-transcriptionally. The transcribed RNA starts folding and once the aptamer domain is fully formed the genetic decision results from the balance between the transcription rate, the folding rate and the ligand-binding rate. There is limited time window for the aptamer to bind the ligand (repression of gene expression), after this time, transcription continues and the alternative competing structure with the anti-terminator is formed resulting in gene expression being down-regulated.

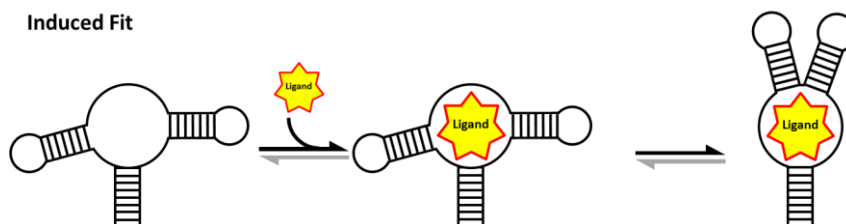
It is known that RNAP pausing sites contribute to influencing the genetic outcome by providing more time for the ligand to bind to the aptamer [124]. For instance, it has been shown that pausing plays an important role in the folding and conformational rearrangement of the *Escherichia coli* *btuB* riboswitch during transcription by the *E. coli* RNA polymerase [124]. Therefore, polymerase pausing at strategic locations may be a general

mechanism that underlies riboswitch regulatory function in response to the environment.

#### **2.6.4. Ligand-binding mechanisms: induced fit vs conformational selection models**

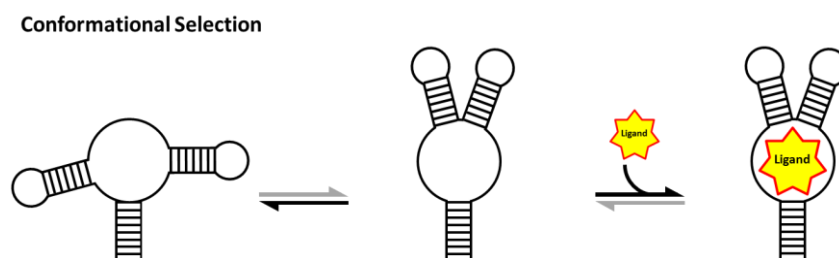
The mechanism by which the ligand is bound to the aptamer core is the most important step in the functionality of a riboswitch. There are two models that offer insight into how this is achieved: induced fit and conformational selection (also known as binding first or folding first) [125]. In the induced fit model, the riboswitch may only form the native state once the ligand has been bound. The aptamer domain exists predominantly in a non-native conformation and remains in this state in order to capture the ligand [126, 127]. Initial interactions between the unfolded non-native aptamer domain and the ligand trigger a series of conformational changes in the aptamer domain that lead to the fully formed native state and the stabilisation of the aptamer-ligand complex. The induced fit mechanism is summarised in Figure 2.8 below [127]. In the case of purine riboswitches, NMR and single molecule studies have shown that the three-way aptamer domain adopts a similar global shape with and without the ligand. However, the aptamer core remains largely unstructured in the absence of ligand. This has been taken as evidence for a partial induced-fit mechanism where ligand binding to the aptamer core induces subtle rearrangements of the local structure around the ligand [128].





**Figure 2.7:** Summary of the induced fit model for ligand capture by a riboswitch. The black arrows denote dominant processes; the grey arrow indicates a less-favoured reaction.

The conformational selection model requires the riboswitch to have adopted a competent structure to which the ligand can bind [127]. In this ‘lock and key’ mechanism, ligand binding is driven by intrinsic complementarities between the ligand and the aptamer domain. In this model, the aptamer domain may dynamically fluctuate between different conformations including ligand-binding competent structures being present at least in minor populations. For instance, it has been shown that the SAM-I riboswitch collapses in the presence of  $Mg^{2+}$  ions in a manifold of folded states, some of which resemble the ligand-bound structure [112]. The mechanism is summarised in Figure 2.8 below.



**Figure 2.8:** The mechanism of a conformational selection fitting process. As for the previous figure, the grey arrows denote less favoured processes, and the black arrows indicate the dominant process.

Though the two models illustrate distinct mechanisms for ligand capture, it has been suggested that for a given sample, environmental factors such as the ratio of ligand to acceptor and the buffer chosen may influence which binding model is followed [129, 130].

#### **2.6.5. Riboswitch classes and structural diversity**

One classifies riboswitches primarily by the metabolite that they sense. There have been over 24 classes or families of riboswitch identified as of 2015 [131]. A small sample of these families, shown along with the name and molecular weight of the ligand sensed, is presented in Table 2.1 below in order to show the diversity of this subspecies of regulatory RNA structure.

**Table 2.1:** A list of all of some of the families of riboswitches, sorted according to the class of metabolite that they sense. Beside each riboswitch is a reference or references related to their discovery. Adapted from [132].

Ligand	Structural variant(s)	Ref(s)
<b>Enzyme cofactors</b>		
Thiamine pyrophosphate (TPP)		[133, 134]
Adenosylcobalamin (B12)		[105]
S-adenosylmethionine (SAM-I)	SAM-II, -III, SAM/SAH	[111, 135, 136]
Flavin mononucleotide (FMN)		[134, 137]
Molybdenum cofactor(Moco)	Tungsten cofactor	[138]
Tetrahydrofolate		[139]
<b>Nucleobases, nucleoside, nucleotides</b>		
Guanine	Adenine, 2'-deoxyguanosine	[140–142]
Prequeosine-1 (preQ1)	preQ2	[143]
Cyclic di-guanosine monophosphate (c-di-GMP-I)	c-di-GMP-II	[144]
Cyclic di-adenosine monophosphate (c-di-AMP)		[145]
<b>Amino acids</b>		
Lysine		[113, 146]
Glycine		[109]
Glutamine		[147]
<b>Ions</b>		
Mg <sup>2+</sup>		[148, 149]
F <sup>-</sup>		[150]
<b>Others</b>		
Glucosamine-6-phosphate		[117]
S-adenosylhomocysteine (SAH)		[151]
Aminoglycoside		[152]

### 2.6.6. Riboswitches as novel antibiotic targets

The binding site in the aptamer domain has a highly specific and selective affinity for its natural ligand. The ligand is bound by forming key tertiary interactions with the riboswitch structure which lock it in place, stabilising the system in its folded conformation [15, 153–155].

However, although naturally tailored to bind this specific ligand structure, there have been extensive studies performed on some riboswitches to investigate the binding affinity of certain riboswitches for a structurally similar molecule acting as a ligand-binding competitor [111, 141, 156].

Because many riboswitches control the expression of proteins that are crucial for the bacterial organism, they have been proposed as a novel antibacterial targeting system [157]. Due to the fact that these structures have thus far not been found in any high-order organisms, this offers encouragement to the pursuit of this collaboration between medicine and biochemistry and some studies have been performed on members of the purine family of riboswitches to test this hypothesis [157].

## **2.7. Single-molecule FRET studies of riboswitches**

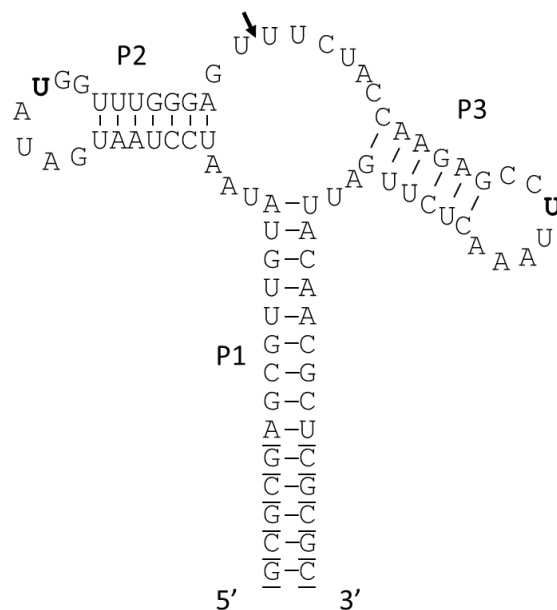
The timescales of the structural evolution of a riboswitch from secondary to tertiary conformations combined with the length scales involved make single-molecule FRET with total internal reflection fluorescence microscopy the perfect tool to examine this progression. The ability to extract not only population statistics but also kinetic characteristics and real-time folding trajectories for individual riboswitches gives an unparalleled understanding of the natural behaviour of the system [14, 30, 158].

Advances in the fields of X-ray crystallography, NMR and other structural determination techniques have shed light on the static conformations of ligand-bound riboswitch structures. By using the high atomic resolution that these techniques offer, one may ensure that any labelling schemes chosen will not interfere with the natural biological functionality of the riboswitch and prevent folding by interfering with the tertiary contacts within the structure [84].

These successes in related structural biophysical fields have allowed several of these riboswitches to be examined in sm-FRET experiments. In particular, extensive studies have been undertaken to characterise the folding pathways of the some members of the SAM family [159, 160], as well as the lysine [161], guanine [162], adenine [128, 155], preQ1-I and TPP riboswitches [125, 163].

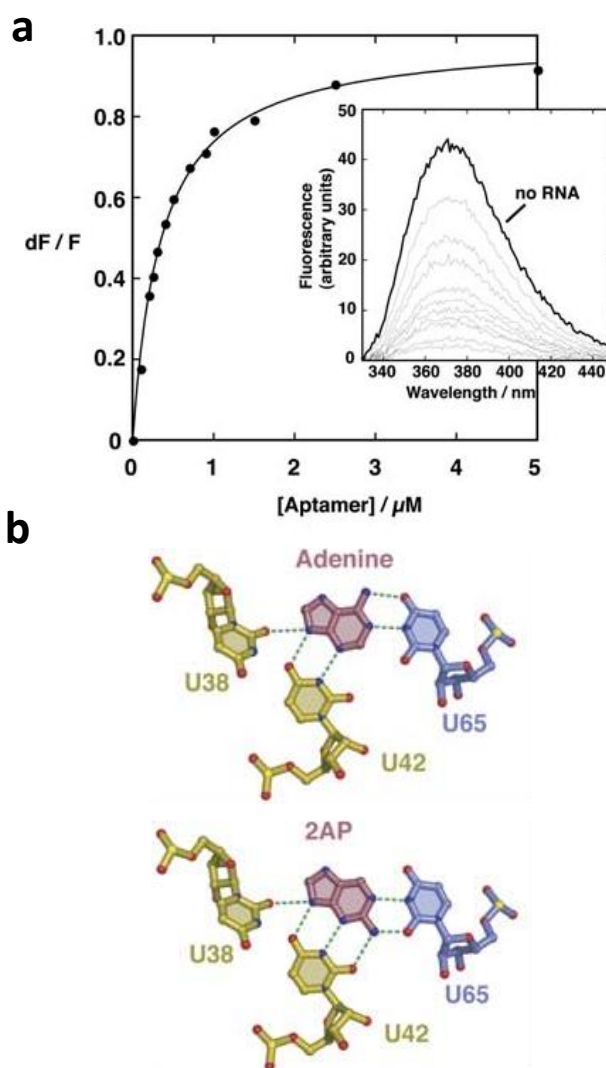
### **The first sm-FRET studies on a riboswitch: structure, function and folding pathway of the adenine riboswitch**

The earliest study of a riboswitch at the single-molecule level was performed on the adenine aptamer in 2006 [155]. In this study, Penedo et al utilised a single-molecule FRET, ensemble FRET and a fluorescence quenching assay in conjunction with measurements in order to characterise the behaviour of the riboswitch in the presence of divalent magnesium ions and the molecular adenine ligand. The aptamers used in each study were labelled differently; the bulk FRET measurements were performed using an aptamer labelled with a fluorescein-Cy3 pair, due to the high quantum yield of the donor, whereas the sm-FRET studies used a Cy3-Cy5 labelling scheme (Figure 2.9) [155]. In the case of the 2-aminopurine (2-AP) binding assay, the ligand analogue was itself the fluorescent species and as such no modifications to the aptamer were necessary [155].



**Figure 2.9:** Sequence structure of the adenine aptamer used in the studies by Penedo et al. The uracil bases in stems P2 and P3, highlighted in bold, were chosen as the sites for dye attachment. For the bulk studies, the fluorescein dye was attached to the P2 stem and the Cy3 to P3. For the sm-FRET experiments, the fluorescein was replaced by a Cy5 dye on P2, chosen to allow the same Cy3-labelled strand to be used in the ligation to create the full structure. The bases underlined in P1 were added beyond the natural sequence to ensure that the stem remained formed allowing the interaction between P2 and P3 to be isolated.

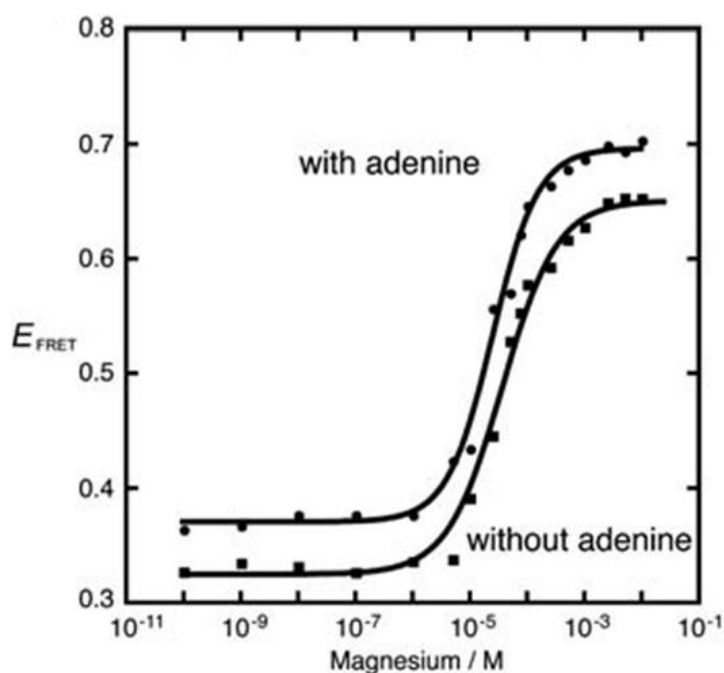
The similarities between the natural ligand and 2-AP (Figure 2.10b), a fluorescent analogue for the purine, mean that it is perfect for investigating the structural rearrangement of the aptamer core in order to accept the ligand into its binding pocket. The 2-AP emits fluorescence in the UV range with a peak wavelength of 375 nm when unbound, and its emission is quenched when captured by the aptamer core. The results shown in Figure 2.10a were indicative of this, showing a marked decrease in the fluorescence changes  $dF/F$  with increasing RNA to 2-AP ratios. The affinity for 2-AP was calculated (Figure 2.10a), giving a  $k_d$  of  $354 \pm 17$  nM, similar to the affinity of the natural ligand of around 300 nM [141].



**Figure 2.10:** a) The normalised changes in fluorescence from 2-AP, plotted against increasing aptamer concentration, showing the binding affinity for the structural analogue. The inset shows the reduction in the 2-AP emission peak caused by increasing RNA concentrations (light grey lines). b) Hydrogen bonding within the aptamer core shown for the natural adenine ligand (upper) and 2-AP (lower), with the key interactions shown as blue dotted lines. Plots and figures taken from [155].

Steady-state analysis showed that the loop-loop interactions between the P2 and P3 stems were able to form without the molecular adenine ligand present (Figure 2.11). Although the interaction was formed, however, it was at a higher magnesium background that with the ligand present; without

the ligand, the transition switched at a concentration of 22  $\mu\text{M}$ , but with a 50  $\mu\text{M}$  adenine background during the same titration the switch occurred at 12  $\mu\text{M}$  [155]. This implied that, although not essential, the ligand does play a role in stabilising the loop-loop interaction [155].



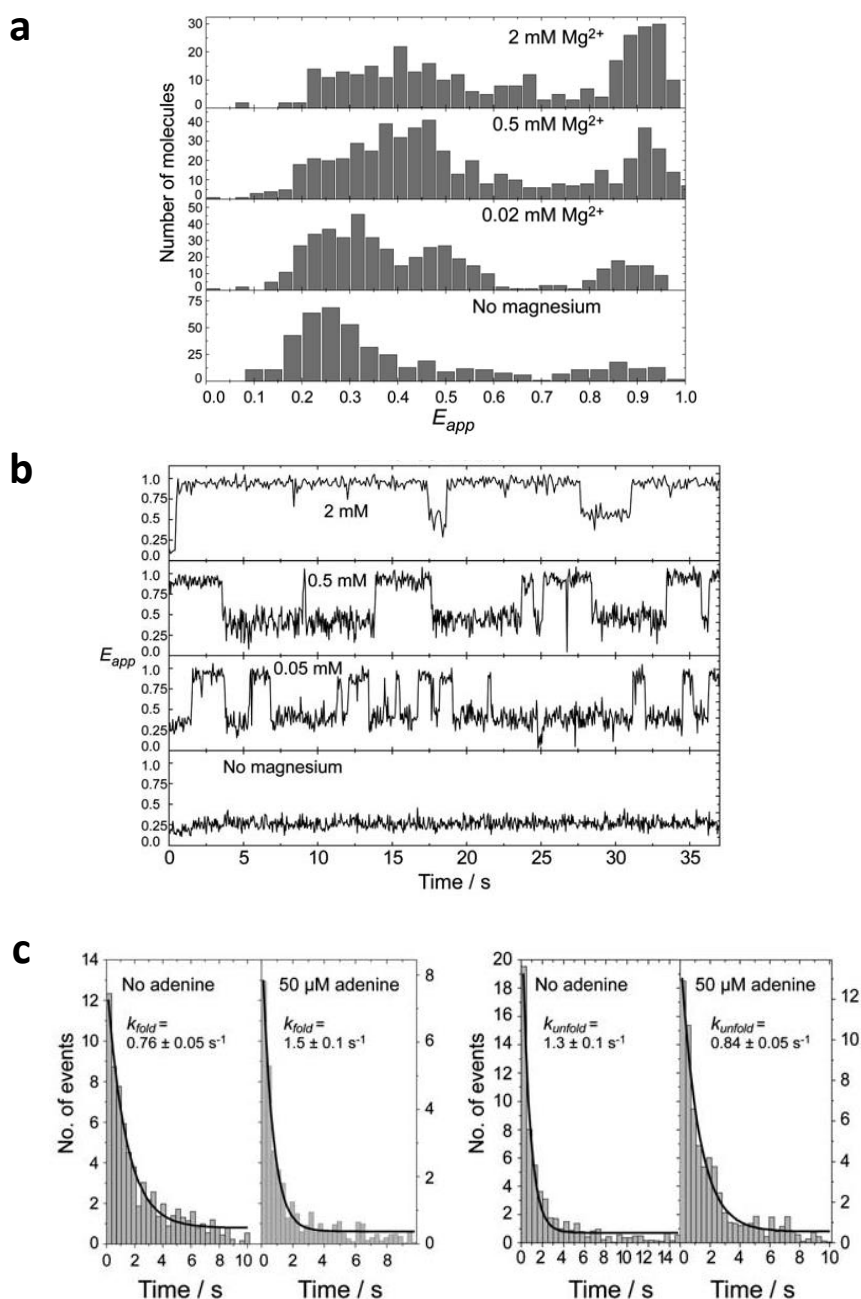
**Figure 2.11:** Results of the magnesium titrations on the adenine riboswitch with and without a background of 50  $\mu\text{M}$  adenine ligand, showing the decreased transition point calculated for the riboswitch where both magnesium and ligand are in the reaction mixture. Plot taken from [155].

The riboswitch was then exposed to different magnesium concentrations, identified from the bulk FRET results, in a sm-FRET experiment. The individual molecules showed a transition between a low apparent FRET state ( $E_{\text{app}} \sim 0.25$ ) and a high state ( $E_{\text{app}} \sim 0.9$ ). The individual molecular traces showed a series of transitions corresponding to folding and unfolding events between these two states (Figure 2.12b). Importantly, there was a third state identified, both in the single-molecule FRET population



histograms (Figure 2.12a) and in the traces, with an apparent FRET of  $E_{\text{app}} \sim 0.45$ , which was significantly populated as the concentration of magnesium was increased from 0 mM. This state, due to its absence in ensemble studies, was therefore hypothesised to be a transient state, sparsely populated while the riboswitch moves between its unfolded and folded conformations.

Kinetic analysis of these traces showed that the rates of folding and unfolding were significantly affected by the presence of the ligand in the imaging buffer. A low ligand concentration of 50  $\mu\text{M}$  was significant to induce a two-fold increase in the folding rate of the riboswitch, from  $k_{\text{fold}} = 0.76 \text{ s}^{-1}$  to  $k_{\text{fold}} = 1.5 \text{ s}^{-1}$ , and to reduce the unfolding rate from  $1.3 \text{ s}^{-1}$  to  $0.84 \text{ s}^{-1}$  (Figure 2.12c).



**Figure 2.12:** a) FRET population histograms showing the conformational arrangements of the riboswitches and the changing populations induced by magnesium. The three conformational states ( $E_{app} \sim 0.25$ ,  $E_{app} \sim 0.45$  and  $E_{app} \sim 0.9$ ) can be identified in each histogram. b) Representative traces, illustrating the same changes in conformational occupation that the histograms present. c) The rate histograms for the folding (left) and unfolding rates (right) showing the effect of the ligand on each. Each plot was fitted with a mono-exponential decay. All plots taken from [155].

## Chapter 3: Materials and Methods

### 3.1. Preparation of the RNA Structures

#### 3.1.1 Adenine Riboswitch

The adenine aptamer is constructed using two RNA strands purchased from Dharmacon (GE Healthcare Life Sciences, Lafayette, CO) post-synthetically labelled with Cy3 and Cy5 dyes. These custom oligonucleotides were purchased with one 5-amino-allyl-uridine base modification in each strand to allow biomarkers to be attached at specific positions in stems P2 and P3. The sequences are listed below in Table 3.1, with the modified uridine bases underlined.

**Table 3.1:** List of sequences and properties of the sequences used in the construction of the adenine aptamer. The extinction coefficients are given at the RNA absorption peak ( $\lambda = 260$  nm).

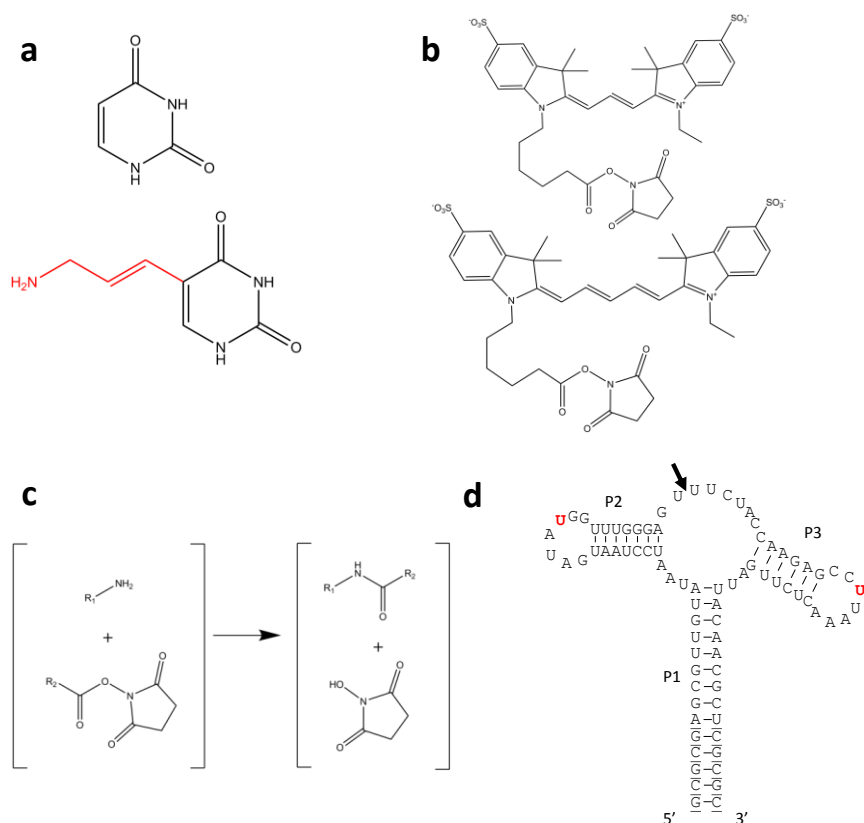
Strand identifier	Sequence (5' to 3')	Molar Mass (g mol <sup>-1</sup> )	Extinction Coefficient (l mol <sup>-1</sup> cm <sup>-1</sup> )
5' strand	Biotin-GCG CGA GCG UUG UAU AAU CCU AAU GAU <u>AUG</u> GUU UGG GAG U	13,349.22	409,000
3' strand	UUC UAC CAA GAG <u>CCU</u> UAA ACU CUU GAU UAC AAC GCU CGC GC	13,420.23	392,400

Fluorescent labelling of the RNA was carried out using N-hydroxyl-succinimide (NHS) ester derivatives of Cy3 and Cy5 to label the 5' and 3' strands respectively, according to the protocol provided by the manufacturer (Amersham, GE Healthcare Life Sciences, Lafayette, CO). The structures of

these dye derivatives are shown in Figure 3.1. After the dye molecules are attached, the RNA strands are deprotected using the buffer provided by the manufacturer (100 mM Acetic acid, adjusted to pH3.8 with Tetramethylethylenediamine (TEMED)) and precipitated in a mixture of 0.1 volumes of 3M sodium acetate and 2.5 volumes of ethanol. The supernatant is removed and the remaining pellet re-suspended in ribonuclease-free water (dd H<sub>2</sub>O) [154]. The concentrations of each strand are determined by UV spectroscopy (Varian Cary Eclipse 50, Agilent Technologies), first by using the extinction coefficients provided by the manufacturer (Table 3.1) to establish the RNA concentration and then by the dye extinction coefficients to estimate the labelling efficiency for each dye-labelled strand. The protocol followed for performing UV spectroscopy is described in more detail in Section 3.5.1 below.

The pre-ligation structure was generated by mixing a 1:1 ratio of each labelling strand, followed by hybridisation in a water bath, first by heating the mixture up to 80°C for 1 minute and then allowing it to cool to room temperature. This reaction is performed in a buffer consisting of 10 mM HEPES (pH 7.5) with a salt concentration of 50 mM NaCl [155]. The solution is then incubated for 4 hours at 37°C with a T4 RNA Ligase (New England BioLabs), added after the return to room temperature, to join the two constituent strands into one. The full structure is shown in Figure 3.1d, including the positions of the modifications for the dyes. The ligation site is also marked with an arrow and was similar as that previously described by

Lemay et al [155]. The labelling positions were chosen using the X-ray structure of the adenine riboswitch as a reference to identify sites that will not interfere with folding or ligand binding.



**Figure 3.1:** a) The structure of the modified 5-(3-amino-allyl)-uridine nucleobase (lower panel) compared to the naturally occurring uracil base. The amino-allyl group crucial for labelling is shown in red. b) The structures of the modified dye molecules, with the NHS group which forms the link with 5-amino-allyl-uridine tethering the dye to the nucleobase. Cy3 mono-NHS is shown in the upper panel, with Cy5 mono-NHS shown below. c) The reaction to covalently attach the dye molecule to the modified nucleobase, where  $R_1$  represents the RNA nucleobase and  $R_2$  represents the dye molecule. d) Full sequence structure of the adenine aptamer, the arrow denotes the T4 RNase ligation site and the red Us show the positions of the modified uracil bases as shown in a).

### **RNA aptamer purification**

The ligation product was purified using polyacrylamide gel electrophoresis (PAGE). A denaturing polyacrylamide gel is prepared (7M urea, 12% acrylamide, 10% TBE in a volume of 50 ml, with 40  $\mu$ l TEMED and 400  $\mu$ l of 10% ammonium persulphate (APS) added to activate polymerisation). In addition to the ligation product, small quantities of each of the individual strands are loaded into separate wells as migration controls for the non-ligated oligonucleotides [154]. The gel is run until clear separation of the fully ligated product and the non-ligated strands can be seen, after which all identifiable bands are removed and the RNA extracted before being precipitated in 0.1 volumes of 3M sodium acetate (NaAc) and 2.5 volumes of ethanol. The supernatant is removed and the pellet is re-suspended in dd H<sub>2</sub>O for the single strands and T50 (50 mM tris hydrochloride at pH 7.8) for the fully-formed aptamer. Finally, the concentration of the riboswitch is determined by UV spectroscopy and the sample stored at 5°C. Single strands can be stored in the freezer.

#### **3.1.2. The Hammerhead ribozyme**

##### **Synthesis of the RNA structures**

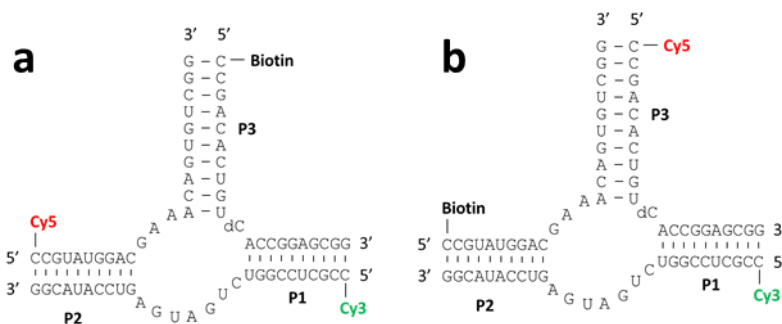
The minimal form of the hammerhead ribozyme was constructed by three-strand hybridisation, where each strand carries a single and different modification. This ensures that only the correctly hybridised constructs are

fully active during single-molecule imaging. Individual strands were purchased from Dharmacon. All necessary modifications, a modified base, the inclusion of Cy3 and Cy5 dye molecules and a biotin linker, were performed by Dharmacon during phosphoramidite synthesis. The strand sequences and extinction coefficients are listed in Table 3.2 below. The strands are first deprotected using the buffer provided as described in the previous section. To prevent cleavage, the substrate strand carries a 2'-deoxyribose nucleotide (indicated by dC in Tables 3.2 and 3.3).

**Table 3.2:** Sequences and properties of the constituent strands of the hammerhead ribozyme. Extinction coefficients are given for  $\lambda = 260$  nm (biotin strands),  $\lambda = 547$  nm (Cy3 strand) and  $\lambda = 647$  nm (Cy5 strands).

Strand identifier	Sequence (5' to 3')	Molar Mass (g mol <sup>-1</sup> )	Extinction Coefficient (l mol <sup>-1</sup> cm <sup>-1</sup> )
Cy3 strand	Cy3-CCG ACA CUG U(dC)A CCG GAG CGG	9,103.42	258,830
Biotin Strand (I-II)	Biotin-CCG ACA CUG U(dC)A CCG GAG CGG	7,114.58	196,800
Biotin Strand (I-III)	Biotin-CCG UAU GGA CGA AAC AGU GUC GG	7,827.51	232,100
Cy5 Strand (I-II)	Cy5-CCG UAU GGA CGA AAC AGU GUC GG	7,958.82	242,100
Cy5 Strand (I-III)	Cy5-CCG ACA CUG UdCA CCG GAG CGG	7,242.39	206,800

The strands are quantified as described in the section 3.1.1, before being annealed in a 1:1:1 ratio by heating to 75°C in a water bath before being left to cool to room temperature. The annealing process is performed in T50 buffer at a salt concentration of 50 mM NaCl. The complete ribozyme structure is illustrated in Figure 3.2.



**Figure 3.2:** The complete sequence and structure of the minimal hammerhead ribozyme as prepared for single-molecule spectroscopy. a) Structure to investigate interactions between stems P1 and P2. b) Structure designed to examine interactions between stems P1 and P3.

The full-length hammerhead ribozyme was generated using a two-strand system. One strand was designed to carry both of the dye molecules, with a 5' Cy5 modification and a 2-aminouridine base substitution at position 22 in the strand to which the succiCy3 base was covalently attached using an NHS-modified Cy3 molecule, as described in the previous section. This strand also carries the 2'-deoxycytosine base modification added to prevent cleavage (dC). The labelled sample was purified using the protocol described in section 3.1.1 before being quantified using UV spectroscopy to estimate the labelling efficiency. The concentrations were calculated for all three peaks (RNA, Cy3 and Cy5 absorption) with the lowest calculated concentration taken as the sample concentration. By doing this, one maximises the number of fully-formed structures generated by hybridisation.

The second strand required no further modifications after purchase, as the biotin modification at the 5' end was added during the synthesis of the



oligomer. The biotin and dye-labelled strands were mixed at a 1:1 ratio and annealed by the same process as used for the minimal structures.

**Table 3.3:** Strands used to generate the natural hammerhead ribozyme, complete with molecular weight and extinction coefficient information. The modified bases are denoted by dC (2'-deoxycytosine) and 2'-amino-U (2'-amino-uracil). The extinction coefficients are given for  $\lambda = 260$  nm (biotin strand) and  $\lambda = 647$  nm (Cy5-Cy3 strand).

Strand identifier	Sequence (5' to 3')	Molar Mass (g mol <sup>-1</sup> )	Extinction Coefficient (l mol <sup>-1</sup> cm <sup>-1</sup> )
Cy5-Cy3 strand	Cy5-CCA UGG CGU (dC)CU GGA UUC CAC (2'-amino-U)UC UCG	9,014.36	255,300
Biotin strand	Biotin-CGA GAA GGU ACA UCC AGC UGA UGA GAC CCA AAU AGG UCG AAA CGC CAU GG	16,578.35	508,400

### Purification of the RNA structures

For this model, a denaturing gel is unsuitable since the structure consists of three individual strands held together by the base pairing. Thus in order to purify the fully hybridised ribozyme from the individual strands we use a 12% native gel (12% acrylamide and 10% TBE in a volume of 50 ml, with 40  $\mu$ l TEMED and 400  $\mu$ l of 10% APS to polymerise the gel) and run to separate the structures, with control lanes of the single strands running in parallel to identify incomplete hybridisation products. The bands corresponding to the full structure are cut and the RNA extracted using the crush and soak method, precipitated, quantified and stored as described in Section 3.1.1.

## 3.2. Channel preparation for single-molecule spectroscopy

### 3.2.1. Cleaning protocol prior to construction

Quartz microscope slides (Quartz Scientific) are recycled and must be thoroughly cleaned to remove all traces of previous biological samples. All new slides must be pre-drilled with two 1.5 mm diameter holes, positioned to allow a diagonal chamber to be built under a coverslip. Slides with chambers from completed experiments are heated in glass slide jars in water to soften the tape and epoxy holding the channel together. The tape and coverslip are removed from the slide, and both surfaces scrubbed first with acetone, and then with methanol, until visibly clean.

These slides are then put into clean glass jars, transferred to a fume cupboard and immersed for two hours in a 50-50 (v/v) of hydrogen peroxide and hydrochloric acid, a powerful cleaning reagent which strips all traces of biomolecules from the quartz slides.

This solution is carefully disposed of and the slides rinsed with ddH<sub>2</sub>O before being washed in a sonic bath following a ten step process as follows:

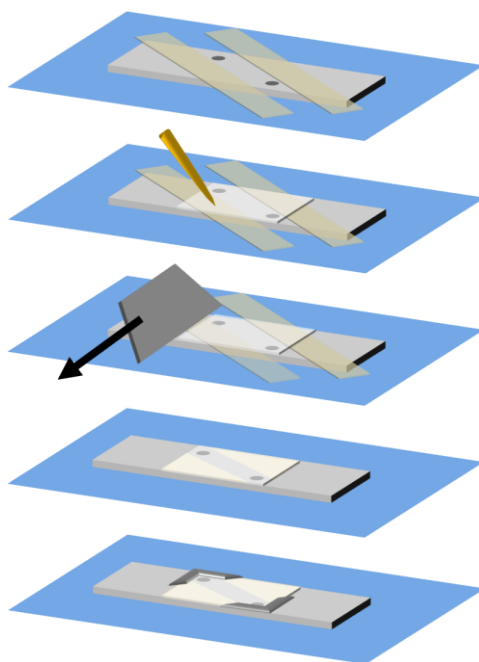
- 15 mins in a solution of Hellmanex detergent
- 5 mins in ddH<sub>2</sub>O
- 15 mins in acetone
- 5 mins in ddH<sub>2</sub>O
- 15 mins in 1M potassium hydroxide (KOH)
- 5 mins in ddH<sub>2</sub>O

- 15 mins in methanol
- 5 mins in ddH<sub>2</sub>O
- 15 mins in 1M KOH
- 5 mins in ddH<sub>2</sub>O

The slides may be stored for up to two weeks, but in this case they need to be regularly refreshed by repeating the final three steps of the sonic bath procedure. [154]

### **3.2.2. Constructing imaging chambers for single-molecule spectroscopy**

A pre-cleaned slide is dried using nitrogen gas before being burned in a Bunsen flame. After cooling with nitrogen, the channel is constructed as shown in Figure 3.3.



**Figure 3.3:** Preparation of a sample chamber for single-molecule spectroscopy. The slide is transferred to a tissue and two lines of double-sided tape (Rymans) are run parallel to the pre-drilled holes. A coverslip (32 x 24 mm, VWR) is placed and the contact sealed using a pipette tip. The excess tape is cut away, both along the edge of the coverslip and the slide, and the ends of the chamber sealed using waterproof epoxy resin (Araldite Instant) [154]

### 3.3. Surface Immobilisation for single-molecule spectroscopy

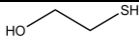
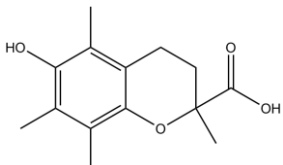


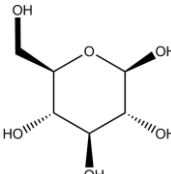
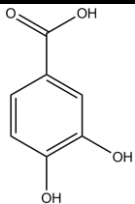
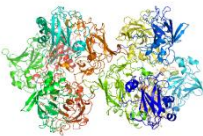
Biotinylated bovine serum albumin (BSA-bi, Sigma-Aldrich) is dissolved first to a stock concentration 10 mg/ml, and then further diluted to 1 mg/ml, in T50 buffer. Neutravidin or Streptavidin (Thermo Scientific) is dissolved to a concentration of 5 mg/ml and then diluted to 0.2 mg/ml in T50. A volume of 60  $\mu$ l of 1 mg/ml BSA-bi is injected into the channel and left to incubate for 15 minutes in a water box, to ensure that the sample won't evaporate. This is then flushed out with 100  $\mu$ l of T50 buffer before 60  $\mu$ l of

0.2 mg/ml Neutravidin is added. This is left to incubate for a further 15 minutes in the water box, before being flushed out with 100  $\mu$ l of T50 buffer as before. The slide is then imaged using the total internal reflection fluorescence microscopy (TIRFM) setup to ensure that it is clear of any fluorescent contaminants before the addition of the RNA sample. Once this is confirmed, the sample is first diluted to a picomolar concentration (10-100 pM) and then 60  $\mu$ l is loaded into the chamber. This is incubated in the water box for a further 5 minutes before the non-immobilised RNA is flushed out with 100  $\mu$ l of T50 buffer [154].

### **3.4. Single-molecule additives to improve dye photostability**

In order to allow measurements to be taken over long timescales, oxygen must be removed. Failure to do so will result in single-molecule observation times being greatly compromised due to excessive blinking and fast photobleaching. To maximise the efficiency of data acquisition, an imaging buffer must be used, containing not only the chemical conditions of interest but also oxygen scavengers known to suppress these two unwanted processes. The components of the imaging buffer depend on the particular experiment being undertaken. Some of the possible choices are listed below in Table 3.4, along with their required concentrations.

**Table 3.4:** A selection of reagents used to minimise problematic fluorescence fluctuations, along with their required concentrations. The stated concentrations are the final required concentrations present in 100  $\mu$ l of prepared imaging buffer [15, 164].

Fluorescence Property	Reagents that limit the property	Structure of the reagent/constituents	Concentration(s) in buffer
Blinking	2-mercaptoethanol		10 mM
	Trolox		1 mM
Photobleaching	Gloxy system: Glucose Oxidase, Catalase and Glucose		100 nM (Glucose Oxidase) 1.5 $\mu$ M (Catalase), 300 mM (Glucose) – dissolved in the T50 buffer
			
			
	PCA-PCD system: Protocatechuic acid and Protocatechuate 3,4-dioxygenase		2.5 mM (PCA) 50 nM (PCD)
			

### Preparation of single-molecule additives

*2-mercaptoethanol:* 3.9 g (0.05 moles) of 2-mercaptoethanol (Sigma-Aldrich) is dissolved in 50 ml of ddH<sub>2</sub>O, to give a final concentration of 1 M. A volume

of 1  $\mu\text{l}$  is required per 100  $\mu\text{l}$  of total volume of imaging buffer to give the optimal concentration.

*Trolox*: 125 mg (5 mmoles) of Trolox (Sigma-Aldrich) is dissolved in 1 ml of methanol, to which 9 ml ddH<sub>2</sub>O is then added. The pH is adjusted to pH 9.5 using sodium hydroxide (NaOH). A volume of 2  $\mu\text{l}$  must be added to the imaging buffer for the concentration required.

*Gloxy*: 8 mg of Glucose Oxidase (Sigma-Aldrich, from the fungus *Aspergillus niger*) is dissolved in 20  $\mu\text{l}$  catalase (Roche, from bovine liver) and mixed with 70  $\mu\text{l}$  of T50 buffer. This solution is spun in a centrifuge until the supernatant is clear and a pellet is visible at the bottom of the tube.

*Glucose*: 3 g D-glucose (Sigma-Aldrich) is dissolved in 50 ml ddH<sub>2</sub>O with 0.394 g Trizma hydrochloride (Sigma-Aldrich) and the pH adjusted to pH 7.8 using NaOH. When gloxy is used, this buffer must be used in the place of regular T50 to make the volume up to 100  $\mu\text{l}$  in order to ensure that an adequate concentration of glucose is present.

*PCA-PCD*: The PCD enzyme (Sigma-Aldrich) is dissolved to a concentration of 150 nM in ddH<sub>2</sub>O. The concentration is calculated from the mass of the supplied sample, molecular mass and the number of units present. A volume of 1  $\mu\text{l}$  of this stock must be added to the imaging buffer.

A mass of 0.48 g of PCA (Sigma-Aldrich) is dissolved in 50 ml ddH<sub>2</sub>O to make a stock concentration of 62.5 mM. A volume of 4  $\mu\text{l}$  of this stock must be added to the imaging buffer.

### 3.5. Bulk Spectroscopic methods for RNA studies

The advantages of choosing to perform single-molecule experiments over bulk methods have already been described in detail in Chapter 1. However, bulk methods still offer very valuable information about the folding mechanism of the RNA structure and constitute a faster approach to initially characterise the system [15].

#### 3.5.1. UV-Vis spectroscopy to determine RNA concentration

In order to determine the concentration of single strands or hybridised RNA aptamers, one must quantify the absorption of the strand across the UV and visible ranges. A 1 cm path length cuvette (Starna/Perkin Elmer) is cleaned by rinsing with ethanol and then ddH<sub>2</sub>O and dried using a nitrogen airline. The 'Scan' software package, a part of the spectrometer software (Varian), is loaded and the baseline correction option chosen. In addition to this, the range must be set from 250 nm to 750 nm to cover the full absorption range of the Cy5 dye. Though it may not always be present, for consistency this range is always chosen.

149  $\mu$ l of ddH<sub>2</sub>O is added to the cuvette and the baseline calculated. Then 1  $\mu$ l of the sample stock is added into the cuvette and the sample scanned. The peaks of the trace are then calculated and the values displayed. The concentration of the sample in the cuvette is calculated by using the Beer-Lambert Law as shown by Equation 3.1 below:



$$[Sample] = \frac{Abs_{\lambda}}{l \cdot \epsilon_{\lambda}} \quad (3.1)$$

Where  $Abs_{\lambda}$  is the peak absorption value at a wavelength ( $\lambda$ ),  $l$  is the optical path length (normally 1 cm for standard microcuvettes) of the sample and  $\epsilon_{\lambda}$  is the molar extinction coefficient for the wavelength ( $\lambda$ ). [9] From this value, the initial stock concentration may be calculated by taking into account the dilution of the volume of stock used in the volume of ddH<sub>2</sub>O in the cuvette.

### 3.5.2. Lifetime measurements using a pulsed-excitation fluorimeter

Fluorescence lifetime measurements are performed using an Edinburgh Instruments FLS980 spectrometer fitted with an MCP-PMT detector. The excitation source is a 470 nm pulsed laser diode head (Picoquant LDH-P-C-470B), driven at a frequency of 20 MHz by a Picoquant PDL-800-B or PDL-800-D laser driver. The components of the TCSPC setup must be cooled throughout, so the water cooling, nitrogen line and detector cooling systems must be on for 15 minutes before any data acquisition.

The instrument response function (IRF) must be calculated before any data acquisition, in order to deconvolve the finite laser pulse from the experimental decay of the sample. Also, the IRF should be retaken after every 15 minutes of imaging, since the pulse profile could vary through

heating. Using the supplied software, the parameters required for recording the IRF and for TCSPC are listed in Table 3.5 below.

**Table 3.5:** The parameters used to record the IRF function and data.

<b>Parameter</b>	<b>Value (IRF)</b>	<b>Value (TCSPC)</b>
Excitation wavelength	467 nm	467 nm
Emission wavelength	467 nm	477 nm
Excitation Slit width	1 nm	1 nm
Emission Slit width	1 nm	Sample-dependent
Max Counts on detector	100,000 counts	100,000 counts
Time Range	50 ns (for fluorescein)	50 ns (for fluorescein)
Channels	4096	4096

In addition to these parameters, the neutral density filter is used to control the laser power and, by extension, the counts on the detector. Before sample changeover and IRF signal recording, this filter must be set to the lowest value to avoid detector damage. Should the signal from the sample be too weak to reach 100,000 counts, then the ND filter is set to allow the maximum laser power in the cuvette and the emission slits are opened out (maximum value possible is 9 nm). After the scan is complete, the data is saved in both the standard Edinburgh Instruments format and as an ASCII file for further analysis.

### **3.6. Analysis of bulk data for dye-labelled RNA Structures**

#### **3.6.1. Calculating fluorescence lifetimes from TCSPC data**

The data is analysed using DecayFit, a package developed by Søren Preus and run through MATLAB [165]. All of the ASCII files saved from the scans

are loaded into the program and divided into IRF and Decay groups. The program pairs up selected decay and IRF files, performs the deconvolution to extract the fluorescence decay components from the decay files, and then fits to the curve the chosen order of exponential decay. In addition to fitting the exponential decay curves to the data, the residuals and chi-squared values are given to allow an indication of how well the data has been fitted. The fits and residuals for each condition examined are then exported in ASCII form to be plotted in origin. To calculate the effective lifetime from a bi-exponential fit or a fit of a higher order, the average lifetime must be calculated. This is calculated using Equation 3.2 below, subject to a fitting equation of the general form shown in Equation 3.3.

$$\tau_{av} = \frac{\sum_{i=1}^n A_i \tau_i^2}{\sum_{i=1}^n A_i \tau_i} \quad (3.2)$$

$$I(t) = \sum_{i=1}^n A_i e^{-\frac{t}{\tau_i}} \quad (3.3)$$

### 3.6.2. Inter-dye distance measurements from TCSPC data

In order to calculate the inter-dye separation for a dual-labelled RNA structure, one must also have an identical structure but with only the donor dye present. With the lifetimes from both of these samples calculated from DecayFit as described above, then the FRET value for the dual-labelled structure may be calculated by Equation 3.4.

$$E = 1 - \frac{\tau_{DA}}{\tau_D} \quad (3.4)$$

The quantities  $\tau_{DA}$  and  $\tau_D$  in the above equation refer to the fluorescence lifetimes of the sample with the donor and acceptor dyes present and with only the donor dye respectively. Using this FRET value, and knowing the Förster radius  $R_0$  for the donor-acceptor dye pair, the separation of the two dyes may be calculated from Equation 3.5.

$$R = R_0 \left( \frac{1 - E}{E} \right)^{\frac{1}{6}} \quad (3.5)$$

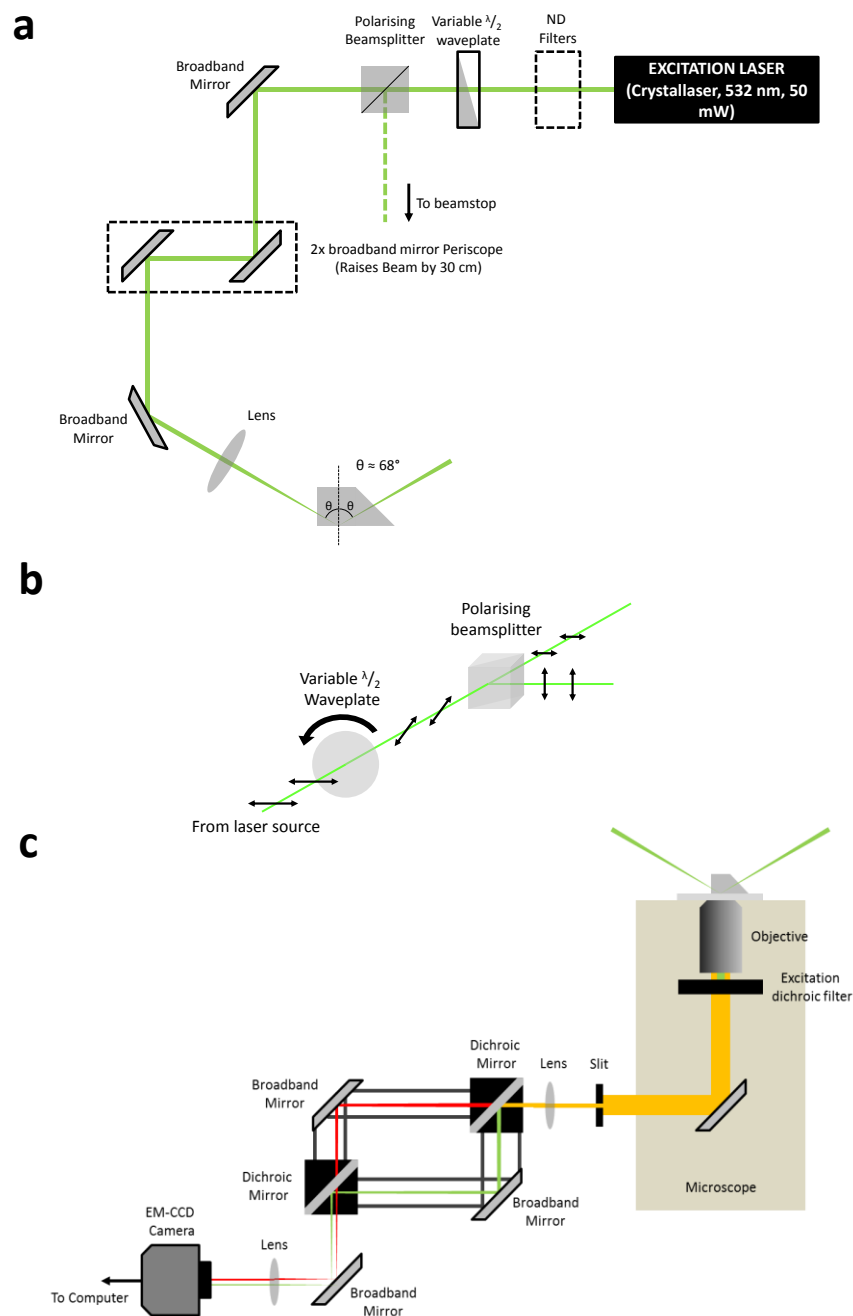
### 3.7. Single-molecule mapping and recording using prism-type total internal reflection fluorescence microscopy (p-TIRFM)

#### 3.7.1. Technical details of the p-TIRFM setup

In order to maximise the signal-to-noise ratio from the single-molecule experiments, the chosen configuration for the setup used in the lab is total internal reflection microscope (TIRFM). More specifically, since no sample access is required whilst measurements are being taken, the configuration chosen was prism-type TIRFM, hereafter denoted by p-TIRFM. Excitation is provided by a 50 mW continuous-wave diode-pumped solid-state laser (Crystalaser) operating at a wavelength of 532 nm. The power output from

the laser is fixed, but attenuation is provided by a half-wavelength waveplate combined with a polarising beam splitter, which provides the benefit of maintaining the linear polarisation of the laser source whilst allowing a high degree of control over the excitation power (Figure 3.4b). Further attenuation may be achieved by placing neutral density (ND) filters in the excitation path. The laser beam is then raised to allow the correct excitation angle by a periscope consisting of two broadband mirrors (Thorlabs) with a third mirror steering the beam through a focussing lens (Melles Griot,  $f = 8$  cm) into the quartz prism. The excitation path is shown with all of these components in Figure 3.4a.

The emission is collected by an Olympus IX71 microscope, fitted with a high NA water-immersion objective (Olympus UPlanSApo, N.A. = 1.2, 60x magnification) to collect as much light as is possible. The fluorescence collected from the microscope is filtered to remove any excitation light by a long pass filter (532 nm, Chroma Technologies) and reshaped into a rectangular slit. It is then focussed by a lens (Melles Griot,  $f = 50$  cm) through an arrangement of broadband mirrors (Thorlabs) and long-pass dichroic beam splitters (647 nm cut-off wavelength, Chroma Technologies). These mirrors create a positional offset between fluorescence signals from the two dye species, and these are then focussed using a further lens (Melles Griot,  $f = 50$  cm) onto separate halves of the detector, an Andor iXon EM-CCD camera. This network of optics is shown in Figure 3.4c.



**Figure 3.4:** a) Schematic of the excitation path and optics for the p-TIRFM system used for all single-molecule FRET experimental work. The positions of all elements are included, including optional power attenuation through the inclusion of ND filters. b) The attenuation mechanism achieved by the half waveplate and beamsplitter. c) The optics used to resolve both dye channels onto different halves of the EM-CCD chip. The elements responsible for the positional offset between the two channels are secured using cage rods, indicated by the dark grey lines.

### 3.7.2. Calibration of the p-TIRFM system and data acquisition

To image the emission of the donor and acceptor dyes on the left- and right-hand sides of the EM-CCD camera respectively, a calibration algorithm needs to be optimised to recognise donor and acceptor spots corresponding to the same single molecule. To do this, a calibration slide is prepared, where 200  $\mu\text{m}$  silica microspheres (Invitrogen) functionalised with Cy3 and Cy5 dyes are added to a channel and non-specifically adsorb to the surface. The high emissivity of these beads in both of the channels makes them perfect for calibrating the system.

An algorithm written in IDL (Exelis Visual Information Solutions) developed by Taekjip Ha is used to perform this calibration. A short movie is recorded by the camera software, ensuring both that the bead density is adequate to allow a thorough calibration and a clear separation between the point-spread function (PSF) from each individual microsphere is maintained ( $\sim 200$  beads in the field of view). After creating an image file from the movie, one PSF is manually chosen in the Cy3 channel. The corresponding position in the Cy5 half of the chip is displayed as a circle, which can be moved manually to be centred on the correct Cy5 PSF and then saved. Three beads are chosen this way, with the choices being spread across the whole chip to maximise the global fitting. The algorithm then reads all PSFs using the three selected coordinates as a reference and generates a mapping file, which is used to match the Cy3 and Cy5 signals for each molecule in an experiment. This mapping process must be performed prior to any

experimental work, as optical drift can lead to variations in the exact position of the light to each side of the chip. Furthermore, it is recommended that a fresh mapping file is taken after every 2 hours of imaging as a precautionary measure to ensure that drift doesn't affect the resolution of the two channels. In reality, however, this change is not enough to compromise the recognition of the emission spots.

Provided that the mapping file provides a good fit, the camera parameters are adjusted for imaging the labelled RNA and the sample-containing chamber can be mounted on the p-TIRFM. The parameters used for both beads and RNA imaging are listed below in Table 3.6. The movies are then analysed in IDL using the mapping file generated before any experimental work has been done, and the intensities of corresponding channels for each single molecule are collated into a file for analysis in MATLAB.

**Table 3.6:** The parameters used to set up the Andor iXon camera for imaging beads and RNA samples. The values listed for the RNA imaging are typical for imaging with an integration time of 50 ms/frame; when using longer or shorter integration times these numbers need to be adjusted.

Parameter	Value for imaging beads	Value for imaging RNA
Integration time (ms/frame)	50	50
Gain	0	180 to 220
Data scalar	>10	2
Background subtraction	Minimum value -10	Minimum value -10
Movie time	100 frames (5 seconds)	1000 frames (50 seconds)



### 3.8. Analysis of single-molecule FRET data

#### 3.8.1. Steady-state FRET histograms

The quickest indication of changes in the conformational state of an RNA structure under different chemical conditions is by generating steady-state histograms for each condition. These plots offer a clear visual indicator of the most common structural arrangement for the RNA for each condition. The traces must first be examined to ensure that no anomalous fluorescence properties are present. Should any of the characteristics listed below be present for a trace, it is not saved for further analysis. These are:

- Intensity steps indicating the presence of more than one dye of each species – in this case, fluctuations cannot be attributed to a single FRET-pair system. These can be a result of multiple RNA molecules binding to a single Neutravidin or incorrect labelling of the RNA.
- Abnormally high trace baseline - background corrections coming from the camera acquisition software should keep the baseline at approximately 10-50 counts; thus this indicates either another molecule bound to the Neutravidin or the presence of some fluorescent contaminant.
- Uncorrelated intensity changes between both dye channels – this suggests the presence of a second molecule on one Neutravidin or an incorrect mapping file.

- Intensities of the channels show an exponential decay as the movie progresses – this is normally caused by microscope drift during the measurement time.
- Intensities for both channels of less than 50 counts for the entire trace – this is background noise and is not to be considered for any analysis.

Should the molecule fit the criteria, it is included in the building of the histogram. This involves taking an average of the first ten frames of the apparent FRET trace, generated from the intensities of each of the dye channels from Equation 3.6 below.

$$E_{app} = \frac{I_{Cy5}}{I_{Cy3} + I_{Cy5}} \quad (3.6)$$

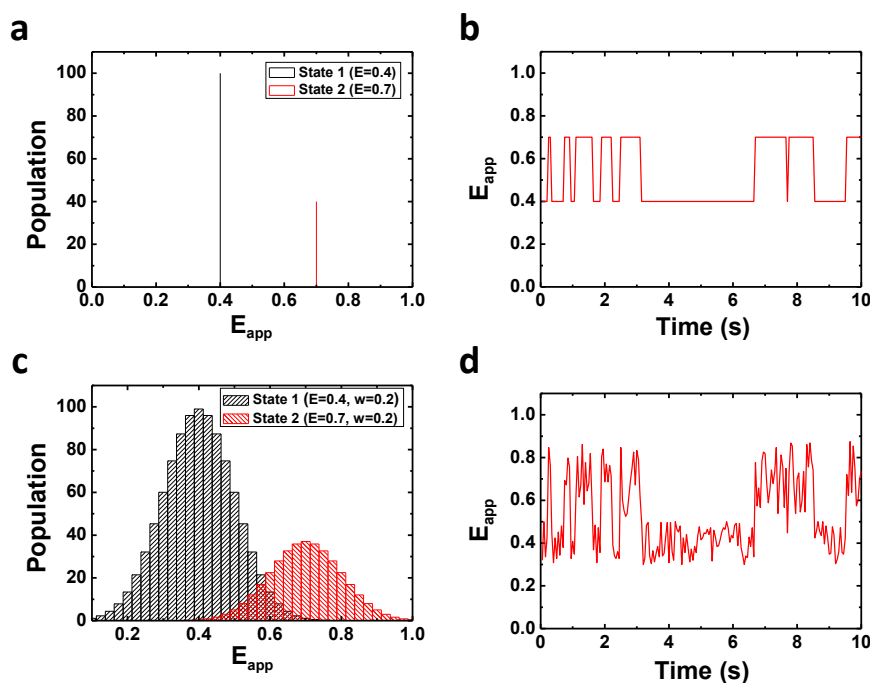
Each contribution is stored in MATLAB before, at the end of each movie, being exported, binned and plotted in Origin, creating the full steady-state histogram for the chemical conditions.

### 3.8.2. Studying dynamics using Hidden Markov Modelling (HMM)

In order to extract kinetic information from single-molecule trajectories, one must fit the traces and extract the time durations for each conformational state. One method widely employed is Markov modelling, where the likelihood of the evolution of a randomly changing sequence between defined states is calculated without considering the previous

behaviour of the system, provided that the kinetics of these transitions may be governed by a mono-exponential relationship [166]. This technique, initially developed to be applied in speech recognition, is now widespread through the field of biophysics [167], particularly in the fields of ion-channel research and single-molecule spectroscopy [166]. The versatility of the technique, requiring simply a stream of data that should move between discrete states, makes it perfect for analysing single-molecule FRET trajectories and fitting a step-wise trend line to view the transitions clearly.

In order to understand the principles of Markov modelling, consider a two-state system, fluctuating between states 1 and 2 which, keeping in the context of sm-FRET, correspond to apparent FRET values of 0.4 and 0.7 respectively (Figure 3.5a). Here, the dynamic behaviour between the two states is perfectly clear; the step-wise plot shown in Figure 3.5b allows rate analysis to be performed on the durations of each transition.



**Figure 3.5:** a) Distribution of FRET values suitable for Markovian analysis, where two discrete states are present (0.4 and 0.7). b) A representative trace which would be associated with this type of data. c) A system unsuitable for conventional Markov modelling, but possible to analyse with Hidden Markov Modelling due to the presence of two states as before but with the definition between the two blurred by instrument and/or sample noise. d) A the same representative trace but with the noise responsible for the broadening of the previously-discrete FRET states.

However, in the case of the analysis of sm-FRET traces, one must employ Hidden Markov Modelling (HMM). This is because when a trace is recorded for a molecule and the FRET trace extracted from the intensity information (Equation 3.6), the stepwise behaviour is obscured by noise, whether this be from the instrument or the sample itself (Figure 3.5d) [166, 168]. As a result, the populations for each state are broadened and can overlap (Figure 3.5c) meaning that additional processing must be done to draw out the same kinetic information.

The first stage of processing is the establishment of a transition probability matrix (Figure 3.6), which states the probabilities for a molecule in an initial state  $\phi$  (1 or 2) to move to a final state  $\psi$  (1 or 2) [166]. The transition probabilities ( $tp_{\phi \rightarrow \psi}$ ) are related by the following equations (Equations 3.7a and 3.7b) [166]:

$$tp_{1 \rightarrow 1} + tp_{1 \rightarrow 2} = 1 \quad (3.7a)$$

$$tp_{2 \rightarrow 2} + tp_{2 \rightarrow 1} = 1 \quad (3.7b)$$

		Final	
		$tp$	1
Initial	1	0.93	0.07
	2	0.02	0.98

**Figure 3.6:** Example of a transition probability matrix. The initial states are listed down the left column and the final state across the top. The rows are representatives of Equations 3.7a and 3.7b.

Next, the emission probability is calculated, which is the probability of observing a particular FRET value ( $FRET_{obs}$ ) when starting in a particular state, as a function of the distribution width ( $w_\phi$ ) [166]. This quantity is calculated for each state (1 and 2) by Equations 3.8a and 3.8b:

$$EP_1(FRET_{obs}) = \frac{1}{w_1} e^{-2 \left( \frac{(FRET_{obs} - FRET_1)}{w_1} \right)^2} \quad (3.8a)$$

$$EP_2(FRET_{obs}) = \frac{1}{w_2} e^{-2 \left( \frac{(FRET_{obs} - FRET_2)}{w_2} \right)^2} \quad (3.8b)$$

Equations 3.8a and 3.8b show that the model factors in a Gaussian distribution of noise around the observed value ( $FRET_{obs}$ ). This can clearly be seen in the plots of single-trace histograms in the data chapters (for example, Figure 4.3). The deviation of the data around this chosen value for the state often gives the impression that the fitting is wrong for heavily noise-affected data, but single-trace histograms confirm that the model has calculated its fitted FRET value at the centre of this distribution. Once these two factors have been calculated, the probability for each state to have emitted at that time point. The probability ( $p(i)$ ) for each individual transition to occur from a particular point ( $i$ ) on the trajectory  $FRET(i)$  is calculated as in Equation 3.9 for a general transition between any starting or final state, defining  $\alpha$  as state 1 or 2:

$$p(i) = EP_{\alpha(i)}(FRET(i)) \times tp_{\alpha(i) \rightarrow \alpha(i+1)} \quad (3.9)$$

The maximum probability recorded for each option for the transitions is taken to represent the sequence of states between point ( $i$ ) and point ( $i+1$ ). Given that this is a probability-based fit, the fit will not record every fluctuation in fluorescence as a transition. This is taken into account in the input data where the program reads the raw intensity data from both FRET dye channels and uses these to determine whether the fluctuation observed in the trace is due to noise or intramolecular rearrangements.

### **Fitting the FRET trajectory using a Hidden Markov Modelling Program**

In addition to the criteria outlined in section 3.7.1, a trace suitable for HMM analysis must have a length of at least 400 frames in order for an accurate idea of the dynamics to be displayed. These are saved in a format suitable for rate analysis software to process. Several research groups have developed software that is either purpose-written or can be used for single-molecule FRET trajectories [166, 168–170]. The one chosen to analyse this work is HaMMY, developed by the Ha Group from the University of Illinois at Urbana-Champaign [166].

The program is launched and the number of states to be fitted is entered. The saved traces are then loaded and fitted, with the fits displayed with the traces as the program proceeds through the list of files. The program generates three output files per trace:

- Dwell files – these files list the durations of all complete transitions in terms of the number of frames. The file shows three columns: starting FRET value; FRET value after the end of the transition; length of the transition in frames.
- Path files – these files are analogous to the input files, listing each intensity, FRET and fitted value per frame. The file contains five columns of data: frame number; donor intensity; acceptor intensity; apparent FRET value; fitted FRET value.

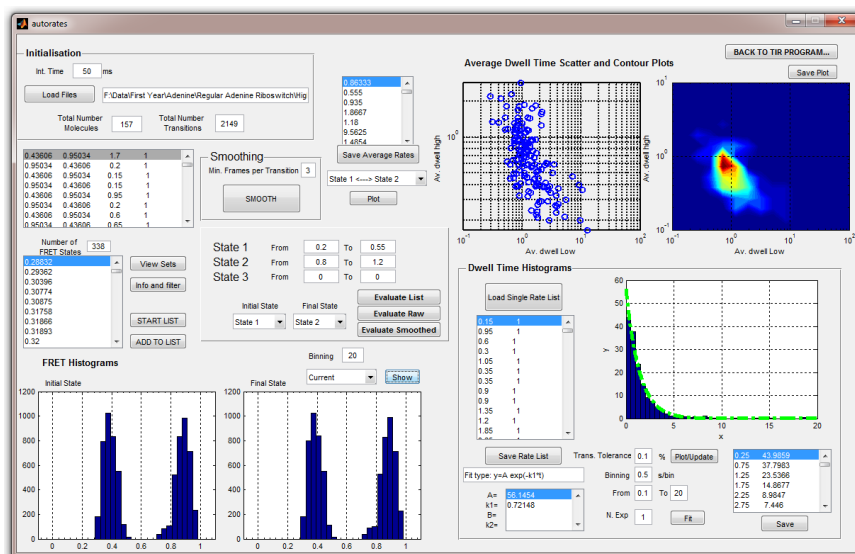
- Report files – these files are used by the transition density plot program that comes with HaMMY and contain the transition probability matrix calculated for the molecule.

### **Extracting dynamic information from HMM-fitted FRET data using MATLAB**

To calculate the folding and unfolding rates, as well as the average dwell times for each state per molecule, I wrote a program in MATLAB to process the output files from HaMMY. The graphical user interface (GUI) created for the program, called ‘autorates’, is shown below in Figure 3.5.

The integration time at which the movies were recorded is entered into the GUI, after which the dwell files are collected from a chosen source folder, a numeric molecular identity grouping all of the transitions associated with each molecule is added to each transition row, and they are compiled into one large matrix. The first and last transitions are then removed from molecule since they cannot be viewed as complete transitions. Finally, the duration of each transition is converted from frames to seconds using the integration time entered before loading the data. At this stage, it is possible to review all fitted molecules and perform manual filtering in case any anomalous molecules have been processed, though this shouldn’t be necessary since the criteria for picking the traces should ensure this doesn’t happen. More often this option is used to identify suitable example traces for each condition.





**Figure 3.5:** Image of the self-written MATLAB auto-rates GUI after full processing of the raw data. All of the plots, with the exception of the FRET histograms, can be saved from the program in an ASCII format. The FRET histograms are plotted used only to define the limits for the states before initialising rate analysis.

A further feature included is the capability to include a second set of data which has been recorded at a different integration time and combine the two sets to analyse them together.

HaMMY fits all transitions, down to a single frame, and although these may well be legitimate transitions it is a widely accepted rule that a transition must span at least three frames in order to be considered a legitimate change of state. As a result, a filtering algorithm was written into the program, where these fast transitions may be smoothed out. A variable number of frames per transition may be chosen (see Chapter 4, Section 3).

In order to identify the boundaries for each state, a histogram is plotted of all unique FRET values fitted by HaMMY. Having these limits, the upper

and lower bounds are entered into the GUI and the transition of interest chosen from the drop-down menu before the program is run. A maximum of three states may be analysed in this way. The dwell times are listed for all transitions falling within the criteria entered before execution. A histogram is generated using the default plot properties, but these may be modified after it has been generated. These properties include the number of seconds per bin and time scale (x axis). In order to ensure that molecules exhibiting more fast transitions don't fully dominate and alter the plot, each molecule is given a weight of 1, and thus for a specific molecule each transition carries the weight:

$$\textit{Transition Weight} = \frac{1}{\textit{Number of transitions}} \quad (3.10)$$

A fitting option is also included, where a mono or bi-exponential decay may be fitted to the plot and the parameters displayed in a dialogue box in the GUI. These histogram plots may be saved in ASCII format to be exported to OriginPro (OriginLab, Northampton, MA, USA) for plotting. The extensions of these ASCII files once saved are indicative of which transition has been processed. The fitted curves are not saved with the data due to the increased fitting accuracy and error calculation offered by OriginPro compared to MATLAB, though the differences between the fitting parameters generated by the two are not significant.

Scatter and contour plots of the average molecular dwell times for each state may be generated and saved in ASCII or SigmaPlot-ready formats respectively. To examine a different transition, the new initial and final states are chosen from the drop-down menu, and the program must then be re-executed. The data saved from the Autorates program is can be broken down into three parts. The first column is the x-axis of the plot. This is calculated from the range of the average molecular dwell times for the low FRET state, where the lower bound is rounded down to the nearest power of ten and the upper bound is rounded up to the nearest power of ten. The number of bins is set in the code but can be easily modified. The same procedure is followed for the data in the second column, which is the y-axis data, though the limits are set by the average molecular dwell times for the high FRET state. The third part is the z-data, which is generated by binning the dwell times into a matrix according to their value. This is a square matrix. This file format allows the data to be imported directly into SigmaPlot, and the contour plots generated and customised.

### **3.9. Data Analysis**

#### **3.9.1. Fitting procedures for single-molecule data**

All of the fitting applied to any plots shown in this was performed in OriginPro. The procedure adopted by the program is a non-linear curve

fitting algorithm. This involves setting up a model of the form shown in Equation 3.11.

$$Y = f(X, \theta) + \epsilon_i \quad (3.11)$$

In this model,  $Y$  represents the data to be fitted,  $X = \{x_1, x_2, \dots, x_n\}$  are the independent variables,  $\theta = \{\theta_1, \theta_2, \dots, \theta_m\}$  are the parameters to fit to and  $\epsilon_i$  are the residuals, the differences between the fit and the data [171].

The first stage is to calculate a quantity known as the chi-square ( $\chi^2$ ), a quantity that quantifies the goodness of fit by taking the residual values for each measurement between the data and fit, defined in Equation 3.12 for the general observation  $i$  [171].

$$\chi^2 = \sum_{i=1}^n \left[ \frac{Y_i - f(x_i', \hat{\theta})}{\sigma_i} \right]^2 \quad (3.12)$$

The goal is to minimise the residuals by adjusting the parameters through an iterative process. This process is known as chi-square minimisation, and in order to do this the partial derivatives of the chi-square with respect to each variable in turn. The general form of this equation is shown below for an equation with  $m$  parameters (Equation 3.13).

$$\frac{\partial \chi^2}{\partial \hat{\theta}_m} = -2 \sum_{i=1}^n \frac{1}{\sigma_i^2} [Y_i - f(x_i', \theta)] \left[ \frac{\partial f(x_i', \theta)}{\partial \hat{\theta}_m} \right] \quad (3.13)$$

The fitting process ends when the difference between successive iterative adjustments causes a negligible change in the chi-square value.

### Steady-state FRET histograms

The steady-state histograms, once plotted, are fitted with Gaussian profiles across each peak. Each function takes the form shown in Equation 3.14.

$$f(x) = \frac{A}{w\sqrt{\frac{\pi}{2}}} e^{-2\left(\frac{x-x_c}{w}\right)^2} \quad (3.14)$$

As can be seen from the equation, each Gaussian has three variables to fit; the area under the curve ( $A$ ), the width of the curve ( $w$ ) and the centre offset of the curve ( $x_c$ ). The number of replicas chosen depends on the number of states deemed present for the sample. For the two samples examined in this thesis, the adenine riboswitch and hammerhead ribozyme, three states have been shown to exist and hence three separate Gaussian functions are fitted to each histogram [128, 155, 172]. In addition to this, in the case of the former, the three states have been characterised not only at the bulk fluorescence level but also at the single-molecule level, and so a starting point for each centre offset is known. This offset is iteratively adjusted to give the best possible fit for each state; the other two parameters are left to be fitted by the software. Cumulative plots are added to the final histogram to show the overall fitting from the three contributions.

## Dwell time histograms

The outputs from the Autorates program are plotted and fitted with exponential decay functions. Where only a single folding or unfolding mechanism is expected to occur, for example where one species of metal ion folds the sample, the process will follow a mono-exponential profile. In the case where a second reagent is introduced which will act in a different way, for example in the case of a riboswitch where both metal ions and ligand molecules are present, each of these species will influence the folding or unfolding rate separately and thus both contributions must be treated accordingly. In this case, a bi-exponential decay is expected to occur, and the two rate constants,  $k_1$  and  $k_2$ , will represent the ligand-docked and ligand-free components of the dwell time histograms. The mono- and bi-exponential fitting equations used are stated below in Equations 3.15 and 3.16 respectively.

$$y = Ae^{-k_1t} \quad (3.15)$$

$$y = Ae^{-k_1t} + Be^{-k_2t} \quad (3.16)$$

### 3.9.2. Error analysis and the propagation of errors

#### Steady-state histograms

The parameters fitted to the histograms by OriginPro have associated errors with them. These standard errors are combined using the equations listed in Appendix II, after being converted to a percentage error where appropriate. The percentage error in each Gaussian is used where any value derived from this fit is applied to any further calculations.

The variance is calculated for the centre values for each of the Gaussian populations, e.g. state 1, state 2 and state 3 for the adenine aptamer. The value of the variance is used as the error in the reported  $E_{\text{app}}$  values in the experimental chapters.

### **Dwell time histograms**

As for the Gaussian curves fitted to the steady-state histograms, the parameters generated from fitting mono- and bi-exponential decays to these histograms have standard errors associated with them. The decay rates are stated with the fitted error in all cases. In addition to this, all further calculations involving these rates have the percentage errors propagated through to the final values (e.g. for the equilibrium constant and free energy calculations performed in Chapters 4 and 5).

## Chapter 4: Influence of monovalent ions on riboswitch folding

### 4.1. Introduction

Nucleic acids play a number of crucial and diverse biological roles within the cell. Deoxyribonucleic acid is, of course, the macromolecular storage for all of our genetic information. Ribonucleic acid, however, is responsible for and involved in many diverse key processes [50]. Among these, the most recently discovered is the metabolite-sensing gene regulatory structures, known as riboswitches, structures found in the non-coding regions of messenger RNA, which activate and deactivate the expression of genes subject to the concentration of a specific metabolite [108].

Single-molecule fluorescence spectroscopy offers the means to resolve the functions and structures of these large molecular switches. In particular, single-molecule Förster Resonance Energy Transfer (sm-FRET) allows not only this insight into the physical conformation of the riboswitch but also offers dynamic information, providing the folding and unfolding rates for any chemical condition considered [15, 153, 154].

For a riboswitch to become biologically active, it must first fold into a stable, tertiary structure, a process that is induced by the presence of mono- and divalent metal cations. It is known that divalent ions, like magnesium,



are bound at specific locations in order to hold the structure in its native state, whereas monovalent ions, such as sodium, non-specifically bind along the RNA backbone to screen the negative charge of the phosphate and allow the secondary and tertiary structures to form without intra-strand repulsion [122, 123].

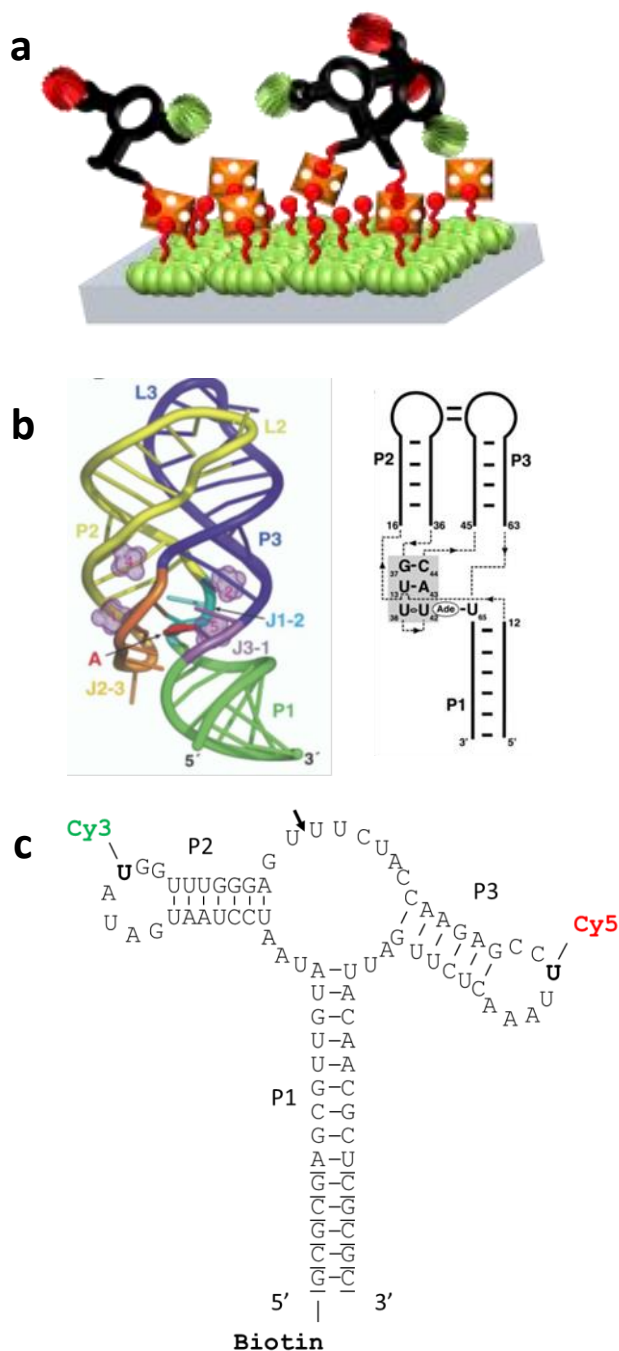
Our group has previously investigated the action of increasing concentrations of divalent magnesium ions, in the presence of a constant and biologically-relevant sodium concentration, upon several classes of these structures; however, there has been very little work into examining the folding pathway due to the presence of monovalent ions alone [108, 112, 128, 155, 173]. The work in this chapter aims to investigate this process and to compare the nature of the monovalent ion-mediated stabilisation of the adenine riboswitch from sodium to that provided by divalent magnesium ions.

#### **4.2. Steady-state monovalent ion-induced folding of the adenine riboswitch**

In order to probe the folding pathway of the adenine riboswitch aptamer induced by monovalent ions, the aptamer must be loaded into one of the slide chambers described in Chapter 3. In addition to this, the surfaces must be functionalised, which serves the dual purpose of fixing the sample in position and forming a protective barrier between the quartz and the RNA sample.

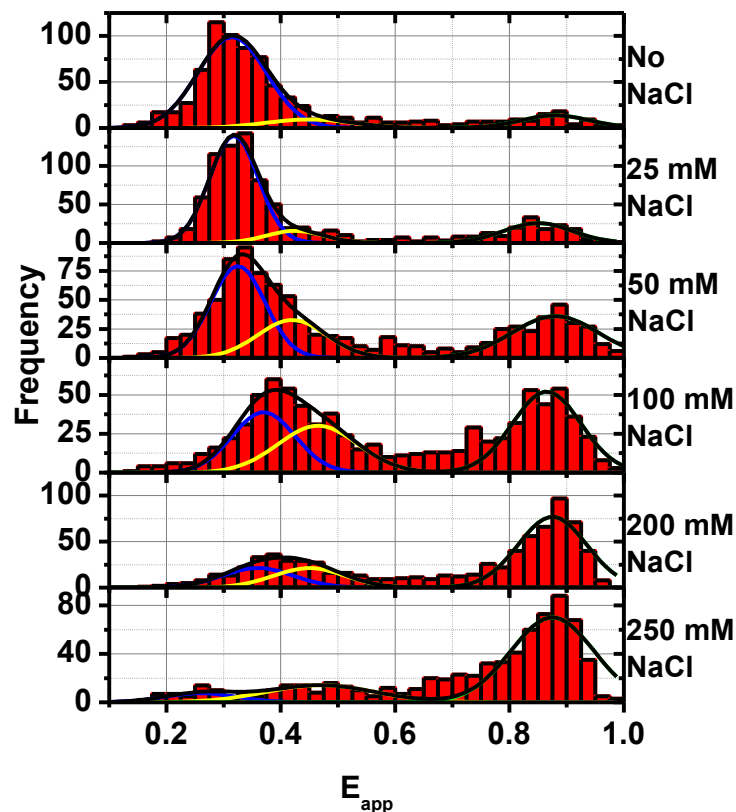
The method of surface immobilisation used in this study employs a small molecular link called biotin. This naturally occurring molecule is used in combination with two proteins, biotinylated bovine serum albumin (BSA-bi), which will coat the surface of the quartz non-specifically, and Neutravidin, one member of a family of proteins with four biotin-specific binding sites. The sample may also be site-specifically modified to allow it to be tagged with a biotin link, thus tethering the entire system to the quartz slide, as shown in Figure 4.1 below.

The chosen model is the adenine riboswitch, responsible for the regulation of the gene coding for adenosine deaminase (ADA), whose main function is in the regulation of the efflux pump controlling the level of purine metabolites in the bacterial cell [174]. To prepare the sample for fluorescence spectroscopy, and specifically for sm-FRET, the sample requires two further site-specific modifications to allow the covalent attachment of the chosen dye pair, Cy3 and Cy5. The sites chosen have bases modified with amino groups, and these are used as target reactive groups for cyanine dyes covalently attached to N-succinimidyl ester groups. This process is covered in greater detail in Chapter 3. The sequence and structure of the riboswitch and the modification sites are illustrated in Figures 4.1b and 4.1c.



**Figure 4.1:** a) The mechanism for BSA-bi-Neutravidin immobilisation. The biotinylated BSA (green) binds non-specifically to the surface, and the Neutravidin (orange) attaches to the biotin linker offering three potential binding sites for the dye-labelled RNA structure (black) to attach. b) The structure of the adenine riboswitch, the crystal form (left panel) and the nucleotide chain showing the bases involved in pairing and loop formation. c) The sequence structure of the adenine riboswitch used, with the locations of the dyes highlighted (green denotes Cy3, red denotes Cy5). In addition to this, the bases involved in the docking of stems P2 and P3 are displayed in bold, with the six additional base pairs in stem P1 added for stability underlined. The biotin modification is at the 5' end of the strand.

The dual-labelled adenine aptamer was diluted to picomolar concentration levels in T50 buffer (50 mM trizma hydrochloride (Sigma-Aldrich), pH 7.8), and then added to the slide chamber previously incubated with BSA-bi and streptavidin. Immediately prior to imaging, a buffer was flushed into the chamber that consisted of D-glucose, glyoxy and  $\beta$ -mercaptoethanol, stabilisation reagents that limit photobleaching and photoblinking respectively, and the chemical reagents necessary to establish the desired sample conditions. The sample was excited by a continuous wave diode-pumped solid-state laser source (Crystalaser, 40 mW,  $\lambda_{\text{excitation}} = 532 \text{ nm}$ ), p-polarised at the quartz-sample interface, in a prism-type total internal reflection fluorescence microscopy setup as introduced in Section 1.5.2 in Chapter 1 and as discussed more specifically in Section 3.7 in Chapter 3. The wide-field capabilities of this technique, combined with the low excitation power due to the evanescent field-induced excitation, allow the fluorescence emissions of up to 250 molecules to be simultaneously imaged for long timescales. Each condition is examined in turn, the chemical conditions being changed by first cleaning the old buffer out by flushing through T50 alone, and then adding the new buffer into the chamber.



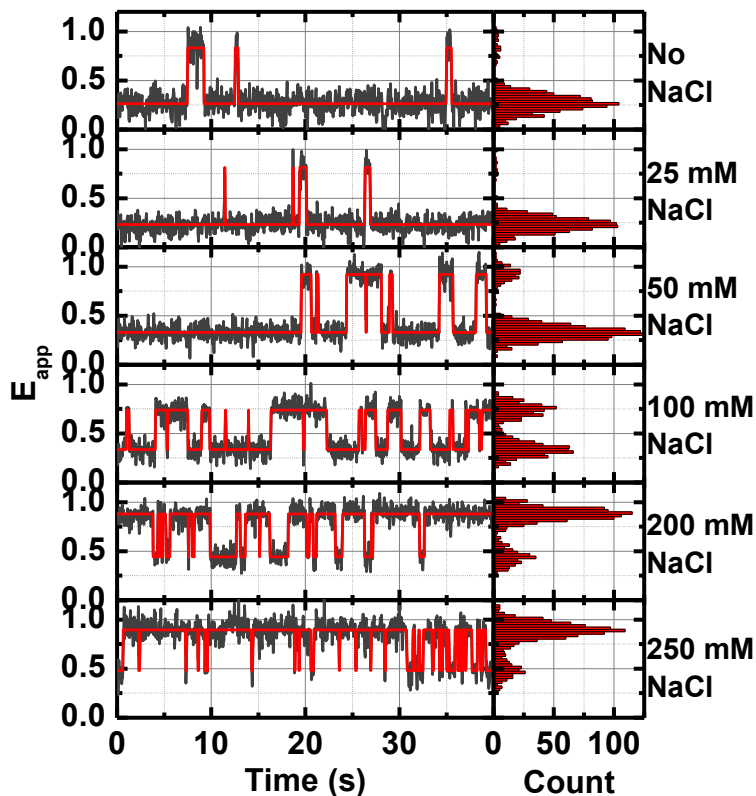
**Figure 4.2:** Single-molecule FRET histograms showing the change in the dominant conformation of the adenine riboswitch induced by increasing the sodium concentration in the imaging buffer. The solid lines represent the result from a fit to three Gaussians corresponding to the unfolded U ( $E_{\text{app}} = 0.32 \pm 0.04$ ), the intermediate state I ( $E_{\text{app}} = 0.45 \pm 0.02$ ) and folded state F ( $E_{\text{app}} = 0.87 \pm 0.01$ ).

In Figure 4.2, I show the results from a single-molecule FRET titration of the adenine riboswitch as a function of  $\text{Na}^+$  ions. The histograms generated were fitted in OriginPro using the progress described in Chapter 3. In the absence of  $\text{Na}^+$ , a main Gaussian centred at  $E_{\text{app}} = 0.32 \pm 0.04$  is observed with a very small contribution from a state located at high FRET ( $E_{\text{app}} = 0.87 \pm 0.01$ ). This suggests that, in the absence of ions to screen the negative charge, stems P2 and P3 of the adenine aptamer are far apart. I

will term this state the unfolded conformation of the adenine riboswitch (**U**). The values obtained here are identical to those obtained previously when investigating the effect of divalent metal ions, particularly  $\text{Mg}^{2+}$  [15, 128, 153, 155]. As I increased the concentration of  $\text{Na}^+$  ions, three main effects are observed: i) a progressive decrease in the contribution of the Gaussian centred at  $E_{\text{app}} = 0.32 \pm 0.04$ , ii) a very significant increase in the contribution of the population of the high FRET state ( $E_{\text{app}} = 0.87 \pm 0.01$ ) and iii) the appearance of an additional population whose contribution initially increases at low  $\text{Na}^+$  ions and then decreases again. Based on the crystal structure of the adenine riboswitch which shows that P2 and P3 stem loops become in close contact and establish hydrogen-bonding interactions between them, I have assigned the high-FRET state to the folded conformation (**F**) that brings the Cy3 and Cy5 dyes in close proximity. The FRET value of this state resembles that obtained at concentrations of  $\text{Mg}^{2+}$  higher than 2 mM, thus supporting this assignation. The FRET state with an  $E_{\text{app}} = 0.45 \pm 0.02$  should be assigned to the intermediate state (**I**). The observation of a bell-shaped distribution for the relative amplitude of this Gaussian supports the hypothesis that it corresponds to a species that is only transiently formed and very poorly populated at all ionic conditions. This is in agreement with previous data using  $\text{Mg}^{2+}$  as a folding agent [155]. The results of this study confirm that monovalent ions do indeed fully fold the adenine riboswitch, but to full fold the structure 250 mM concentrations are required. Comparing this to the figures from previous studies on

magnesium, which overpopulates the fully folded state at concentrations of 4 mM, this corresponds to a 50-fold increase in the required salt concentration [128].

Overall, the data supports a model where  $\text{Mg}^{2+}$  and  $\text{Na}^+$  ions induce a similar folding pathway. However, as expected and previously observed for other riboswitches and ribozymes, the concentration of monovalent ions required to induce a similar degree of folding is much higher for  $\text{Na}^+$  than for  $\text{Mg}^{2+}$ . This is partially due to the higher charge of  $\text{Mg}^{2+}$  ions but also due to the inability of  $\text{Na}^+$  ions to replace  $\text{Mg}^{2+}$  and bind specific pockets of the RNA that are required to stabilise the folded conformation as observed in the crystal structure. The origins of this FRET transition could be confirmed as distance dependent as opposed to a dipolar rearrangement by performing polarisation fluorescence measurements such as those mentioned in Chapter 1 by modifying the p-TIRFM emission optics to work with a single dye but to image different polarisations on each half of the EM-CCD chip.



**Figure 4.3:** Representative single-molecule FRET trajectories (black) obtained at the specified concentrations of  $\text{Na}^+$  ions showing the change in the dominant conformation of the adenine riboswitch. The red lines represent the result from a Hidden Markov simulation of the experimental trace (see main text for details). The corresponding single-molecule FRET histograms for each individual trace and the corresponding Gaussian fits are also shown.

The interconversion dynamics of single adenine aptamers as a function of monovalent ions is shown in Figure 4.3. In the absence of  $\text{Na}^+$ , the single-molecule FRET trajectory extracted as described in Chapter 3 indicates that the aptamer is mostly in the U state ( $E_{app} = 0.32 \pm 0.04$ ) with very occasional fluctuations to the folded conformation ( $E_{app} = 0.87 \pm 0.01$ ). The folded conformation at these conditions exhibits a very short lifetime ( $< 1\text{s}$ ). Upon increasing the concentration of  $\text{Na}^+$  ions, it can be seen that the frequency of fluctuations between the low- and high-FRET states progressively increases. For instance, at 100 mM  $\text{Na}^+$  ions, the dwell times in the low- and high-



FRET state are similar, suggesting the adenine aptamer spends equal time in each conformation. At high concentrations of  $\text{Na}^+$  ions ( $\sim 250$  mM), the single-molecule trajectories are characterised by a long-lived high-FRET state with occasional and short-lived transitions to the low-FRET state.

In addition to this, the presence of an intermediate state at  $E_{\text{app}} = 0.45 \pm 0.01$  is also confirmed (Figure 4.3, single-trace histograms). For instance, whereas in the absence of  $\text{Na}^+$  ions the fluctuations take place from a high-FRET to low FRET state that has a value of  $0.32 \pm 0.04$ , at intermediate concentrations of  $\text{Na}^+$  this takes place from a similar high FRET state ( $0.87 \pm 0.01$ ) to a low state with a value of  $0.45 \pm 0.02$ . Thus, it appears that the intermediate state is a transient conformation given that, although significantly populated, it is at no point during the titration the dominant macromolecular arrangement. This is consistent with the observations from previous studies on the action of divalent magnesium ions upon the adenine riboswitch [128, 155, 173].

### 4.3. Investigating the sodium-induced folding dynamics using Hidden

#### Markov modelling

The raw traces obtained from the EM-CCD camera offer a visual indication of the dynamic behaviour of the riboswitch at each salt concentration, and an overall idea of the occupancies of each state can be inferred from the histograms built up by analysis of many hundreds of

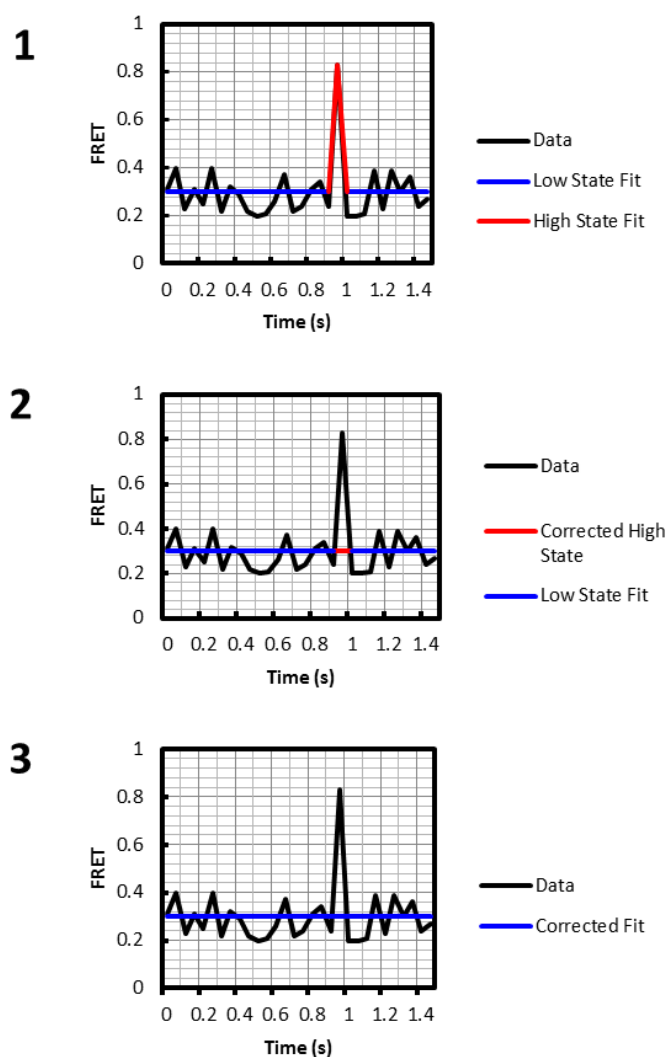
traces. However, due to the low light levels recorded and the sensitivity of the camera the traces are often very noisy, and so instead of one FRET value per state one tends to obtain a distribution, centred on a peak value. This is not an issue when the experimental objective is to identify the state of highest occupancy, but in order to gain some quantitative idea of the two-state kinetic behaviour of the riboswitch at each condition this by eye approach is insufficient and depends on the experience of the researcher. Moreover, when the fluctuations take place between more than two states it becomes very difficult to extract the kinetic information by hand; therefore a more consistent and mathematically reproducible approach is required.

In order to overcome this issue, Hidden Markov Modelling-based programs are widely utilised [166, 168–170]. In our group, the package chosen was the program developed by the Ha group based in Illinois, called HaMMy, which fits a predefined maximum number of possible states and forms output files compatible with graphical software [166]. The selection of traces was carefully regulated; any traces of insufficient length and exhibiting photobleaching, photoblinking or any other irregularities such as multiple dyes in either of the channels were omitted so to obtain the most accurate model possible from the sample. The folding rates were then calculated using the self-written program in MATLAB described in Chapter 3, Section 8 and displayed in graphical form.

### **Post-HaMMy filtering of “fast transitions”**

Due to the low signal intensities associated with single-molecule experiments, and the short integration times dealt with, there is a risk that false transitions can be detected. For example, should a non-tethered dye cross through the excitation volume and create an intensity for only one frame, this can be wrongly interpreted as a conformational change when using automated fitting programs, such as HaMMy. As a result, it is widely accepted that for a true transition, the state must be held for at least three frames.

In order to examine the effect of these so-called “fast transitions”, a small program was written in MATLAB to smooth out these spikes subject to a predefined threshold. Should a fitted state fall below the required duration, its value is changed to that of the previous state and the length of the dwell time summed. This is illustrated below in Figure 4.4.



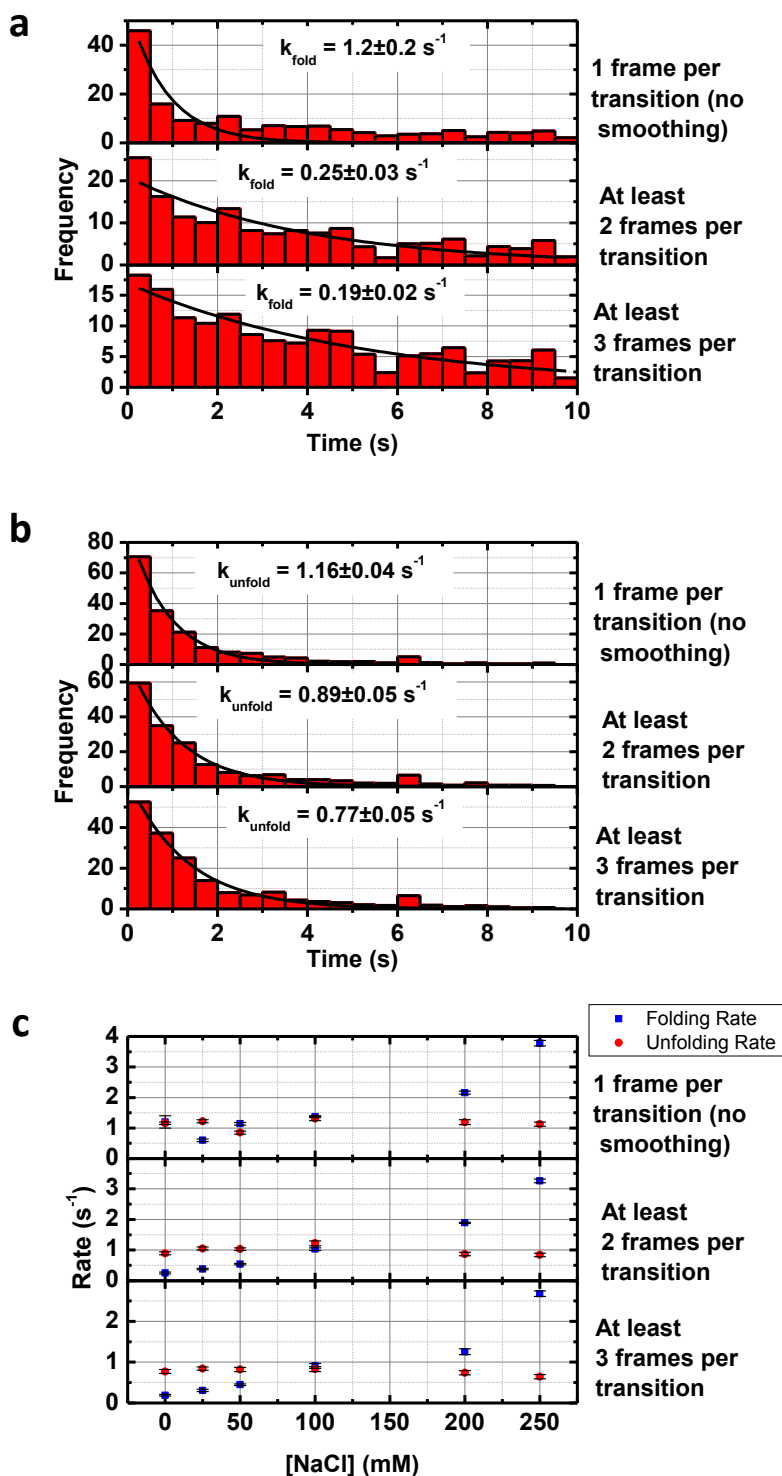
**Figure 4.4:** Diagrammatic illustration of the action of the smoothing program written for the HaMMY output files. 1. The program reads the duration, starting and finishing FRET values for a transition. 2. If the duration falls below the stated threshold (as it does in 1.) then, provided that they belong to the same molecule, it reads the FRET values of the previous transition and changes the FRET value of the fast transition to match. 3. The program then joins the original transitions and the smoothed fast transition into one single long transition.

The conditions chosen to be examined were 0 mM sodium chloride, primarily due to the fact that this condition clearly displayed a large spike due to these fast transitions. Initial analysis of the raw, unsmoothed data had shown that this peak hindered the accurate fitting of an exponential

decay curve, particularly to the high dwell time populations in the histogram, as is shown in Figure 4.5 below. The program was aiming to minimise this spike, and in doing so maximise the accuracy of the fit across the full range of the dwell times. Furthermore, the effect of smoothing on all of the folding and unfolding rates was examined in order to quantify the effect of smoothing the traces on the overall folding picture presented in section 4.2.

The effect of the fast transitions is most apparent for conditions where a long dwell time component is significant. It can clearly be seen that the quality of the exponential fit improves markedly with the smoothing of the traces to eliminate them, particularly for the higher times (Figure 4.5a and Table 4.1). The visual improvement is backed up by the mathematics of the fit, where the error in the folding rate is reduced by an order of magnitude between the raw, unsmoothed data and the establishment of a three-frame threshold.

Though it is most clear for the conditions that promote a longer dwell time, it can also be observed that the fitting is improved for shorter decay times, as can be seen in the unfolding rate dwell time histograms shown in Figure 4.5b. Again, it is clear to see that the fitting to the longest times of the histograms improves significantly with smoothing, as shown in both Figure 4.5b and statistically in Table 4.1.



**Figure 4.5:** a) Dwell time histograms for the low FRET state fitted with a mono-exponential decay to give the folding rate after setting the transition threshold to one, two and finally three frames. b) Dwell time histograms for the high FRET state, fitted with mono-exponential decay curves, after smoothing the traces to give a minimum of one, two and three frames per transition. c) Scatter plots showing the folding and unfolding rates as a function of sodium concentration. The error bars are calculated by OriginPro during the fitting process.

**Table 4.1:** Table summarising the distribution of the residuals from the fitting of the dwell time plots with different transition length thresholds for the high and low FRET states. The percentages falling within the ranges of  $\pm 1$  and  $\pm 2$  from the fitted curve show that there is an improvement where a limit is imposed on the number of frames per transition. This improvement is optimised when the limit is set to 3 frames per transition.

Min. transition length (frames)	Low FRET state		High FRET state	
	Residuals  < 1	Residuals  < 2	Residuals  < 1	Residuals  < 2
1	5%	5%	35%	90%
2	25%	60%	30%	65%
3	25%	70%	50%	80%

When this procedure was then performed for all conditions examined in section 4.2, it was observed that whilst the numerical values for the folding and unfolding rates change, the overall picture does not; the smoothing reaffirms that the folding rate increases with sodium concentration whereas the unfolding rate remains constant. The removal of fast transitions doesn't affect the longer dwell times as significantly unless there are many such events in a trace. However, with the initial filtering technique, traces are chosen to eliminate this as much as possible. As a result, the portion of the histogram beyond the first bin is largely unaffected and all that changes is the effectiveness of the fit across the full region, resulting in an increase in the calculated decay constant.

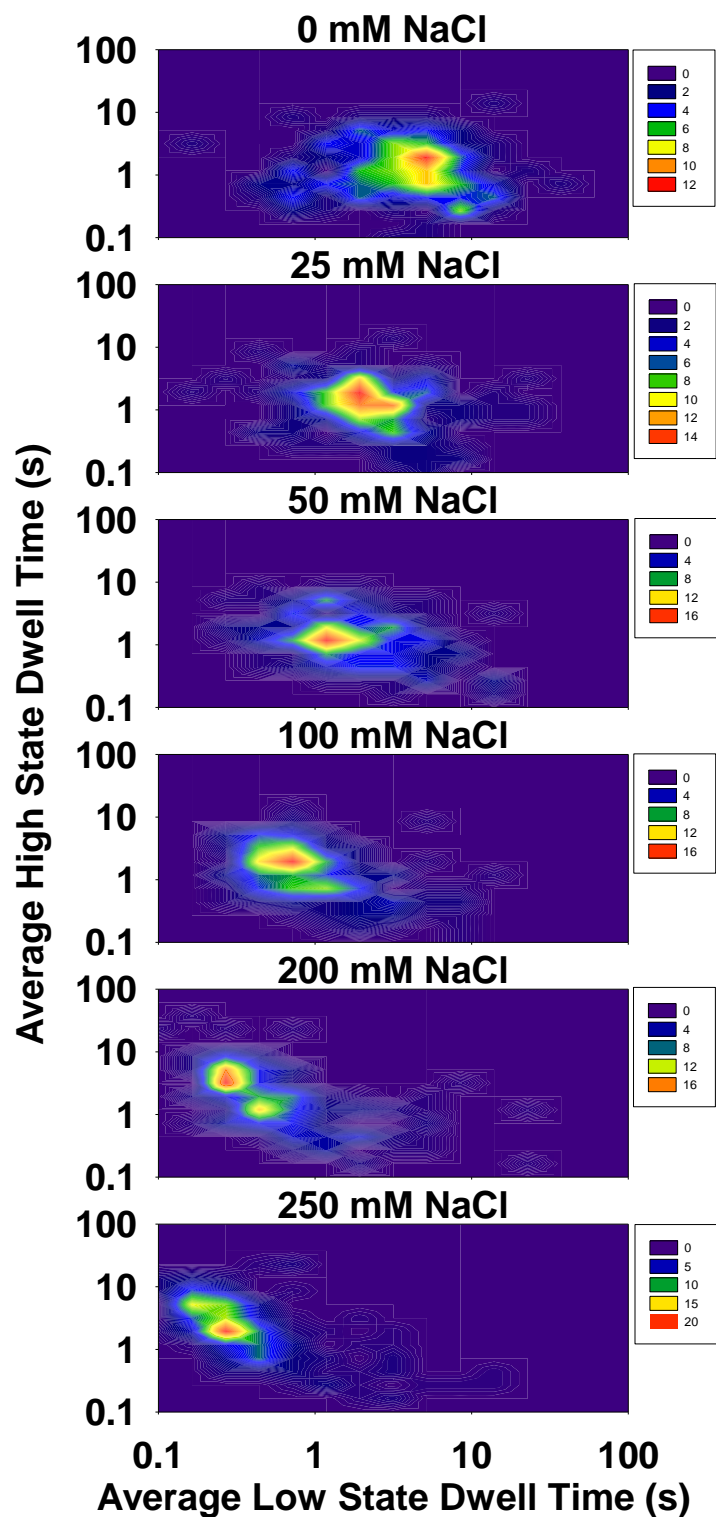
#### 4.4 Na<sup>+</sup>-induced folding dynamics of the adenine riboswitch aptamer

The kinetic behaviour of the adenine riboswitch was, as is normal in the single-molecule RNA folding field, analysed in two ways to extract all available information. First, the average folding and unfolding rates per

molecule are calculated and represented as a contour plot, where each point (Figure 4.6). This means of representation enables us to quantify the degree of dynamic heterogeneity across the sample. Secondly, I extracted the kinetic behaviour based on individual transitions, independent of from which molecule they come. For this, dwell times in each state are extracted and plotted as cumulative histograms as shown in Figures 4.7 and 4.8.

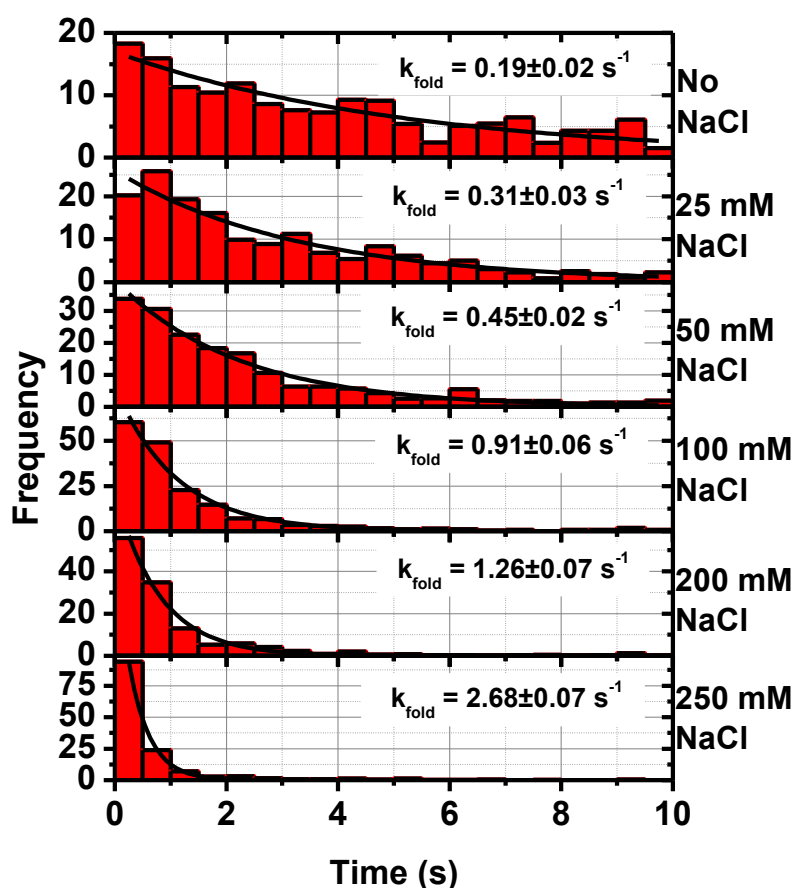
From the contour plot we observe that as the concentration of  $\text{Na}^+$  ions is increased from 0 to 250 mM, the adenine aptamer folds progressively faster, whereas the unfolding rate (dwell time on the folded, high FRET state) remains almost unchanged. Interestingly, as the concentration of ions is increased, the kinetic heterogeneity of the sample decreases substantially. At 0 mM  $\text{Na}^+$ , the average low-state dwell times for each molecule span the range of 0.2 s to 50 s, though the majority of the population falls within 0.5 – 20 s. When one increases the salt concentration, this span immediately shortens; at 25 mM  $\text{Na}^+$  ions this range is reduced to 1 – 10 s. At the highest  $\text{Na}^+$  ion concentration (250 mM) examined in this study, the unfolded state dwell times have reduced further, to within 0.2 – 2 s. On the other hand, the average high-state dwell times remained constant – every plot exhibited the molecular averages to span 0.1 – 20 s, with the majority in all conditions between 0.3 s and 10 s. This implies that although it can fold the riboswitch, it does not stabilise it entirely; rather the weak interaction characterised by  $\text{Na}^+$  induced structural collapse.





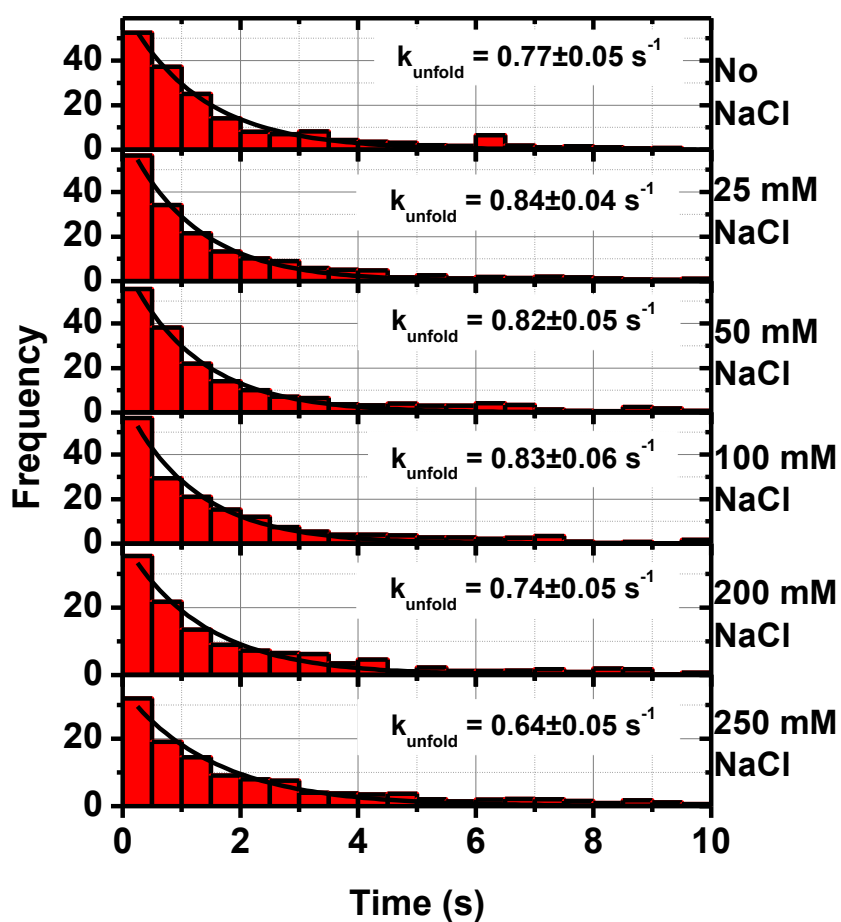
**Figure 4.6:** Contour plot showing the changes in the average dwell times for each individual molecule in the high FRET state versus the low FRET states for each sodium concentration examined. The plot highlights that whilst the average molecular dwell time for the low FRET state decreases with increasing sodium concentration, the distribution of times spent in the high FRET state remains unchanged.

Analysis of the single-molecule FRET trajectories provided two key observations. Firstly, the time that the riboswitch spends in the unfolded conformation decreases progressively with the increasing sodium concentration in the imaging buffer, with the unfolding rate increasing from  $0.19 \pm 0.02 \text{ s}^{-1}$  at 0 mM sodium to  $2.68 \pm 0.07 \text{ s}^{-1}$  at 250 mM sodium (Figure 4.7). This is in agreement with the steady state picture presented in section 4.2, as the riboswitch structure folds up entirely in the presence of a high sodium background.

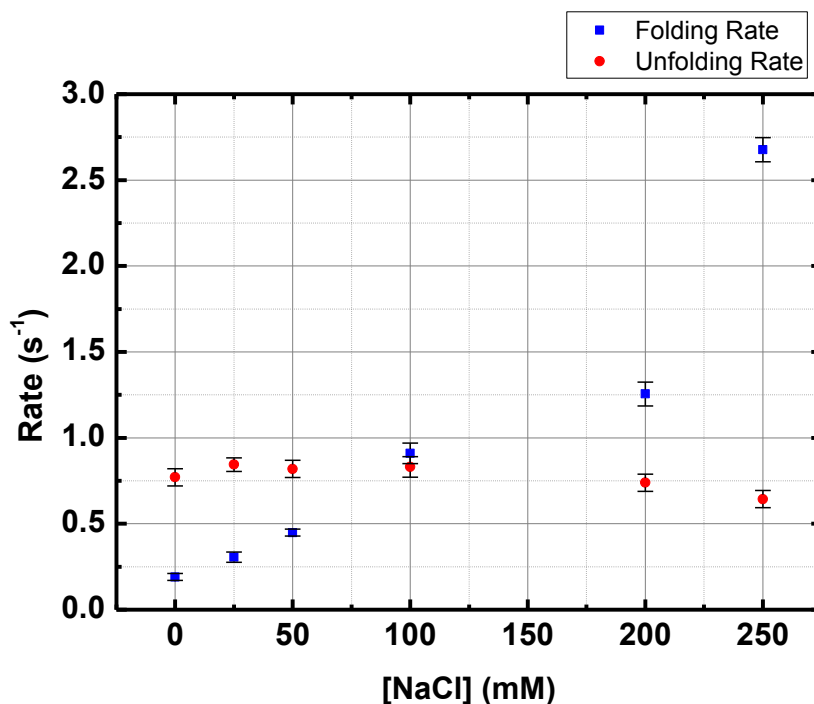


**Figure 4.7:** Single-molecule dwell time histograms of the low FRET state as a function of  $\text{Na}^+$  ions concentration. These histograms were fitted to a mono-exponential decay function to extract the folding rate value shown at each condition.

The second observation is that the unfolding rate is independent of the concentration of  $\text{Na}^+$  ions present in the imaging buffer. This phenomenon is observed in both the average dwell times per molecule as shown in Figure 4.6 and the single-molecule dwell time histograms presented in Figure 4.8.



**Figure 4.8.** Single-molecule dwell time histograms of the high FRET state as a function of  $\text{Na}^+$  ions concentration. These histograms were fitted to a mono-exponential decay function to extract the folding rate value shown at each condition.



**Figure 4.9:** Scatter plot showing the average folding rates for all molecules as a function of the concentration of sodium. These values are extracted from the histograms shown in Figures 4.7 and 4.8, and are presented with the error bars corresponding to the error of the fit

The unfolding rate had a value of  $0.77 \pm 0.05 \text{ s}^{-1}$  in the absence of sodium ions and a value of  $0.64 \pm 0.05 \text{ s}^{-1}$  at 250 mM concentration of  $\text{Na}^+$  ions. A scatter plot of summarising the folding and unfolding rates is shown in Figure 4.9. Based on the knowledge that monovalent ions screen the negative charges along the phosphate backbone, as explained earlier, the finding that increasing the concentration of  $\text{Na}^+$  accelerates folding is expected. Interestingly, the folding rate obtained in this work at 250 mM concentration of  $\text{Na}^+$  ions,  $2.68 \pm 0.07 \text{ s}^{-1}$ , is 4-5 fold faster than the folding rate obtained from previous studies at a saturating concentration of  $\text{Mg}^{2+}$  ions (4 mM) ( $0.71 \pm 0.07 \text{ s}^{-1}$ ) [128]. Also, in contrast to the observed independence of the unfolding rate with  $\text{Na}^+$  ions, the addition of  $\text{Mg}^{2+}$  ions is known to decrease the unfolding rate. A previous study reported a 2-fold

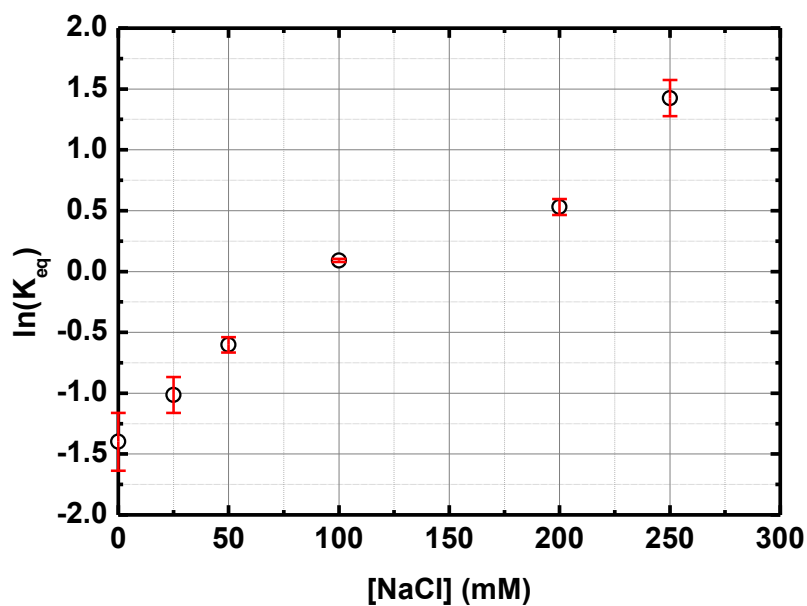
decrease in the unfolding rate from  $0.48 \text{ s}^{-1}$  at  $0.1 \text{ mM}$  to  $0.19 \text{ s}^{-1}$  at  $4 \text{ mM}$  concentration of  $\text{Mg}^{2+}$  ions [128]. From the values of  $k_{fold}$  and  $k_{unfold}$ , I calculated the values for the equilibrium constant ( $K_{eq}$ ) of the folding process as a function of  $\text{Na}^+$  ions using the expression below:

$$K_{eq} = \frac{k_{fold}}{k_{unfold}} \quad (4.1)$$

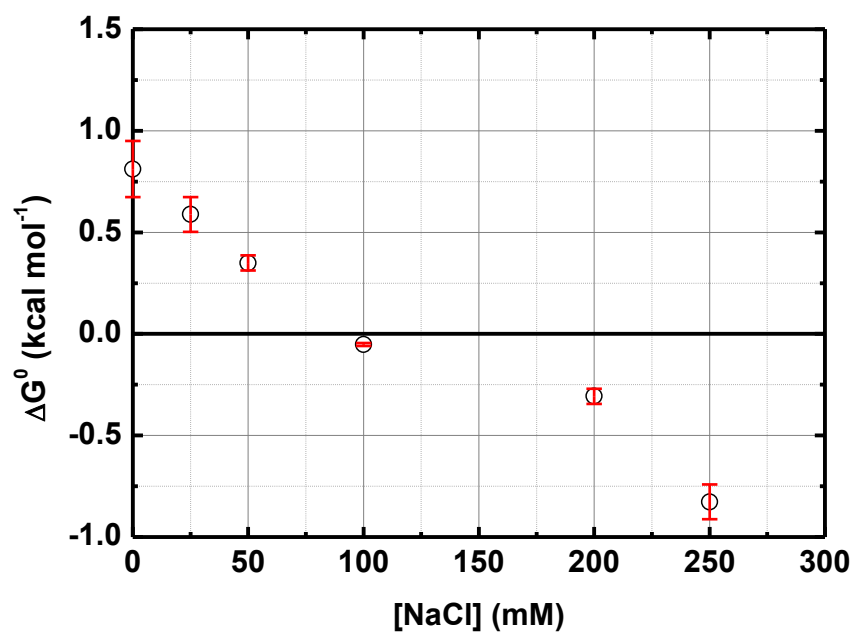
The calculated equilibrium constants expressed as Napierian logarithmic values are shown in Figure 4.10. It is clear that increasing the concentration of  $\text{Na}^+$  ions favours the formation of products (F state) over reactants (U state) as reflected in the increase of  $K_{eq}$ . I also derived the free energy ( $\Delta G^0$ ) in  $\text{kcal mol}^{-1}$  for the folding process using the expression below:

$$\Delta G^0 = -RT \ln K_{eq} \quad (4.2)$$

where R is the gas constant ( $1.98 \times 10^{-3} \text{ kcal mol}^{-1}$ ) and T is the temperature in K (293 K).



**Figure 4.10:** Scatter plot showing the variation of the equilibrium constant for the  $U \leftrightarrow F$  process with the concentration of  $\text{Na}^+$  ions. These values were extracted from the  $k_{fold}$  and  $k_{unfold}$  rates plotted in Figure 4.7 and Figure 4.8 using Equation 4.1.



**Figure 4.11:** Scatter plot showing the variation of free energy for the  $U \leftrightarrow F$  process as a function of the concentration of  $\text{Na}^+$  ions, calculated using Equation 4.2.

As the concentration of monovalent ions is increased, the folding process becomes more favourable, as reflected by  $\Delta G^0 < 0$  (Figure 4.11). For instance, at 50 mM  $\text{Na}^+$ , folding is slightly unfavourable ( $\Delta G^0 = 0.35 \pm 0.04$  kcal mol<sup>-1</sup>),

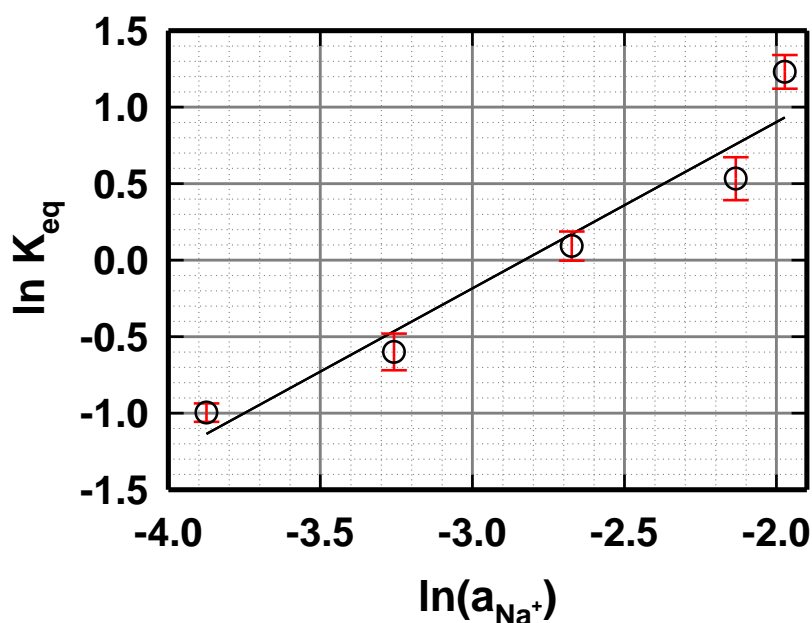
whereas, at 250 mM, folding becomes significantly favoured ( $\Delta G^0 = -0.83 \pm 0.09$  kcal mol<sup>-1</sup>). From this analysis, it is also clear that at near physiological concentrations of Na<sup>+</sup> ions (~100 mM), the formation of the folded tertiary structure results in no free energy change.

A model-free description of ion-RNA interactions has been formulated by Draper et al [175]. According to this model, ion-induced conformational changes in RNA can be described by the so-called single-ion interaction coefficients,  $\Gamma_+$  (excess number of cations/RNA phosphate) and  $\Gamma_-$  (the deficiency of anions/phosphate). Single-ion interaction coefficients are difficult to extract experimentally but changes in  $\Gamma_+$  or  $\Gamma_-$  induced during RNA folding can be derived from the expression:

$$\frac{\partial \ln K_{obs}}{\partial \ln a_{\pm}} = \Delta \Gamma_+ + \Delta \Gamma_- = 2\Delta \Gamma_{\pm} \quad (4.3)$$

where  $K_{obs}$  is the equilibrium constant for a two-state system and  $a_{\pm}$  is the mean ionic activity of the monovalent ions. The interaction coefficients describe how the negative charge on the RNA is screened by ions. The important principle is that any RNA transition to a more compact state with a higher charge density should increase  $\Gamma_+$  and  $\Gamma_-$  (become less negative). Because added Na<sup>+</sup> stabilises more compact (**F**) states of the adenine aptamer riboswitch as shown in Figures 4.2 and Figures 4.3, the term  $2\Delta \Gamma_{\pm}$  is expected to be positive and reflect the relative uptake of Na<sup>+</sup> ions during that transition. From this model, a plot of logarithm of the equilibrium

constant against the logarithm of the thermodynamic activity of  $\text{Na}^+$  ions should show a linear relationship and the slope,  $2\Delta\Gamma_{\pm}$ , corresponds to the net change in the number of ion pairs associated with the transition. As shown in Figure 4.12, this plot does indeed exhibit this linear dependence with a slope  $2\Delta\Gamma_{\pm}$  of  $1.1 \pm 0.2$ , suggesting that the formation of the tertiary structure of the adenine aptamer involves the net uptake of 1.1 ion pairs. For comparison, a  $2\Delta\Gamma_{\pm}$  value of  $2.4 \pm 1.0$  has been recently reported for the docking of the GAAA-tetraloop receptor structure induced by  $\text{Na}^+$  ions [176].



**Figure 4.12:** Scatter plot showing the linear dependence of  $\ln K_{\text{eq}}$  with the logarithm of the activity of  $\text{Na}^+$  ions. The slope of the plot corresponding to  $2\Delta\Gamma_{\pm}$  gives a value of  $1.1 \pm 0.2$ .



#### 4.5. Ligand-induced stabilisation of the folded adenine riboswitch complex using Na<sup>+</sup> ions as tertiary folding agents

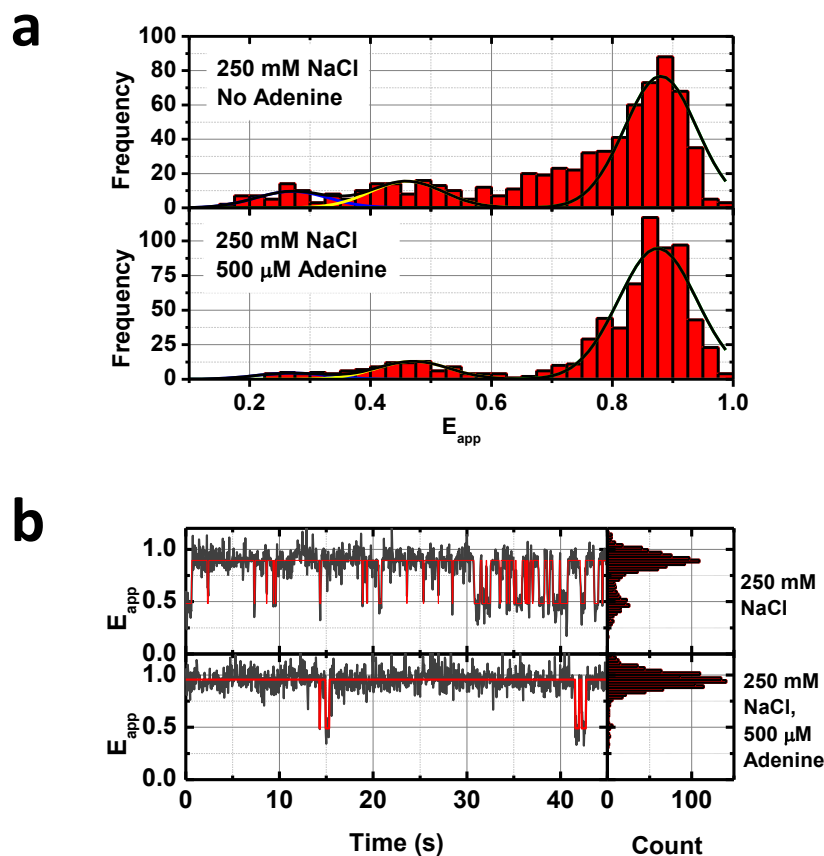
The previous sections demonstrated that monovalent sodium ions can efficiently promote the formation of the tertiary structure of the adenine aptamer and allow the loop-loop docking interaction between stems P2 and P3, though not with the stability provided by Mg<sup>2+</sup> ions. The binding site in a riboswitch, as discussed in Chapter 2, is highly selective, showing high affinity for one specific naturally occurring molecule. For the adenine aptamer, the crystal structure of the aptamer bound to the adenine ligand shows the presence of several Mg<sup>2+</sup>-bound ions. Thus, it is worth speculating whether the folded structure formed in the presence of only Na<sup>+</sup> ions is able to bind and accommodate the ligand efficiently.

In order to verify that a ligand-compatible folded structure was also formed in the presence of monovalent ions, a single-molecule FRET study was performed on the adenine riboswitch. A saturating concentration of adenine ligand (500 μM) was added in a background of 250 mM sodium, where the folded structure is favoured as was shown in Section 4.2 (Figure 4.2). Previous work done in the group and by our collaborators has shown for this model that the ligand, when bound, resulted in further stability in the folded state in the presence of magnesium, thus any similar change observed for sodium would be attributed to correct binding [128, 155]. The aim of this study is to compare the decay of the dwell times with

and without ligand added in order to verify whether the ligand-binding stabilises the folded structure formed in the presence of monovalent ions and how this compares to the stabilisation obtained in a background of  $\text{Mg}^{2+}$  ions. As observed previously with  $\text{Mg}^{2+}$  ions, depending on the ligand concentration, the equilibrium between ligand-bound and ligand-free populations should be reflected in two dwell times for the folded state (F), one short-lived dwell time corresponding to the ligand-free state ( $F_{\text{LF}}$ ) and one long-lived dwell time corresponding to the stabilised ligand-bound state ( $F_{\text{LB}}$ ). Under these conditions, the single-molecule histograms should exhibit a bi-exponential decay (Equation 4.4).

$$y = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} \quad (4.4)$$

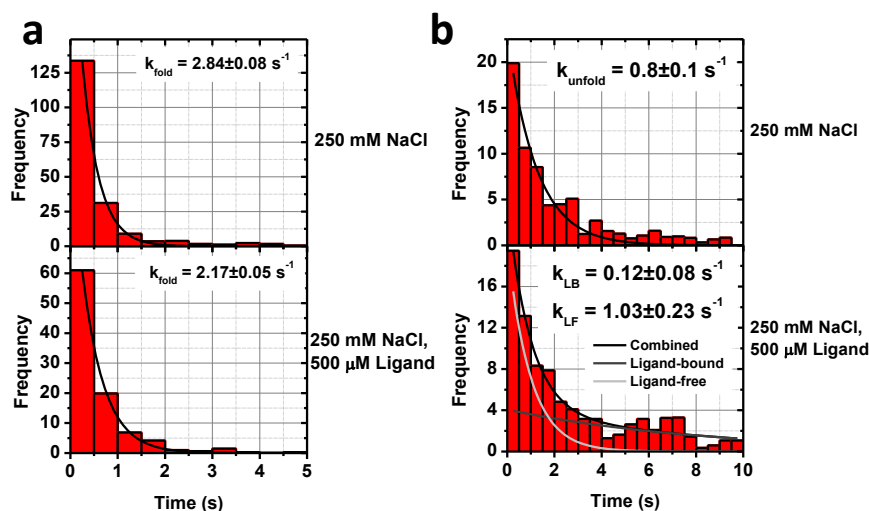
Thus, efficient ligand binding and population of the  $F_{\text{LB}}$  state should be reflected by a decrease in the unfolding rate upon the addition of adenine ligand. In Figures 4.13a and 4.13b, the single-molecule FRET histograms obtained at 250 mM without ligand added and with 500  $\mu\text{M}$  concentration of adenine ligand are compared. From these histograms, we observe that the addition of 500  $\mu\text{M}$  adenine ligand shifts further the equilibrium towards the folded state ( $E_{\text{app}} = 0.87 \pm 0.01$ ) with only a very small contribution from the intermediate state and a practically negligible contribution from the unfolded state.



**Figure 4.13:** a) Steady-state single-molecule FRET histograms for the adenine riboswitch at a constant sodium background of 250 mM without (upper panel) and at a saturating concentration of 500  $\mu$ M adenine ligand. b) Sample traces for each of the conditions presented with the corresponding single-trace histograms.

The influence of the addition of adenine ligand can also be observed by comparing the single-molecule FRET trajectories (Figure 4.13b). In the absence of adenine ligand, we observe fast fluctuations between the folded and unfolded state (Figure 4.13b). Upon addition of 500  $\mu$ M adenine ligand the frequency of fluctuations decreases considerably and only occasionally a transition between the two states is observed. A longer dwell time of the adenine aptamer in the presence of the adenine ligand confirms that the adenine aptamer can bind the ligand even in the absence of divalent metal ions. To quantify the degree of stabilisation induced by ligand binding, we

calculated the single-molecule dwell time histograms with and without added ligand and extracted the corresponding kinetic rates (Figures 4.14a and 4.14b).



**Figure 4.14:** a) Histograms showing the weighted distribution of the dwell times for the low FRET state for salt only (upper panel) and for salt and ligand, fitted with mono-exponential decays curves. The decay constants from the fitted parameters are indicated in each case. b) The dwell time histograms for the high FRET state, with the salt only decay (upper panel) and the ligand with salt (lower panel), fitted with mono- and bi-exponential decay curves respectively. The decay constants are shown for each condition, with the ligand docked and undocked components identified in the lower panel.

The plots in Figure 4.14a show that the folding rates in the absence and presence of adenine ligand are relatively similar. I obtained values of  $2.84 \pm 0.08 \text{ s}^{-1}$  with no ligand added and  $2.17 \pm 0.05 \text{ s}^{-1}$  for the decay in the presence of  $500 \mu\text{M}$  adenine ligand. This is in contrast to the observed 2-fold acceleration of the folding rate reported in the presence of  $0.5 \text{ mM Mg}^{2+}$  ions. These results seem to suggest that the observed acceleration of the folding rate in  $\text{Mg}^{2+}$  may be caused by the appropriate positioning of  $\text{Mg}^{2+}$  ions at specific locations within the aptamer structure. In the absence of

such site-specific bound  $\text{Mg}^{2+}$  ions, the initial interaction of the adenine ligand with the aptamer core may require conformational rearrangements that are slower without  $\text{Mg}^{2+}$  ions. Indeed, it is known through NMR [177] and biochemical probing that although the peripheral structure of the adenine aptamer (i.e. loop-loop interaction) is formed in the absence of ligand, ligand binding induces significant conformational changes within the aptamer core.

With regard to the unfolding process, the single-molecule analysis of the dwell time on the folded state showed that in the presence of adenine ligand the unfolding rate has a bi-exponential decay with values for each component of  $1.03 \pm 0.23 \text{ s}^{-1}$  and  $0.12 \pm 0.08 \text{ s}^{-1}$  and relative amplitudes given from the pre-exponential factors of  $20 \pm 2$  and  $4 \pm 2$  respectively (Figure 4.14b). When this is compared with the unfolding rate of  $0.8 \pm 0.1 \text{ s}^{-1}$  in the absence of the ligand, this confirms that the ligand-bound complex indeed forms in the presence of monovalent sodium ions alone, thus stabilising the folded structure.

It is interesting to note that in the ligand-free situation, the folded conformation is not locked into the native state and as a result, there is a fast interconversion between the folded structure and the unfolded structures, near our time-resolution limit. Because of this constant state of dynamism observed for some molecules, when the first ten frames are averaged out to add the apparent FRET contribution to the histogram there is often a transition within this time. As a result, there is a significant

number of molecules registering an apparent FRET value in between  $E_{\text{app}} = 0.5$  and  $E_{\text{app}} = 0.75$ . The ligand stabilises the high FRET state and one can see clearly that this averaged population is vastly reduced, leaving distinct Gaussian distributions for each of the three conformational states.

#### 4.6. Conclusions

In this chapter, I have focused on characterising the folding pathway of the adenine aptamer in monovalent ions and compare it with that previously reported in the presence of divalent metal ions. Although it was qualitatively known that  $\text{Na}^+$  can promote the formation of the native state from previous biochemical data, a detailed biophysical analysis of whether the aptamer undergoes similar transitions to the  $\text{Mg}^{2+}$ -induced pathway was lacking. Using an RNA construct with the FRET pair placed at identical positions as those used to investigate the influence of divalent metal ions, I have demonstrated that  $\text{Na}^+$  ions induce a similar folding pathway as that observed using  $\text{Mg}^{2+}$  ions. The observed FRET states, unfolded (U,  $E_{\text{app}} = 0.32 \pm 0.04$ ), intermediate state (I,  $E_{\text{app}} = 0.45 \pm 0.02$ ) and folded ligand-free ( $F_{\text{LF}}$ ,  $E_{\text{app}} = 0.87 \pm 0.01$ ) and ligand-bound ( $F_{\text{LB}}$ ,  $E_{\text{app}} = 0.88 \pm 0.01$ ) structures, are within the error, identical to those observed using divalent metal ions. As expected from previous studies on the folding of simpler RNA structures, a higher concentration of  $\text{Na}^+$  ions ( $\sim 250$  mM) is required to stabilise the native state compared to  $\text{Mg}^{2+}$  ions ( $\sim 4$  mM). I further showed

that the folded structure formed in a background of  $\text{Na}^+$  ions is able to bind the ligand, as reflected by decrease in the unfolding rate in the presence of adenine ligand compared to the aptamer alone, which is a consequence of the additional stabilisation of the aptamer structure provided by ligand binding.

A detailed analysis of the folding and unfolding kinetics of the ligand-free state revealed subtle differences with respect to that observed in the presence of  $\text{Mg}^{2+}$  ions. Whereas the folding rate is progressively accelerated as the concentration of  $\text{Na}^+$  ion is increased, as observed for  $\text{Mg}^{2+}$  ions, the unfolding remained independent on the concentration of  $\text{Na}^+$  ions. This is in marked contrast to the 2-fold decrease in the unfolding rate reported using divalent metal ions as folding agents. We speculate that the lack of bound  $\text{Mg}^{2+}$  ions stabilising the folded state might be the reason for the observed independence of the unfolding rate to  $\text{Na}^+$  ions. The fact that  $\text{Na}^+$  ions only influence the folding rate in combination with the use of unfolding agents such as urea will be exploited in the next chapter to overpopulate the transient intermediate state.

From the folding and unfolding rates we extracted the equilibrium constant and the corresponding free energy for the folding transition at each concentration of monovalent ions. As expected, we observed that folding becomes favourable as we increase the concentration of  $\text{Na}^+$  and interestingly at physiological concentrations of  $\text{Na}^+$  ions ( $\sim 100\text{-}150$  mM), the folding transition takes place with no free energy change. Assuming the

single-ion interaction coefficients formalism developed by the Draper lab, we demonstrate that the log-log plot of the  $K_{eq}$  is linear with the activity coefficient of  $\text{Na}^+$  ions and the slope ( $2\Delta\Gamma_{\pm}$ ) of  $1.1\pm 0.2$  reflects the net uptake of ion pairs associated to the folding transition in the absence of adenine ligand. Recently, using a melting assay, an uptake of  $3.8\pm 1$  ion pairs has been reported for the same folding transition in the presence of the adenine ligand using  $\text{Na}^+$  ions as folding agents [178]. It is known that the ligand-folded state is more compact than the ligand-free folded state. Indeed, a conformational rearrangement of the aptamer core has been proposed upon ligand binding based on NMR and biochemical data [177]. Thus, the variation between values of  $2\Delta\Gamma_{\pm}$  of  $\sim 1$  and  $3.8$  agrees with a more compact aptamer core structure and suggests that reaching the ligand-bound native state requires the additional uptake of additional ion pairs compared to the ligand-free state.



# Chapter 5: Manipulating RNA dynamics to isolate transient folding states

## 5.1. Introduction

### 5.1.1. Chemical Denaturation as a method to manipulate RNA dynamics

The majority of single-molecule studies on RNA to date have focussed on the folding pathway, i.e. the evolution from a primary structure through the formation of secondary motifs until the final tertiary contacts are made, giving the native RNA structure [179]. The collapse is mediated by metal ions, monovalent ions such as sodium bind non-specifically across the negatively-charged phosphate backbone or the RNA strand and divalent ions are taken in at precise locations to stabilise the structure in its tertiary configuration [122].

This is the usual method of probing the functional dynamics of most nucleic acid structures. Upon examining the analogous field of single-molecule protein dynamics, however, one finds that, rather than investigating the fully-folded protein, the native structure is forced back to its primary form [67, 180–182]. This ‘unfolding’ process of forcing the macromolecule to return to its initial primary conformation is known as denaturation, and may be achieved through mechanical means, such as using optical or magnetic tweezers, or thermally, through heating the

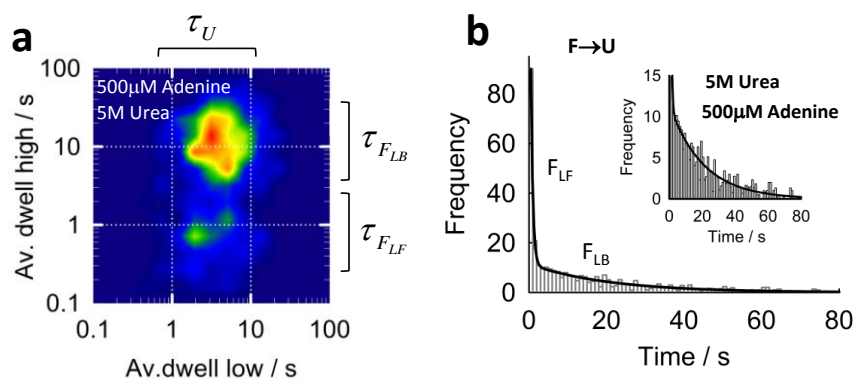
sample for a sustained period of time, or chemically through altering pH or by the addition of some denaturant [67, 180, 183].

Mechanical denaturation for RNA structures has been undertaken for many years, but for AFM and magnetic tweezers, only a single molecule can be investigated at a time, and additionally there are concerns over the hysteresis effects observed for refolding the structures after full denaturation [67, 184]. Chemical denaturation of RNA structures has been examined at the bulk level, but until recently had not been realised at the single-molecule level [66, 185]. The first detailed analysis on the influence of chemical denaturants on an RNA structure was reported by our group in 2013 [128]. Here, the folding and unfolding kinetics of the adenine riboswitch using  $Mg^{2+}$  and urea as folding and unfolding agents, respectively, was observed using single-molecule FRET and total internal reflection microscopy.

### **5.1.2. Previous single-molecule FRET studies on the adenine aptamer using single-molecule chemical denaturation**

The work by Dalgarno et al highlighted the advantages of combining the traditional ion-assisted folding approach with urea-induced unfolding to extract detailed information that otherwise will remain hidden when looking exclusively at the folding pathway. Importantly, it was found that at concentrations of up to 4 M urea, the effects of the chemical denaturant are strictly restricted to the unfolding of the tertiary structure (i.e., disruption of

the loop-loop interaction). No effect on the secondary structure was reported provided that the experiments were carried out in a background of at least 10 mM Na<sup>+</sup> ions. Moreover, it was proven that concentrations up to 5 M urea do not affect the immobilisation method based on the Neutravidin-biotin interaction and the Neutravidin remains in the native state at these concentrations of urea. In addition to this, the changes in the refractive index of the solution in the range of urea concentrations used in the study had no significant effect on the performance of the total-internal reflection fluorescence microscope. In this work, the influence of urea on the folding and unfolding rates in the presence of different concentrations of Mg<sup>2+</sup> ions was determined. For instance, analysis of the folding rates allowed the extraction of information about the nature of the rate limiting step for folding into the ligand-free and ligand-bound states, whereas analysis of the unfolding rates allowed quantification of the degree of stabilisation of the native folded state induced by ligand binding. In the absence of urea, quantifying the degree of stabilisation promoted by binding of the adenine ligand is a difficult task. This is due to the fact that at saturating ligand concentrations, the lifetime of the bound complex is too long (several minutes) and the dyes used in the sm-FRET studies usually photobleach before the ligand is released and the corresponding transition to the unfolded state can be observed.

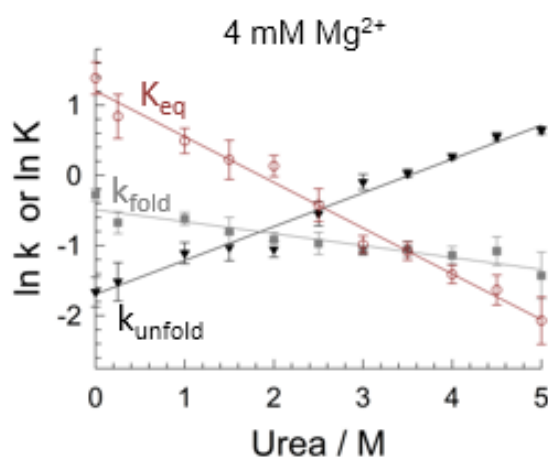


**Figure 5.1:** a) Two-dimensional contour plot showing the distribution of average (per molecule) folding (low-FRET $\rightarrow$ high-FRET) and unfolding (high-FRET $\rightarrow$ low-FRET) rates at 4 mM Mg<sup>2+</sup> in the presence of 5 M urea and 0.5 mM adenine ligand. b) Single-molecule dwell time histogram obtained for the unfolding process at the same conditions as described in a). The solid line represents the fit to a biexponential function with values of  $2.1\pm 0.1$  s<sup>-1</sup> (minor component), corresponding to the rate of urea-induced unfolding of ligand-free states, and  $0.045\pm 0.1$  s<sup>-1</sup> (major component) corresponding to the rate of urea-induced unfolding events taking place from the ligand-bound state. Both plots adapted from Dalgarno et al [128].

It was found that by using urea as a denaturing agent, it is possible to decrease the lifetime of the ligand-bound complex to a time-window compatible with sm-FRET. Comparing this with the effect of an equivalent concentration of urea on the ligand-free state, the relative degree of stabilisation could be easily extracted. To illustrate this, in Figure 5.1 taken from [128] we showed the single-molecule kinetic contour plots (Figure 5.1a) and dwell time histograms (Figure 5.1b) obtained at 5 M urea and 0.5 mM concentration of adenine ligand in a background of 4 mM Mg<sup>2+</sup>. A fast ( $2.1\pm 0.1$  s<sup>-1</sup>) and a slow ( $0.045\pm 0.1$  s<sup>-1</sup>) unfolding rate component were obtained and assigned to events taking place from ligand-free and ligand-bound states respectively. Thus, the difference between both unfolding rates constitutes a measure of the degree of additional stabilisation of the folded

state provided by ligand binding. A value of  $\sim 50$ -fold stabilisation was obtained. This value is close to that reported using optical force microscopy [186]. It is important to emphasise that single-molecule chemical denaturation provides a method to differentiate, based on their kinetic behaviour, conformational states that otherwise will be identical in FRET values (ligand-free and ligand-bound states exhibit the same high-FRET value of  $\sim 0.85$ ). In summary, the analysis of the unfolding rates provided a method to discriminate between ligand-free and ligand-bound states and to quantify the degree of RNA stabilisation resulting from ligand-binding.

The analysis of the folding rates also shed new insights into the behaviour of the adenine aptamer. In a background of  $4\text{ mM Mg}^{2+}$ , it was found that urea decreases the folding rate for ligand-free aptamers from  $0.71\text{ s}^{-1}$  with no urea to  $0.24\text{ s}^{-1}$  at  $5\text{ M}$  urea, and that this magnitude was independent of background concentration of  $\text{Mg}^{2+}$  ions (Figure 5.2).



**Figure 5.2:** Influence of urea on the folding ( $k_{\text{fold}}$ ), unfolding ( $k_{\text{unfold}}$ ) and equilibrium constant ( $K_{\text{eq}}$ ) of ligand-free aptamers at  $4\text{ mM}$  concentration of  $\text{Mg}^{2+}$  ions. Taken from [128].

In contrast, the folding rate in the presence of adenine ligand was also decelerated by the addition of urea, but the magnitude of this decrease showed a strong dependence with the background concentration of  $Mg^{2+}$  ions. This difference was taken as evidence for a different rate-limiting step for folding with and without ligand [128]. From the data, it was suggested that with no ligand, the rate-limiting step is dominated by the formation of initial tertiary contacts between the two loops on stems P2 and P3. In the presence of ligand, the interaction of the adenine ligand with a partially folded aptamer requires further stabilisation by specifically-bound metal ions. Thus, the rate-limiting step is determined by the ability to trap these metal ions.

This newly-developed protocol for chemical denaturation at the single-molecule level uses the competition between folding and unfolding agents to extract additional information. Because the adenine aptamer is known to fold through an intermediate step, we hypothesised that by fine-tuning the balance between concentration of folding agents ( $Na^+$ ,  $Mg^{2+}$ ) and that of unfolding agents such as urea, it may be possible to overpopulate the transient intermediate state. This manipulation of the aptamer dynamics to lock it into otherwise a very short-lived and low-populated but obligatory state could have a significant impact in our understanding of RNA function and folding and how they relate. For instance, being able to overpopulate and lock the aptamer in the intermediate state will allow the determination of its structure using not only single-molecule approaches but also

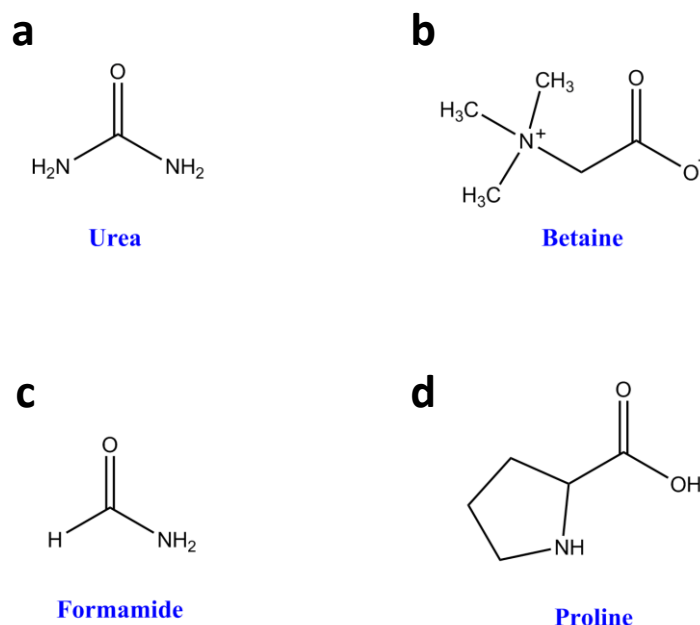
conventional ensemble averaging methods such as NMR and chemical footprinting. The short-lived nature of the transient intermediate makes current methods unsuitable and its structure can only be inferred indirectly. Even at the single-molecule level, the intermediate state is never the most populated state at any experimental condition investigated using  $\text{Mg}^{2+}$ ,  $\text{Na}^+$  or a combination of both ions. The conclusion drawn from the previous chapter, that monovalent metal ions promote folding but do not influence the unfolding rate, offers a perfect platform to build upon for this investigation. Here, by using a combination of  $\text{Na}^+$  ions that act exclusively on the folding rate and a denaturing agent acting on the unfolding step, in contrast to  $\text{Mg}^{2+}$  ions that act on both the folding and unfolding rates, we establish an extremely flexible parameter landscape upon which to manipulate RNA dynamics.

## **5.2. Comparative analysis of the relative unfolding efficiency of different chemical denaturants at the single-molecule level**

Urea as a chemical denaturant is one of a number of options to consider for a single-molecule chemical denaturation assay. For instance, in the protein unfolding field other molecular denaturants such as guanidinium chloride and formamide are commonly used instead of urea. Furthermore, there are a number of examples where the influence of osmolytes such as proline, formamide and betaine on RNA folding at ensemble level has been

determined, however to date their characterisation as unfolding agents at single-molecule level has not been reported [187, 188].

My first aim in this chapter is to explore the denaturation properties of different denaturants to determine their potential to unfold RNA and manipulate its structure in a controlled manner in combination with folding agents. The chemical structures of the four denaturants investigated (urea, proline, betaine and formamide) are shown in Figure 5.1 below and their physicochemical properties listed in Table 5.1.



**Figure 5.3:** Chemical structures of the chemical denaturants tested in this work: a) urea, b) betaine, c) formamide and d) proline.

The denaturant buffers used for each condition were prepared in 50 mM trizma hydrochloride (T50 buffer) with glucose, specifically by dissolving 0.157 g trizma hydrochloride and 3 g D-glucose (both Sigma-Aldrich) and the appropriate amounts of denaturant before adding ribonuclease-free water (ddH<sub>2</sub>O) to make a final volume of 50 ml. For each buffer, the mass or

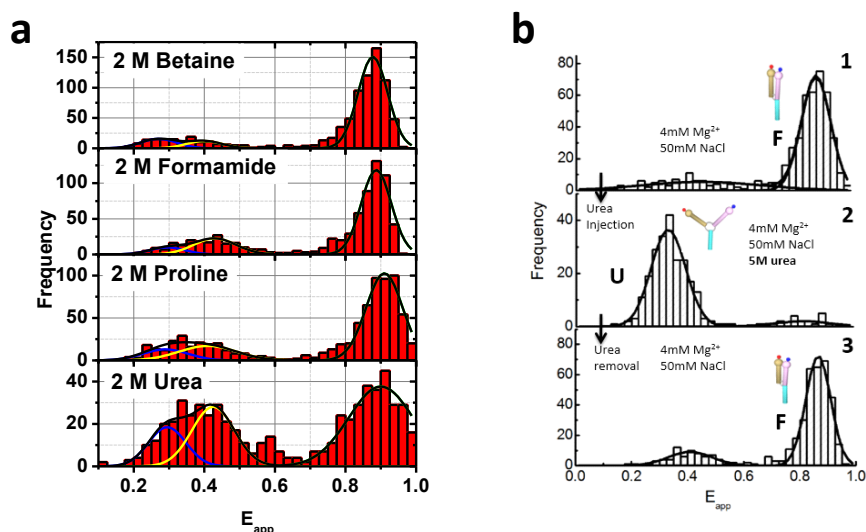


volume of denaturant and the corresponding stock concentrations are presented in Table 5.1 below. The pH of each buffer is then adjusted to 7.8 to match the cellular conditions for the riboswitch.

**Table 5.1:** A list of the denaturant composition of each of the buffers used in the experiment to determine their comparative strengths.

<b>Denaturant</b>	<b>g mol<sup>-1</sup></b>	<b>Mass/Volume add to buffer</b>	<b>Stock buffer concentration (M)</b>
<b>Betaine</b>	117.15	29.3 g	5 M
<b>Formamide</b>	45.04	10 ml	5 M
<b>Proline</b>	115.13	28.8 g	5 M
<b>Urea</b>	60.06	15 g	5 M

Preparing the denaturants in this way ensures that, for high experimental concentrations of denaturant, the glucose levels are maintained in the imaging buffer and, therefore, the effectiveness of the scavenging system is not compromised. Single-molecule FRET with wide-field total internal reflection fluorescence microscopy was used to give the most detailed picture possible of the distribution of the conformational states for each denaturant. The steady-state histograms obtained using 2M concentration of different denaturant agents in a background of 4 mM Mg<sup>2+</sup> ions are presented together with the corresponding traces and single-trace FRET histograms in Figure 5.4a.



**Figure 5.4:** a) Steady-state single-molecule FRET histograms showing the FRET populations obtained for the adenine aptamer at each set of experimental conditions. The solid lines represent fits to Gaussian functions describing the contribution of each FRET population. b) Single-molecule FRET histograms reported for the folded state (F) at 4 mM saturating concentration of  $Mg^{2+}$  ions (panel 1), unfolded state (U) after addition of 5 M urea (panel 2) and upon refolding the same molecules by removing the urea and injecting 4 mM concentration of  $Mg^{2+}$  ions (panel 3). Taken from Dalgarno et al [128].

One immediately notes from Figure 5.2a that at 2 M concentrations of chemical denaturant, the folded state ( $E_{app} = 0.90 \pm 0.01$ ) is the most populated state for all denaturing agents investigated. However, urea induces a much higher population of unfolded states (U + I) than any of the other agents. For comparison, the single-molecule FRET histograms obtained at a saturating 4 mM concentration of  $Mg^{2+}$  ions with (unfolded conformation) and without (folded state) 5 M urea are shown in Figure 5.4b [128]. Additionally, to prove that urea-induced denaturation is a reversible process and molecules within the same field of view can be folded and refolded as demonstrated by Dalgarno et al, a final histogram obtained after replacing urea by 4 mM  $Mg^{2+}$  ions alone is also shown [128]. In order to

quantify the difference in the relative population of the different conformational states as a function of denaturant agent, one would ordinarily for a two-state system look at the heights of the Gaussian peaks fitted to each population. However, since the unfolded population is a combination of the unfolded and intermediate states, in order to investigate the folded vs unfolded populations the areas under each portion of the histogram were used. These parameters were generated during the fitting of the populations in OriginPro. The results are listed in Table 5.2.

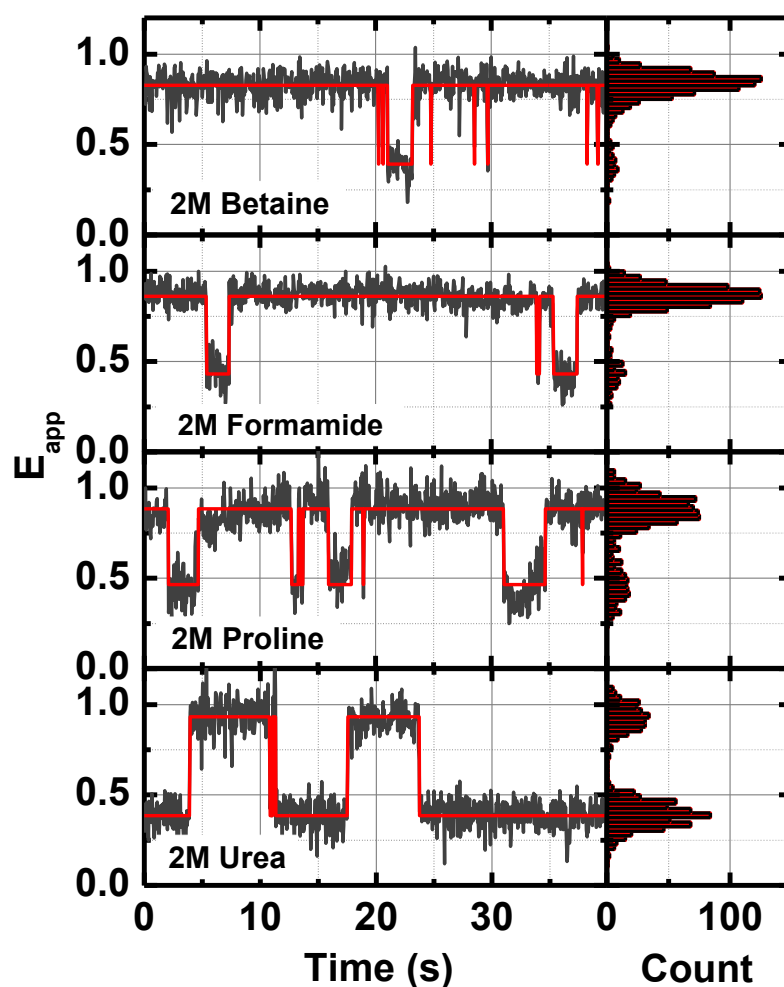
**Table 5.2:** Relative areas for each of the conformational populations obtained from the Gaussian fitting of the single-molecule steady-state histograms presented in Figure 5.4 for each of the denaturant agents investigated. The relative population of the unfolded states (U+I) is also given together with the average FRET value.

Denaturant	Area of unfolded population	Area of intermediate population	Area of folded population	% population unfolded	Average FRET value
Betaine	2.0±0.7	1.5±0.6	15.9±0.6	18±7	0.77±0.29
Formamide	1.2±0.7	3.4±0.6	12.6±0.5	27±8	0.74±0.22
Proline	2.1±0.7	3.5±0.9	13.3±0.5	30±9	0.74±0.23
Urea	2.4±0.4	4.2±0.4	8.6±0.6	43±6	0.64±0.10

It is clear from the final column of Table 5.2 that urea behaves significantly stronger as a denaturant than betaine, formamide or proline. The relative ability of the tested denaturant agents follows the trend: urea > proline > formamide > betaine.

Additional experimental confirmation of this aspect comes from the analysis of the single-molecule trajectories shown in Figure 5.5 obtained for the different denaturant agents at identical conditions. The trajectories obtained in the presence of formamide, betaine and proline are

characterised by long-lived folded states with occasional short-lived fluctuations to an unfolded conformation. In contrast, using urea as chemical denaturant induces much longer-lived transitions to the unfolded state lasting for several seconds before returning to the folded state.



**Figure 5.5:** Representative single-molecule FRET trajectories obtained at 2 M concentration of the indicated denaturant agents in a similar 4 mM background of  $Mg^{2+}$  ions. The solid red line represents the Hidden Markov modelling of the experimental trajectory. The corresponding single-molecule FRET histogram for each trace is also shown on the right.

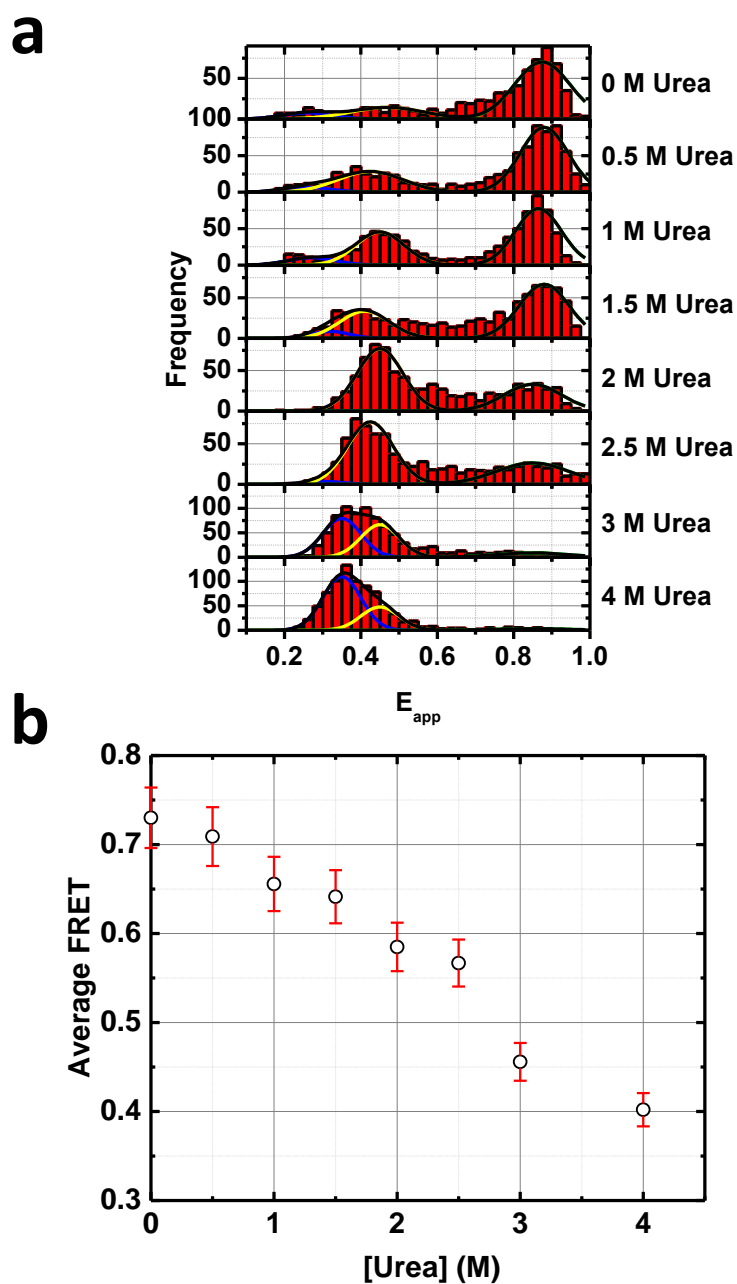
From these studies it is clear that urea acts more efficiently as denaturant agent than any of the other investigated chemical structures and therefore we have chosen urea to chemically manipulate the dynamics of the adenine aptamer with the aim of overpopulating the intermediate state. A detailed understanding of the molecular basis of this difference between the chemical denaturant is beyond the scope of this thesis, but it is currently under investigation by other members of the group. Here, the observed higher denaturation potential of urea compared to the other compounds offers several practical advantages when aiming to chemically manipulating the RNA dynamics at the single-molecule level. In requiring a lower concentration of urea denaturant to alter the population of the RNA sample, we avoid significantly altering the refraction index of the solution and subsequently compromising the critical angle condition required for total internal reflection.

### **5.3. Influence of urea in the monovalent ion-induced folding pathway of the adenine riboswitch**

Having determined that urea is the most efficient chemical denaturant of all the compounds tested, I next carried out a single-molecule FRET titration as function of urea concentration in a constant background of  $\text{Na}^+$  ions sufficient to ensure that all RNA molecules are initially (with no urea added) in the folded state. It was shown in the previous chapter that a of

250 mM of  $\text{Na}^+$  ions is enough to place the aptamer in the folded conformation. This titration was set up and carried out in the same manner as described in Chapter 4, Section 2. The urea buffer was prepared by dissolving 15 g urea (Sigma-Aldrich) in the T50 with glucose buffer described in Chapter 3, section 4, giving a final buffer concentration of 5 M urea. The competition between the denaturing effect of increasing concentrations of urea and the folding promoted by the sodium ions at this constant high concentration was then investigated to determine whether the transient intermediate state could be overpopulated.

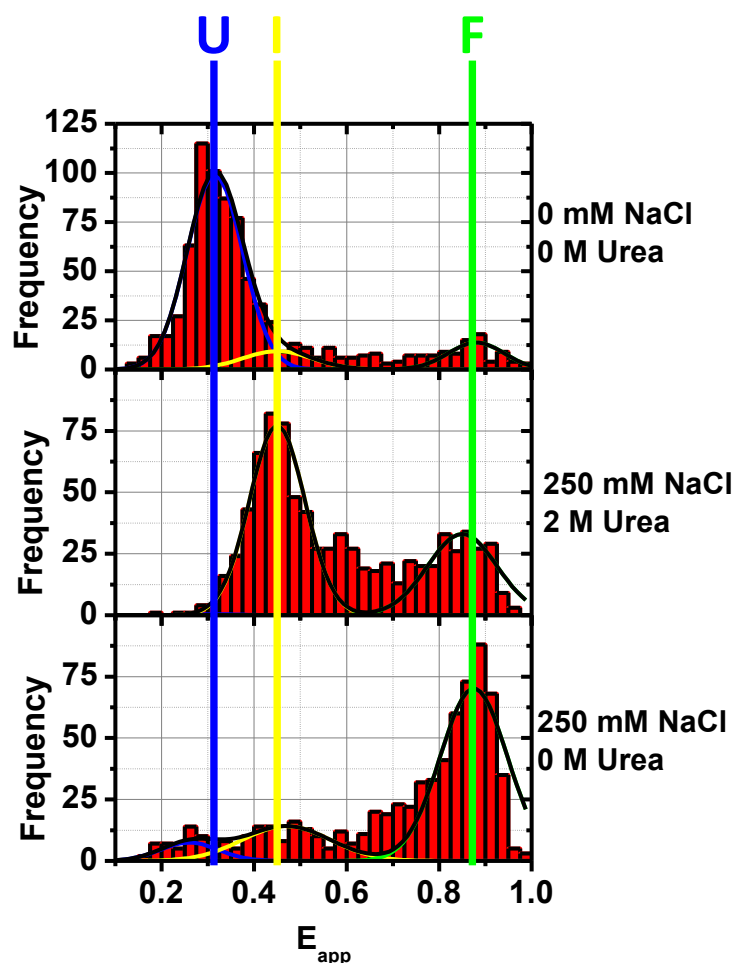
The steady state histograms shown in Figure 5.6 demonstrate the complete denaturation of the adenine aptamer from its folded conformation ( $E_{\text{app}} = 0.86 \pm 0.01$ ) to the unfolded state ( $E_{\text{app}} = 0.30 \pm 0.04$ ) at 4 M urea. As the concentration of urea is increased, the contribution of the folded state decreases and a concomitant increase in the population of the unfolded states is observed (U+I). At concentrations higher than 2 M urea, unfolded and intermediate states are the predominant species. This contrasts with the study by Dalgarno et al [128], where only at 3.5 M did the dominant state change from the folded conformation ( $E_{\text{app}} = 0.86 \pm 0.01$ ) to the unfolded conformations U and I ( $E_{\text{app}} \sim 0.25-0.55$ ), suggesting that in the presence of  $\text{Na}^+$  ions urea acts as a more efficient denaturing agent.



**Figure 5.6:** a) Single-molecule FRET histograms of the adenine aptamer obtained at the indicated concentrations of urea in background of 250 mM Na<sup>+</sup> ions. The solid lines represent the fit to Gaussian distributions for each FRET population: (black, folded; blue, unfolded and yellow corresponding to the intermediate. b) Scatter plot showing how the average FRET value (taken from each histogram in Figure 5.6a) changes as a function of the concentration of urea in the buffer. The plot shows a linear decrease in the average FRET value with increasing urea conditions.

From the analysis of the single-molecule FRET histograms shown in Figure 5.6a we can see that at intermediate concentrations of urea ranging from 2 to 2.5 M, the most populated conformation of the adenine aptamer corresponds to a FRET state with a Gaussian peak centred at  $E_{\text{app}}=0.44\pm 0.02$ . This corresponds to the typical value assigned to the intermediate folded state (I). This finding confirms that by balancing the concentration of folding and unfolding agents, it is indeed possible to overpopulate an otherwise low-populated state. A comparative plot summarising the single-molecule FRET histograms, with the corresponding dominant species and the experimental conditions at which they are favoured, is presented below in Figure 5.7. It is clear from this plot that the Gaussian centred at  $E_{\text{app}}=0.44\pm 0.02$  does not correspond to that of the unfolded state ( $E_{\text{app}}=0.30\pm 0.04$ ) obtained in the absence of mono- or divalent metal ions, or to the folded conformation obtained in the presence of saturating concentrations of monovalent ions.

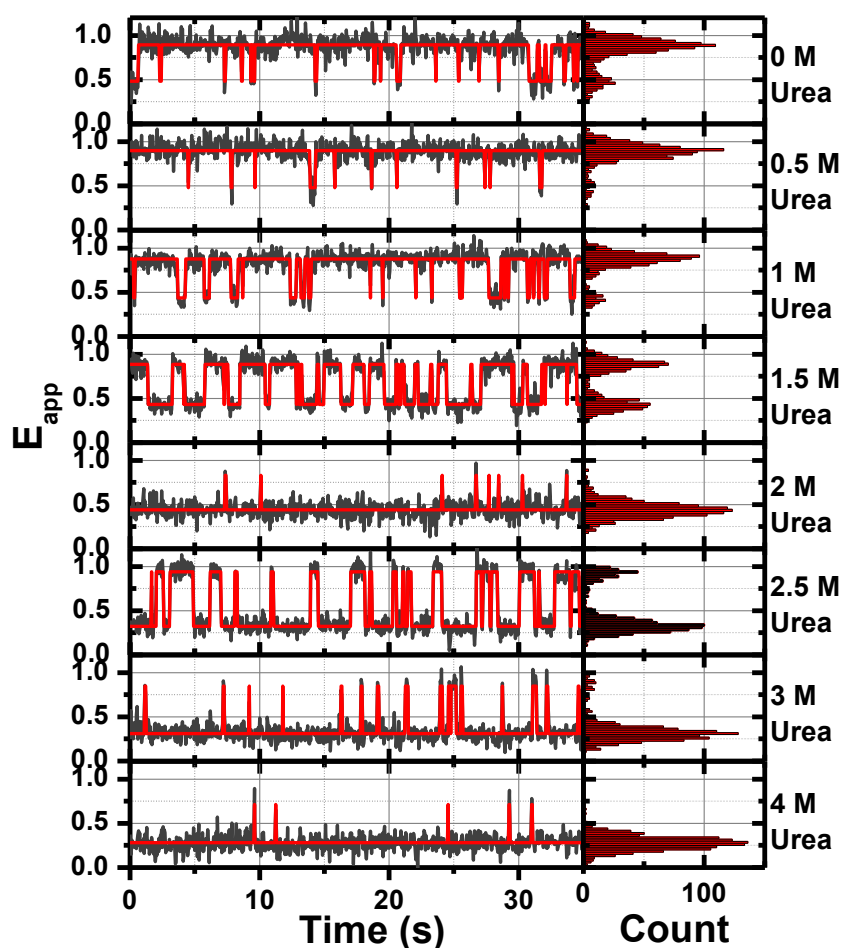




**Figure 5.7:** Comparison of the single-molecule FRET histograms obtained at experimental conditions that select one of the possible three conformational states of the adenine riboswitch; green (fully folded state, 250 mM Na<sup>+</sup> ions,  $E_{app} = 0.86 \pm 0.01$ ), yellow (intermediate state, 2 M urea and 250 mM Na<sup>+</sup> ions,  $E_{app} = 0.44 \pm 0.02$ ) and blue (unfolded state, no Mg<sup>2+</sup> ions and no Na<sup>+</sup> ions,  $E_{app} = 0.30 \pm 0.04$ ).

Additional evidence supporting this assignment and giving insights into the influence of urea modulating the conformational dynamics of the adenine aptamer are obtained from the analysis of the single-molecule FRET trajectories as a function of urea concentration in a constant 250 mM background of Na<sup>+</sup> ions. As shown in Figure 5.8, in the absence of urea, the adenine aptamer is mostly in the folded conformation with very occasional short-lived transitions to the unfolded conformations. As the concentration

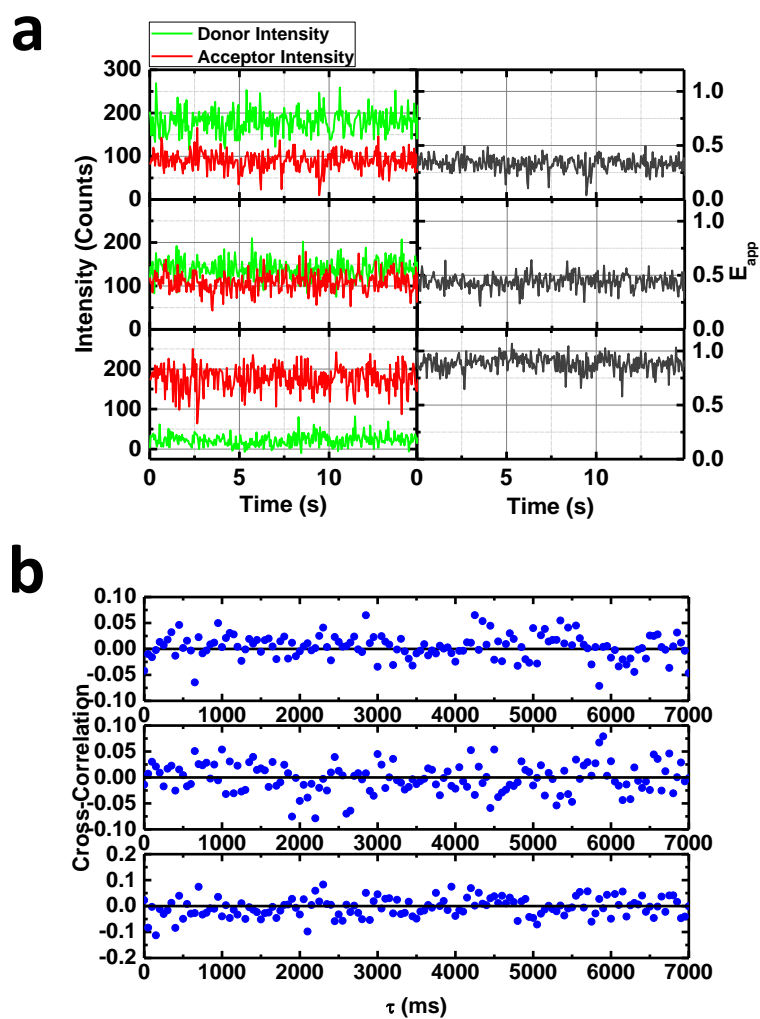
of urea is progressively increased, the frequency and lifetime of these fluctuations to the low-FRET state increases accordingly. In this background of  $\text{Na}^+$  ions, at concentrations between 1.5 M and 2 M, the aptamer fluctuates between folded and unfolded states exhibiting similar lifetimes. A further increase in the concentration of urea beyond 2 M shifts the adenine aptamer to the unfolded state with occasional and short-lived incursions into the folded conformation.



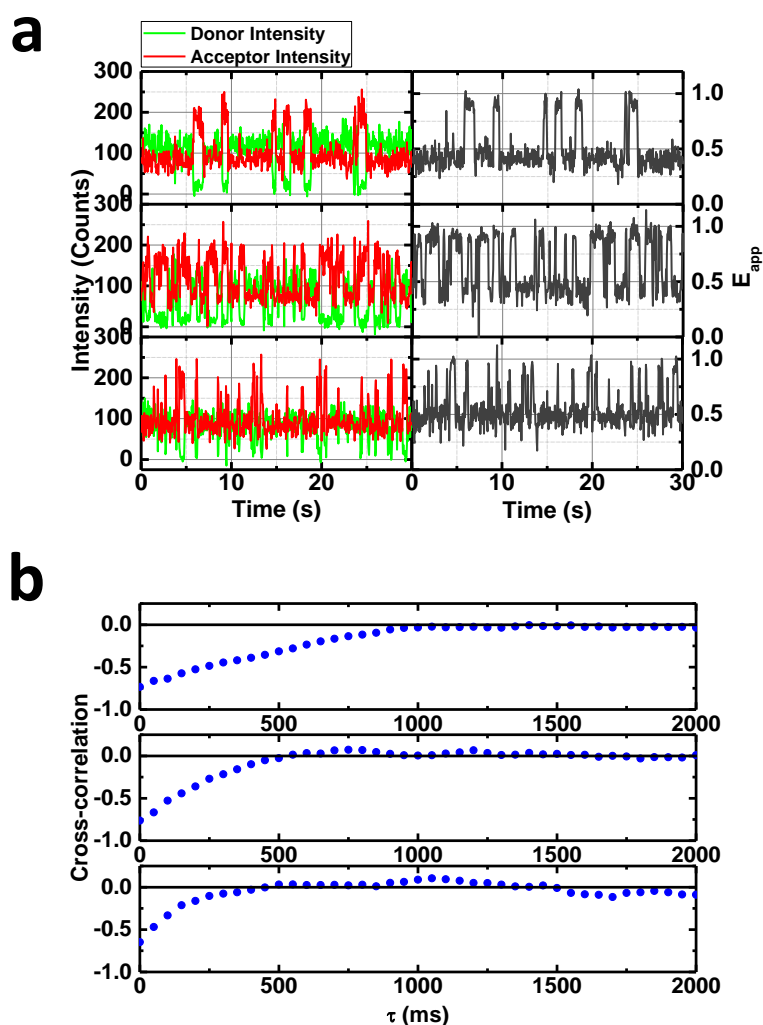
**Figure 5.8:** Representative single-molecule FRET trajectories and corresponding single-molecule trace histograms obtained for the adenine aptamer at the indicated concentrations of urea in a background of 250 mM  $\text{Na}^+$  ions. The solid lines represent the idealised FRET trace resulting from the Hidden Markov analysis of the experimental trajectory.

To rule out the possibility that the Gaussian centred at  $E_{\text{app}} = 0.44 \pm 0.02$  and assigned to the intermediate state contains contributions from fast interconverting unfolded (U) and intermediate (I) states, I carried out a cross-correlation analysis of the FRET trajectory for representative traces at 2M urea. The results are shown in Figure 5.9. Although the state could be a consequence of interconversion dynamics that are faster than our measureable time resolution, the absence of cross-correlation even at the highest acquisition rate of our CCD camera (16 ms) suggests that the state with  $E_{\text{app}}$  of 0.45 corresponds to a single state and not the averaging of two fast interconverting species.

As a means of validation of the cross-correlation analysis, the same procedure was carried out on three traces exhibiting clear dynamics with various kinetic properties (Figures 5.10a and 5.10b). The cross-correlation functions show this trend, and identify the general behaviour of the FRET traces in each case (Figure 5.10b). One may extract an approximation of the folded population lifetimes by fitting the plots with exponential curves, though for the purpose of this analysis a visual indication is sufficient.



**Figure 5.9:** a) Sample intensity (left panels) and  $E_{app}$  traces for each of the three conformational states of adenine: unfolded (upper row), intermediate (middle row) and folded state (bottom row). b) Cross-correlation plots for each of the three conditions shown in Figure 5.9a. The three plots show a complete lack of correlation for the regions plotted, indicative of a lack of dynamic behaviour and thus three distinct conformational states.

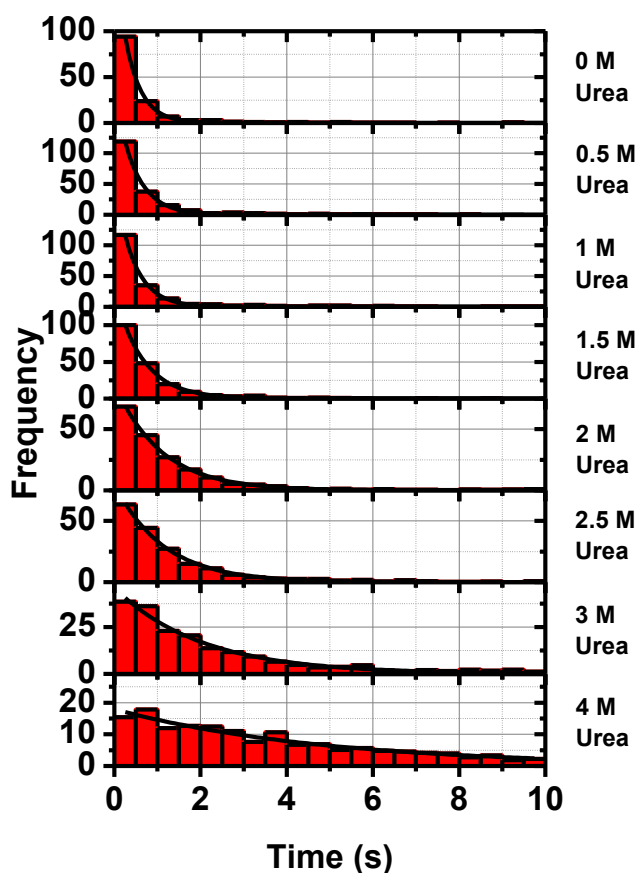


**Figure 5.10:** a) Plots showing representative traces exhibiting various dynamics, plotted in terms of donor and acceptor intensities (left column) and  $E_{app}$  (right column). b) Cross-correlation plots for each of the three sample traces, showing the variation in the cross-correlation decay with respect to the shift time fitted. These may be fitted with a mono-exponential curve but, as mentioned in the text, this is beyond the scope of this study.

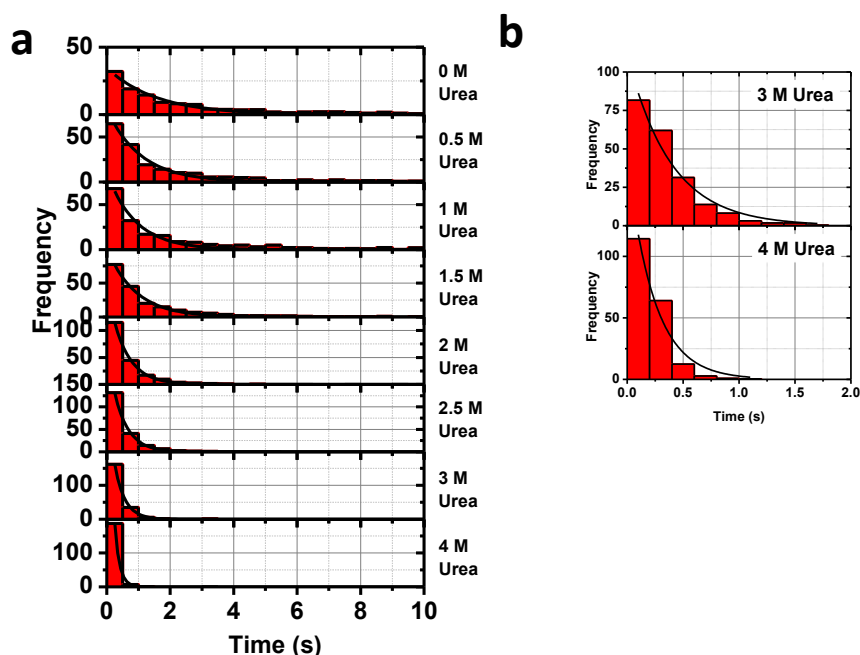
#### 5.4. Single-molecule dynamics of the adenine riboswitch under chemical denaturation

The changes in folding and unfolding rates of the adenine aptamer as a function of urea were examined in more detail in an identical manner to the

analysis performed in chapter 4 for the sodium-induced folding of the aptamer. The selected traces were smoothed to give a minimum of three frames per transition to eliminate the effects of fast transitions that could be photoblinking events falsely identified as conformational changes by the Hidden Markov algorithm. Single-molecule dwell time histograms fitted to mono-exponential decay curves and scatter plots showing the average dwell time in the high FRET state versus that of the low FRET state are shown in Figures 5.11 to 5.14.



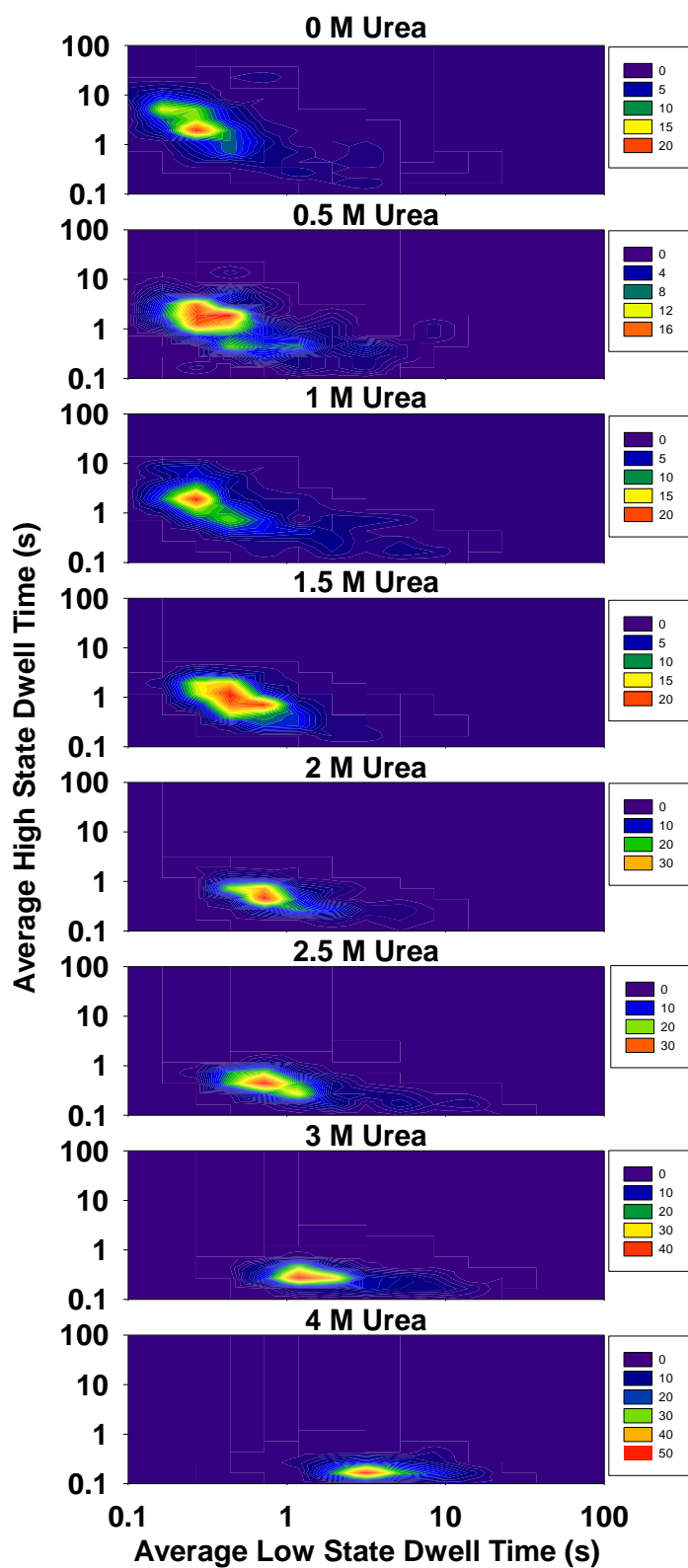
**Figure 5.11:** Single-molecule dwell time histograms for the folding process at the indicated concentrations of urea. The solid line represents the fit to a mono-exponential decay function. Corresponding values of the rates and associated errors from the fit are also shown.



**Figure 5.12:** a) Single-molecule dwell time histograms for the unfolding process at the indicated concentrations of urea. The solid line represents the fit to a mono-exponential decay function. Corresponding values of the rates and associated errors from the fit are shown in Table 5.3. b) Plots made at a higher bin rate and a lower time range to illustrate the behaviour of the unfolding rate at high urea concentrations remains mono-exponential in form.

**Table 5.3:** List of the folding, unfolding and equilibrium rates for the HMM-fitted traces analysed from the 250 mM Na<sup>+</sup> urea titration. The values for  $K_{eq}$  were calculated as described in the previous chapter. All errors were generated from the mono-exponential fitting parameters calculated by Origin.

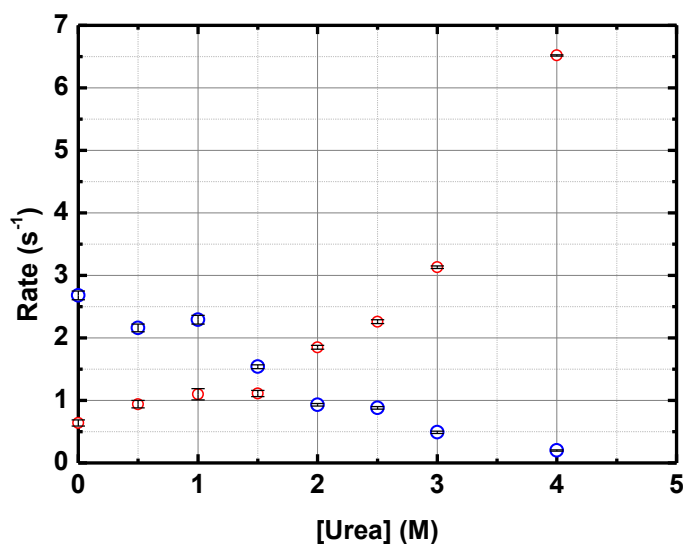
[Urea] (M)	$k_{fold}$ (s <sup>-1</sup> )	$k_{unfold}$ (s <sup>-1</sup> )	$K_{eq}$ (s <sup>-1</sup> )
0	2.68±0.07	0.64±0.05	4.2±0.4
0.5	2.16±0.06	0.94±0.06	2.3±0.2
1	2.29±0.07	1.10±0.09	2.1±0.2
1.5	1.54±0.03	1.11±0.05	1.4±0.1
2	0.93±0.02	1.85±0.03	0.50±0.02
2.5	0.88±0.02	2.26±0.03	0.39±0.01
3	0.49±0.02	3.13±0.02	0.157±0.008
4	0.20±0.01	6.52±0.01	0.031±0.002



**Figure 5.13:** Contour plots showing variation in the the average folding and unfolding rate per molecule at he indicated concentrations of urea. Each plot is composed from the average low and high FRET state dwell times for >150 molecules.

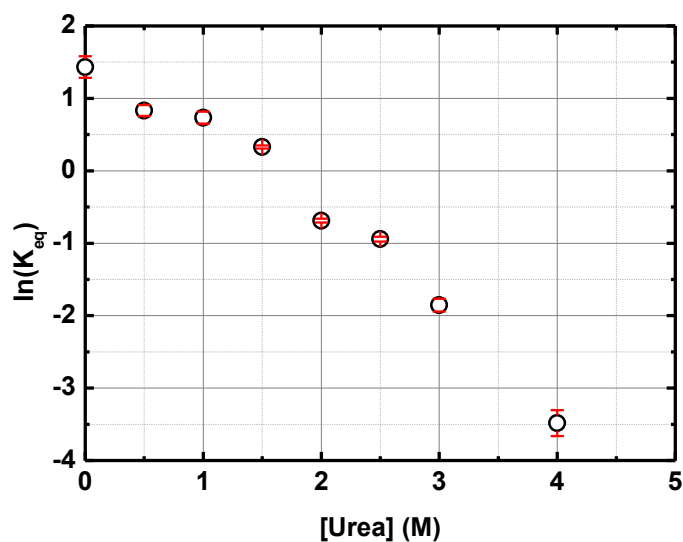


The dwell time plots shown in Figure 5.11 and Figure 5.12, together with the contour plots of the folding and unfolding rates as a function of urea concentration shown in Figure 5.13, indicate that the conformational dynamics of the adenine aptamer are strongly influenced by urea. As shown in Figure 5.14, and summarised in Table 5.3, for the folding rates we observed a  $\sim 14$ -fold decrease from  $2.68 \pm 0.07 \text{ s}^{-1}$  to  $0.20 \pm 0.01 \text{ s}^{-1}$ , whereas the unfolding rate showed a 10-fold increase from  $0.64 \pm 0.05 \text{ s}^{-1}$  to  $6.52 \pm 0.01 \text{ s}^{-1}$  with increasing urea concentration. This fits with the steady state and single trace histograms presented in Figure 5.6, where it can be seen clearly that the molecules spend longer time periods in the low FRET state as the urea background increases in concentration.

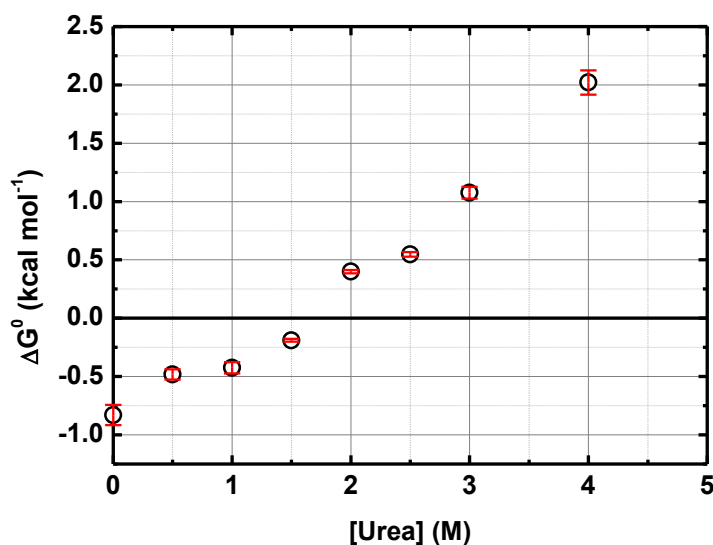


**Figure 5.14:** Folding (blue) and unfolding (red) rates obtained from single-molecule FRET for the adenine aptamer as a function of urea concentration in a background of 250 mM Na<sup>+</sup> ions. Errors are from the absolute uncertainties from the dwell time histograms.

The equilibrium rate constant ( $K_{eq}$ ) was calculated in the manner described in the previous chapter, and its value shown in Table 5.3. This column shows that the constant clearly declines, indicative of a reducing probability of a folding event to occur. This backs up the results presented in the steady state histograms in Figure 5.6a where the population shifts to the unfolded states with increasing urea concentrations in the imaging buffer. As before, calculation of the natural logarithm of this equilibrium rate constant allows the free energy  $\Delta G^0$  to be calculated. These quantities are plotted as a function of urea concentration in Figures 5.15 and 5.16.



**Figure 5.15:** Scatter plot showing the change in equilibrium constant for the folding process as a function of the urea concentration. The values of  $K_{eq}$  are calculated through Equation 4.1 from the previous chapter. Error bars are from the propagation of the uncertainties from the dwell time histograms.



**Figure 5.16:** Free-energy changes related to the urea concentration in the imaging buffer, as calculated by Equation 4.2. The zero line is shown to indicate the switching point for folding to be favourable ( $\Delta G^0 < 0$ ) and unfavourable ( $\Delta G^0 > 0$ ). Errors from the propagation of uncertainties in the equilibrium constant.

The change in the free energy exhibits a linear relation, as in the picture presented in the previous chapter for monovalent-induced folding. The plot illustrates that although initially a favoured process, aptamer folding becomes increasingly unfavourable as the urea background increases. This is in agreement with the observations drawn from Figures 5.6 and 5.11-5.15.

This trend also fits well with that observed for the urea-induced denaturation of the adenine aptamer in the presence of 4 mM magnesium and 50 mM sodium, where the folding rate changes from  $0.71 \pm 0.04 \text{ s}^{-1}$  at 0 M urea to  $0.24 \pm 0.1 \text{ s}^{-1}$  at 5 M urea, and the unfolding rate exhibits a strong ten-fold change from  $0.19 \pm 0.04 \text{ s}^{-1}$  with no urea to  $1.9 \pm 0.1 \text{ s}^{-1}$  at 5 M urea [128]. However, it is clear that urea has a stronger denaturation

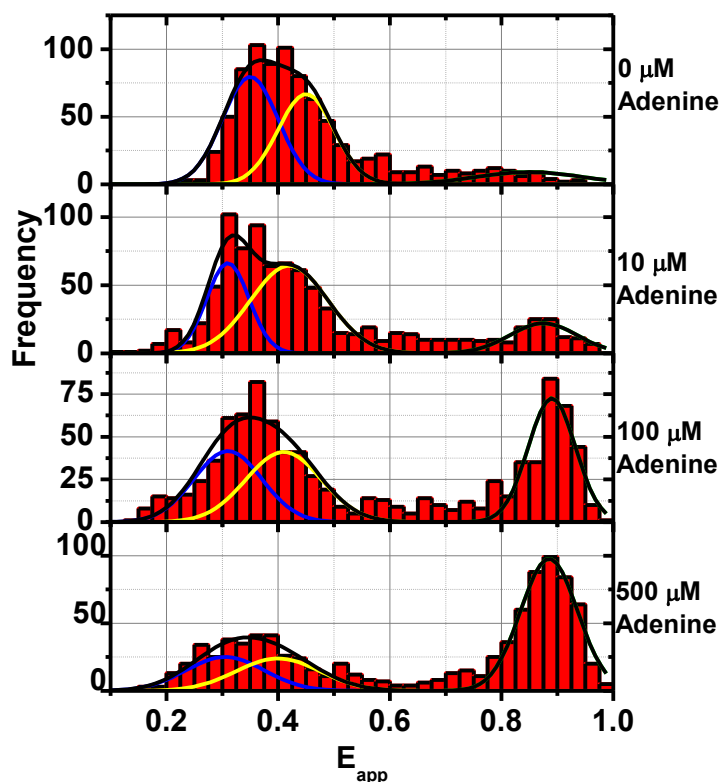
potential in the presence of monovalent ions compared to when  $\text{Mg}^{2+}$  ions are used as folding agents. Although the molecular basis of this difference needs further investigation, we interpreted this as evidence of the conformational stabilisation offered by the site-specific binding of the divalent magnesium ions in creating and stabilising the folded state of the aptamer. Where the urea has only to overcome the non-specific neutralisation of the phosphate backbone as opposed to this strongly bound divalent ion, the denaturing effects are observed at lower concentrations of urea. This can be seen clearly from the steady-state histogram plotted for 4 mM  $\text{Mg}^{2+}$  and 5 mM  $\text{Na}^+$  at 2 M urea, where the dominant conformational state is still the folded structure as opposed to the intermediate [128].

### **5.5. Investigating the ligand-induced stability of a chemically denatured adenine riboswitch**

The additional structural stability of the folded aptamer upon ligand binding in the presence of sodium ions has been investigated and shown in the previous chapter. The formation of the binding pocket can be induced by monovalent sodium ions alone, but lacks the structural stability gained by the capture of  $\text{Mg}^{2+}$  ions at specific positions. It was also shown in the previous section that the balancing of urea and  $\text{Na}^+$  ions overpopulates the transient intermediate conformational state of the adenine riboswitch.

In the case of a magnesium-stabilised folded aptamer, the work of Dalgarno et al showed that, although taking place with a lower affinity constant, the presence of urea does not prevent the binding of the adenine ligand [128]. Here, we wish to confirm whether the ligand can bind to a partially unfolded adenine aptamer significantly populated in the intermediate state due to the competition between  $\text{Na}^+$  ions and urea. For this, we carried out a ligand titration in a background of 3M urea and 250 mM  $\text{Na}^+$  ions. By using single-molecule FRET at chemical conditions which favour the intermediate and unfolded conformations almost equally, and increasing the concentration of the ligand in the buffer, the ligand binding dynamics can be further probed in order to obtain a more complete knowledge of the ligand recognition pathway for the adenine riboswitch.

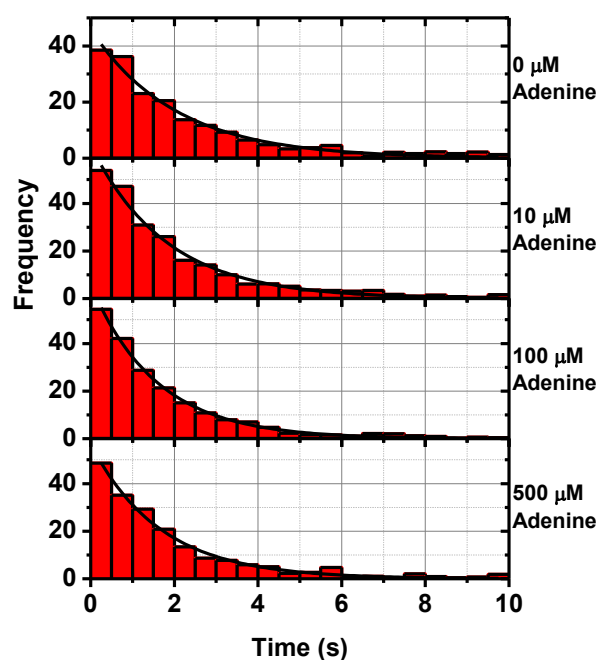
The ligand was prepared in solution by dissolving 6.8 mg adenine (Sigma-Aldrich) in 10 ml dd  $\text{H}_2\text{O}$ , which gives a 5 mM stock. From this stock, dilutions were prepared in T50 in order to maintain the background of trizma hydrochloride at 50 mM. As for all other sm-FRET experiments performed in this chapter, the imaging buffers used contained the scavenging reagents glucose oxidase, catalase, glucose  $\beta$ -mercaptoethanol to preserve the structural integrity of the sample and fluorescent labels. The steady-state histograms obtained at increasing concentrations of adenine ligand are presented in Figure 5.17.



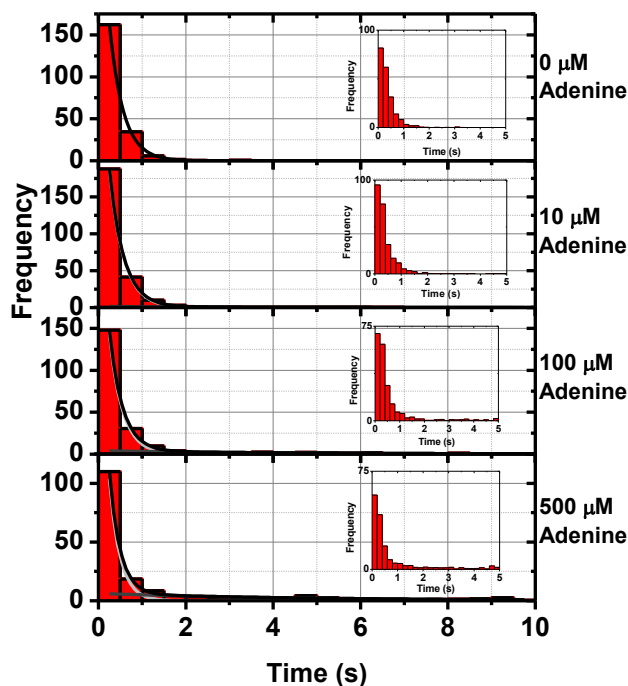
**Figure 5.17:** Single-molecule steady-state FRET histograms obtained for the adenine aptamer at increasing concentration of adenine ligand in a background of 250 mM Na<sup>+</sup> ions, 3 M urea. The solid lines represent the fitting to a Gaussian distribution for each state (blue: unfolded, yellow: intermediate and black: folded)

It is clear that as we increase the concentration of adenine ligand, the adenine aptamer progressively adopts the folded population ( $E_{app} = 0.86 \pm 0.01$ ). It is also interesting to note that the ratio between the relative population of the intermediate and unfolded states are independent of the concentration of ligand. We interpreted this as evidence that ligand binding does not affect the equilibrium between the unfolded and intermediate states. The stabilising effect of the adenine ligand can be easily observed by comparing the single-molecule FRET histogram obtained at 250 mM concentration of Na<sup>+</sup> ions and 3 M urea without ligand (Figure 5.6) to that in

the presence of 500  $\mu\text{M}$  adenine ligand (Figure 5.17), where the folded state is the predominant conformation. The dynamic behaviour was also studied; traces clearly showing single molecules with one donor and one acceptor label were saved and analysed using HaMMY [166], before the rates were extracted using MATLAB. The results of this further analysis are shown in Figures 5.18 and 5.19.



**Figure 5.18:** Single-molecule dwell time histograms for the folding of the adenine aptamer ( $\text{U}+\text{I}\rightarrow\text{F}$ ) obtained at the indicated concentrations of adenine ligand in a background of 3M urea and 250 mM concentration of  $\text{Na}^+$  ions. The solid black line represents the fit to a mono-exponential function. Fit parameters are listed in Table 5.4.



**Figure 5.19:** Single-molecule dwell time histograms for the unfolding of the adenine aptamer (F→U+I) obtained at the indicated concentrations of adenine ligand in a background of 3M urea and 250 mM concentration of Na<sup>+</sup> ions. The solid line represents the fit to a biexponential function. The values extracted from the fit at each concentration of ligand are listed in Table 5.4 below. The inset plots show the same data plotted at a finer binning ratio of 0.2 s/bin, instead of 0.5 s/bin, due to the fast decay rate of the ligand-free component.

**Table 5.4:** The folding and unfolding rates for the adenine riboswitch calculated under different ligand concentrations. The ligand bound (LB) and ligand free (LF) components of the unfolding rates are calculated for each relevant condition. Error analysis is described in Chapter 3, section 9.

[Adenine] ( $\mu\text{M}$ )	Unfolding Rate ( $\text{s}^{-1}$ )				$k_{\text{fold}}$ ( $\text{s}^{-1}$ )
	$k_{\text{LB}}$	% $k_{\text{LB}}$	$k_{\text{LF}}$	% $k_{\text{LF}}$	
0	-	-	$3.13 \pm 0.04$	100	$0.49 \pm 0.02$
10	$0.8 \pm 0.4$	$0.4 \pm 0.2$	$2.44 \pm 0.08$	$99 \pm 1$	$0.55 \pm 0.02$
100	$0.4 \pm 0.1$	$1.2 \pm 0.3$	$3.07 \pm 0.02$	$99 \pm 3$	$0.63 \pm 0.01$
500	$0.21 \pm 0.05$	$2.1 \pm 0.5$	$4.1 \pm 0.2$	$98 \pm 9$	$0.6 \pm 0.02$

The single-molecule kinetic analysis showed that the folding rate was, within the error, unaffected by the presence of the adenine ligand (Figure 5.18). The unfolding rate, however, does change. In the absence of ligand, the single-molecule dwell time histogram is well fitted using a



monoexponential decay function and we obtained a value of  $3.13 \pm 0.04 \text{ s}^{-1}$ . In the presence of adenine ligand, the single-molecule dwell time histograms required a double exponential function to fit the data. It must be emphasised that due to the low population of the folded state at low concentrations of adenine ligand, the rates extracted at these conditions have a much higher intrinsic error in their determination than those observed at concentrations of adenine ligand  $> 100 \text{ }\mu\text{M}$ , where the folded state is significantly populated. At these conditions where the folded state is substantially populated, the fast component decay displayed values close to that obtained in the absence of ligand and ranging from  $2.44 \pm 0.08 \text{ s}^{-1}$  to  $4.1 \pm 0.04 \text{ s}^{-1}$ . The amplitude of this lifetime component decreases with increasing the concentration of ligand; therefore we assigned this component to unfolding events taking place from a ligand-free state. The other component exhibited very similar values at  $100 \text{ }\mu\text{M}$  ( $0.22 \pm 0.06 \text{ s}^{-1}$ ) and  $500 \text{ }\mu\text{M}$  ( $0.21 \pm 0.05 \text{ s}^{-1}$ ) concentrations of adenine ligand. Because the relative amplitude increased with the addition of ligand (Table 5.4, columns 3 and 5), I assigned this component to unfolding events taking place from a ligand-bound state. Thus, the difference between the unfolding rate for the ligand-free and ligand-bound states represents the degree of stabilisation induced by ligand binding at these experimental conditions ( $\sim 14$ -fold).

## 5.6. Conclusions

In this chapter I have used the previous knowledge on the folding pathway of the adenine aptamer induced by monovalent ions to explore the combination of folding and unfolding agents as a method to manipulate the folding dynamics of the aptamer. Our main goal was to demonstrate for the first time that the competing interplay between both driving forces can be harnessed to place the aptamer structure in particular states on its conformational landscape and, in particular, to overpopulate obligatory intermediate states that are otherwise transient and poorly-populated.

In the absence of previous data at single-molecule level regarding the relative denaturant efficiency of different chemical compounds known to act as denaturant agents, we decided to first perform a comparative study on the unfolding efficiency of betaine, proline, formamide and urea. The latter has been previously used to unfold the adenine aptamer in a background of divalent metal ions, but no data for the others had been reported. Using a similar concentration of denaturant agents (2 M), our findings clearly indicate that urea is the most efficient unfolding agent for the adenine aptamer, achieving almost a 77% relative population of the unfolded states (U+I) compared to values of 22%, 37% and 42% for betaine, formamide and proline, respectively. Although the exact nature of the interactions between RNA and denaturant agents are not still entirely understood [61], it has been suggested that urea disrupts the network of hydrogen bonds between

the base pairs and, more recently, evidence for intercalation of urea between adjacent base pairs has also been reported [189]. In this context, the lack of a primary amino group in the case of formamide may explain the observed reduced ability to denature RNA. For betaine and proline, in addition to the lower ability to engage in hydrogen-bonding interactions, we hypothesise that the steric hindrance caused by the presence of multiple methyl groups in the case of betaine or a bulky pyrrolidine substituent in the case of proline may rule out the possibility of these compounds intercalating between the base pairs and disrupting base-stacking interactions. Although only qualitatively, these effects explained reasonably well the trend of denaturation efficiency observed for the tested compounds: urea>formamide>betaine and proline.

Having established that urea is the most efficient denaturant of all the compounds analysed, we next explored the influence of urea on the folding pathway induced by  $\text{Na}^+$  ions. The first aspect to emphasise is that the unfolding pathway induced by urea follows the same discrete FRET states (U, I and F) as observed in the folding pathway reported in the previous chapter. Thus, the urea-induced unfolding pathway is virtually a mirror image of the corresponding folding route in  $\text{Na}^+$  or  $\text{Mg}^{2+}$  ions in terms of structural states, and no evidence was found for urea promoting the formation of misfolded conformations. However, a comparison of the data obtained using a background of  $\text{Na}^+$  ions with those previously reported using  $\text{Mg}^{2+}$  ions, both in the absence of adenine ligand, reveals a

significantly stronger denaturation effect in  $\text{Na}^+$  than in  $\text{Mg}^{2+}$  for a similar concentration of urea. This difference is mostly due to a much more pronounced deceleration of the folding process in  $\text{Na}^+$  ions (14-fold) than in  $\text{Mg}^{2+}$  ions (2-fold) [11]. In contrast, the unfolding rate increases by a 10-fold factor independent of whether  $\text{Na}^+$  or  $\text{Mg}^{2+}$  ions are used as folding agents. The molecular basis of these findings requires further investigation. However, it seems plausible to suggest that urea can efficiently compete and displace diffusive monovalent ions loosely placed near the phosphate backbone, whereas divalent metal ions trapped at specific locations are more difficult to displace.

Finally, regarding the interaction of urea with a  $\text{Na}^+$ -stabilised RNA structure, our data confirms the feasibility of overpopulating an on-route intermediate state (I), simply by scanning and balancing the relative concentrations of folding and unfolding agents. This is a new finding only enabled by using a single-molecule FRET approach that allows the unambiguous identification of the relative contributions of the different conformers. The implications of this new method are far reaching. Being able to isolate intermediate states in an RNA folding landscape should allow its characterisation at the single-molecule level via FRET. Importantly, simply by transferring these single-molecule experimental conditions to ensemble averaging techniques such as NMR, it should be possible to probe in detail the structural architecture of these intermediate states.

## Chapter 6: From folding to catalysis: Studies of small RNA enzymes

### 6.1. Introduction to catalytic RNA structures

As mentioned in previous chapters, initially, the function of RNA sequences was thought to be limited to the transportation of genetic material between the genome and the ribosome. This role, although crucial, is only one of several fulfilled by various RNA structural motifs, from regulation of gene expression to encoding the information itself [50]. The first functional role discovered beyond transportation and genetic storage was in the catalysis of chemical reactions within the cell. This catalytic role was identified by Cech et al while investigating the removal of non-coding regions from mRNA strands [190] and Cech coined the term *ribozyme* to describe these structures. This novel role for RNA displaced proteins as the only biomolecules capable of catalysing chemical reactions[51]. Further functions beyond the removal of these non-coding regions, known as introns, include processing of transfer RNAs, gene regulation and controlling the formation of individual transcripts from a single-stranded viral genome [75, 76, 191].

The smallest of these catalytic structures are the self-cleaving small nucleolytic ribozymes [191]. Two of the most thoroughly-studied ribozymes

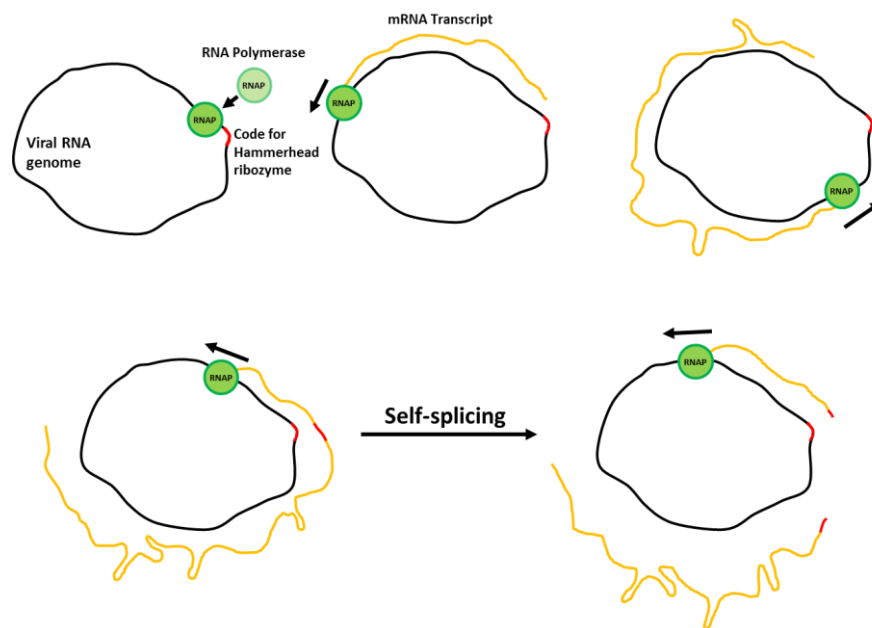
are the hairpin ribozyme and the hammerhead ribozyme, which are also the smallest members of the ribozyme family.

The hairpin ribozyme has been extensively studied through many experimental techniques, including studies detailing the folding dynamics and conformational changes at the single-molecule level using sm-FRET. The structure, although a simple two-domain system, was shown to travel through an intermediate state on its folding path, and the intramolecular interactions which mediate the folding pathway and stabilise the folded states thoroughly characterised [89, 90, 103]. A detailed analysis of the folding dynamics of the hairpin ribozyme was described in Chapter 2 and we refer the reader to that chapter and to excellent reviews by Tremblay and others [84, 93–95, 192]. The hairpin ribozyme has become the paradigm of single-molecule studies of ribozymes. In contrast, although widely studied using ensemble averaging techniques, very little is known at single-molecule level regarding the folding landscape of the hammerhead ribozyme (HHRz). To our knowledge, only one paper has briefly investigated the folded conformation [92]. This is striking if we consider the current technological uses of HHRz for medical purposes including cancer, viral infections, rheumatoid arthritis and cardiovascular diseases [193]. In this chapter, we aim to fill this gap by characterising the complete folding pathway of the minimal and natural forms of the HHRz using a single-molecule FRET approach.

## 6.2. The Hammerhead Ribozyme: a three-way RNA junction with catalytic function

### 6.2.1. Identification and action of the hammerhead ribozyme

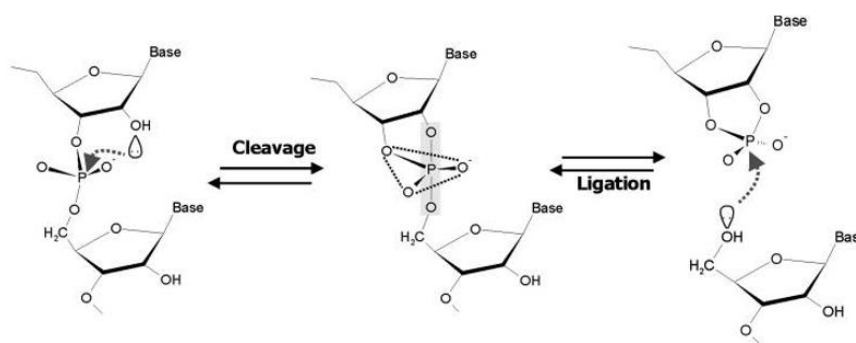
The smallest of the catalytic RNA motifs, the hammerhead ribozyme, was first identified in 1986, separately published by Prody and Hutchins [194, 195]. It was found to be present in the genome of small RNA structures called viroids, which are single-stranded pathogens of typically around 250-375 nucleotides in length [196]. The structure sequence was identified as the self-splicing element of the viroid transcript in the process of replication. The basic process for this role is summarised below in Figure 6.1.



**Figure 6.1:** The function of the hammerhead ribozyme in the replication cycle of an RNA viroid genome. The sequence, shown in red, cuts itself, dividing the otherwise infinitely-long transcript into individual copies of the RNA genome.

The cycle proceeds in the following way: the RNA polymerase (RNAP-II) binds to the viroid and starts to transcribe the genome, building the mRNA copy as it moves. It circulates the genome, passing the binding site and continues to circulate generating the transcript. The sequencing for the hammerhead ribozyme, marked in red in Figure 6.1, is generated and, once the full secondary and tertiary structures form, the ribozyme cleaves itself and separates a complete, single linear copy of the genome [191].

The mechanism that this cleavage process follows is shown in Figure 6.2 [197]. The reaction proceeds via an SN2 mechanism where the hydroxyl group at C2' attacks the 3'-5' phosphodiester bond leading to bipyramidal transition state common to the forward (cleavage) and backward (ligation) reaction where the newly cleaved linear copy of the genome is linked into a circle by joining the chain from both ends [50, 51, 76, 198].



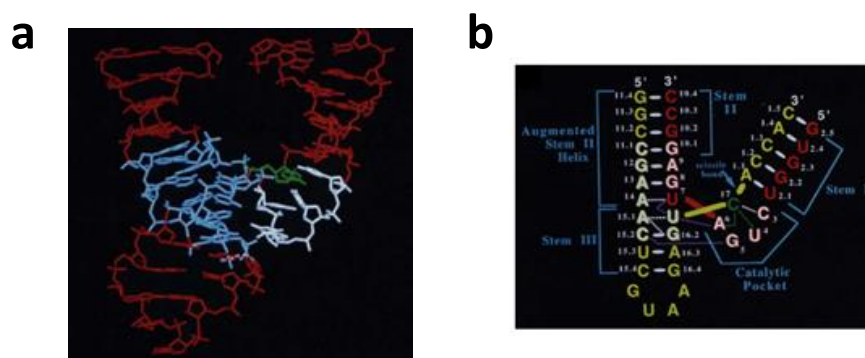
**Figure 6.2:** Schematic of the self-cleavage and ligation mechanism associated with the hammerhead ribozyme. Taken from [198].

### 6.2.2. Structural studies of the hammerhead ribozyme

The crystal structure of the minimal form of the hammerhead ribozyme (mHHRz) was resolved by Scott et al in 1995, following on from an



experiment by Pley et al which resolved the structure whilst bound to a DNA substrate [199, 200]. The ribozyme-only model showed the three arms of the complex, clearly resolved, in a stacked conformation. This crystal structure is presented in the first panel of Figure 6.3, along with the sequence of the ribozyme. With the structure identified, the structure became of great interest to biochemists and biophysicists as a base for investigation on the mechanism by which the mHHRz folds into the correct 3D structure to allow catalysis to take place. The mHHRz has been studied in great detail using spin-based studies, particularly electron paramagnetic resonance (EPR). The use of a paramagnetic divalent metal ion, manganese (II), in the place of the usual magnesium ions in the folding of the structure allowed a detailed study of the conformational changes invoked by site-specific divalent metal ion capture [201–205].

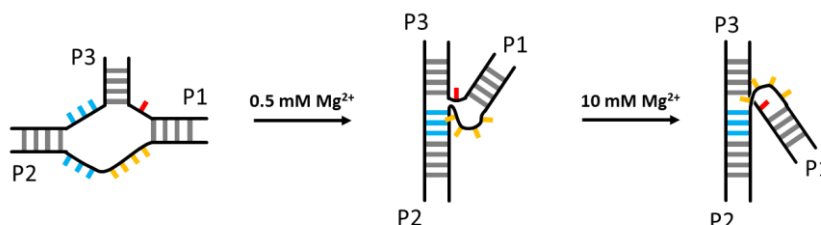


**Figure 6.3:** a) Crystal structure of the minimal hammerhead ribozyme, as resolved by Scott et al in 1995 [199]. PDB ID: 1MME. The cleavage-site nucleotide shown in green in a cleft created by a CGUA turn shown in white. b) Diagram of the backbone connectivity with the enzyme strand shown in red, the substrate strand in yellow and the cleaved cytosine in green. Taken from [199].

For structural studies, it is required to remove the possibility of self-cleavage so that one may work with an intact ribozyme for long periods of time. Because the cleavage mechanism, as stated above and shown in Figure 6.2, relies upon the hydroxyl group on the sugar moiety, by simply replacing the cytidine with the corresponding deoxyribonucleoside deoxycytidine variant, the 2'-hydroxyl is removed and catalysis is no longer possible [192]. An alternative also commonly used is to replace the 2'-hydroxyl by a 2'-methoxy group. Both substitutions avoid cleavage while preserving the ability of the ribozyme to fold into the native state. The crystal structure of the mHHRz indicates that the mHHRz adopts a shape like a wishbone with stems I and II forming the arms and stem III at the base of the ribozyme central core (Figure 6.3a and Figure 6.3b). All stems are A-helices but the central core involves non-canonical base-pairing interactions. The highly conserved CUGA between stems I and II forms a tight uridine turn (U-turn) and the CUGA sequence interacts strongly with the cleaved cytosine [199]. Interestingly, native crystals soaked in a solution of manganese or magnesium ions revealed the presence of up to five hydrated magnesium ions. Importantly, one of these hydrated magnesium ions is located near the catalytic pocket and it was shown later to be positioned closer to the phosphodiester bond to facilitate the in-line attack mechanism [199].

### 6.2.3. Folding pathway of the hammerhead ribozyme at the ensemble-FRET level

Fluorescence and FRET-based studies of the structure and conformational changes of the minimal form hammerhead ribozyme have been mostly performed by David Lilley's lab, building upon their earlier work using comparative gel electrophoresis [206, 207]. The FRET studies performed on the modified hammerhead ribozyme as a function of divalent metal ions showed three conformational states, with the final folded state corresponding to the catalytically-active structure [192]. The  $Mg^{2+}$ -induced folding pathway of the mHHRz shown in Figure 6.4 was deduced by analysing the conformational changes taking place between the three distance vectors involving the three stems.

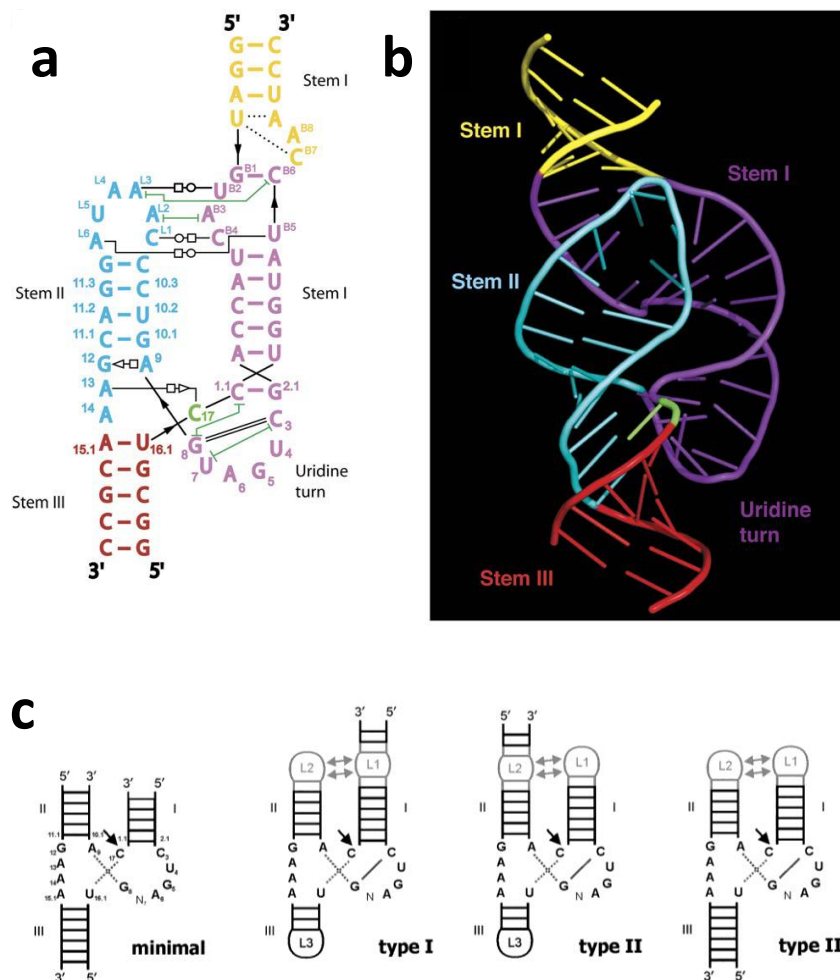


**Figure 6.4:** Two-stage folding mechanism proposed for the minimal form of the hammerhead ribozyme mediated by  $Mg^{2+}$  ions showing the conformational changes leading to the formation of the catalytic site. Adapted from [208].

According to the ensemble FRET model (Figure 6.4), in the absence of metal ions, the mHHRz exists in an extended conformation and folding by the addition of  $Mg^{2+}$  ions takes place following a two-stage process. The first step takes place at concentrations of  $Mg^{2+}$  ions around 0.1-0.5 mM and involves the helical stacking of stems P2 and P3. The U-turn and the

catalytic pocket form in the second stage, which requires a 20-fold higher concentration of  $Mg^{2+}$  ions. It was shown that the catalytic activity of the ribozyme as a function of  $Mg^{2+}$  ions follows the same dependence as the folding with half-maximum activity at around 5 mM  $Mg^{2+}$  ions and no detectable cleavage below 1 mM. However, the concentration of magnesium required to fully fold the structure obtained from these studies is much higher (>10 mM) than the physiological concentration of magnesium in a cell, which is typically assumed to be around 2 mM [163].

The solution to this discrepancy came from the analysis of the natural sequence of the hammerhead ribozyme and finding that by preserving the terminal loops at the end of stems P1 and P2, the natural hammerhead ribozyme (nHHRz) was able to cleave efficiently at concentrations of  $Mg^{2+}$  ions 2 orders of magnitude lower than the mHHRz. De la Pena et al, using the nHHRz sequence from *Schistosoma mansoni*, suggested that this marked reduction in the concentration of  $Mg^{2+}$  ions was due to a crucial interaction between both terminal loops that was missing in the minimal ribozyme [209]. This hypothesis was later confirmed by Khvorova et al [210], and by Penedo et al using ensemble FRET [172]. The crystal structure of the natural form of the hammerhead ribozyme was resolved two years later [211], after the FRET study had confirmed the presence of a tight interaction between the peripheral loops, as shown in Figure 6.5.



**Figure 6.5:** a) Secondary structure and connectivity of the natural hammerhead from *Schistosoma mansoni*, colour-coded to complement the crystal structure (b). c) Comparison of the minimal and natural forms of the hammerhead ribozyme. Natural forms occur in three types, named after the open stem. Interactions between the L1 and L2 loops take place in all three natural forms. An additional base pair is formed between the C3 and G8 nucleotides. a) and b) taken from [211], (c) taken from [198].

#### 6.2.4. Single-molecule FRET studies of the hammerhead ribozyme

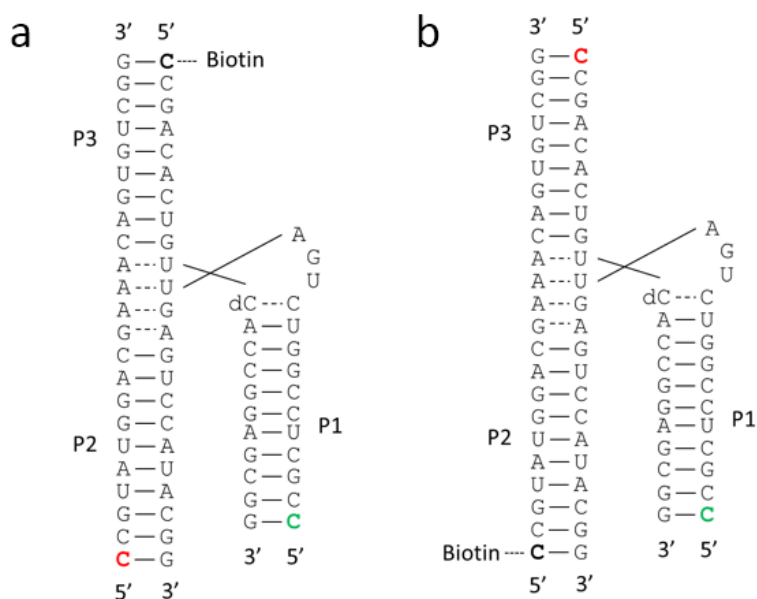
Although extensively studied at the ensemble level, there haven't been any detailed single-molecule studies of the folding dynamics of the hammerhead ribozyme. The only study to date focussed on the effects of mutational substitutions on the folding pathway itself, not on the effects of different metal ion concentrations on the natural structure itself [92]. In

order to shed light on the folding mechanism of the hammerhead ribozyme, a single-molecule FRET titration was performed in order to fully characterise at single-molecule level the behaviour of the structure under increasing concentrations of divalent metal ions. Furthermore, the differences between the natural and minimal derivatives were studied in order to quantify the kinetic behaviour of the folding pathway of both structures.

### **6.3. Results and discussion**

#### **6.3.1 Single-molecule folding dynamics of the minimal hammerhead ribozyme**

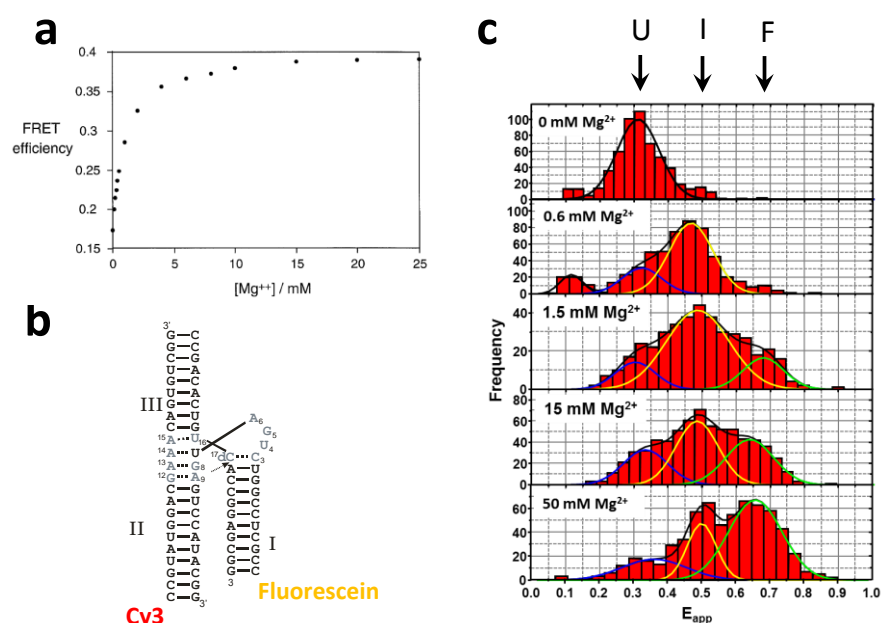
Two synthetic constructs were designed and prepared to probe the folding mechanism, shown in Figure 6.4. The FRET pair chosen is Cy3 as the donor and Cy5 as the acceptor, and the dyes positioned as shown in Figure 6.6.



**Figure 6.6:** Sequence of the synthetic minimal hammerhead ribozyme constructs used to investigate its single-molecule folding dynamics. Each of the constructs for the minimal hammerhead is built from three strands. The preparation is covered in detail in Chapter 3. The positions of the dyes are shown in green (Cy3) and red (Cy5). a) The construct prepared to study the folding between stems P1 and P2. b) The construct used to investigate the folding dynamics between stems P1 and P3. Cytosine at the catalytic site was replaced by deoxycytosine (dC) to avoid cleavage.

### 6.3.2. Single-molecule analysis of conformational changes involving P1 and P2 stems in the minimal hammerhead ribozyme

The first structure that was investigated was the P1-P2 vector, hereafter referred to as mHHRz12. The sample was diluted in T50 (50 mM trizma hydrochloride, pH 7.9) and incubated on the slide. I carried out a single-molecule FRET titration as a function of  $Mg^{2+}$  ions and the results are summarised in Figure 6.7.

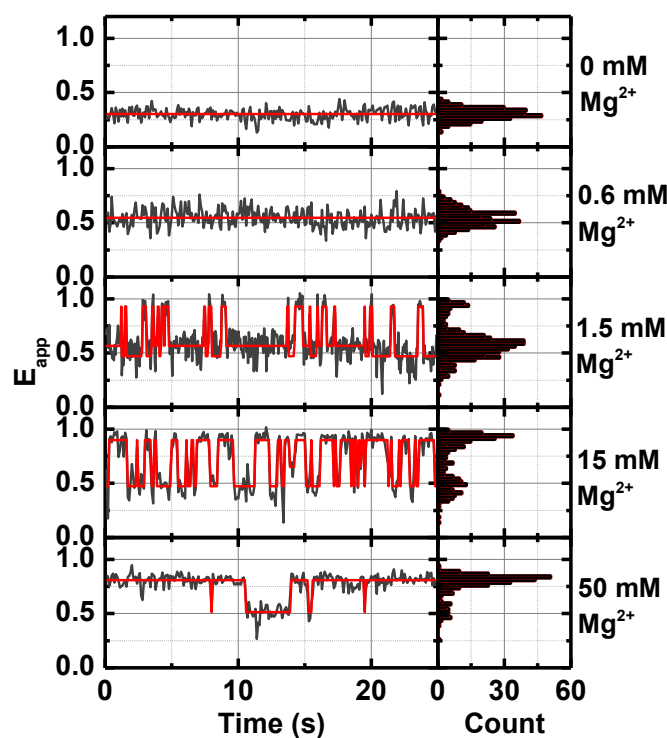


**Figure 6.7:** a) Ensemble FRET titration carried out for mHHRZ12 as a function of divalent metal ions taken from [172]. b) mHHRZ12 construct showing the position of the fluorescein-Cy3 pair used in the ensemble FRET studies [208]. c) Single-molecule FRET histograms obtained for mHHRz12 at the indicated concentrations of  $Mg^{2+}$  ions. The solid lines represent the result from fitting each population to a Gaussian distribution.

The folding pathway shows a progression through three conformational states. In the absence of metal ions, we obtained a single Gaussian distribution centred at  $E_{app} = 0.31 \pm 0.02$ , and we assigned this to the unfolded state. Increasing the concentration of  $Mg^{2+}$  ions to 0.6 mM promoted a significant shift of most molecules to a higher FRET state ( $E_{app} = 0.48 \pm 0.04$ ). This agrees with the observations from the bulk data, where the FRET signal showed an increase from 0.17 to 0.24 [172] (Figures 6.7a and 6.7b). An additional increase in the magnesium concentration to 1.5 mM introduces a new population with a higher FRET value ( $E_{app} = 0.67 \pm 0.02$ ) (Figure 6.7c). The relative population of this high-FRET state increased as we increased the concentration of  $Mg^{2+}$  ions and became



the predominant conformation at concentrations of divalent metal ions around 50 mM. This behaviour is also in excellent agreement with the previously reported observations from ensemble studies, where a further increase is observed, which begins to plateau >10 mM and approaches a value of 0.39 [163] (Figure 6.7a). A gradual increase in the relative contribution of FRET populations exhibiting a high FRET value agrees with the folding scheme shown in Figure 6.4. According to this model, the state with a FRET efficiency value of  $\sim 0.48$  should be assigned to the stacking of stems P1 and P2 that brings the donor and acceptor in closer proximity. Similarly, the state with the highest FRET value corresponds to the native state involving the formation of the U-turn and the catalytic pocket (Figure 6.4).

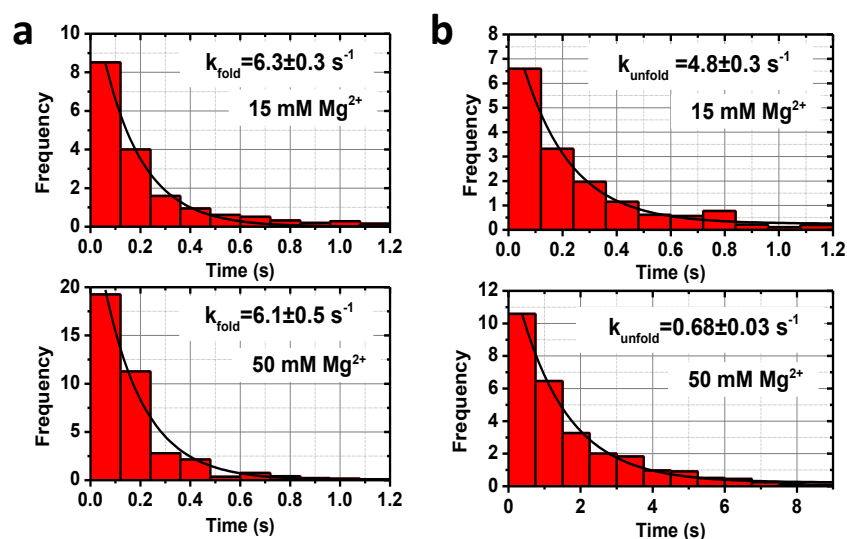


**Figure 6.8:** Representative single-molecule FRET trajectories obtained at the indicated concentrations of  $\text{Mg}^{2+}$  ions for the mHHRz12 construct. The red line represents the idealised trajectory obtained by applying Hidden Markov modelling.

The single-molecule FRET trajectories obtained from the P1-P2 vector (Figure 6.8) mirror the findings from the single-molecule FRET histograms shown in Figure 6.7c. In the absence of  $\text{Mg}^{2+}$  ions, we observed a single FRET level at  $E_{\text{app}} = 0.3 \pm 0.04$  with no fluctuations. At concentrations that favour the intermediate FRET state (0.6 mM) as the predominant conformation, we observed mostly single FRET level trajectories remaining at  $E_{\text{app}} = 0.54 \pm 0.02$  for long periods of time before photobleaching. At concentrations of  $\text{Mg}^{2+}$  ions where the high-FRET level starts to be populated (1.5 mM), the FRET trajectories display fast fluctuations between a long-lived  $E_{\text{app}} = 0.54 \pm 0.02$  state and a short-lived state at  $E_{\text{app}} = 0.90 \pm 0.08$ . As the concentration of  $\text{Mg}^{2+}$  increases (15 mM), the high-FRET state became stabilised with longer dwell times. At concentrations of  $\text{Mg}^{2+}$  ions (50 mM) where the high-FRET state is the most populated, the FRET trajectories showed a very long lived folded state ( $\sim 10$  s) with occasional fluctuations to the intermediate FRET level. It is important to emphasise that we have only observed the formation of the high-FRET folded state from the intermediate conformation. The lack of a significant number of direct transitions from the unfolded to the folded conformation suggests that the intermediate state, representing the stacking of the P1 and P2 stems, is indeed an obligatory step in the folding pathway.

The single-molecule dwell time analysis of the folding dynamics between the intermediate and high-FRET levels corresponding to the I $\rightarrow$ F transition

are shown in Figure 6.9. The folding rate for this construct remained constant within the error with values of  $6.3 \pm 0.3 \text{ s}^{-1}$  at 15 mM and  $6.1 \pm 0.5 \text{ s}^{-1}$  at 50 mM. The unfolding rate, however, was significantly altered by the addition of  $\text{Mg}^{2+}$  ions, showing a change from  $4.8 \pm 0.3 \text{ s}^{-1}$  at 15 mM magnesium to a value of  $0.68 \pm 0.03 \text{ s}^{-1}$  at 50 mM magnesium. This  $\sim 10$ -fold decrease in the unfolding rate represents the degree of stabilisation of the catalytically active native state by divalent metal ions and provides a functional role for the hydrated ions observed at specific locations in the crystal structure of the mHHRz [211].



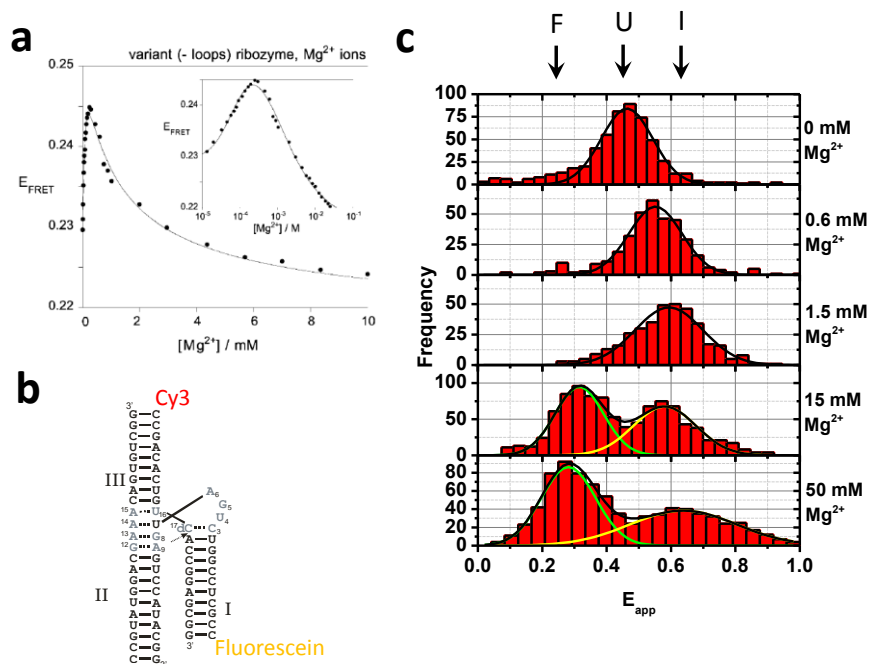
**Figure 6.9:** Single-molecule dwell time histograms of a) the folding rate, and b) the unfolding rate, obtained for the mHHRz12 construct at the indicated concentrations of  $\text{Mg}^{2+}$  ions. The solid line represents the fit to a mono-exponential decay function and the extracted values with the associated error from the fit are also shown.

The observation that the rate for the  $\text{I} \rightarrow \text{F}$  transition is independent of the concentration of  $\text{Mg}^{2+}$  ions suggests that the uptake of the most trapped divalent metal ions may take place in the first stage ( $\text{U} \rightarrow \text{I}$ ) of the folding

pathway. An additional explanation could be that at least some of these  $\text{Mg}^{2+}$  ions are trapped after the formation of the U-turn to stabilise the catalytic pocket. Our current data does not allow us to differentiate between both possibilities, but experiments are being carried out to investigate the dependence of the transition rate for the (U→I) with the concentration of  $\text{Mg}^{2+}$  ions. Also, additional experiments using the competition between  $\text{Na}^+$  and  $\text{Mg}^{2+}$  ions should provide additional insights.

### **6.3.3. Single-molecule analysis of conformational changes involving P1 and P3 stems in the minimal hammerhead ribozyme**

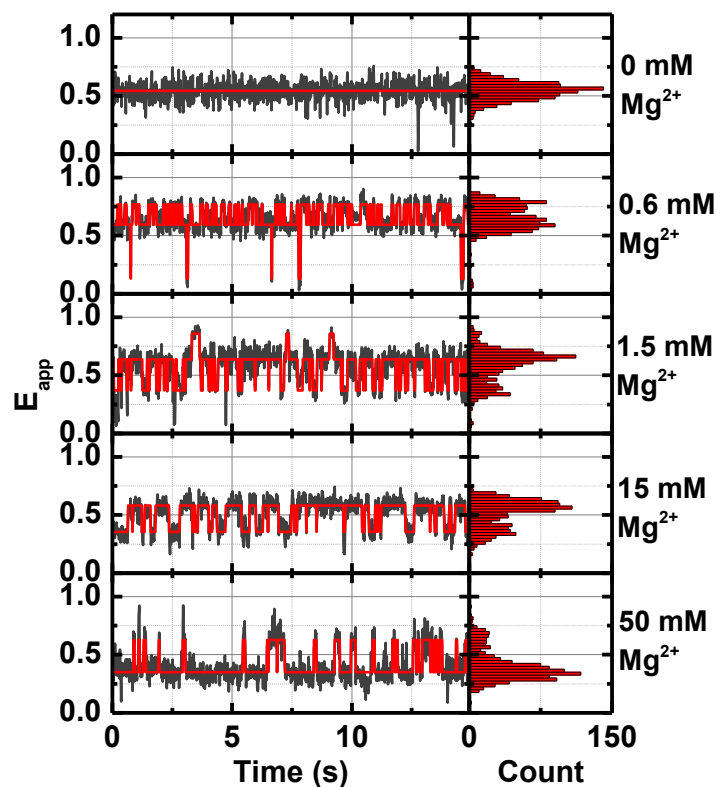
The ensemble FRET studies using a fluorescein-Cy3 FRET pair located on P1 and P3 showed an increase in FRET from 0.23 to 0.25, followed by a drop to a low FRET value of 0.22 [172] as shown in Figures 6.10a and 6.10b. The construct shown in Figure 6.6b, referred to as mHHRz13, was designed to investigate at single-molecule level the conformational changes taking place between stems P1 and P3. The mHHRz13 construct was examined at the same chemical conditions as the mHHRz12 structure. The distribution of single-molecule FRET populations as a function of  $\text{Mg}^{2+}$  ions are presented in Figure 6.10c.



**Figure 6.10:** a) Ensemble FRET titration carried out for mHHRz13 as a function of divalent metal ions taken from [172]. b) mHHRz13 construct showing the position of the fluorescein-Cy3 pair used in [172] for ensemble FRET studies. c) Single-molecule FRET histograms obtained for mHHRz13 at the indicated concentrations of  $\text{Mg}^{2+}$  ions. The solid lines represent the result from fitting each population to a Gaussian distribution.

The single-molecule FRET histograms obtained for mHHRz13 (Figure 6.10c) broadly confirm the observations noted from the bulk studies (Figure 6.10a) [163]. In the absence of  $\text{Mg}^{2+}$  ions, we observed a single Gaussian distribution centred at  $E_{\text{app}} = 0.46 \pm 0.03$ . Increasing the concentration to 1.5 mM stimulated the structure to move into a slightly lower FRET state ( $E_{\text{app}} = 0.60 \pm 0.09$ ). At higher concentrations of  $\text{Mg}^{2+}$  ions, a new population appears at lower FRET efficiency values ( $E_{\text{app}} = 0.31 \pm 0.04$ ). This population becomes the predominant conformer at 50 mM concentration of  $\text{Mg}^{2+}$  ions. Overall, the observation of an initial shift in the FRET efficiency to a higher value, which then reverts to a lower

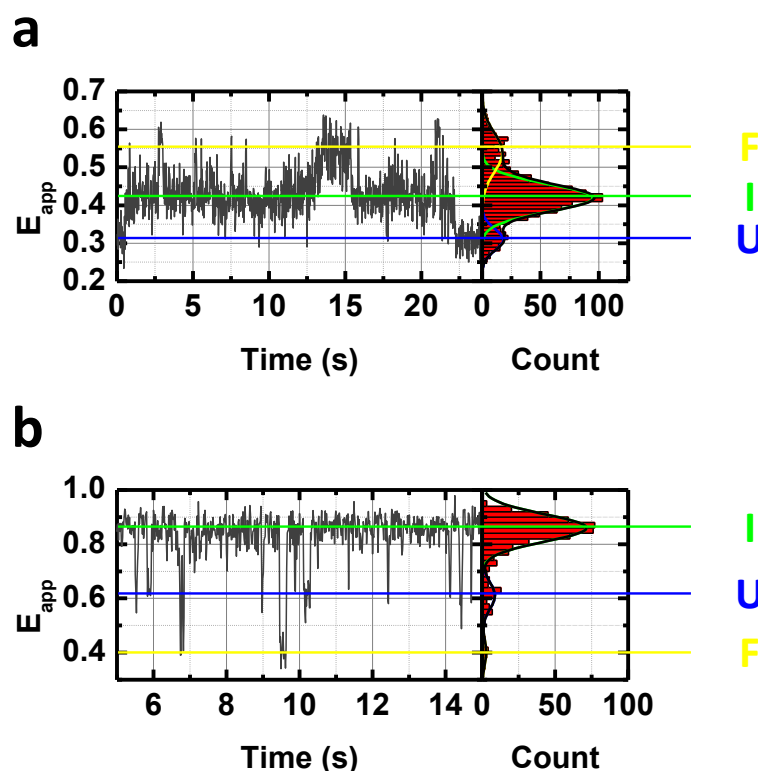
FRET state, is consistent with the results obtained previously using ensemble FRET.



**Figure 6.11:** Representative single-molecule FRET trajectories obtained at the indicated concentrations of  $\text{Mg}^{2+}$  ions for the mHHRz13 construct. The red line represents the idealised trajectory obtained by applying Hidden Markov modelling.

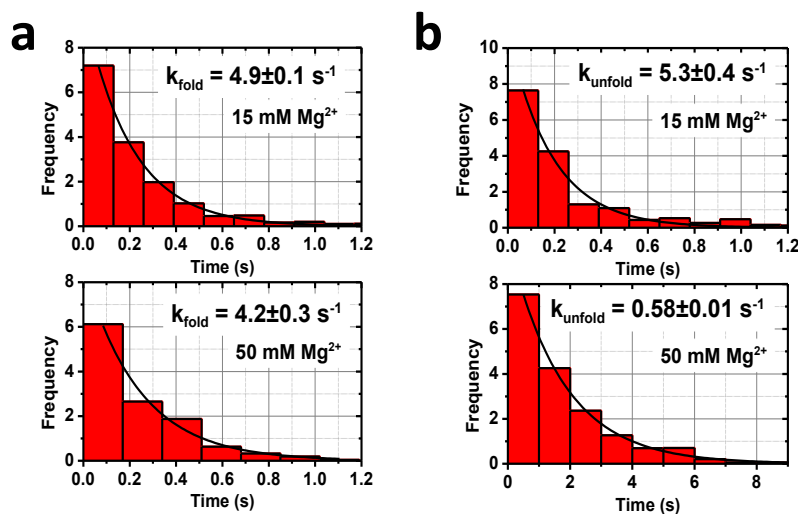
Representative FRET trajectories obtained for mHHRz13 are shown in Figure 6.11. At low concentrations of  $\text{Mg}^{2+}$  ions (0.6 mM) we observed very fast fluctuations between the unfolded ( $E_{\text{app}} = 0.52 \pm 0.03$ ) and intermediate states ( $E_{\text{app}} = 0.78 \pm 0.10$ ). This transition corresponds to the stacking of P2 and P3, a conformational change that brings stems P1 and P3 in closer proximity. At higher concentrations of  $\text{Mg}^{2+}$  ions we observed fast fluctuations between  $E_{\text{app}} = 0.78 \pm 0.10$  (I state) and  $E_{\text{app}} = 0.37 \pm 0.05$

(F state), which became less frequent as we approached saturating concentrations of  $Mg^{2+}$  ions. Importantly, as shown in Figure 6.12, single-molecule studies of mHHRz13 and mHHRz12 constructs at 16 ms confirmed that single hammerhead molecules fluctuate between the three states within a single trajectory. The noise fluctuations around the peak centre are consistent with the observations for HMM analysis (Chapter 3).



**Figure 6.12:** Representative single-molecule FRET trajectories showing the interconversion between the three FRET states (U, I and F) for both constructs investigated; a) mHHRz12, and b) mHHRz13.

This demonstrates that the intermediate conformation is an on-path state. Moreover, traces showing interconversion between these states indicate that the formation of the native state takes place from the intermediate conformation suggesting that this is an obligatory state in the folding pathway.



**Figure 6.13:** Single-molecule dwell time histograms of a) the folding rate, and b) the unfolding rate, obtained for the mHHRz13 construct at the indicated concentrations of  $\text{Mg}^{2+}$  ions. The solid line represents the fit to a mono-exponential decay function and the extracted values with the associated error from the fit are also shown.

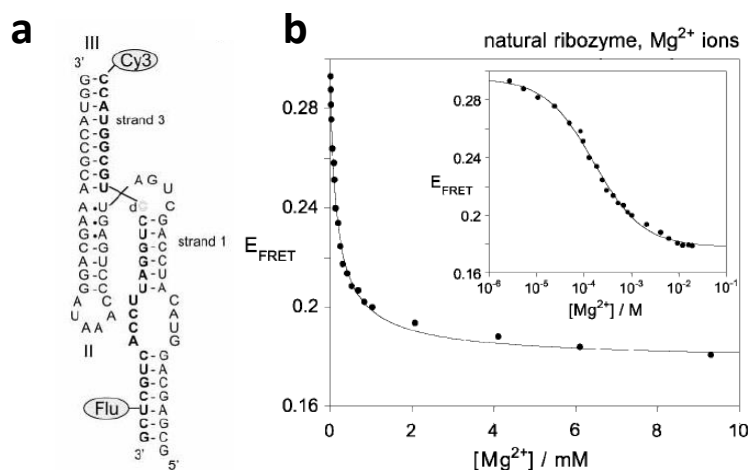
The single-molecule analysis of the dwell times for mHHRz13 is shown for the (I→F) (Figure 6.13a) and the reverse unfolding transition (Figure 6.13b). As previously reported for mHHRz12, the folding rate had a value that was independent of the concentration of  $\text{Mg}^{2+}$  ions ( $4.2 \pm 0.3 \text{ s}^{-1}$  at 50 mM). The unfolding rate also showed a similar 10-fold decrease from  $5.3 \pm 0.4 \text{ s}^{-1}$  to  $0.58 \pm 0.01 \text{ s}^{-1}$ , and therefore, an equivalent degree of stabilisation of the catalytically active conformation. It is interesting to note that, within the error, the folding and unfolding rates obtained for both constructs exhibited very similar values, confirming that simultaneous movement of the three stems and a similar dependence on the uptake and trapping of metal ions.



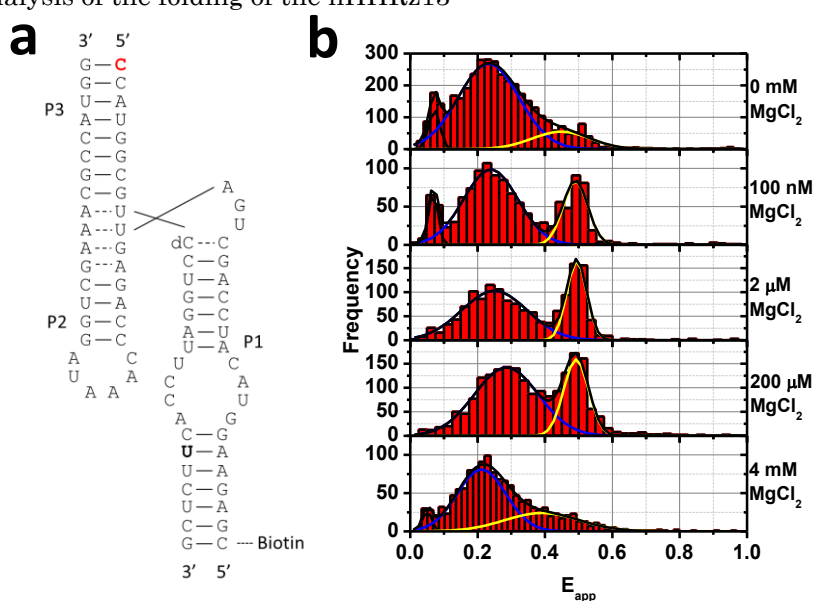
#### 6.4. Using sm-FRET to investigate the folding dynamics of the natural hammerhead ribozyme

When the peripheral loops within the natural hammerhead sequence (nHHRz) were shown to have a significant effect on the catalytic activity of the hammerhead ribozyme, the questions regarding its functionality at physiological conditions were answered [172, 210]. Subsequently, a bulk fluorescence study demonstrated that a significantly reduced concentration of near physiological concentrations of  $Mg^{2+}$  ( $\sim 2$  mM) was required to fully fold the structure [172]. Surprisingly, in this bulk FRET study the rise in FRET attributed to the initial formation of the intermediate state, assigned in the minimal form to the stacking of P2 and P3 stems, was not observed in the natural form (Figure 6.14a and Figure 6.14b) [172, 192, 209, 210]. The lack of evidence for this stacking at the ensemble level for the natural form raises the question of whether both structures follow the same folding pathway. To answer this question and compare the folding of the natural form with our observations on the minimal form, a natural construct (nHHRz13) was designed as shown in Figure 6.15a. I chose to investigate the conformational changes taking place between stems P1 and P3, as this is the one that showed an initial increase followed by a decrease in FRET and thus provides an easy test of the existence of a two-stage folding as observed in the minimal form. The nHHRz13 was built from two strands including an amino base modification to which covalently attach the Cy3

label as shown in Figure 6.15a. The sm-FRET titration for this construct is shown in Figure 6.15b.



**Figure 6.14:** a) Sequence and modifications used by Penedo et al to investigate the folding transition of the natural form of the hammerhead ribozyme (nHHRz13) [172]. The location of the FRET pair and the dC substitution to avoid cleavage are also shown. b) FRET analysis of the folding of the nHHRz13



**Figure 6.15:** a) Sequence and structure of the natural hammerhead ribozyme structure, nHHRz13, used for the sm-FRET titration. The acceptor (Cy5) position is shown in red; the position of the amino modification to which the donor is attached is highlighted in bold type. b) Steady-state histograms from the titration, showing the evolution of the population with increasing magnesium concentration.

The steady-state histograms show that the structure adopts a low FRET

conformation ( $E_{\text{app}} = 0.27 \pm 0.04$ ) in the absence of added  $\text{Mg}^{2+}$  ions. Although this state is the most populated, there is a significant fraction in the intermediate state ( $E_{\text{app}} = 0.48 \pm 0.07$ ). The relative contribution of this population is significantly increased at 100 nM magnesium. At 2  $\mu\text{M}$   $\text{Mg}^{2+}$ , the fraction of nHHRz13 molecules in this higher FRET is further increased. However, at 200  $\mu\text{M}$  concentration of  $\text{Mg}^{2+}$  ions one observes an increase in a lower FRET state with the majority of the population occupying a state of  $E_{\text{app}} = 0.21 \pm 0.02$  at 4 mM  $\text{Mg}^{2+}$  ions. The close values of the U ( $E_{\text{app}} \sim 0.27$ ) and F states ( $E_{\text{app}} \sim 0.21$ ) have precluded us from analysing the folding and unfolding rates as both states cannot be differentiated. Alternative constructs with the FRET pair located at different positions trying to differentiate between both states are currently under study. Nevertheless, our single-molecule data on the natural form clearly confirms a two-stage folding process, as observed for the minimal form. However, the first step of the folding transition in the natural form saturates at a concentration of  $\text{Mg}^{2+}$  ions around 1-2  $\mu\text{M}$ , which is about 500-fold lower than the concentration of  $\text{Mg}^{2+}$  ions required for first folding step in the minimal form. We noticed that the bulk titration only investigated the influence of  $\text{Mg}^{2+}$  ions for concentrations higher than 2-3  $\mu\text{M}$  (Figure 6.14b), a concentration at which, according to our single-molecule titration, the first folding stage has already reached saturation. Thus, it is likely that the observed decrease in FRET efficiency observed in bulk corresponds mostly to the second step in the folding pathway.

## 6.5. Conclusions

The hammerhead ribozyme and the hairpin ribozyme are the most extensively investigated small RNA enzymes. However, in contrast to the folding and catalytic activity of the latter, which have been widely studied at the single-molecule level, an equivalent understanding of the hammerhead ribozyme is still lacking. The hammerhead ribozyme has attracted considerable attention in the last years since it has been discovered to be ubiquitous across a large number of genomes [198]. As for the hairpin ribozyme, which was initially studied in a minimal form lacking the 4-way junction from which it inherits a significant part of its functional dynamics, initial studies on the hammerhead ribozyme were carried out on constructs lacking peripheral loops that were later discovered to be essential for activity at physiological concentrations of  $Mg^{2+}$  ions. Unfortunately, the majority of our knowledge of the function of the hammerhead ribozyme continues, even at this stage, to be based on findings extracted from the non-physiologically active minimal form. Whilst a substantial body of single-molecule FRET work has been recently carried out on the full-length natural hairpin ribozyme, only one single-molecule study addressing the influence of point mutations on the formation of the catalytic site has been reported to date. In this chapter, I aimed to provide a first comparative analysis of the folding, specifically on the influence of divalent metal ions, of the structures of the minimal and natural hammerhead ribozymes. I started

by probing the folding pathway of the minimal structure using two constructs that target the conformational changes taking place between different motifs, mHHRz12, reporting distance changes between the P1 and P2 helical stems, and mHHRz13 that targets P1 and P3 stems. The analysis of the FRET changes as a function of  $\text{Mg}^{2+}$  ions confirmed the two-stage folding mechanism previously suggested by ensemble FRET and comparative gel electrophoresis. According to this model, an initial folding step involving the stacking of P2 and P3 ( $< 0.5 \text{ mM Mg}^{2+}$ ) is followed by the formation of an U-turn motif that induces a rotation of the P1 stem to become closer to the P2 stem ( $> 20 \text{ mM Mg}^{2+}$  ions). The single-molecule FRET trajectories confirmed the coexistence of the three states within a single molecule and have proven that the formation of the helical stack is an obligatory step in the folding pathway of the minimal hammerhead. Interestingly, the second transition involving the formation of the U-turn and the catalytically active species was found to be mostly insensitive to the concentration of metal ions. In contrast, once formed,  $\text{Mg}^{2+}$  ions helped to stabilise the native conformation as observed by the 10-fold reduction observed in the unfolding rates. These findings suggest that most of the trapped  $\text{Mg}^{2+}$  ions observed in the crystal structure of the mHHRz are bound during the first transition. Although this observation needs further confirmation using a broader range of  $\text{Mg}^{2+}$  ions between 0.5 mM and 50 mM, it agrees well with phosphorothioate interference experiments that identify three candidate phosphates at positions A<sub>9</sub>, A<sub>13</sub> and A<sub>14</sub> as potential

metal ion binding sites around the stacked P2-P3 helices that induce the first folding stage [212]. In contrast, only one  $Mg^{2+}$  metal ion near the catalytic core that folds in the second step has been found using rapid freezing of hammerhead crystals [213].

For the natural form of the hammerhead ribozyme it was previously shown using ensemble FRET that the observed increased catalytic activity obtained at physiological concentrations of  $Mg^{2+}$  ions was due to a reduction in the amount of  $Mg^{2+}$  ions required to form the catalytically active native state, and that this was indeed due to the formation of specific interactions between the peripheral loops [172]. However, in contrast to the minimal form of the hammerhead ribozyme, the folding transition showed only a single step. Our single-molecule FRET analysis of the natural form suggests that, as in the minimal form, a two-stage transition is indeed present. However, the initial step takes place at a low micromolar concentration of  $Mg^{2+}$  ions, which corresponds to a 500-fold decrease with respect to the minimal form. The second folding step takes place in our single-molecule FRET experiments in a range of  $Mg^{2+}$  concentrations similar to that reported by ensemble FRET (0.1-1 mM). If the nature of the first folding step in the natural form of the HHRz is similar to that in the minimal form, involving the stacking of the P2 and P3 stems, it is difficult to think how the loop-loop interaction that is formed in the second step could affect the first folding transition and decrease its  $Mg^{2+}$  requirement. A possible explanation is that a tight tertiary interaction between the loops could increase the

folding rate for the second step and certainly decrease the unfolding rate, thus quickly shifting the equilibrium from the intermediate to the final native state. Further studies using point mutations that can lock the folding pathway into the first step are currently underway to shed light on this.

# Chapter 7: Implementation of single-molecule techniques with freely-diffusing molecules

## 7.1. Introduction

### 7.1.1. Fluorescence imaging without surface immobilisation

In order to realise single-molecule spectroscopy, there are two crucial experimental criteria that must be satisfied: the use of a scheme which employs a low excitation volume, minimising background signals, and low sample concentrations [30]. In Chapter 1, two of the most common means of achieving this low background excitation were introduced. The first was through selective excitation, where an evanescent field is generated and used to produce a strongly distance-dependent excitation volume. This is the principle behind imaging techniques such as total internal reflection fluorescence microscopy (TIRFM) [37, 39], a wide-field excitation technique, and near-field scanning optical microscopy (NSOM), a point excitation technique [29].

The second low-volume excitation scheme involves focussing the excitation beam to a diffraction-limited confocal spot [214]. Confocal microscopy is a very flexible base to build in imaging setup upon and has been applied to diffraction-limited and super-resolution techniques [30]. A further advantage is the planar selectivity that may be achieved by confocal



microscopy; by adjusting the focal spot position, one can excite molecules moving within the buffer, thus moving away from the need to tether the sample to the surface of the substrate.

By expanding into solution-based imaging, one expands the focus of the biophysical properties being probed from intramolecular dynamics and docking of two species to macroscopic effects, such as diffusion mechanics. This change opens up the opportunity to investigate reactions involving changes in size or mass at the single-molecule level [215, 216].

### 7.1.2. The property of diffusion

A molecule within a system in the liquid or gaseous state is not confined in space and time. Instead, thermal fluctuations and concentration gradients within the medium result in a constant molecular rearrangement. This rearrangement is characterised by a random walk, or Brownian motion, which gives a measure of how far the molecule is expected to travel in a period of time ( $t$ ) [49].

Due to the non-linear motion exhibited by a molecule while moving in solution, it is more prudent to calculate the properties of motion in terms of an area covered. One does so by looking at the mean squared displacement of the molecule, as shown by Equation 7.1 shown below:

$$\langle r(t)^2 \rangle = 6Dt \tag{7.1}$$

In this equation, ( $D$ ) is defined as the diffusion coefficient, which is an estimate of how far the molecule will travel in one second. This property will clearly depend on the size and mass of the molecule. The early work on diffusion by Fick explained the macroscopic property of diffusion, yet the microscopic behaviour wasn't fully characterised until Einstein's work in 1905 [49]. He derived the expression for the diffusion coefficient in terms of the properties of both the solvent and sample, shown in Equation 7.2, where  $k_B$  is Boltzmann's constant ( $1.38 \times 10^{-23} \text{ JK}^{-1}$ ),  $T$  is the temperature of the system in kelvin,  $\eta$  is the viscosity of the solution and  $R_H$  is the hydrodynamic radius of the sample. Knowing that the mass of a sample is directly proportional to its volume (linked through the density), then one may infer that the diffusion coefficient is inversely proportional to the mass of the sample ( $m$ ) cubed (Equation 7.3).

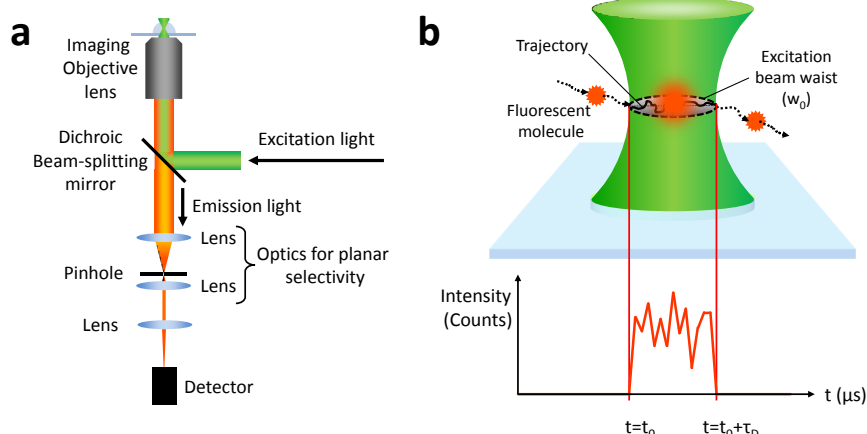
$$D = \frac{k_B T}{6\pi\eta R_H} \quad (7.2)$$

$$D \propto \frac{1}{\sqrt[3]{m}} \quad (7.3)$$

The combination of Equations 7.2 and 7.3 allows the mass of the sample to be directly linked to its mobility within a buffer of known viscosity, subject to the correct calculation or knowledge of the diffusion coefficient, information not readily available from surface-immobilised schemes [9].

### 7.1.3. Single-molecule fluorescence techniques to measure diffusion

The measurement of diffusion at the single-molecule level is typically through the measurements of fluorescence burst analysis [217]. One records the emission of a fluorescently labelled molecule while diffusing through a confocal excitation spot with a volume below 1 femtolitre (1 fl), and through averaging over these events a spread of diffusion times will be obtained. A possible optical setup to record bursts of fluorescence from a suitably diluted sample is presented in Figure 7.1a.

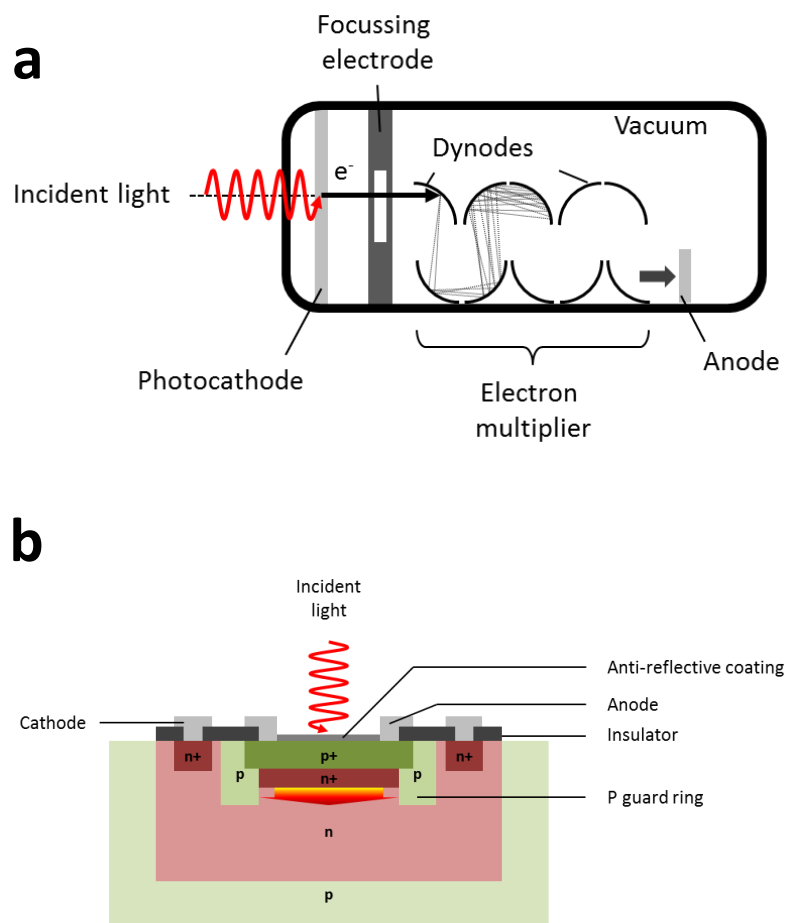


**Figure 7.1:** a) Optical setup to record the bursts of fluorescence from a labelled molecule diffusing through a confocal volume. b) Illustration of how the emission of a molecule is recorded as it traverses a confocal laser spot. The light for this molecule is collected only when the molecule is in the excitation volume; otherwise the intensity is zero.

The fluorescence is recorded over time and the intensity plotted against time, as illustrated in Figure 7.1b. The duration of each region of intensity emission is the diffusion time for that particular molecule.

#### 7.1.4. Detection of single photon emission

The high sensitivity of detection required calls for detectors capable of recording single-photon events. The two most common choices for detectors are photomultiplier tubes (PMTs) and single-photon avalanche diodes (SPADs). The basic modes of operation are summarised in Figures 7.2a and 7.2b.



**Figure 7.2:** a) Schematic of how a photomultiplier tube works, showing the internal components. b) Cross-section of a SPAD chip, showing the diode dopant regions and the region of increasing avalanche charge build-up (red arrow).

In a PMT, an incident photon is absorbed by a photocathode, which generates an electron. This electron is then focussed into a series of

dynodes, which have a multiplying effect generating multiple charge carriers which are finally collected by an anode, thus completing the circuit and generating a photocurrent which is associated with the initial photon [218–220].

A SPAD is based around a reverse biased p-n junction, which is biased at a voltage in excess of the breakdown level. This means that a single photon can trigger a huge number of charge carriers, improving the counting efficiency of the device due to the non-linear gain at in this unstable regime [218–220].

The choice between these two options is generally made at the operator's discretion due to the similar timing resolutions and sensitivities of detection. The quantum detection efficiency of SPADs is greater than that of PMTs in the near-infrared regime [218]. In addition to this, modern advances in the development of SPAD units have pushed the efficiency peak further into the visible regime. Furthermore, the small detection area further increases the planar selectivity achieved by a confocal system.

#### **7.1.5. Fluorescence correlation spectroscopy and cross-correlation spectroscopy**

##### **Single detector: Autocorrelation**

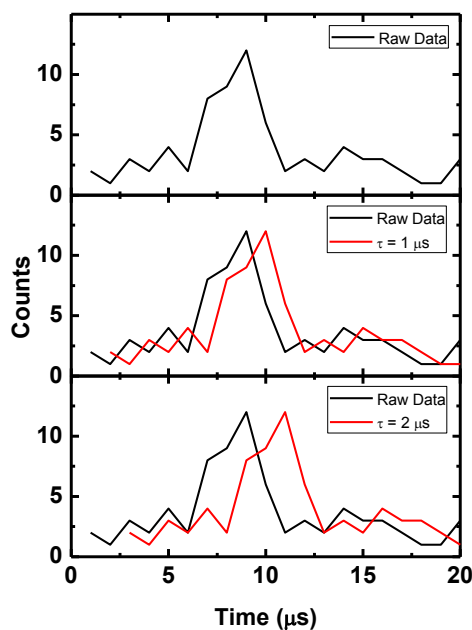
The fluorescence bursts collected by these single-photon counters may be analysed in different ways but the most frequently employed of these

techniques is through fluorescence correlation spectroscopy (FCS). This technique was introduced by Magde et al in 1972, where binding events between DNA and ethidium bromide were monitored by correlating the signal generated by a photocathode detector [221], and was eventually reduced down to the single-molecule level in 1993 [222].

The basic process for this analysis procedure is as follows. The raw fluorescence burst data is collected by a correlator unit and the average intensity value is calculated. This allows the intensity at any given time to be stated as in Equation 7.4 [9], in terms of the average intensity ( $\langle I_{AV} \rangle$ ) and the deviation from this value ( $\delta I(t)$ ).

$$I(t) = \langle I_{AV} \rangle + \delta I(t) \quad (7.4)$$

Having redefined the intensity in terms of this deviation from the average value, one calculates the autocorrelation function ( $G(\tau)$ ), a mathematical tool used to identify periodic fluctuations within an intensity trace. This function is generated by introducing a small shift time ( $\tau$ ) to a duplicate of the bursts recorded by the detector. Intensity points at corresponding times are multiplied together and the average of these products is computed and plotted as a function of the shift time applied to the duplicate trace [223]. A summary of this process is shown in Figure 7.3 below.



**Figure 7.3:** Summary of the process of shifting the duplicate trace as a part of the procedure of calculating the correlation in an intensity trace.

Once this function has been calculated, one may equate two equations to fit the experimental data (Equations 7.5 and 7.6). These curves offer insight into the translational kinetics of the sample under investigation and determination of properties of the sample, including the concentration ( $c$ ), mobility and size ( $D$ , and, through Equations 7.2 and 7.3, the hydrodynamic radius ( $R_H$ ) and mass) and stoichiometry for intermolecular reactions [9, 224]. One may also obtain information on the characteristics of the experimental setup itself, namely the dimensions of the confocal volume; the beam waist ( $w_0$ ) and the eccentricity of the Gaussian ellipsoid ( $k$ ). The derivation of these expressions and the conversion between the two is covered in Appendix III.

$$G(\tau) = \frac{G(0)}{\left(1 + \frac{4Dt}{w_0^2}\right) \left(1 + \frac{4Dt}{(kw_0)^2}\right)^{1/2}} \quad (7.5)$$

$$G(\tau) = \frac{G(0)}{\left(1 + \frac{t}{\tau_D}\right) \left(1 + \frac{t}{\tau_D k^2}\right)^{1/2}} \quad (7.6)$$

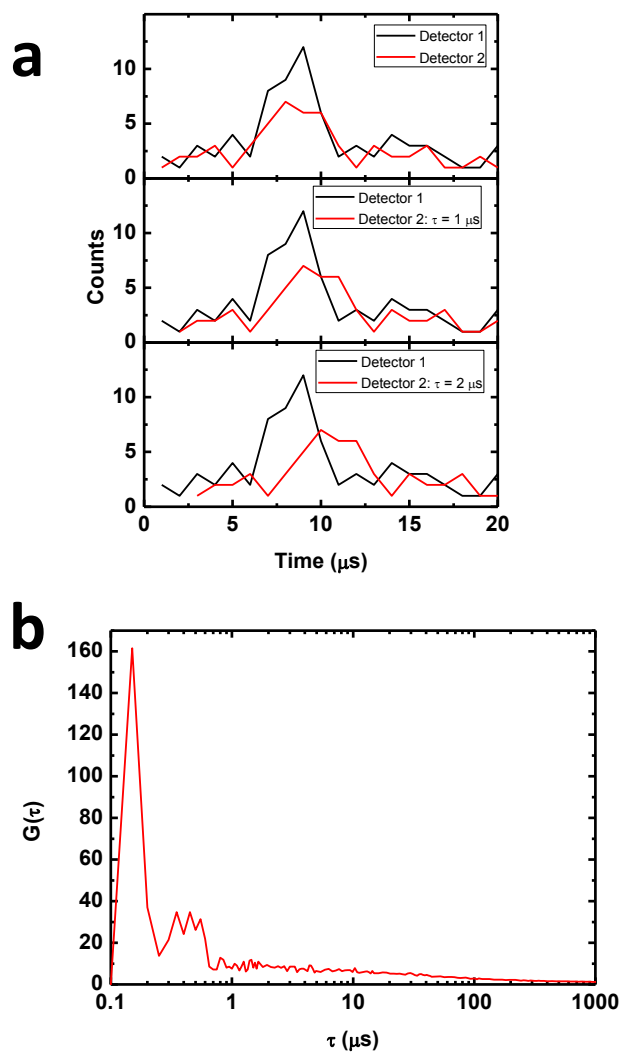
### Two detectors: Cross-correlation

One may perform the correlation between traces recorded by two separate detectors. This is known as fluorescence cross-correlation spectroscopy (FCCS), a technique first defined in 1994 by Manfred Eigen and Rudolf Rigler [225], before being realised experimentally by Petra Schwille three years later [215]. Though typically performed using a dual excitation scheme, with two separate dye species in the sample, the term may also be applied to a single excitation system so long as the experimental setup uses two detectors. In these cases, this function is not autocorrelation but cross-correlation. The shift ( $\tau$ ) is introduced to one of the traces, as shown by the red trace in Figure 7.4a below.

In this experimental situation, the equations must be rewritten to account for the intensities from two separate detectors, ( $I_1$  and  $I_2$ ). In the cross-correlation scenario, one must define the correlation factor ( $G(\tau)$ ) as in Equation 7.7.

$$G_{12}(\tau) = \frac{\langle I_1(t)I_2(t + \tau) \rangle}{\langle I_1(t) \rangle \langle I_2(t) \rangle} = 1 + \frac{\langle \delta I_1(t) \delta I_2(t + \tau) \rangle}{\langle I_1(t) \rangle \langle I_2(t) \rangle} \quad (7.7)$$





**Figure 7.4:** a) Summary of the shifting process for cross-correlation, showing the shift introduced to one of the detector traces. b) FCS trace for single-detector correlation where the after-pulsing peak is present at low correlation times ( $<1 \mu\text{s}$ ).

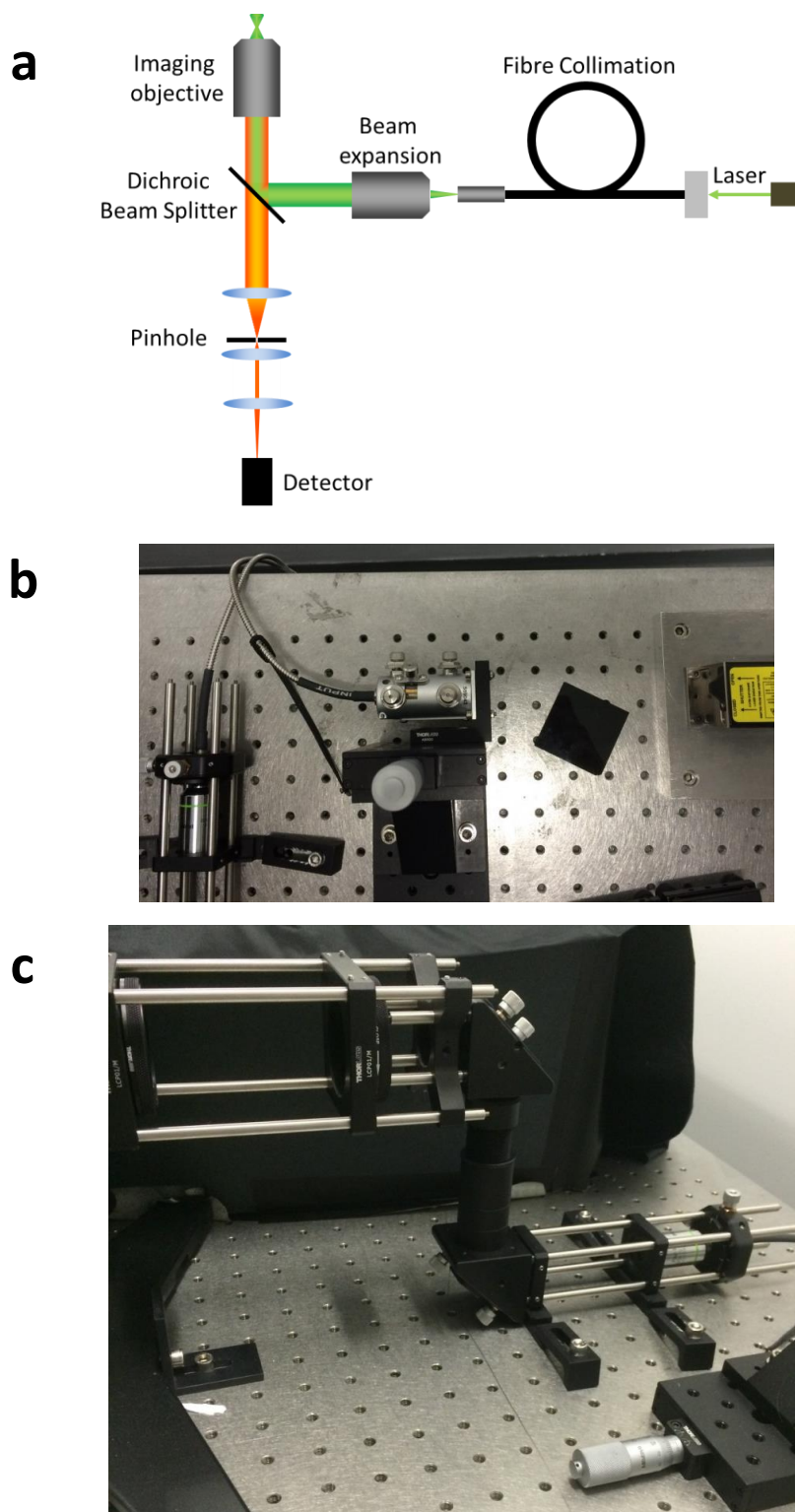
The fitting equations remain the same because the fluorescence comes from a single excitation volume. The main advantage of using two detectors is to eliminate the effect of after-pulsing from a single detector, an issue that causes a significant peak to occur at low correlation timescales (Figure 7.4b).

## 7.2. Experimental realisation of single-molecule FCS

### 7.2.1. Excitation: designing a single-molecule confocal microscope

With the background established, my first aim was to design and construct a single-molecule FCS system and then upgraded with FRET and single-molecule multi-parameter fluorescence detection (sm-MFD) capabilities. The excitation scheme chosen for such a system is generally a continuous excitation source, although a pulsed laser is absolutely required for sm-MFD. In confocal configuration, a Gaussian-profiled collimated beam is required to overfill the back aperture of the imaging objective.

The source chosen was a diode-pumped solid state laser (Vortran Stradus 532-40,  $P_{\text{peak}} = 40 \text{ mW}$ ,  $\lambda_{\text{excitation}} = 532 \text{ nm}$ ). The beam was shaped using an optical fibre and coupler (Qioptiq Kineflex coupling unit, broadband optical fibre (405-640 nm)), and expanded and collimated using a microscope objective (20X Olympus Plan Achromat Objective,  $\text{NA} = 0.4$ , 1.2 mm working depth). The collimated beam was passed into the back aperture of an Olympus IX-73 inverted microscope via a periscope system. This collimated beam is reduced in profile by two irises, and directed to overfill the back aperture of the imaging objective (Olympus UPLSAPO 60XW) by a dichroic mirror (Chroma Technologies,  $\lambda_{\text{cut-off}} = 532 \text{ nm}$ ,  $\text{AOI} = 45^\circ$ ). A schematic of the setup is presented below, in Figure 7.5, along with photos of the optics as they are set up.



**Figure 7.5:** a) Schematic of the optics used to set up the confocal excitation required to achieve single-molecule fluorescence correlation spectroscopy. b) and c) Photographs illustrating the arrangement of the components used to excite the sample. The cage structure is used to fix the optics in position and offer additional rigidity to the setup.

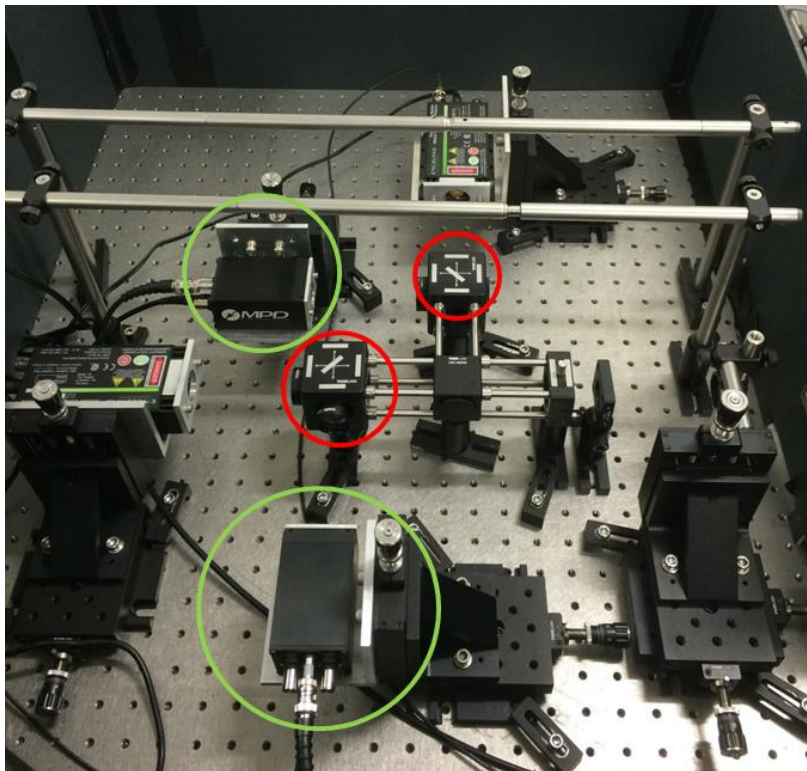
### 7.2.2. Resolving the emission from a confocal spot onto a point detector

The emission must be collected by the imaging objective, which is as stated in the previous section also responsible for the excitation of the sample. The dichroic mirror serves the dual purpose of directing the excitation beam into the objective and filtering the scattered light from reaching the detectors. The tube lens in the microscope ( $f = 180$  mm) collects the emission and focusses it to a diffraction-limited spot.

By positioning the pinhole (100  $\mu\text{m}$  diameter, Thorlabs) at the correct location within the emission pathway, the detection plane can be carefully chosen to correspond to the focal spot of the excitation beam, resulting in the greatest signal to noise ratio and lowest background intensity. This pinhole is mounted upon an X-Y translation stage (Thorlabs) which offers fine-tuning capabilities to maintain the system in its greatest operating efficiency. The emission leaving the pinhole is re-collimated using a lens (Thorlabs,  $f = 100$  mm) mounted in an X-Y-Z translation stage to aid aligning the detection scheme.

An additional long pass filter (Semrock,  $\lambda_{\text{cut-off}} = 532$  nm) is in place to filter any additional excitation light which passed through the dichroic and pinhole and prevent any unnecessary photodamage to the detector. The emission is split by polarisation using a polarising beam splitter cube (Thorlabs, 10 mm size), before it is focussed with another lens (Thorlabs,  $f=100$  mm) onto the detector itself, which is mounted on an X-Y translation stage for ease of alignment.

The detectors chosen were SPAD units (Micro Photon Devices PDM series PD-050-CTE, 50  $\mu\text{m}$  active area, 25 dark counts per second) chosen due to their excellent quantum efficiency in the upper range of the visible spectrum ( $\Phi > 40\%$  for  $475 \text{ nm} < \lambda < 625 \text{ nm}$ ,  $\lambda_{\text{peak}} = 550 \text{ nm}$ ) and the low dark count rate. This setup that I designed and built is photographed in Figure 7.6 below.



**Figure 7.6:** A photograph showing the arrangement of the optics used in order to realise FCS and FCCS.

The photograph in Figure 7.6 shows the detectors used circled in green, in a setup which has four detectors in place. This is a photograph of the extended setup, introduced in the following sections. The dichroic beam splitters (Chroma Technologies, T595lpxr), circled in red function, as mirrors here; with a cut-off wavelength of 595 nm they ensure that almost

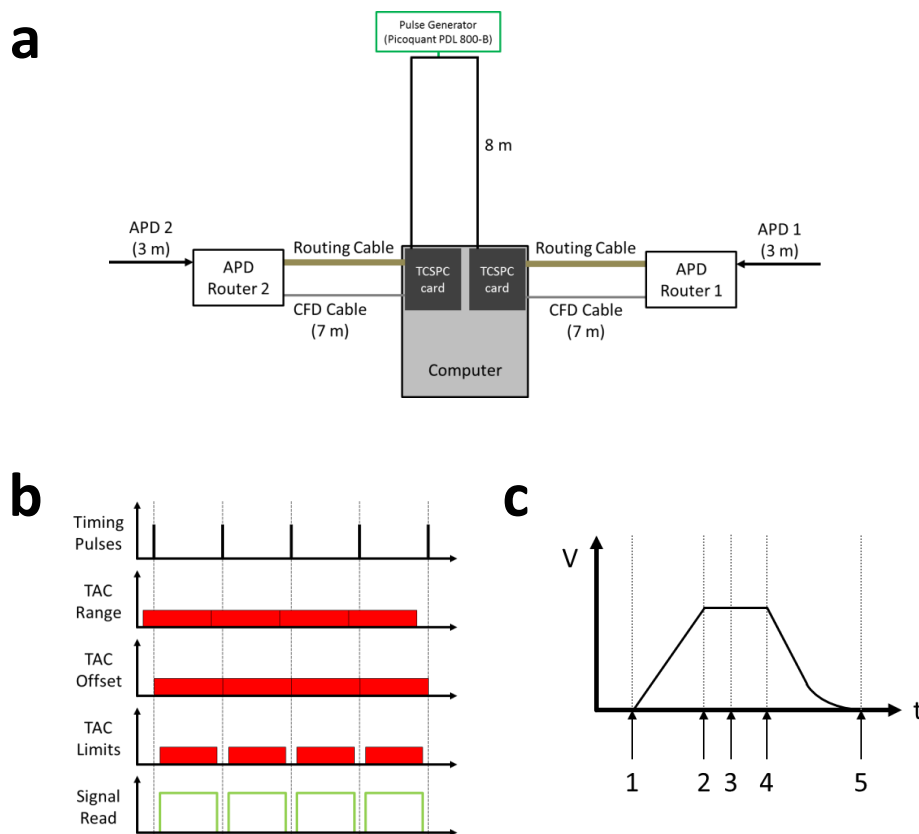
the entire spectrum of a dye excited by the 532 nm laser line will be directed into the MPD SPAD units.

### **7.2.3. Electrical diagram to collect emission signals and allow FCS resolution**

The signals connected by the SPAD units must be collected in order to allow the computer card to perform the FCS and FCCS calculations on the incoming photon bursts. I built the timing circuit used to facilitate this collection process, in conjunction with electrical components which will also allow easy system expansion to a setup with more detectors, as will be discussed in section 8 of this chapter.

The electrical system is shown in Figure 7.7a below. The signal from each detector is connected into a router (Becker & Hickl HRT-41 or HRT-41-A), which is in turn connected to a TCSPC computer card (Becker & Hickl SPCM-130-EM) by two cables. The first cable, the CFD (constant fraction discriminator) contains all of the intensity information from all detectors connected to the router. The second cable, the routing cable, contains specific information of the detector upon which a photon is incident. In order to ensure that the light emitted from each excited molecule is collected at the same time in each channel, it is crucial that all of the cable lengths for each detector path are kept the same. The system that I constructed uses two of these computer cards, connecting one router, and therefore one detector, into each.

The timing of the circuit is regulated by the TCSPC card and a pulse generator which is, in this setup, a Pulsed laser driver (Picoquant PDL 800-B). The timing window is set using the software provided with the computer card. This window is known as the time-amplitude conversion or TAC range, and sets the maximum period of time over which the signal will be amplified, until the stop pulse is read [220]. The TAC range is the period in which photons are recorded by the detector and processed by the computer card. The period is started by a photon read through the CFD cable. The pulse generator provides the stop pulse for the photon acquisition for each cycle, and is connected to the SYNC input to the TCSPC card. One may tune the acquisition window using the offset and start/stop limits in the software. This is shown in Figure 7.7b.



**Figure 7.7:** a) Schematic of the electrical layout of the connections between detector, pulse generator and computer in order to realise single-molecule FCS and FCCS. b) The tuning of the TAC range to line up with the sync pulses. The Text beside each plot shows which parameter controls each property these being the length of the range (TAC range), the displacement in time (TAC offset) and the start and end points of the signal (TAC limits). c) Operational mechanism of the TAC, showing the voltage changes in time. A photon is picked up by the computer card (1) and the TAC builds up a linear amplification until the stop pulse is detected through the SYNC connection on the card (2). After a short delay the ADC begins, and completes the process of assigning timing to each photon's arrival (4). The system then relaxes to zero, ready for a new detection event (5).

The TAC signal is then converted by an analogue-digital converter (ADC) which assigns arrival times to each photon, according to the time binning resolution specified by the software. Once this process is complete, the card relaxes to zero voltage, in preparation for a new cycle to begin. This whole process is summarised by Figure 7.7c, where each stage is indicated by a numerical marker.

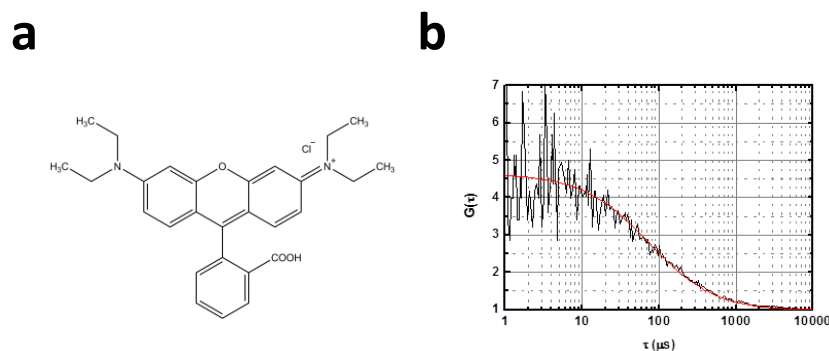


### 7.3. System verification using different fluorescently-labelled species

#### 7.3.1. Calculating the confocal volume using a sample with known diffusive properties

In order to calculate the excitation volume, one must eliminate some of the fit variables in a mathematically-analysed FCS or FCCS curve. One may achieve this objective by choosing a sample where the diffusion coefficient is known. By eliminating this variable and by performing an FCS scan on a suitably diffuse concentration, one may extract the diffusion time ( $\tau_D$ ) from the curve, and hence, through Equation 7.8, the beam waist of the confocal excitation volume may be calculated.

Knowing this, I performed an experiment using a dye whose diffusive properties were well characterised. The dye chosen was Rhodamine B, shown in Figure 7.8a, a molecule whose diffusion coefficient in water was calculated to be  $(4.2 \pm 0.3) \times 10^{-10} \text{ cm}^2\text{s}^{-1}$  in water [226]. A fluorescence cross-correlation spectroscopy scan was obtained from a sample of concentration 1 nM for a time of 1000 s. This scan is plotted in Figure 7.8b below. In addition to the sample, 0.1% by volume of Tween-20 was added to the buffer in order to prevent the dye from sticking to the coverslip.



**Figure 7.8:** a) Molecular structure of Rhodamine B. b) Plot of the FCS curve obtained from the scan of Rhodamine B, with the experimental data shown in black and the fitted curve in red.

The curve was fitted with Equation 7.7 through Quick Fit 3, developed by Dr Jörg Langowski of the German Cancer Research Centre [227], which yielded fit parameters of  $\tau_D = 75 \pm 10 \mu\text{s}$  and  $G(0) = 3.6 \pm 0.1$ . The list of calculated parameters and their associated standard errors are presented in Table 7.1 below.

**Table 7.1:** List of the fitted and calculated parameters derived from the single-dye FCCS scan of Rhodamine B.

Parameter	Fitted/Calculated Value	Standard Error	Percentage Error
$\tau_D$	75 $\mu\text{s}$	10 $\mu\text{s}$	14%
$G(0)$	3.6	0.1	2.7%
$w_0$	355 nm	56 nm	15.7%
$V_{\text{eff}}$	0.46 fl	0.01	2.7%
$w_z$	656	105	16%
$k$	1.85	0.4	22.4%

One can see that the effective volume is  $0.46 \pm 0.01$  fl, comfortably below the 1 fl upper bound for such an experimental setup. This volume, now calculated, shall be used where needed for all subsequent calculations in this chapter in the further characterisation of the system performed on different samples.

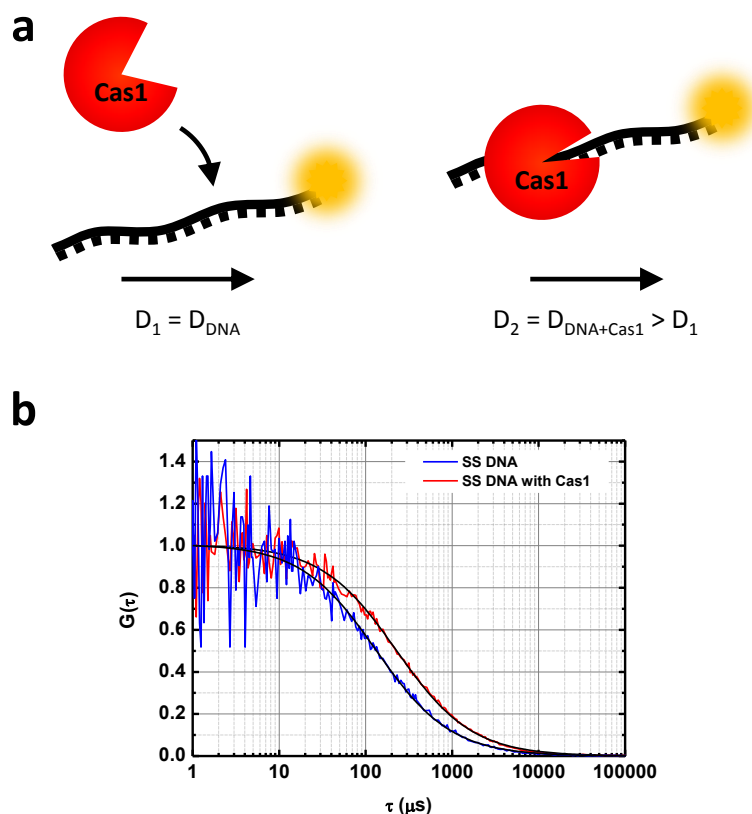
The largest source of error arises from the fit to the diffusion time, and one could reduce this error by scanning a lower concentration of sample, a change which, as can be seen from Equation 7.8, would result in an increase in the correlation amplitude  $G(0)$  and thus an increase in signal to noise ratio. This is a preferable choice to increasing laser power, as although this would increase the fluorescent emission from the dye molecules, it would also affect the confocal volume size. Also, due to the increased excitation power, the sample would photobleach very quickly when in the volume.

### 7.3.2. Protein-DNA interactions observed using FCS

One of the most powerful observations possible with FCS is the ability to track changes in the mass of a diffusing species through the increased diffusion time through the excitation volume. The association of two or more biomolecules to form complexes underpins every aspect and function within the cell [50, 51]. In this context, using the increase in the diffusion time of a labelled species to monitor its association to other biomolecules offers a crucial advantage because it only requires one of the interacting molecules to be fluorescently labelled.

The binding of two species, specifically the interaction between the protein Cas1 and a Cy3-labelled single-stranded DNA structure (ssDNA) 45 nucleotides in length, was examined as a means of testing the system. Cas1 is part of a recently discovered immune response present in bacterial and archaeal organisms, so-called CRISPR (clustered regularly interspaced

short palindromic repeats) immune system [228]. Cas1 is a DNA binding protein that is currently being studied in our group using TIRFM-based sm-FRET and therefore we have used it here as a model for testing our sm-FCS in the context of protein-DNA interactions. A summary of the Cas1-binding process is shown below in Figure 7.9a. The Cy3-labelled ssDNA was scanned first, in order to determine the diffusion dynamics of the unbound structure. As for the free dye scan performed in the previous chapter, Tween-20 was added to prevent unwanted surface adhesion [229], a property that the protein is particularly prone to exhibiting. The results are shown in Figure 7.9b below.



**Figure 7.9:** a) Summary illustration of the binding of the Cas1 protein to a ssDNA construct. The effect on the diffusion coefficient is shown qualitatively. b) Results from the scans, showing the normalised fitted FCS curves of the ssDNA alone (blue) and the ssDNA and Cas1 complex (red).

The results, plotted in Figure 7.9b, show a clear shift in the diffusion time between the two experiments, with an increase in diffusion time noted for the Cas1-ssDNA complex over the ssDNA alone. The fitting of this data confirms that the diffusion time is significantly increased;  $\tau_{\text{DNA}} = 131 \pm 13 \mu\text{s}$ , whereas  $\tau_{\text{DNA}+\text{Cas1}} = 227 \pm 17 \mu\text{s}$ .

When this diffusion time is translated into a coefficient, using the value for  $w_0$  calculated in the previous section, one obtains a value of  $D_{\text{DNA}} = (2.41 \pm 0.06) \times 10^{-10} \text{ m}^2\text{s}^{-1}$  and  $D_{\text{DNA}+\text{Cas1}} = (1.39 \pm 0.04) \times 10^{-10} \text{ m}^2\text{s}^{-1}$ . This corresponds to a reduction in the motion of a factor of  $1.73 \pm 0.06$ .

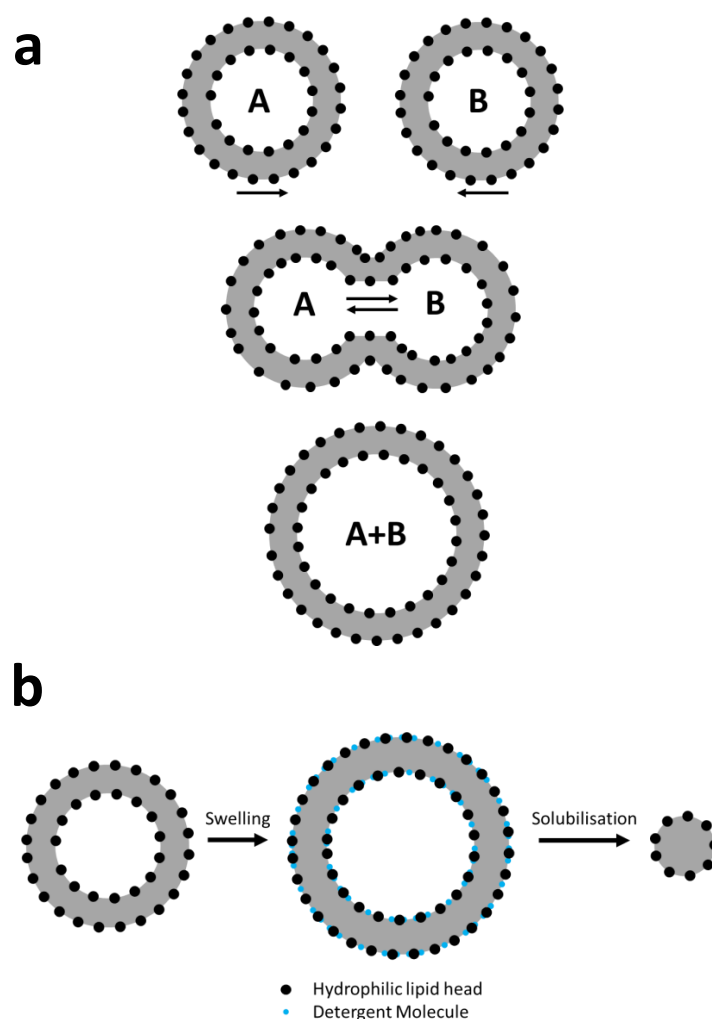
### 7.3.3. Monitoring the swelling and solubilisation of lipid vesicles with Triton X-100 using FCS

Phospholipid bilayers are important structures in life, representing, together with various embedded protein structures, the boundary of a cell in forming cell membrane [50]. This selectively permeable structure regulates the flux of essential cellular constituents and building blocks in and out of the cell. They consist of pairs of sheets of lipid molecules, which may be divided into two parts: a hydrophilic head and a hydrophobic tail. The sheets are arranged to shield the hydrophobic tails from exposure to water [50, 51].

In the field of pharmacological research, however, liposomes are synthetically manufactured and used as drug delivery vehicles [230]. The

fusion of these two structures, the transport liposome and the cell membrane, allows the contents to mix [230]. This process is illustrated in Figure 7.10a below.

A second property of liposomes is the fluidity of the surface. The lipids aren't bound in position, and, as a result, the structure is constantly changing [51]. The lack of bonding means that small molecules may intercalate and expand the structure, a process known as swelling. Previous work performed by our group has shown that detergent molecules such as Triton X-100 can cause this expansion, which we tracked using ensemble FRET. Moreover, detergents are commonly used in cellular biology to break up the cell walls and release the components. At high concentrations, the bilayer structure becomes unstable and breaks up, forming micelles, small lipid structures arranged with the hydrophobic tails contained inside, and the hydrophilic heads arranged exposed to the buffer. This whole process is summarised in Figure 7.10b.



**Figure 7.10:** a) Illustration of the process of fusion, showing the mixing of the contents of two initially distinct liposomes, A and B. The structure of the liposomes is summarised by the black circles which indicate the location of the hydrophilic lipid heads and the grey region indicating the location of the hydrophilic lipid tails. b) Summary of the solubilisation process, showing the intercalation of a detergent molecule (blue) into the lipid bilayer before the structures disintegrate and form small lipid micelles.

The process we chose to track was the swelling and solubilisation of dye-labelled liposomes. Liposomes were prepared using a standard extrusion protocol, using 0.1% Cy3-labelled lipid molecules in the preparation volume [154]. The maximum diameter of the liposomes was set by the filter to be 100 nm. Measurements were performed in T50 buffer, prepared as in

Chapter 3. The Triton X-100 was prepared in water, to a stock concentration of 100 mM. The measurements were performed with a vesicle concentration of 10 nM, with increasing concentrations of Triton X-100. The results of the scans are shown in Figure 7.11a.

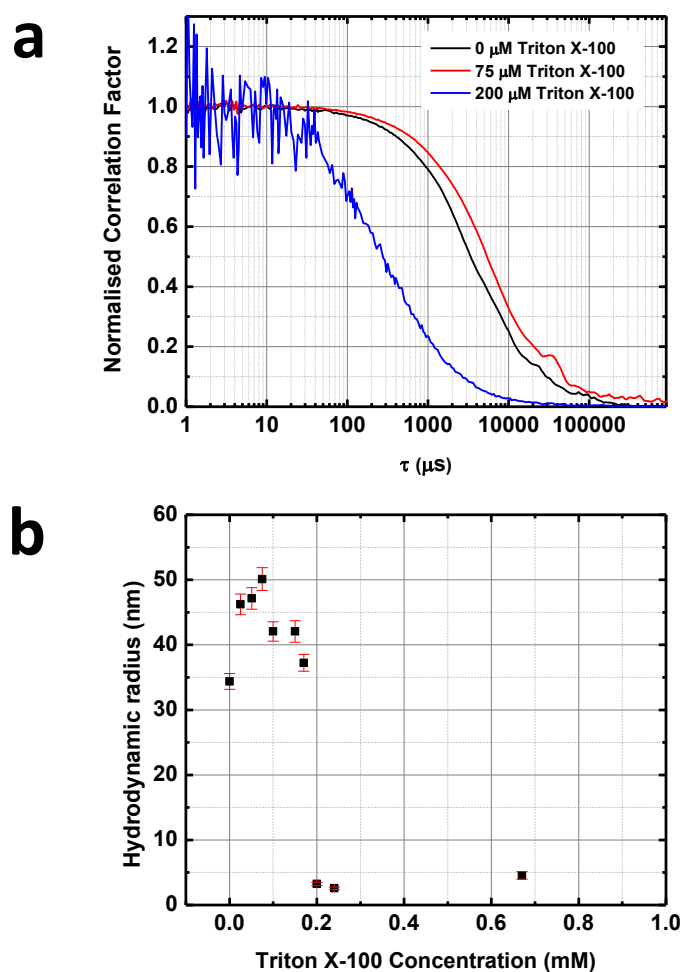


Figure 7.11: a) FCS traces illustrating the dynamic differences associated with the three stages in the process of liposome solubilisation: normal diffusion (black), swelling, resulting in slowed diffusion (red) and fully solubilised micelles (blue). b) Changes in the hydrodynamic radius of the liposomes induced by the action of the Triton X-100 intercalating the bilayer. The uncertainties in the value of the radius are shown in red.



The results shown in Figure 7.11a show the dynamics at three chemical conditions that represent the three stages illustrated in Figure 7.10b. The diffusion time is 3.55 ms for the first stage, where the liposomes are unperturbed by the detergent. The maximum extent of swelling occurs at a concentration of 75  $\mu\text{M}$ , which results in a diffusion time of 5.18 ms. When the concentration is increased further to 200  $\mu\text{M}$ , the liposome bilayer structure breaks apart, forming micelles. This is characterised by a sharp drop in the diffusion time to 335  $\mu\text{s}$ .

The diffusion coefficient fitted using Equation 7.6 are presented in Table 7.2, along with the errors and concentrations (derived from the  $G(0)$  term). The diffusion coefficients mirror the trend shown by the diffusion times. These values were used to calculate the hydrodynamic radius, using Equation 7.2. These values are plotted in Figure 7.11b.

**Table 7.2:** List of the diffusion coefficients and sample concentrations for liposomes under exposure to different concentrations of Triton X-100

[Triton X-100] ( $\mu\text{M}$ )	D ( $\mu\text{m}^2\text{s}^{-1}$ )	Standard error in D ( $\mu\text{m}^2\text{s}^{-1}$ )	[Sample] ( $\mu\text{M}$ )	Standard error in [Sample] (mM)
0	6.3	0.2	0.054	2.7e-4
0.025	4.6	0.2	0.044	2.2E-4
0.051	4.6	0.2	0.042	2.1E-4
0.075	4.3	0.2	0.048	2.4E-4
0.1	5.1	0.2	0.030	1.5E-4
0.15	5.1	0.2	0.36	2E-3
0.17	5.8	0.2	0.1	5.4E-6
0.2	66	5	1.53	0.02
0.24	83	6	1.35	0.02
0.67	47	6	8.1	0.2

The filter itself sets the maximum radius of each vesicle to 50 nm. Through repeated extrusions, larger structures that have formed are broken up and eventually the maximum structure size will be limited. Structures with radii below 50 nm will remain formed. Thus, this process does not regulate the minimum size of the liposomes, which is the reason why the diffusion coefficient characterises a smaller hydrodynamic radius than expected.

The large change in sample concentration is easily explained by understanding the mechanics of the reaction that is taking place. Each vesicle is broken up into multiple small micellar units; hence the number of sample units is vastly increased. This breakdown is also the cause of the noise in the trace for the micelles in Figure 7.11a. Due to the high reduction in the number of dye molecules per structure, the signal is much weaker than for the full vesicle structures and, as a result, the data appears much noisier.

#### **7.4. Moving from single- to two-colour FCCS**

The power of single-molecule FCS and FCCS to the field of biophysics is particularly relevant to the areas of molecular diffusion, tracking changes in motion due to changes in size, binding events and aggregation. Where one examines binding events, in particular, the technique focussed on one dye-

labelled species; the second species need not be labelled, a situation that enhances the biological validity of the technique.

The fluorescence is collected in a series of bursts as the molecules drift through the confocal excitation volume. Provided that the sample is prepared at a suitably low concentration, one may attribute each burst to a single molecule or single interacting pair when the experiment is set up to examine such samples. These bursts last for, on average, the diffusion time of the sample, a time which typically ranges from around 70  $\mu\text{s}$  for a single free dye to times in the millisecond regime, depending on the sample size. Mathematically, this relation between size and diffusion time was covered in the previous section through Equations 7.1 to 7.3.

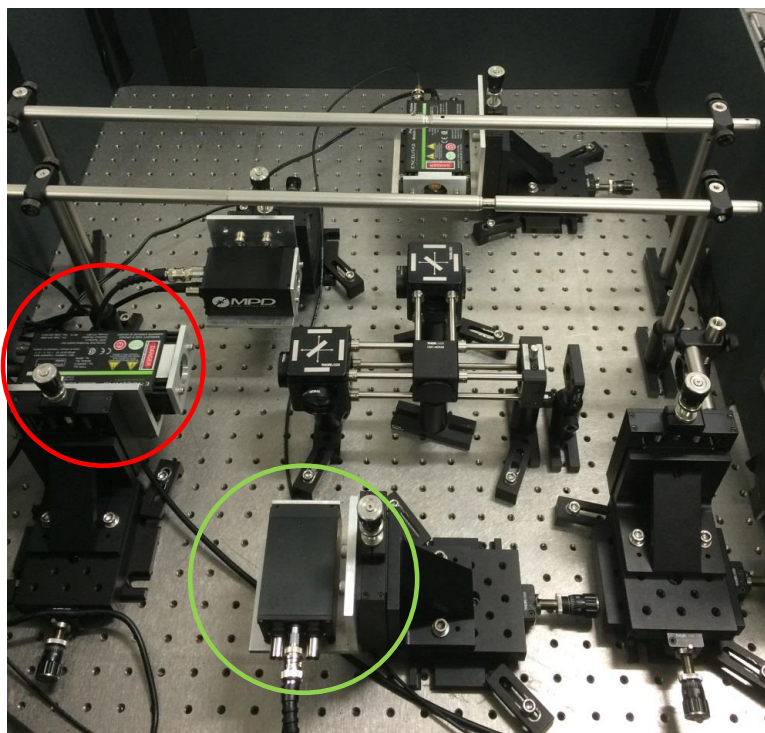
A logical extension of this is the introduction of a doubly-labelled sample, with a FRET dye pair. In this situation, two detectors must be employed and the emission from each dye resolved onto separate detector units. By doing this, one may monitor the changes in proximity of these two dyes using sm-FRET during each burst of emission. This allows the resolution of dynamics at a time resolution in the microsecond regime, significantly lower than with the TIRFM setup used to probe intramolecular dynamics at the single-molecule level in the previous chapters.

### 7.5. System modifications to optimise for two-colour detection

The detectors used in the sm-FCS and sm-FCCS experiments performed for the previous chapter were by Micro Photon Devices and have a quantum efficiency of detection peaking at 49% at approximately 550 nm. This efficiency, however, drops significantly from this peak. In particular, for a standard FRET pair of Cy3 and Cy5, detecting the acceptor emission which peaks at around 645 nm reduces this efficiency of detection down to only 37%, and at the upper bound of the wavelength of the spectrum of 700 nm this drops further to 30%. It is clear that a more suitable detector is required.

In light of this, a different SPAD unit, manufactured by Excelitas Technologies, was chosen to receive the emission from the acceptor channel (Excelitas SPCM-AQRH-14). This unit has a detection efficiency of approximately 72% at 700 nm, and more importantly this efficiency is in excess of 70% within the range of 650 nm to 750 nm, making it perfect for detection of Cy5 dye emission.

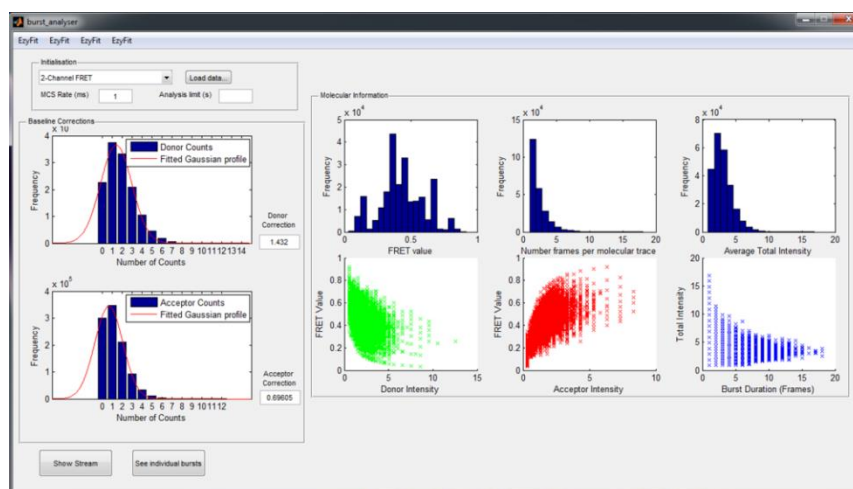
The setup for the detection optics is shown below in Figure 7.12. The detectors used for sm-FRET are shown circled, with the donor detector in green and the acceptor detector in red. The detection was set at a single polarisation in order to ensure that the bursts would be properly correlated, thus ensuring that only the raw intensity transfer between the dyes was probed and not the orientations relative to each other.



**Figure 7.12:** Photo of the optics used to resolve the emission for a free-diffusion sm-FRET experiment. The detectors used within the schematic are circled in green (donor) and red (acceptor).

## 7.6. Extracting FRET information from individual intensity bursts

The photon streams from the detectors are recorded with the native software for the TCSPC cards, as in the case of single-colour FCCS. For the two-colour situation, however, the software has a limitation in that it cannot be used to extract any FRET information. With this in mind, I wrote a program in MATLAB to extract this information from the raw intensity recorded from the program. Given the temporal sensitivity of the data, this offers the scope of recording dynamic behaviour at a resolution of  $10 \mu\text{s}$ , three orders of magnitude greater than the wide field EM-CCD detectors used in the TIRFM-based setup.



**Figure 7.13:** Screenshot of the program designed to analyse the fluorescence bursts, with sample data loaded. The plots illustrating the calculation of the baseline for the donor and acceptor signals are shown in the ‘Baseline corrections’ panel. In addition, in the molecular information panel, the following plots are presented (clockwise from top left): average molecular FRET histogram, burst length histogram, average total intensity histogram for each burst, total intensity per burst vs burst duration, average acceptor intensity vs the average molecular FRET and average donor intensity vs average molecular FRET.

### 7.6.1. Extracting the background count level from each channel

The data recorded was saved in a two-columned ASCII format directly from the Becker&Hickl acquisition software, the first representing the time and the second the intensity information. The program loads in the donor stream first, followed by the acceptor stream, after the integration time for the streams is entered. The low count rates from these experiments mean that the data takes a pixelated appearance, as can be seen for the lower three scatterplot panels in Figure 7.13 above.

In order to establish an estimate of the FRET values, one must baseline both spectra. The average count rate was calculated for each spectrum

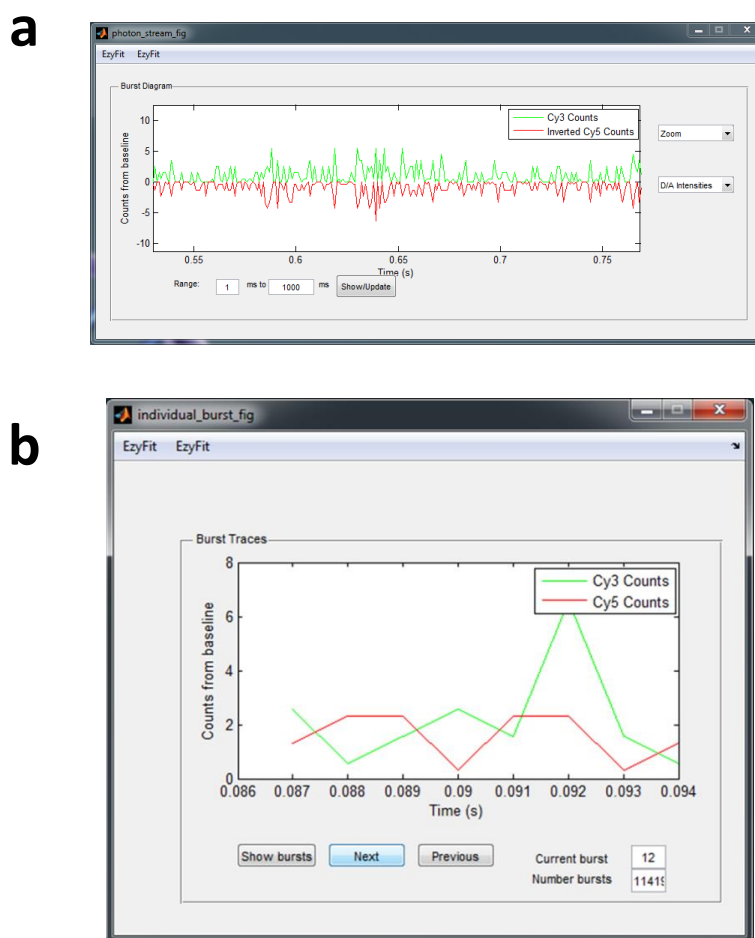
separately by the following procedure. Each trace was binned by the number of photons detected per frame, and the results plotted in a histogram. The photon distribution was fitted with a Gaussian curve, as shown in Equation 7.11, and the peak position ( $B_1$ ) was taken to be the background level. For both the donor and acceptor spectra, the respective background values were subtracted from every photon counting data point, with any negative instances set to zero. This baselined intensity information is plotted in a separate figure, known as the Photon Stream plot in the program. A sample screenshot of this window is shown in Figure 7.14a below.

Once the baselines are established, the FRET efficiency is calculated, through the ratios of the emission intensities, taken in terms of the photon count rate (Equation 7.12):

$$N(x) = A_1 e^{-\left(\frac{x-B_1}{C_1}\right)^2} \quad (7.11)$$

$$E_{app} = \frac{I_A}{I_A + I_B} \quad (7.12)$$

The individual bursts may be viewed in a separate figure, launched by the ‘See Individual Bursts’ command button. A sample of this figure, with data loaded, is shown below in Figure 7.14b.



**Figure 7.14:** a) Screenshot of the full photon stream plot, with a zoom taken within a specific region. The donor signal is shown in green, and the acceptor signal inverted and plotted in red to allow a clear visualisation of the dynamics. The range of view may be chosen using text inputs below the plot, and the plot itself panned and zoomed to examine regions of interest. In addition, one may view the total intensity as opposed to the separate donor and acceptor channels. b) Screenshot of the figure for viewing individual photon bursts. The donor and acceptor channels are coloured as for the previous figure, but the acceptor is not inverted. The total number of bursts is shown as well as the index for the current burst within the list.

### 7.6.2. Extracting and analysing the individual fluorescence bursts

The analysis of dynamic behaviour hinges upon being able to isolate the individual fluorescence bursts from within the full photon stream. In MATLAB, I wrote code which achieves this goal in the following way. A zero is introduced at the start and end of the FRET vector calculated in the



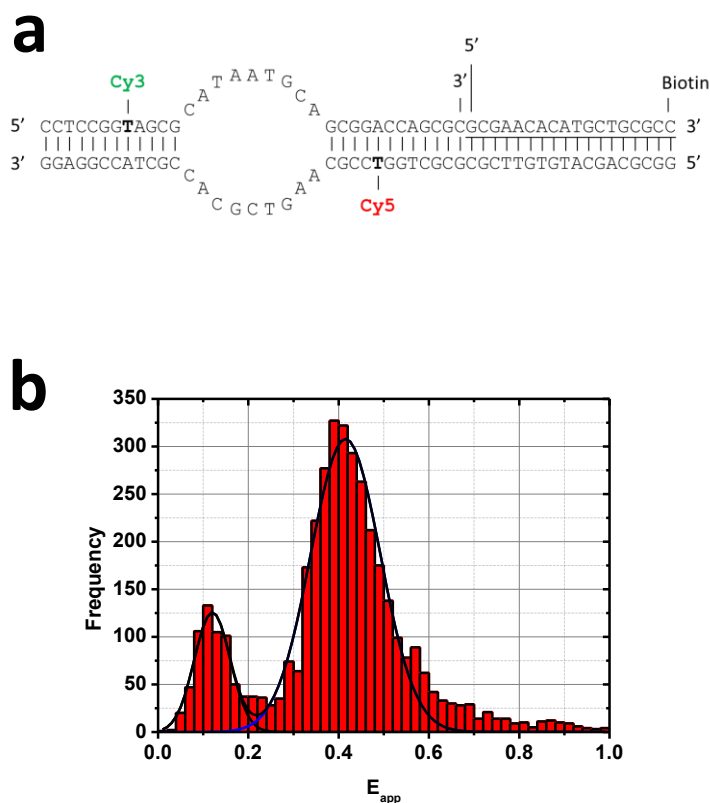
previous part of the code. This allows the initial burst calculation to occur at row 1 of the vector, and to finish if necessary at the last row.

A difference test is performed upon this extended FRET vector; at the start of a non-zero FRET burst, a value of one is registered, and at the start of a zero FRET section, a value of negative one is returned. At all other positions, the status is not changing and thus a zero is entered in the edge array vector. Having determined the positions for which the molecules are present and absent in the excitation volume, the indices for each molecule are determined. The first photon emitted by a molecule entering the confocal volume and recorded by the detector is indicated by a one in the edge array. As a result, all instances of this value are searched and their indices tabulated. In the same way, a value of negative one is indicative of the start of a region where no molecules are present or emitting. As a result, the index for the last emission frame of the burst is given by one less than the location of all occurrences of negative one. These modified indices, too, are tabulated, showing a complete list of the starting and finishing indices for each molecular burst.

Having extracted these indices, the average FRET value per burst is calculated and plotted in a histogram, along with histograms of the burst durations (in frames) and total intensity for each burst. In addition to this, for each burst the average donor and acceptor intensities are plotted against their calculated average FRET value. Finally, the burst duration is plotted against the total intensity for each molecule.

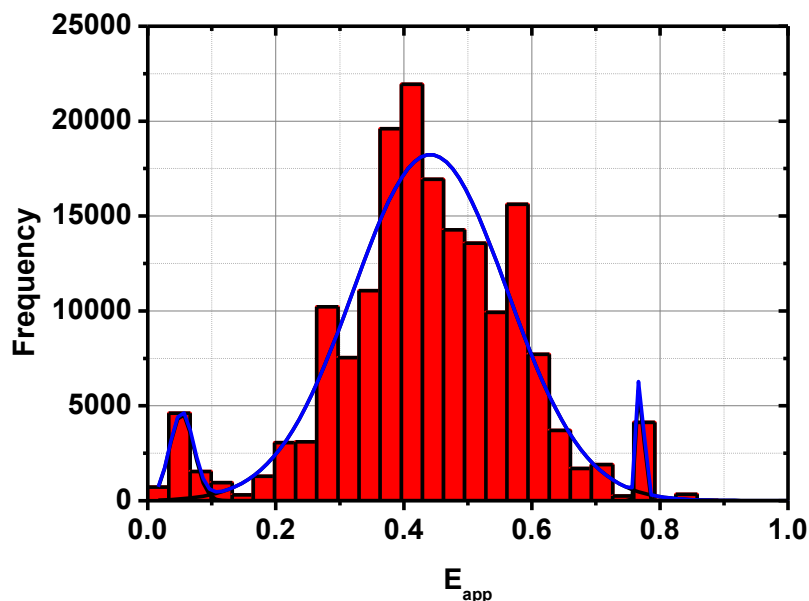
## 7.7. Experimental test of non-immobilised sm-FRET for a doubly-labelled DNA structure

To validate the accuracy and effectiveness of the system, a model structure was examined of known FRET. The chosen model was a double-stranded DNA bubble, as shown in Figure 7.15a below. The sample was constructed from three separate strands, purchased from (PrimeTech, Minsk); a Cy3-labelled strand 33 nucleotides long, a Cy5-labelled strand 51 nucleotides long and a biotin-labelled handle to allow surface immobilisation using conventional TIRFM techniques (see Chapter 3 for more detail). The bubble itself was 9 nucleotides in length. This sample, when investigated in a conventional surface-immobilised sm-FRET setup, presented an apparent FRET value of  $E_{\text{app}} = 0.415 \pm 0.003$  (Figure 7.15b).



**Figure 7.15:** a) Structure of the DNA bubble used in the free-diffusion sm-FRET experiments. The biotin-tagged linker is shown as the underlined sequence, and the positions of the dyes indicated by a bold nucleotide base, to which the labelled dye (Cy3 or Cy5) is attached. b) Results from a TIRFM sm-FRET measurement on the sample, showing the peak value of  $E_{app} = 0.415$ . The lower peak is the donor-only peak resulting from incomplete hybridisation or a photobleached acceptor dye.

With this value recorded, the structure was analysed using the non-surface immobilised sm-FRET setup in order to check the validity of the system. The scans were performed for a time of 1000 s, in order to allow a full FCCS scan to be taken in addition to the sm-FRET populations. The background subtraction was performed as described in the previous section, and the plots of the data retrieved and replotted in OriginPro (Figure 7.16).



**Figure 7.16:** Histogram showing the FRET distribution for the DNA bubble as retrieved by the free-diffusion sm-FRET measurement. The Gaussian fitted to this data has a peak centre of  $E_{app} = 0.441 \pm 0.008$ .

The population histogram shown in Figure 7.16 shows the average FRET for each molecular burst isolated from the full photon stream. A Gaussian fit to this data shows a peak value of  $E_{app} = 0.442 \pm 0.009$ , which is in reasonable agreement with the surface-tethered study.

One thing to note is the population number itself: the TIRFM data is built from 3914 molecular contributions, which were recorded from 15 separate movies. The free-diffusion data came from one single scan of 1000 s but presented a total of 176010 molecular contributions. One can clearly see that the volume of data yielded from a single scan using the non-immobilised system is far in excess of that from conventional TIRFM and provides a consistent result. Although the free-diffusion scan itself takes 4 minutes longer to complete than all of the TIRFM movies combined, when one takes into account the additional preparation time for channel

preparation and immobilisation reagent incubation it is, in fact, a faster experiment to perform.

## **7.8. Towards single-molecule multi-parameter fluorescence spectroscopy (sm-MFD)**

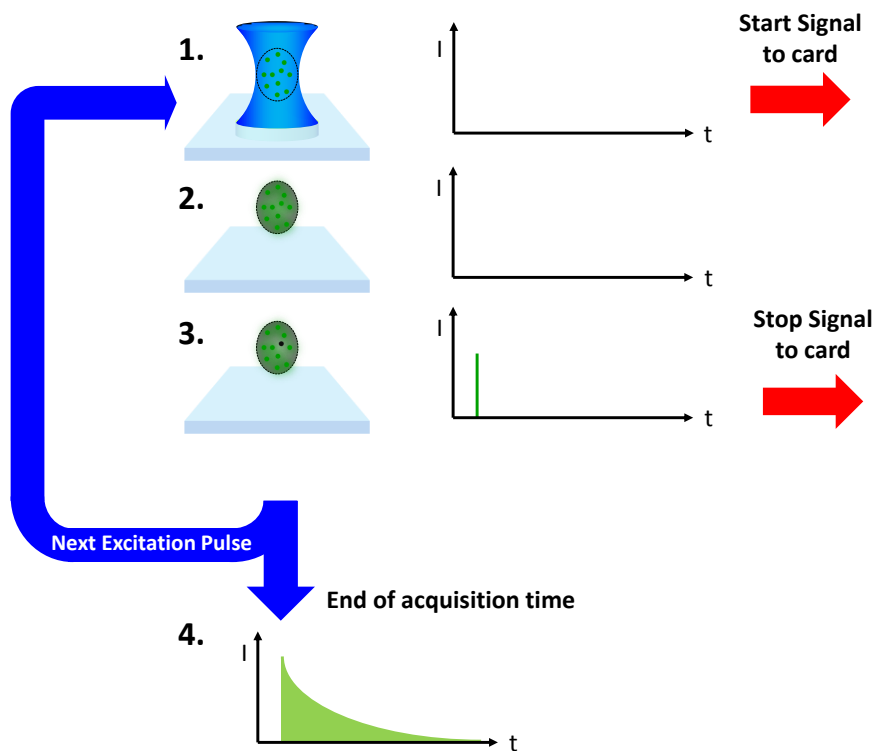
### **7.8.1. System modifications to measure fluorescence lifetimes**

If one considers the diffusion of the molecules in the sample to be an isotropic process, then at any given time there should be approximately the same number of molecules within the excitation volume at any moment in time.

The continuous-wave (cw) laser source offers a consistent and high level of fluorescence output, though due to the nature of this repeated molecular excitation if one considers the time-dependent intensity collected from the excitation volume then this should be approximately constant, though with small fluctuations due to deviations in the numbers of molecules from this average and bleaching events, for example. This is since as soon as a molecule emits a photon and returns to its ground state, it is immediately re-excited and thus no global change in the emission intensity can be observed.

When the source is a pulsed laser, the pulse frequency plays a significant role. A single high-energy pulse will fully excite the sample. However, in the time between excitation pulses the fluorescence intensity recorded by the

detectors will decrease. For high pulse frequencies ( $>250$  MHz), with a temporal separation below the fluorescent lifetime of the dye-labelled sample, a result similar to that recorded for cw excitation is observed. If, however, the pulse frequency is reduced to a temporal separation in excess of the fluorescent lifetime of the dye, then an exponential decay in the fluorescent intensity is recorded, as described in Chapter 1, Section 3.4. In the situation where the sample carries a FRET pair, the decay time is further reduced, as discussed in Section 4.3 of Chapter 1. A summary of the pulsed excitation lifetime measurement for a confocal microscope is presented in Figure 7.17 below.



**Figure 7.17:** Summary schematic of the process of obtaining the fluorescence lifetime for a dye-labelled sample using a confocal microscopy setup. The excitation pulse illuminates the sample within the confocal volume (1), simultaneously starting the timing on the computer card, and fully excites all molecules present (2). Each molecule decays in some time, emitting a photon. The first detection event stops the timing, adding the photon to the lifetime histogram (shown at stage 3). This process repeats for the duration of the experiment, until the final photon is recorded, building the full profile shown in (4).

By taking advantage of the nanosecond-scale time recording capabilities offered by SPAD units, one can record fluorescent lifetimes for these freely-diffusing samples through the standard software offered with the computer cards by a simple switch of the excitation source. For the purposes of our work we switched to a Picoquant LDH-P-C-470B pulsed laser diode head with a peak excitation wavelength  $\lambda_{\text{exc}} = 470 \text{ nm}$ , and maximum pulse widths of 70 ps at 0.8 mW (40 MHz repetition rate) and 500 ps at 4 mW (40 MHz repetition rate).

### 7.8.2. Extracting polarisation information from single-molecule emission: fluorescence anisotropy

#### Fluorescence Anisotropy

When in solution, fluorescent molecules have random orientations and are permanently in a state of translational and rotational motion. As a result, when such a sample is illuminated with an unpolarised light source, for example, a flash lamp, all of the molecules should be excited and have the capacity to emit.

When a photon is emitted, it not only has energy but a polarisation associated with it. Therefore, when such an unpolarised source excites all of the fluorescent molecules in a sample, the emission too is unpolarised.

However, if the excitation source is polarised then the situation is quite different: only those molecules whose position at the moment of photon incidence will allow the absorption to take place and promote an electron to an excited state. This selectivity also has an effect on the emission of light upon relaxation. The emitted light will have a polarisation that falls within a range determined by the fluorescence lifetime of the dye molecule ( $\tau$ ) and the rotational lifetime of the molecule ( $\phi$ ). The spread of the emission polarisation is known as the anisotropy of the sample ( $r$ ) and is defined as in Equation 7.15:



$$r = \frac{r_0}{1 + \frac{\tau}{\phi}} \quad (7.15)$$

Where  $r_0$  is the fundamental anisotropy of the molecule, which would be the value recorded for the sample in the absence of any motion [231].

### **Experimentally determining fluorescence anisotropy**

The two stages in the evolution of this system have used two different means of separating the fluorescence emission onto separate detectors. In the sm-FCS system discussed initially, the light was split by polarisation using a cubic beam splitter, directing the vertical and horizontal components onto two identical SPAD detectors. When the system was modified to investigate sm-FCCS with a two-colour system, the cross-correlation was performed for fluorescence emission with the same polarisation with the light separated by wavelength onto two different SPAD detectors (MPD PDL for Cy3 emission, Excelitas SPCM for Cy5 emission).

When one records data on all four detectors simultaneously, the system has the capability to extract polarisation information from the fluorescence emission. There are two quantities that may be calculated: the polarisation ( $P$ ) and the anisotropy ( $r$ ), calculated using Equations 7.13 and 7.14 respectively.

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \quad (7.13)$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad (7.14)$$

The system constructed for the purpose of this work consists of two MPD and two Excelitas SPAD units, for detecting the donor and acceptor emission respectively. These were, as stated previously in this chapter, initially chosen for their excellent quantum yields in the emission ranges for Cy3 and Cy5 respectively. Crucially, in the case of the MPD units, this quantum yield is still high for detecting Alexa488 emission ( $\lambda_{em}(\max) = 519$  nm), one possible donor suitable for excitation with the pulsed laser described above.

### **Correction for effects of emission polarisation on detection efficiencies**

In practice, it is necessary to take into account not only the polarisation dependence of the absorption of the sample on but also the effect of polarisation on the detection efficiency. As a result, it is important to introduce a correction factor ( $G$ ) which compensates for the discrepancy between detection efficiencies for vertically and horizontally polarised light. By including this factor, the true anisotropy measured by a fluorescence experiment is calculated through Equation 7.16.

$$r = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} + 2GI_{\perp}} \quad (7.16)$$

$$\text{where } G = \frac{I_{\parallel}}{I_{\perp}} \quad (7.17)$$

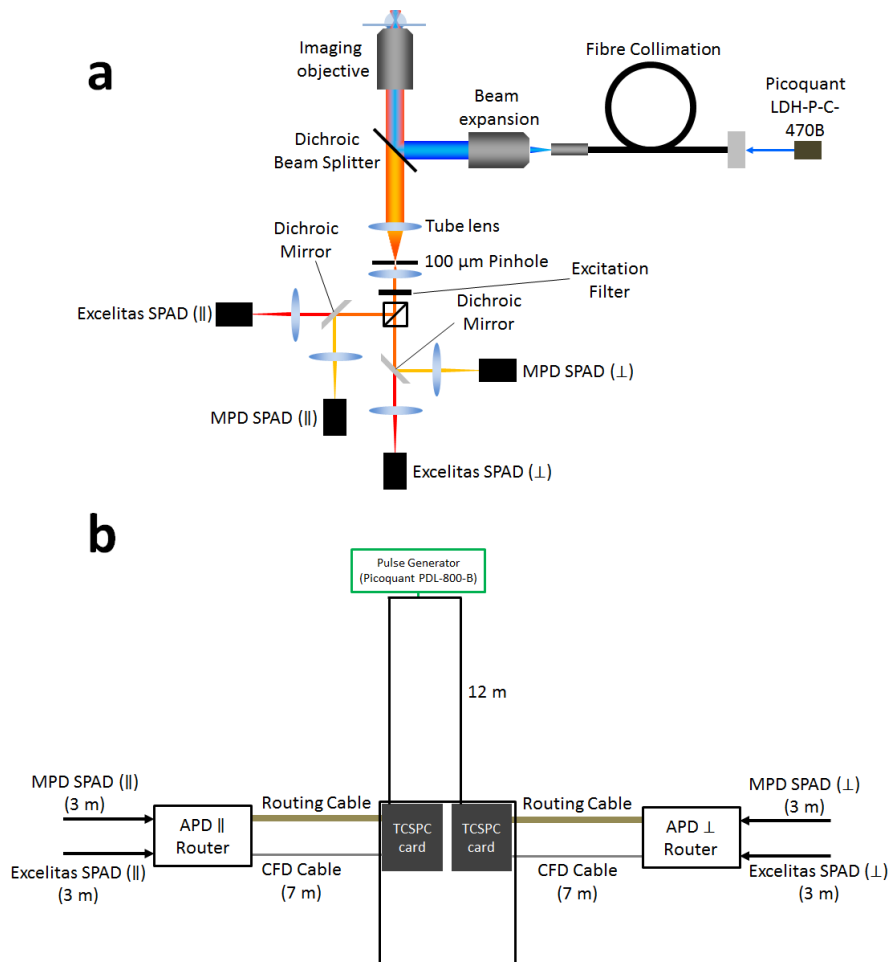
One typically calculates this factor by illuminating a known concentration of dye molecules using the polarised experimental excitation scheme and taking the ratio of the recorded intensities as shown in Equation 7.17. When a two-colour setup is employed, both colour channels must be calibrated separately since the emission detection efficiency is also determined by the wavelength of the emission.

### 7.8.3. Electrical circuitry for sm-MFD

The use of a pulsed excitation source requires no additional consideration into the timing sequence of the data acquisition cycle; in the cw situation, the driver itself was used to drive the system, with the regular pulses used to stop the data acquisition through the TCSPC card stop input. Given that these pulses are now concurrent with the excitation (due to both being provided by the laser driver) the system is stopped by the laser pulse. The signal is, as in the previous stages of development, delayed using an additional 2 m of cable length than the photon's electrical path (Figure 7.18b), equating to a difference of approximately 4 ns.

The only difference is that all four SPAD units must be connected, two to each router. The routers are divided by polarisation to maintain the possibility of performing sm-FCCS and sm-FCS without disconnecting

cables and risking misalignment. The optical arrangement that I designed and built is presented in Figure 7.18a. This was built around the previous arrangements from sections 7.2 and 7.4.



**Figure 7.18:** a) Optical layout constructed to realise sm-MFD. The four detectors are labelled to aid the electrical diagram shown in b). All components common to the previous stages of the evolution of the apparatus are unchanged; this helps retain the flexibility of the system and the possibility to take a step back in complexity to perform sm-FCS and FCCS measurements without disruption of the full setup. b) Electrical diagram for the sm-MFD system, including cable sources and lengths where appropriate.

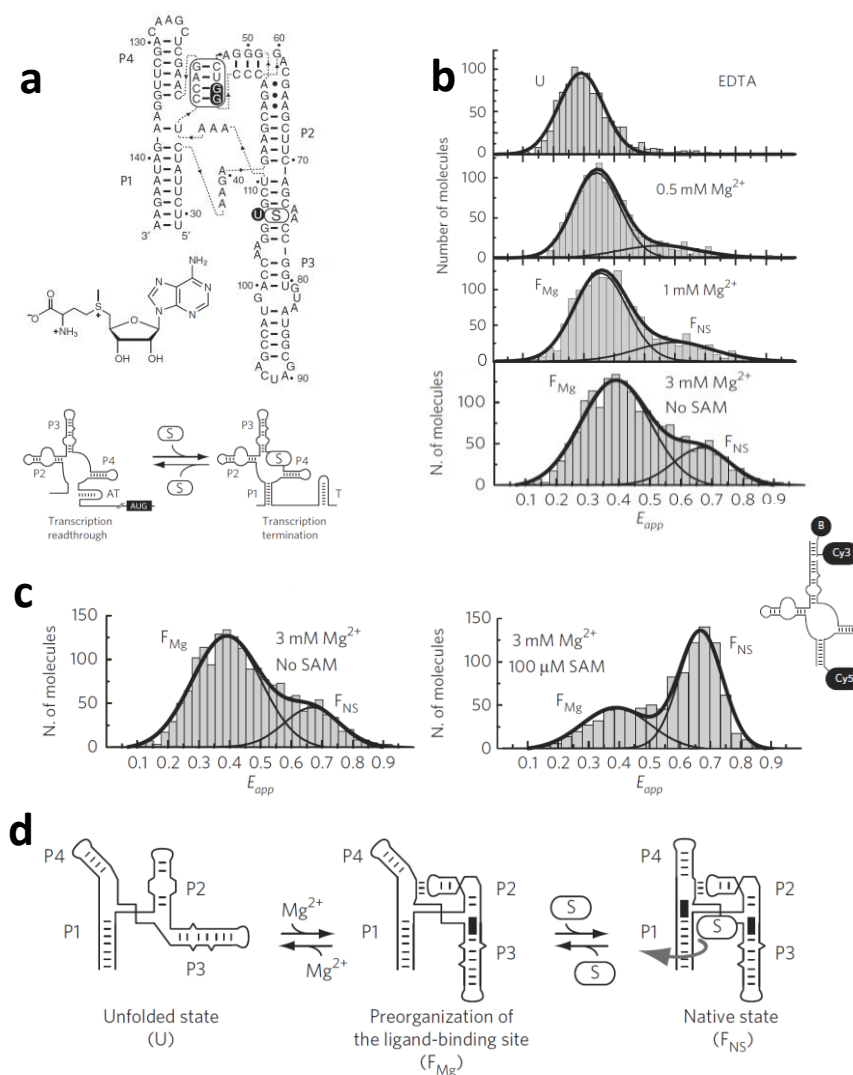
## 7.9. Experimental verification of the sm-MFD system: Examining a previously-characterised structure

### 7.9.1. SAM-I riboswitch studied using sm-FRET shows three conformational states

The sm-MFD system was tested using a known riboswitch, the SAM-I aptamer, which had been fully characterised at the single-molecule using sm-FRET level by our group [159]. The aptamer structure examined in this study is shown in Figure 7.19a. The studies indicated three conformational states: with only EDTA in the buffer to remove all metal ions from solution, the aptamer existed in an unfolded state of  $E_{\text{app}} \sim 0.28$  [159]; with the addition of a small concentration of  $\text{Mg}^{2+}$  (0.5-1 mM) induced a shift in this population to an intermediate conformation with an apparent FRET of  $E_{\text{app}} \sim 0.4$ ; higher concentrations of  $\text{Mg}^{2+}$  promoted a small fraction of the structures to the native, folded state ( $E_{\text{app}} \sim 0.62$ ) (Figure 7.19b); and finally, at 3 mM  $\text{Mg}^{2+}$  and 100  $\mu\text{M}$  SAM ligand, the structure exists predominantly in the folded conformation ( $E_{\text{app}} \sim 0.62$ ) (Figure 7.19c). These states were calculated by using an aptamer labelled with the donor in stem P3 and the acceptor on stem P1.

This leads to the conclusion that the aptamer core is pre-formed before ligand binding can occur, and this intermediate state was a result of site-specific binding of  $\text{Mg}^{2+}$  to stabilise this aptamer core. The folding model was thus proposed as described in Figure 7.19d. In the absence of any  $\text{Mg}^{2+}$ ,

the structure is fully unfolded, with no tertiary contacts made between any of the stems. With the addition of 0.5-1 mM  $Mg^{2+}$ , the ligand binding pocket is coordinated, resulting in tertiary interactions between the loop motif at the end of stem P2 and the bulge in stem P4. Finally, at higher concentrations of  $Mg^{2+}$  (3 mM) and with the ligand present, the structure is stabilised in this conformation, adding further rigidity and drawing the dyes closer still.



**Figure 7.19:** Summary of the results and conclusions from the sm-FRET study of the SAM-I riboswitch, all panels taken from or adapted from [159]. a) Sequence of the aptamer used (main), with the ligand S-adenosylmethionine (SAM) structure shown below, as well as the gene regulatory action induced by ligand binding. The riboswitch is a negatively-controlled structure. b) FRET histograms showing the evolution of the conformational populations with increasing magnesium concentrations. The shift from the unfolded state (U) to the magnesium-only intermediate ( $F_{Mg}$ ) to the folded native structure ( $F_{NS}$ ) can be seen. c) Influence of the ligand molecule in stabilising the folded state, clearly shown by the changes in populations of  $F_{Mg}$  and  $F_{NS}$  with the ligand is introduced. d) Proposed 2-stage folding mechanism for the riboswitch, with the first induced by the  $Mg^{2+}$  ions alone and the second stage due to ligand capture.

With these three distinct conformational states observed to exist, a construct was designed to be studied using sm-MFD in order to observe the

presence of these three states and thus test the performance of the built smMFD microscope. The structure of the riboswitch was kept the same; the sequence was as shown in Figure 7.19a, and the labelling positions were also retained, but in order to work with the 470 nm excitation source the donor dye was switched for Alexa488. Based on the results of this scan, three sets of chemical conditions were examined, all at 50 mM Na<sup>+</sup>: 0 mM Mg<sup>2+</sup>, 3 mM Mg<sup>2+</sup> and 3 mM Mg<sup>2+</sup> with 100 μM SAM ligand.

### 7.9.2. Investigating SAM-I riboswitch population changes using sm-MFD

The samples were prepared from the hybridisation of four constituent strands, whose sequences are listed below in Table 7.3. The two strands without any chemical modifications were prepared by transcription as described in [115], where a DNA template is used to generate an mRNA strand. The modified strands were purchased from IDT and deprotected and purified as described in Chapter 3. The riboswitch was constructed by the same procedure reported by Heppell et al [159], with equimolar quantities added to a hybridisation buffer.

All imaging was performed using samples suspended in 50 mM trizma hydrochloride and with 1 mM trolox in the solution to aid photostability, in order to keep chemical conditions consistent with those from the TIRFM-based sm-FRET results.

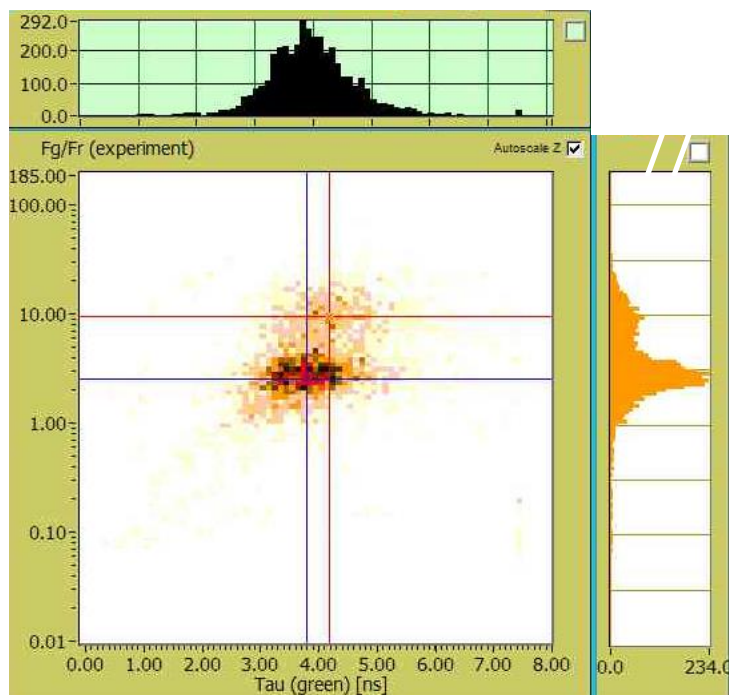
There is no open source analysis program available to process sm-MFD data. Although some companies, namely Becker & Hickl and Picoquant, and



some research groups such as Claus Seidel's group in Düsseldorf offer packages the price in all cases in the region of £15,000. The high commercial price of the software in addition to the time limitations for writing our own software, we contacted Dr Steven Magennis at the University of Glasgow, with whom we are establishing a collaborative effort to write an open source program capable of processing and presenting sm-MFD data efficiently, and he performed the processing of the data. We should mention that the setup at the University of Glasgow is to the best of our knowledge the only currently existing smMFD microscope in the UK in addition to the one implemented now at St Andrews. The results of the scans at the stated chemical conditions are presented in Figures 7.20 to 7.28 below.

### **Examining the unfolded SAM-I aptamer structure using sm-MFD**

The first condition examined was the  $Mg^{2+}$ -free case, where the molecules exist in the unfolded population which, using surface-immobilised sm-FRET, yielded an apparent FRET value of  $E_{app} \sim 0.28$ . With this in mind, a scan was taken under these conditions and the fluorescence lifetime and fluorescence anisotropy results are plotted in Figures 7.20 and 7.21, with a further plot showing the intensity emissions over the full scan duration confirming the presence of a single two-dye system (Figure 7.22).



**Figure 7.20:** Plot of the fluorescence lifetime obtained as a function of the ratio of donor emission ( $F_g$ ) to acceptor emission ( $F_r$ ). There are two populations visible; this is due to incomplete labelling of the aptamers in the sample, where a donor-only population is visible. The data is binned in x, y and z and the histograms plotted used to calculate the peaks of the lifetime and intensity distributions.

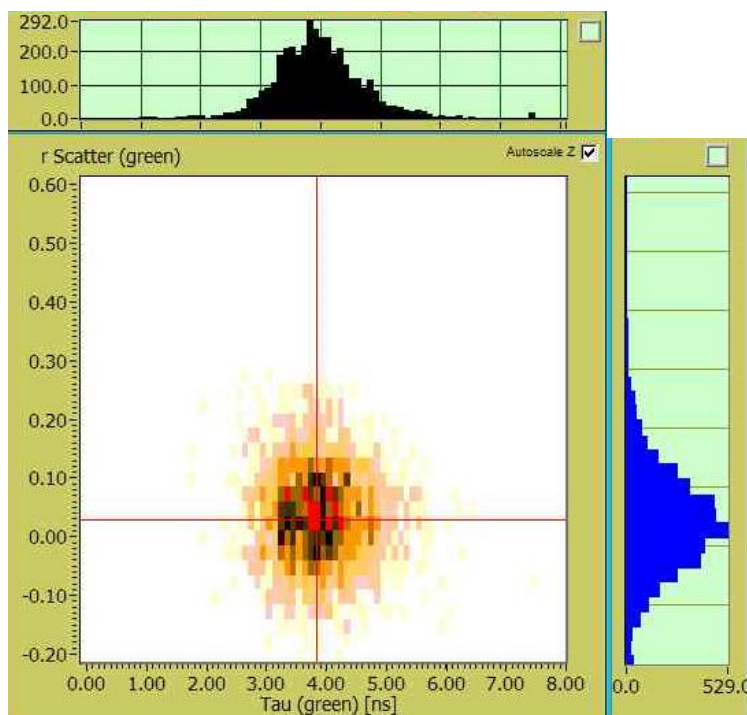
Through a simple rearrangement of Equation 7.12 we obtain an expression for the FRET in terms of the ratio of donor to acceptor emission (Equation 7.18).

$$E = \frac{1}{1 + \left(\frac{I_D}{I_A}\right)} \equiv \frac{1}{1 + \left(\frac{F_g}{F_r}\right)} \quad (7.18)$$

By applying this equation we can obtain a FRET value for this population of  $E = 0.28 \pm 0.06$ , which is identical to that determined for the unfolded state using sm-FRET. This offers an early indication that the technique is consistent with the observations made using conventional sm-FRET.

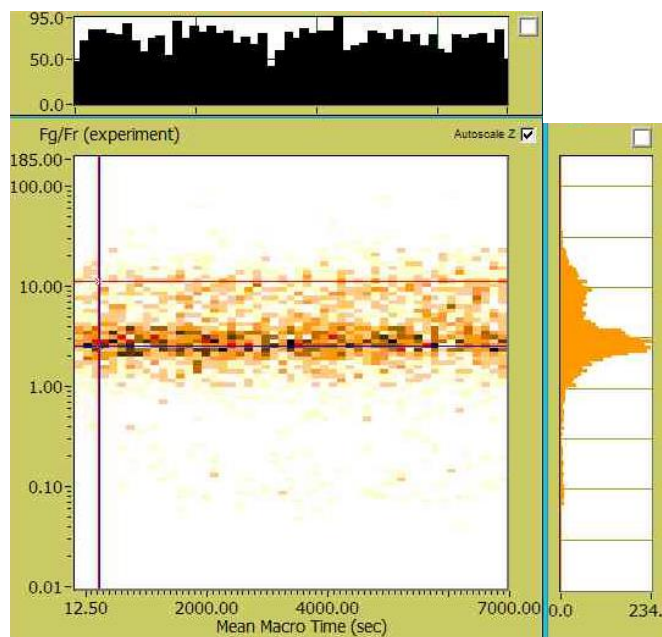
It is clear visually that there are two distinct populations. The largest population is a result of the unfolded conformation of the dual labelled aptamer, which exhibits a lifetime of  $3.85 \pm 0.34$  ns, detected with a ratio of donor to acceptor emission of  $2.52 \pm 0.51$ . The smaller population is due to incomplete labelling, where the aptamers have only the donor dye present. This is confirmed by two observations from the plot; first the  $F_g/F_r$  ratio indicates that the donor is the only emitting species, and second the lifetime for this species is  $4.24 \pm 0.08$  ns, and the lifetime of the free dye is 4.1 ns [232]. The ability to separate donor-only labelled molecules from doubly-labelled species constitutes one the advantages of sm-MFD over TIRFM-FRET. In the latter, the leakage from the donor channel into the acceptor channel introduces an apparent FRET peak, which depending on the particular TIR microscope can range between 0.1 (10% leakage) and 0.15 (15% leakage). This apparent FRET, known as ‘zero peak’, limits the lowest FRET value that can reliably be assigned to a structural conformation. In contrast, the ability of MFD to separate the molecules according to their lifetime allows us to filter ‘donor-only’ species, and thus, be able to measure much longer inter-dye distances.

The fluorescence anisotropy was then calculated as in Equation 7.16, with the correction factor (Equation 7.17) included. This yielded data shown in Figure 7.21, where the anisotropy is plotted as a function of the fluorescence lifetime recorded.



**Figure 7.21:** Plot showing the recorded fluorescence anisotropy as a function of the lifetime recorded from the sample. Histograms in x and y are fitted to calculate the peak values in each dimension.

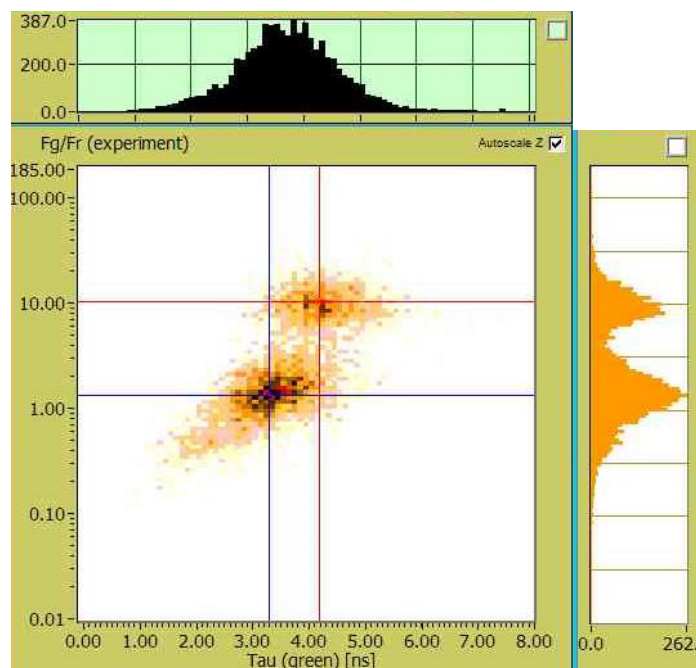
The plot shown in Figure 7.21 yields a value of  $0.0275 \pm 0.0051$  for the recorded anisotropy ( $r$ ). This low value for the anisotropy is indicative of a freely rotating single dye. To assess the both the photostability of the dyes in the sample and also the concentration fluctuations in time, a third plot is created which plots the ratio of green to red fluorescence against the macro time. This is plotted in Figure 7.22.



**Figure 7.22:** The ratio of the intensities of the donor ( $F_g$ ) and acceptor ( $F_r$ ) intensities plotted against time. The histogram plotted above the plot illustrates the fluctuations due changes in molecular numbers in the excitation volume. It can be seen that the intensity deviates around an average value. The  $F_g/F_r$  histogram shows the same two populations as evident in Figure 7.20, with the high  $F_g/F_r$  ratio indicating donor-only structures and the lower ratio to the structure with both dyes present.

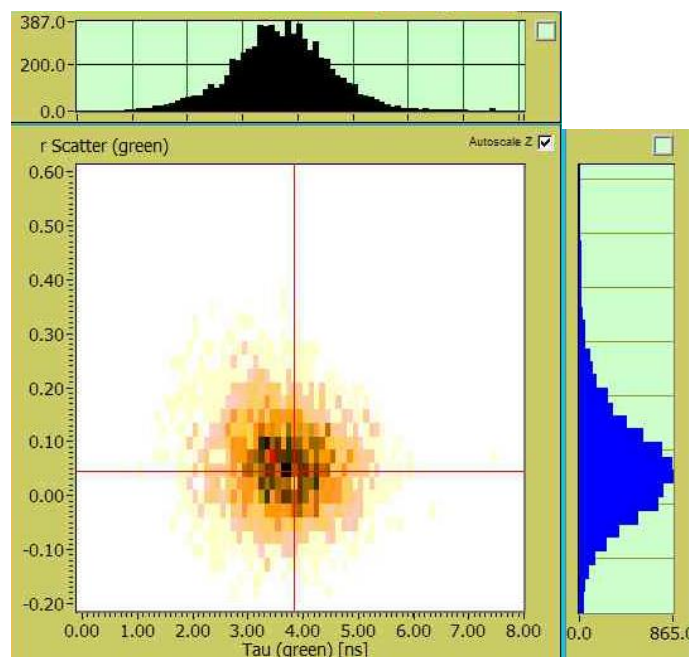
### Examining the magnesium-induced intermediate state using sm-MFD

The same experiment was performed with  $Mg^{2+}$  in the buffer, in order to promote the folding of the structure into the intermediate and folded conformations, as shown in Figure 7.19b. The sm-FRET populations for these two conformations are  $E_{app} \sim 0.4$  and  $E_{app} \sim 0.62$  respectively, with the aptamer predominantly in the intermediate state at these chemical conditions [159]. The results are presented in Figure 7.23 below.



**Figure 7.23:** Plot showing the relation between donor lifetime and donor and acceptor ratio ( $F_g/F_r$ ). Two populations are still visible, representing the donor-only and donor-acceptor populations as for the  $\text{Na}^+$ -only situation. Histograms illustrate the two populations, clearly distinguishable in the intensity axis but due to the proximity of the lifetimes the overlap is too strong to clearly resolve both populations separately.

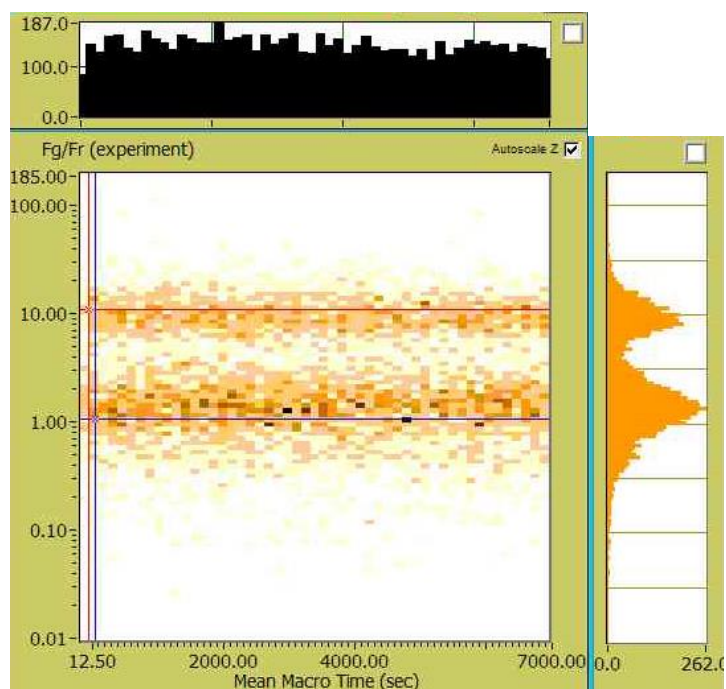
The plot in Figure 7.23 shows, as before, two distinct populations. As in the  $\text{Na}^+$ -only scan, both the intensity ratio  $F_g/F_r$  and the lifetime ( $4.24 \pm 0.07$  ns) indicate that the high lifetime population (red lines) is the donor-only population, and the population with the lower lifetime is the dual-labelled species. This population has a lifetime of  $3.35 \pm 0.32$  ns, and the  $F_g/F_r$  ratio of  $1.29 \pm 0.11$  leads to an apparent FRET of  $E_{\text{app}} = 0.44 \pm 0.04$ . This value compares favourably with the value obtained for the magnesium-induced intermediate from sm-FRET, which gave  $E_{\text{app}} \sim 0.4$ .



**Figure 7.24:** Plot of the anisotropy calculated as a function of the fluorescent lifetime. Histograms for each dimension are plotted, showing the spread of values with respect to the maximum for each dimension.

The anisotropy plot, shown in Figure 7.24, when fitted gives a value for the fluorescence anisotropy of  $0.0439 \pm 0.0021$ . The increase in anisotropy is due to the decrease in the fluorescent lifetime of the donor species induced by FRET. With the donor relaxation enhanced by the additional depopulation through energy transfer, the lifetime decreases and hence the anisotropy will be higher (see Equation 7.15).

In order to assess the consistency of the signal recorded for the duration of the scan, the  $F_g/F_r$  ratio was plotted against the acquisition time, which shows as it did for the  $\text{Na}^+$ -only that the intensity fluctuates around an average value and remains consistent throughout.



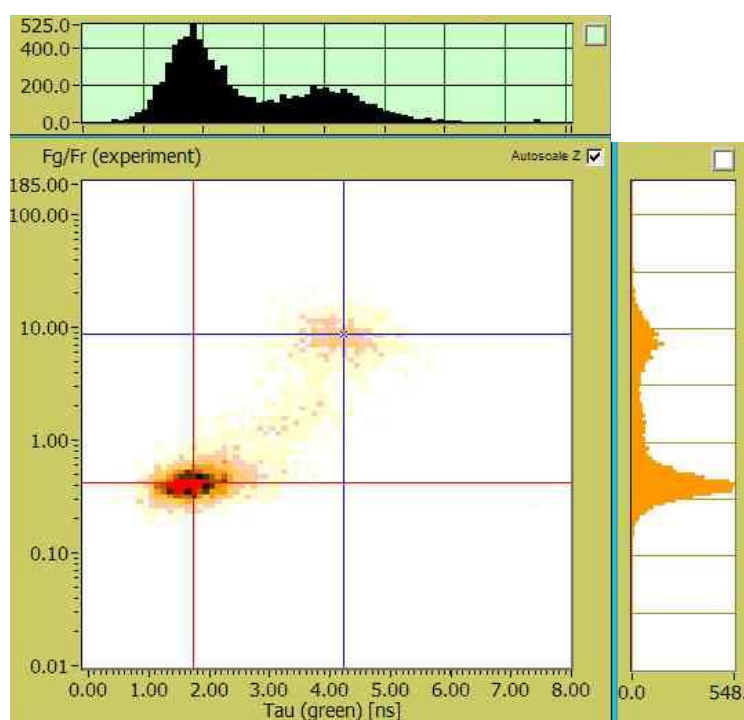
**Figure 7.25:** Plot showing the variation in the emission intensities for the dyes through the ratio  $F_g/F_r$ . Two populations are indicated, the donor only ( $F_g/F_r = 10.8$ ) and donor-acceptor pair ( $F_g/F_r = 1.03$ ). The histogram plotted for the time axis indicates a consistent fluorescence output for the duration of the scan.

In addition to the verification of the fluorescence emission remaining at a consistent level throughout the scan, the histogram of  $F_g/F_r$  shows a distinct tail leading to the lower values. This offers the indication of a second FRET conformation with a lower  $F_g/F_r$  ratio, corresponding to a higher FRET state. This additional observation is consistent with the TIRF-based sm-FRET data, which at these chemical conditions showed conformational populations predominantly at the intermediate conformation ( $E_{app} \sim 0.4$ ) with a minority population of the native, folded state ( $E_{app} \sim 0.62$ ).



### Using sm-MFD to study the ligand-bound state of the SAM-I riboswitch

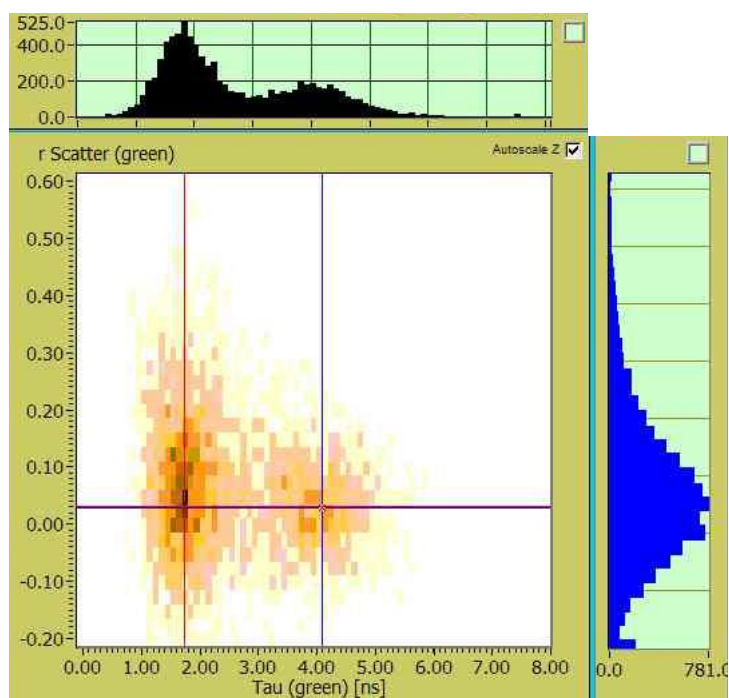
The final chemical condition examined was the ligand-bound state. The sm-FRET data collected by the surface-immobilised system reported that, at conditions of 3 mM  $Mg^{2+}$  and 100  $\mu M$  SAM, the structure exhibited a small population at the intermediate conformation ( $E_{app} \sim 0.4$ ) and the majority in the folded state ( $E_{app} \sim 0.62$ ) (Figure 7.19c). A scan was performed at these chemical conditions with a view to replicate these observations and the results are shown below in Figure 7.26.



**Figure 7.26:** The fluorescence lifetimes fitted against the ratio of the donor to acceptor intensities ( $F_g/F_r$ ). Three populations are visible; the donor-only population, the magnesium induced intermediate (dominant conformation in the previous section) and the folded conformation.

The plot indicates that the dual-labelled aptamer exhibits a fluorescence lifetime of 1.78 ns, with a donor to acceptor intensity ratio of

$F_g/F_r = 0.41 \pm 0.01$ , a ratio that corresponds to a FRET of  $E_{app} \sim 0.71 \pm 0.02$ . The donor-only population is visible, with a fitted lifetime of  $4.29 \pm 0.16$  ns ( $F_g/F_r = 8.65 \pm 0.62$ ). Furthermore, a very small population of the intermediate state is visible, which is representative of the surface-immobilised FRET data, shown in Figure 7.19c.

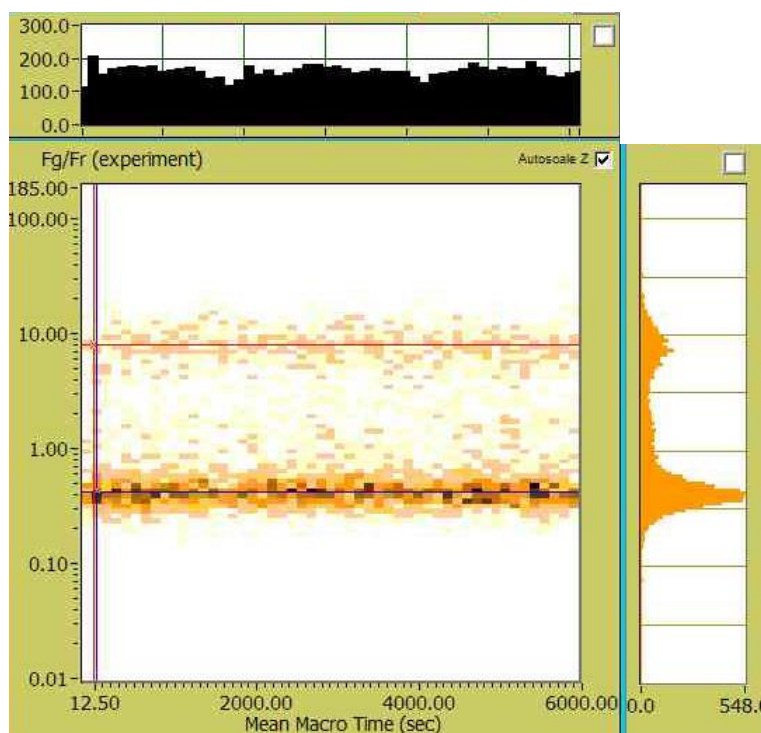


**Figure 7.27:** Fluorescence anisotropy plotted versus the donor lifetime, and histograms showing the distribution spread across both axes; anisotropy (blue) and lifetimes (black).

The anisotropy, plotted against the fluorescence lifetime (Figure 7.27), shows the two populations (single- and dual-labelled) separated only by lifetime; the fitted anisotropy for the folded conformation was shown to be  $0.0299 \pm 0.0067$ . Though this is the peak value, the variation in the anisotropy is significantly larger than that observed for any other conformational state, with a significant population still present at  $r = 0.3$ . These high anisotropy contributions are indicative of a greatly reduced

fluorescence lifetime due to a high degree of inter-dye energy transfer (FRET).

The intensity remains consistent across time, fluctuating around an average value, which signifies that the concentration fluctuations are relatively small. The plot also shows the presence of a small population of the intermediate state ( $F_g/F_r = 1.29 \pm 0.11$ ), though only a very small proportion. Also, the plot shows that the sample retains its photostability for the duration of the scan.



**Figure 7.28:** The  $F_g/F_r$  ratio plotted against the scan time. The histograms for both the time and intensity ratio are shown, showing the different conformational populations present, namely the intermediate ( $F_g/F_r = 1.29 \pm 0.11$ ) and folded states ( $F_g/F_r = 0.41 \pm 0.01$ ).

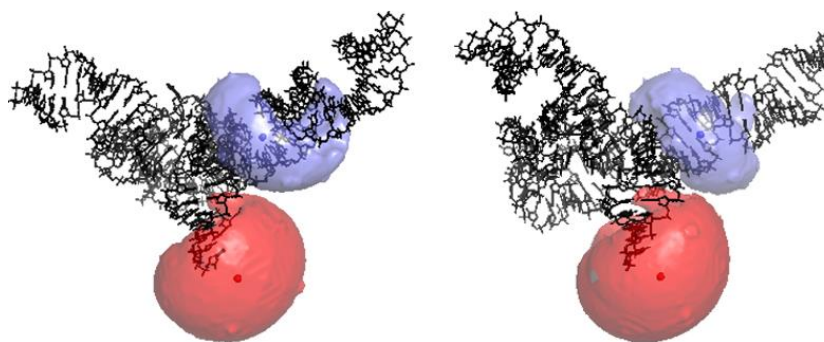
The fluorescence lifetime information allows the extraction of accurate inter-dye distances (described in detail in Chapter 1) [219], specifically

through combinations of the equations shown in Chapter 1 (Equations 1.11-1.13) with prior knowledge of the Förster radius ( $R_0$ ). By using the Förster radius quoted by Wozniak et al [22], the inter-dye distance derived from both the intensity ratios and lifetimes was calculated for the ligand-bound state for which there is a crystal structure.

From the intensity ratio, I obtained a value of  $R \sim 47 \text{ \AA}$  and from the lifetimes the value obtained is  $R \sim 49 \text{ \AA}$ . The relative discrepancy between the values obtained by sm-MFD using the raw intensity data and the lifetime is expected as the detectors have different efficiencies at both wavelengths (520 nm emission maxima for Alexa488 compared to 660 nm for Cy5) and this has not been accounted for in our analysis of the FRET efficiency via intensity data. This difference in efficiency is further enhanced by the fact that I am using two different APDs (a faster one for the donor from Micro Photon Devices and a slower one from Excelitas for the acceptor) with slightly different responses depending on the spectral region. On the other hand, the FRET efficiency and distance values extracted from the analysis of the MFD lifetimes are not affected by this difference in detector efficiency, or by other artefacts such as different background levels in both regions of emission, and therefore should provide a more accurate measure of the distance.

To compare the inter-dye distance extracted from sm-MFD, molecular models of the FNS state were built in Pymol using the crystal structure of the SAM-I Riboswitch [233] as a template for the Alexa488-Cy5 construct

(Figure 7.29). The desired sequences of stems P1 and P3 were built with the UCSF Chimera package [234] and were aligned by overlapping two base pairs with their counterpart in the crystal structure. The dyes were attached to specific bases via C6-amino linkers and their positions were modelled by the accessible volume (AV) approach with the software provided by the Seidel lab [235]. Then, this information can be used to calculate the FRET-averaged distance between the two dyes (Figure 7.29). The model predicts an inter-dye distance value of 43 Å, which is in very good agreement with the values obtained by sm-MFD of 47 Å by intensity ratio and 49 Å by using the fluorescence lifetimes.



**Figure 7.29:** Alexa488 (red)-Cy5 (blue) accessible distances calculated for the  $F_{NS}$  state of the SAM riboswitch using the crystal structure of the ligand-bound state [233] and the modelling software provided by Seidel's group [234, 235]. The structures shown differ only by rotation and were made using Pymol.

## 7.10. Conclusions

In this chapter I have expanded the range of single-molecule microscopy techniques at St Andrews, by implementing single-molecule FCS, FCCS and MFD. The implementation of these techniques was carried out from the very first steps, including the design of the optical, mechanical and electrical pathways, the choice of the different components and their assembly. I also developed the analysis software for sm-FCS and sm-FCCS combined with FRET where required. Due to time limitations I did not complete the development of the sm-MFD analysis software and this currently constitutes a joined project between St Andrews and the group of Dr. Steve Magennis at Glasgow University.

I started with the implementation of sm-FCS for freely diffusing molecules using an after-pulse correction configuration that involves the use of two APDs detecting vertically and horizontally polarised light of the same wavelength. I next tested and calibrated the performance and the geometrical dimensions of the FCS volume using Rhodamine B with a known diffusion coefficient. I determine a volume of 0.4 fl and an axial to lateral ratio ( $z$ ) of 1.85, both of which are well within the common values reported for commercial FCS instruments. The performance of the FCS setup was determined across a time window of correlation times spanning from  $\sim 50 \mu\text{s}$  to  $\sim 50 \text{ms}$  using different singly-labelled biomolecular test systems, including protein-DNA interactions to access the sub-millisecond

correlation time, and liposome dynamics to target the tens of milliseconds regime. This configuration was then upgraded for fluorescence cross-correlation (sm-FCS) which gives access to the characterisation of FRET populations from freely diffusing doubly-labelled species. In this case, I used a DNA bubble labelled with the Cy3-Cy5 FRET pair. A good agreement between the values extracted from sm-FCS-FRET and TIRFM-FRET was obtained.

In the third phase of the implementation, I upgraded the sm-FCCS with multi-parameter-fluorescence detection by incorporation a picosecond pulsed laser source and the simultaneous operation of four avalanche photodiodes and two TCSPC computer cards. To test the performance of this system, I examine a SAM riboswitch extensively characterised in our lab by TIRFM-FRET, but with the Cy3-Cy5 FRET pair replaced by an Alexa488-Cy5 pair. The reason for this change is two-fold. Firstly, Alexa488 dye in the absence of acceptor has a relative long fluorescence lifetime ( $\sim 4.2$  ns) compared to Cy3 dye ( $\sim 0.3$  ns), thus providing a higher dynamic range to determine the influence of the acceptor. Secondly, the emission spectral region of Alexa488 and the emission of the acceptor Cy5 are much better separated ( $>120$  nm between emission maxima) than the Alexa488-Cy3 pair ( $<40$  nm), thus decreasing any potential cross-talk between both channels. The analysis of the folding pathway of the SAM riboswitch as a function of  $Mg^{2+}$  and ligand yielded broadly similar global variations in FRET efficiency as those obtained using TIRFM-FRET. Because there is a crystal structure available

for the ligand-bound state of the SAM riboswitch, I carried out a comparison of the theoretical inter-distances available with that obtained for the same FRET state via sm-MFD. From the modelling I determine a value of 43 Å, which is in very good agreement with the values obtained by smMFD of 47 Å by intensity ratio and 49 Å by using the fluorescence lifetimes.



## 8. Conclusions and future directions

### Conclusions from experimental work

The primary focus of the work performed in this thesis has been on the interrogation of the behaviour of three-way RNA junctions involved in either gene regulation, in the case of the adenine riboswitch, and in catalytic processes for the hammerhead ribozyme. The importance of these structures, both in terms of the roles played within the cell and their prevalence in nature, makes them a very important class of RNA structures to investigate and as such their potential as therapeutic targets has been speculated. Our main experimental tool to date in the group has been single-molecule total internal reflection fluorescence microscopy combined with fluorescence resonance energy transfer. However, this technique, although very powerful, is only applicable to surface-immobilised molecules. Thus, as part of this thesis, I implemented from the very first steps a range of single-molecule techniques tailored for freely diffusing molecules including fluorescence correlation spectroscopy (sm-FCS), dual colour fluorescence cross-correlation spectroscopy (sm-FCCS) and multi-parameter fluorescence detection (sm-MFD).

In the first experimental chapter, I used single-molecule TIRFM and FRET to investigate in detail the influence of monovalent ions on the folding

and structure of adenine riboswitch and compare it with a previous single-molecule FRET study using divalent  $\text{Mg}^{2+}$  ions.

I demonstrated that the folding pathway induced by  $\text{Na}^+$  ions alone follows the same discrete structural states as that previously reported in the presence of  $\text{Mg}^{2+}$  ions [128, 155]; that is the four states exhibited were  $E_{\text{app}} = 0.32 \pm 0.04$  (unfolded),  $E_{\text{app}} = 0.45 \pm 0.02$  (intermediate),  $E_{\text{app}} = 0.87 \pm 0.01$  (folded, ligand-free) and  $E_{\text{app}} = 0.87 \pm 0.01$  (folded, ligand bound). Both the values recorded for the apparent FRET and the progression through the states induced by increasing  $\text{Na}^+$  concentrations were similar to that reported using  $\text{Mg}^{2+}$  ions as folding agents. Furthermore, despite the lack of site-specifically bound  $\text{Mg}^{2+}$  ions that coordinate the binding site in the aptamer core of the adenine riboswitch, I demonstrate that the  $\text{Na}^+$ -induced folded structure is still perfectly competent to bind the adenine ligand. However, while the  $\text{Mg}^{2+}$  concentration required to fully fold the aptamer was 4 mM, the adenine riboswitch was not in a fully folded population until the  $\text{Na}^+$  background was at 250 mM, slightly above physiological conditions. Significant differences were also observed when the folding and unfolding kinetics were taken into consideration. In particular, although the folding rate was shown to increase with  $\text{Na}^+$  as it did with  $\text{Mg}^{2+}$ , the unfolding rate was shown to be independent of the concentration of  $\text{Na}^+$  present in the imaging buffer. I interpreted this difference in terms of the aptamer stabilisation in the folded conformation due to the trapping of  $\text{Mg}^{2+}$  ions at specific locations as observed in the crystal structure of the ligand-bound

complex. By plotting the activity of  $\text{Na}^+$  ions against the equilibrium constant I have shown that the ligand-free structure requires a net uptake of 1.1 ionic pairs per aptamer to fold. By comparison, recent studies in the presence of adenine ligand indicated that 3.8  $\text{Na}^+$  ion pairs are required for the formation of the ligand-bound structure. We reasoned that this difference in ion uptake arises from the increased core compaction required for ligand binding and ligand stabilisation.

In chapter 5, I explored the use of the competing interplay between folding and unfolding agents to manipulate the folding dynamics of the adenine aptamer. Building on the knowledge of the behaviour of the adenine aptamer using  $\text{Na}^+$  as folding agents, I next explored what unfolding/denaturant agents would offer the best properties (i.e. the range of concentration, solubility, refraction index compatible with TIRF microscopy) to fine tune the dynamics of the adenine aptamer. I, therefore, characterised the ability of four commonly used chemical denaturants (betaine, formamide, proline and urea) to compete with  $\text{Na}^+$  ions and inhibit or reverse the folding of the adenine aptamer. The data demonstrated that urea is the strongest denaturant of the tertiary structure of riboswitch aptamers. I offered that this stronger effect may be arising from a combination of two factors: the first factor is the ability of urea to intercalate between the base pairs, which may be restricted in the other denaturants due to the presence of bulky groups. The second factor relates to the stronger ability of urea to act as a hydrogen-bond donor and acceptor and

thus disrupt more efficiently the hydrogen-bonding network present in the structure of the native RNA. In particular, disruption of the hydrogen bond interactions between the bases connecting the terminal loops may have a key role in the denaturation ability of urea. Having chosen urea as the chemical denaturant for the study, the titration of urea using the previously characterised Na<sup>+</sup> induced folded state showed a mirrored reversal of the folding pathway induced by Na<sup>+</sup> alone. Significantly, the effect of urea on a folded sample mediated by Na<sup>+</sup> ions alone was shown to be far greater than that observed previously in a background of divalent ions. I showed that urea induces a 14-fold decrease in the folding rate for the adenine aptamer, as opposed to the 2-fold decrease calculated for the Mg<sup>2+</sup> mediated folded state.

I then demonstrated, for the first time, that by balancing the unfolding effect generated by the urea and the Na<sup>+</sup>-induced folding it is possible to overpopulate and isolate the structural intermediate observed in both mono- and divalent ion-induced folding pathways. Because of the low-population and transient character of these folding intermediates, their characterisation at both ensemble and single-molecule level has remained elusive. I envisage that the methodology developed in this thesis will fill this gap and open up the opportunity to 'freeze' and characterise the structure of this intermediates using NMR, EPR and X-ray crystallisation techniques. Moreover, the possibility of determining the structure of these obligatory on-path intermediates expands the target range to develop suitable small

molecule compounds that interfere with the function of the riboswitch. This is crucial given the enormous threat that antibiotic resistance constitutes to global health.

Chapter 6 saw a shift in focus from a gene regulatory to a catalytic 3-way RNA junction, specifically to the hammerhead ribozyme. Although extensively studied in bulk techniques based on both fluorescence and EPR, the only single-molecule studies on the structure focussed not on characterising natural folding pathway of the structure but on the effect of point mutations on the cleavage efficiency of the ribozyme [92]. The motivation for this work is based on the growing interest in the hammerhead ribozyme for biotechnological applications and also from more fundamental knowledge perspective since the discovery that is present in a much larger number of genomes than initially speculated.

Although it is known that the presence of the loops in the natural form decreases the requirement for  $Mg^{2+}$  ions to physiological concentrations, a complete folding pathway of both ribozymes at the single-molecule level has not been reported. Therefore, in this chapter, I carried out a comparative study of the minimal (no terminal loops) and natural forms of the hammerhead ribozyme at the single-molecule level using sm-FRET. The folding pathway under the action of divalent ions was examined. For the minimal hammerhead, the sm-FRET data confirmed the two-stage folding process proposed from bulk fluorescence studies involving the stacking of stems P2 and P3 (occurring at  $[Mg^{2+}] < 0.5$  mM) followed by the formation of

the U-turn rearrangement of P1 (for  $[\text{Mg}^{2+}] > 20 \text{ mM}$ ). Importantly, it was also observed that all three states would occur within a single trace, and the order of occurrence suggested that P2-P3 stacking process is an obligatory initial step that must occur before folding of the U-turn in P1. Interestingly, the folding rate corresponding to the formation of the U-turn showed very little influence by  $\text{Mg}^{2+}$  ions, suggesting that specific trapping of  $\text{Mg}^{2+}$  ions observed in the X-ray structure of the minimal form may take place predominantly during the stacking step.

The natural hammerhead ribozyme, when investigated at the single-molecule level, also exhibited a two-stage folding process. Specifically, the population initially at  $E_{\text{app}} \sim 0.25$  was promoted to a high FRET state of  $E_{\text{app}} \sim 0.6$  at  $100 \text{ nM Mg}^{2+}$  before returning to a low state of  $E_{\text{app}} \sim 0.2$ . The similarity between the  $E_{\text{app}}$  values for the unfolded and fully-folded states made the kinetic study impossible for this particular design of RNA structure. Nevertheless, the confirmation that the natural form also follows a two-step folding pathway is a new finding allowed by single-molecule FRET. Our sm-FRET study confirms that the significant reduction in the  $\text{Mg}^{2+}$  concentration required to stimulate catalytic activity in the natural form ribozyme is due to an increase in the population of the native folded state. In the natural form, the intermediate state is populated at a concentration of  $\text{Mg}^{2+}$  ions 500 times less than for the minimal form. This is in agreement with the observation that the presence of the peripheral loop motifs on the stems of the natural form of the ribozyme lead to a significant

reduction in the concentration of magnesium required to fold the structure. Though the mechanism of how exactly this reduction arises is not entirely clear, this could be explained by the interactions between these peripheral elements stimulating an increase in folding rate, and an associated decrease in the unfolding rate, resulting in a much faster transition between the unfolded and folded conformations. A further understanding of the influence of the loop-loop interaction is currently underway using a different distance vector where each folding step should have a different FRET value.

The final chapter documents the stages in the design and development of a single-molecule multi-parameter fluorescence detection (sm-MFD) spectrometer by first constructing a single-molecule fluorescence correlation spectroscopy system (sm-FCS) and expanding the setup to extract its full potential. In developing this system, one may obtain simultaneous information on conformation in terms of both inter-dye distances, by monitoring intensity changes and fluorescence lifetimes and equating the relevant FRET expressions, and in terms of dye orientation and confinement through fluorescence anisotropy information. Furthermore, one may calculate stoichiometric properties of the sample where two interacting species are present in the solution.

The initial stages focussed on developing a sm-FCS system, primarily due to the desire to investigate biological samples freely diffusing through a buffer solution free from interactions with the quartz substrate which compromise both the emissive properties of the sample and its natural

biological function. A system using two identical detectors was developed based on a continuous wave laser source. Through the calculation of the diffusion time of a free dye of known diffusive properties through the excitation volume, an effective excitation volume of 0.4 fl was established, with a lateral to axial size ratio of 1.58. Both of these values fall comfortably within the accepted ranges for such setups reported in the literature, where an elliptical volume of less than 1 fl is considered normal.

To move towards sm-MFD, the second developmental stage was the simultaneous detection of the emission of two dyes located on a DNA structure, through which the FRET could be calculated. The double-stranded DNA structure was not designed to exhibit significant dynamic behaviour, allowing a straight comparison between the observations from TIRFM-FRET and freely diffusion sm-FRET to be made. In order to access the FRET recorded from the fluorescence bursts, a program was written in MATLAB to investigate the photon stream data recorded by the detectors. The results confirmed the accuracy of the system, with recorded FRET efficiencies of  $E_{\text{app}} \sim 0.41$  from TIRFM-FRET and  $E_{\text{app}} \sim 0.44$  from the freely diffusing system. The discrepancies come from the different detection efficiencies of the two SPAD units across the emissive range of each dye, a correction not included in the analysis of the data.

Having established a system capable of calculating FRET accurately for a two-dye sample, the final stage was to introduce a pulsed excitation scheme and four detectors, with the emission light separated by both wavelength



and polarisation. In doing so, a fully functional sm-MFD setup was realised. In order to verify the accuracy of the system, as for the two-colour FCS a sample characterised using TIRFM-FRET was chosen. Due to the availability of samples in the lab suitably labelled for such an experiment, the model chosen was the SAM-I riboswitch, whose sm-FRET behaviour had been characterised using a TIRFM system by our group previously [159]. Having identified the three chemical conditions that populated each of the three conformations in turn, scans were taken to investigate the homogeneity of the two sm-FRET detection schemes.

It was found that both systems offered consistent conformational results; the unfolded state of the riboswitch (U) was calculated to have a FRET of  $E \sim 0.28$  from the sm-MFD FRET data, taken at 50 mM Na<sup>+</sup>. The TIRFM-FRET experiments reported the exact same value ( $E_{\text{app}} \sim 0.28$ ). The fluorescent lifetime for this conformation was shown to be 3.85 ns, which when compared to the donor-only lifetime of 4.24 ns indicates a low FRET rate between the two dyes. Exposure to millimolar concentrations of Mg<sup>2+</sup> promotes the structure to a structural intermediate ( $F_{\text{Mg}}$ ). The scan performed at 3 mM Mg<sup>2+</sup> showed an increase in the FRET, and the intensity ratio of both dyes confirmed an increase in FRET to  $E \sim 0.44$ , again consistent with the TIRFM-based FRET of  $E_{\text{app}} \sim 0.4$ . This increase in FRET was confirmed by both the fluorescent lifetime, which decreased to 3.35 ns, and the anisotropy which increased to 0.044, both of which are indicative of a greater rate of transfer from the excited donor molecules. The addition of

the SAM ligand into the buffer stabilises the structure in the folded conformation ( $F_{NS}$ ). The TIRFM-FRET measurements at 3 mM  $Mg^{2+}$  and 100  $\mu$ M SAM indicated that this conformation had a FRET of  $E_{app} \sim 0.62$ . When exposed to this chemical condition and using sm-MFD this FRET was shown to be  $E \sim 0.7$ . As for the  $F_{Mg}$  state, the increase in FRET is confirmed by the decrease in the fluorescence lifetime, significantly reduced to 1.78 ns. The anisotropy, in this case, exhibited a significant spread, with the population distributed up to values in excess of 0.3, which is further proof of a far higher significant degree of transfer between the dyes. The values for the FRET extracted from both the lifetimes and intensity data were used to calculate the distances between the dyes, and resulted in calculated values of  $R \sim 49 \text{ \AA}$  and  $R \sim 47 \text{ \AA}$  respectively, which compare very favourably to the computationally calculated value of  $R = 43 \text{ \AA}$ .

### **Future directions**

Now that the conditions to achieve this overpopulation of transient states have been determined, the next step will be to determine the actual peripheral organisation of the aptamer by FRET by using different vectors. By complementing this with ensemble techniques such as EPR, which have a higher precision on the determination of intramolecular distances, then, when using probes located at the same positions, a more detailed picture of the structure and associated conformational rearrangements can be established.

I could try to look at the local structure by inserting fluorescent nucleotide analogues like 2-AP or pyrrolo-dC. The ability to detect 2-AP has already been established [236]. 2-AP is quenched when stacked but becomes fluorescent when exposed to the solvent, so by locating 2-AP at certain positions within the aptamer core by replacing native adenine and look for conformational rearrangements.

Additional vectors of the hammerhead ribozyme could offer a more detailed insight into the folding and unfolding pathways, in particular probing the interaction between stems P1 and P2 of the natural form of the structure. The natural function of the structure involves cleavage and ligation, processes eliminated by the introduction of the deoxy-nucleobase in the artificial constructs probed in the sm-FRET experiments in Chapter 6. By combining single-molecule FRET with vesicle encapsulation, described in [154], the natural form of the ribozyme including the natural nucleobase could be studied including the natural catalytic functions, since the liposome will confine all of the RNA strands and allow full recombination.

The development of the software to analyse the photon streams obtained from the sm-FCS and sm-MFD is currently ongoing. This collaboration, with Dr Steven Magennis of the University of Glasgow, will result in a freely-distributed package capable of presenting and processing these streams in real-time.

## Appendices

### I: Deriving the Förster radius ( $R_0$ ) from first principles

The transfer of energy between two dyes in close proximity is a consequence of the interaction between the electric field ( $\vec{E}_D$ ) generated by the magnetic dipole of the donor ( $\vec{\mu}_D$ ) and the acceptor dipole moment ( $\vec{\mu}_A$ ), separated by a distance vector ( $\vec{r}$ ). By re-expressing the vectors in terms of a magnitude and a unit vector, one obtains the following expression for the interaction between  $\vec{E}_D$  and  $\vec{\mu}_A$ :

$$E_{int} = \vec{\mu}_A \cdot \vec{E}_D = \frac{\mu_D}{n^2 r^3} [(3\hat{\mu}_D \cdot \hat{r})(\hat{\mu}_A \cdot \hat{r}) + \hat{\mu}_D \cdot \hat{\mu}_A] = \frac{\mu_D \mu_A}{n^2 r^3} \quad (\text{I-1})$$

The fluorescence lifetime of an excited molecule may be broken into its radiative ( $k_r$ ) and non-radiative ( $k_{nr}$ ) components (I-2). Where this molecule is the donor in a FRET pair, this lifetime has a further component, the transfer rate ( $k_{tr}$ ) (I-3):

$$\frac{1}{\tau_D} = k_r + k_{nr} = k_D \quad (\text{I-2})$$

$$\frac{1}{\tau_{DA}} = k_r + k_{nr} + k_{tr} = \tau_D^{-1} + k_{tr} \quad (\text{I-3})$$

One may think of this transfer rate as the rate of absorption of energy to the acceptor. This may be expressed by the Beer Lambert Law (I-4), by considering the absorption of a thin layer of the acceptor:

$$I(z) = I_0 10^{-C\epsilon z} \quad (\text{I-4})$$

This represents the absorption of radiation initially of intensity ( $I_0$ ) by an acceptor concentration ( $C$ ) with an absorption coefficient ( $\epsilon$ ) as a function of distance ( $z$ ). By calculating the rate of change of intensity with respect to distance, and by normalising this quantity by the number of molecules present in the volume and introducing a factor of 3 assuming that the transition dipole moment and polarisation of the donor electric field are completely aligned, one obtains the transfer rate for the pair (I-5):

$$k_{tr} = -\frac{3}{N_A C} \cdot \frac{dI}{dz} = -\frac{3}{N_A C} \cdot (-I\epsilon C \ln 10) = \frac{3I\epsilon \ln 10}{N_A} \quad (\text{I-5})$$

Combining equations (I-1) and (I-5) and using the relation between the electric field and intensity (I-6), one obtains the combined final equation for the transfer rate (I-7):

$$I = \frac{cn}{8\pi} |E_{int}|^2 \quad (\text{I-6})$$

$$k_{tr} = \frac{3 \ln 10 c \epsilon \mu^2 \kappa^2}{8\pi N_A n^3 r^6} \quad (\text{I-7})$$

The power emitted by a free dipole in the far field for a frequency ( $\omega$ ) is given by equation (I-8). This power relates to the donor transfer rate since this transfer rate multiplied by the donor quantum yield, giving the transfer rate for the donor alone given in equation (I-9).

$$P_{dipole} = \frac{\mu^2 \omega^4 n}{3c^3} \quad (\text{I-8})$$

$$k_D = \frac{P_{dipole}}{\phi_D} = \frac{\mu^2 \omega^4 n}{3\phi_D c^3} = \frac{16\pi^4 \mu^2 c n}{3\phi_D \lambda^4} \quad (\text{I-9})$$

Having established equations for both the rates of decay for the donor-only and FRET pair-labelled systems, one may calculate the FRET from equation (I-10). Substituting equations (I-6) and (I-8) into (I-10) gives the final expression for the energy transfer (I-11).

$$E = 1 - \frac{\tau_{DA}}{\tau_D} = \frac{1}{1 + \frac{k_D}{k_{tr}}} \quad (\text{I-10})$$

$$E = \frac{1}{1 + \frac{16\pi^4 \mu^2 c n}{3\phi_D \lambda^4} \cdot \frac{8\pi N_A n^3 r^6}{3 \ln 10 c \epsilon \mu^2 \kappa^2}} = \frac{1}{1 + \frac{128\pi^5 \mu^2 N_A n^4 r^6}{9 \ln 10 \phi_D \lambda^4 \epsilon \mu^2 \kappa^2}} \quad (\text{I-11})$$

Equation (I-11) can be generalised to any donor and acceptor spectra with an overlap integral ( $J(\lambda)$ ) (I-12). This term depends on the wavelength ( $\lambda$ ), the emission spectrum of the donor ( $F_D(\lambda)$ ) and the absorption spectrum of the acceptor ( $\epsilon_A(\lambda)$ ). These equations lead finally to the expression for the

FRET between two dyes separated by a distance  $r$  given in equation (I-13), where the term  $R_0$  is known as the Förster radius.

$$J(\lambda) = \int \varepsilon_A(\lambda)\lambda^4 F_D(\lambda) d\lambda \quad (\text{I-12})$$

$$E = \frac{1}{1 + \frac{128\pi^5 N_A n^4 r^6}{9 \ln 10 J(\nu) \phi_D c^4 \varepsilon \kappa^2}} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad (\text{I-13})$$

$$\text{where } R_0^6 = \frac{9 \ln 10 J(\nu) \phi_D c^4 \varepsilon \kappa^2}{128 \pi^5 N_A n^4}$$

## II: Uncertainty calculations and combinations

Suppose one wishes to calculate some quantity  $z$ , a function of variables  $x$  and  $y$  with associated errors  $\Delta x$  and  $\Delta y$ . For the general case, the error  $\Delta z$  is calculated from the square root sum of the products of the squares of the partial derivatives of  $z$  with respect to each variable and the standard errors associated with each variable.

Listed here are the standard equations and rules followed when combining standard errors derived from fitting algorithms for use in further calculations. Adapted from reference [237].

Function ( $z = f(x, y)$ )	Uncertainty ( $\Delta z$ )	
$z = x \pm y$	$\Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2}$	(II-1)
$z = x \times y$ or $z = \frac{x}{y}$	$\frac{\Delta z}{ z } = \sqrt{\left(\frac{\Delta x}{ x }\right)^2 + \left(\frac{\Delta y}{ y }\right)^2}$	(II-2)
$z = Ax$	$\Delta z =  A \Delta x$	(II-3)
$z = x^n$	$\frac{\Delta z}{ z } =  n \frac{\Delta x}{ x }$	(II-4)
$z = f(a, \dots, n)$	$\Delta z = \sqrt{\left(\frac{\partial z}{\partial a}\Delta a\right)^2 + \dots + \left(\frac{\partial z}{\partial n}\Delta n\right)^2}$	(II-5)



### III: Deriving the fluorescence correlation function $G(\tau)$

The number of molecules in a confocal volume is governed by Poissonian distribution (III-1):

$$\frac{\sqrt{\langle(\delta N)^2\rangle}}{\langle N\rangle} = \frac{\sqrt{\langle(N - \langle N\rangle)^2\rangle}}{\langle N\rangle} = \frac{1}{\sqrt{\langle N\rangle}} \quad (\text{III-1})$$

Fluorescence fluctuations are given by (III-2, III-3):

$$\delta I(t) = I(t) - \langle I(t) \rangle \quad (\text{III-2})$$

$$\langle I(t) \rangle = \frac{1}{T} \int_0^T I(t) dt \quad (\text{III-3})$$

If all fluctuations arise only from changes in concentration in the confocal volume  $V_{eff}$  then

$$\delta I(t) = \kappa \int_V I_{ex}(\vec{r}, t) \cdot S(\vec{r}) \delta(\sigma \cdot q \cdot C(\vec{r}, t)) dV \quad (\text{III-4})$$

Where:

- $\kappa$  is overall detection efficiency
- $I_{ex}(\vec{r}, t)$  spatial distribution of excitation energy with a maximum amplitude of  $I_0$

- $S(\vec{r})$  is the transfer function of the objective-pinhole combination, like the optical efficiency of the setup
- $\delta\sigma$  is change in the absorption cross section of the dye
- $\delta q$  is represents fluctuations in the quantum yield
- $\delta C(\vec{r}, t)$  is change in local concentrations in time

In order to simplify this, define the function  $W(\vec{r})$  as the spatial distribution of emitted light, approximated by a 3D Gaussian:

$$W(\vec{r}) = \frac{1}{I_0} I_{ex}(\vec{r}, t) \cdot S(\vec{r}) = e^{-2\left(\frac{x^2+y^2}{w_0^2} + \frac{z^2}{z_0^2}\right)} \quad (\text{III-5})$$

And also define the photon count rate per molecule per second  $\eta$  (III-6):

$$\eta = I_0 \kappa \sigma q \quad (\text{III-6})$$

This allows us to re-express the fluorescence fluctuations  $\delta I(t)$  as (III-7):

$$\delta I(t) = \int_V W(\vec{r}) \delta(\eta C(\vec{r}, t)) dV \quad (\text{III-7})$$

Now, the autocorrelation function  $G(\tau)$  is defined by the following relation (III-8):

$$G(\tau) = \frac{\langle \delta I(t) \cdot \delta I(t + \tau) \rangle}{\langle I(t) \rangle^2} \quad (\text{III-8})$$

And after substituting in Equations III-5 to III-7:

$$G(\tau) = \frac{\iint W(\vec{r})W(\vec{r}')\delta(\eta C(\vec{r}, t))\delta(\eta C(\vec{r}', t + \tau)) dV dV'}{(\int W(\vec{r})dV)^2} \quad (\text{III-9})$$

With this expression (III-9) one can isolate the term which relates to the fluctuations (III-10):

$$\delta(\eta C(\vec{r}, t)) = C\delta\eta + \eta\delta C \quad (\text{III-10})$$

Some assumptions must now be made:

1. Fluorescence properties don't change within observation time i.e.  $\delta\eta = 0$ .
2. Consider only particles that are freely diffusing in 3D with diffusion coefficient D.

First assumption:

Leads to (III-11)

$$G(\tau) = \frac{\iint W(\vec{r})W(\vec{r}')\delta(\eta C(\vec{r}, 0))\delta(\eta C(\vec{r}', \tau)) dV dV'}{(\int W(\vec{r})dV)^2} \quad (\text{III-11})$$

Second assumption: This allows us to substitute (III-12) into our expression for  $G(\tau)$  (III-11):

$$\langle \delta C(\vec{r}, t) \delta C(\vec{r}', t + \tau) \rangle = \langle C \rangle \frac{1}{(4\pi D\tau)^{-3/2}} e^{-\left(\frac{(\vec{r}-\vec{r}')^2}{4D\tau}\right)} \quad (\text{III-12})$$

So the following expression for  $G(\tau)$  is obtained (III-14):

$$G(\tau) = \frac{\iint W(\vec{r}) W(\vec{r}') \langle C \rangle \frac{1}{(4\pi D\tau)^{-3/2}} e^{-\left(\frac{(\vec{r}-\vec{r}')^2}{4D\tau}\right)} dV dV'}{\langle C \rangle \int W(\vec{r}) dV)^2} \quad (\text{III-13})$$

$$G(\tau) = \frac{1}{\langle C \rangle (4\pi D\tau)^{-3/2}} \cdot \frac{\iint W(\vec{r}) W(\vec{r}') e^{-\left(\frac{(\vec{r}-\vec{r}')^2}{4D\tau}\right)} dV dV'}{\left(\int W(\vec{r}) dV\right)^2} \quad (\text{III-14})$$

Including two further simplifications (III-15 and III-16):

$$\tau_D = \frac{w_0^2}{4D} \quad (\text{III-15})$$

$$V_{eff} = \frac{\left(\int W(\vec{r}) dV\right)^2}{\int W^2(\vec{r}) dV} = \frac{\int e^{-2\left(\frac{x^2+y^2+z^2}{w_0^2+z_0^2}\right)} dV}{\int e^{-4\left(\frac{x^2+y^2+z^2}{w_0^2+z_0^2}\right)} dV} = \pi^{3/2} w_0^2 z_0 \quad (\text{III-16})$$

We are left with the final result (III-17):

$$G(\tau) = \frac{1}{\langle C \rangle V_{eff}} \cdot \frac{1}{\left(1 + \frac{\tau}{\tau_D}\right) \left(1 + \left(\frac{w_0}{z_0}\right)^2 \cdot \frac{\tau}{\tau_D}\right)^{1/2}} \quad (\text{III-17})$$

By defining the correlation amplitude (III-20) and using the eccentricity of the Gaussian of the confocal volume (III-19) we reach the alternative form:

$$G(\tau) = \frac{G(0)}{\left(1 + \frac{\tau}{\tau_D}\right) \left(1 + \frac{\tau}{k^2 \tau_D}\right)^{1/2}} \quad (\text{III-18})$$

Where

$$k = \frac{z_0}{w_0} \quad (\text{III-19})$$

And

$$G(0) = \frac{1}{V_{eff}\langle C \rangle} \quad (\text{III-20})$$

#### IV: List of publications

##### Published papers:

K. McCluskey, E. Shaw, D. A. Lafontaine, J. C. Penedo, "Single-Molecule Fluorescence of Nucleic Acids", *Fluorescence Spectroscopy and Microscopy, Methods in Molecular Biology*, 1076 759-791 (2013)

E. Shaw, K. McCluskey, P. St-Pierre, D. A. Lafontaine\* and J. C. Penedo, "Using smFRET and denaturants to reveal folding landscapes" *Methods in Enzymology* Vol. 549, 313-341 (2014)

P. St-Pierre, K. McCluskey, E. Shaw, J. C. Penedo\* and D. A. Lafontaine, "Fluorescence tools to investigate riboswitch structural dynamics", *BBA- Gene regulatory* 1839 (10), 1005-1019 (2014)

**To be submitted:**

E. Shaw, P. Dalgarno, P. St-Pierre, D. A. Lafontaine and J. C. Penedo, “Influence of monovalent ions in the folding mechanism of the A-box riboswitch”. Written; to be submitted to RNA Biology

E. Shaw, P. Dalgarno, P. St-Pierre, D. A. Lafontaine and J. C. Penedo, “Overpopulating on-path folding intermediates using the competing interplay between folding and unfolding agents”. Written; to be submitted to Nucleic Acids Research

E. S. Shaw and J. C. Penedo, “A comparative single-molecule study of the folding dynamics of the minimal and full-length hammerhead ribozyme”. Writing in progress; to be submitted to Nucleic Acids Res.

## References

- [1] B. Valeur and M. N. Berberan-Santos, "A Brief History of Fluorescence and Phosphorescence before the Emergence of Quantum Theory," *J. Chem. Educ.*, vol. 88, no. 6, pp. 731–738, 2011.
- [2] D. Brewster, "XIX. On the Colours of Natural Bodies," *Earth Environ. Sci. Trans. R. Soc. Edinburgh*, vol. 12, no. 02, pp. 538–545, 1834.
- [3] J. F. W. Herschel, "[image] No. I. On a Case of Superficial Colour Presented by a Homogeneous Liquid Internally Colourless," *Philos. Trans. R. Soc. London*, vol. 135, pp. 143–145, 1845.
- [4] G. G. Stokes, "On the Change of Refrangibility of Light," *Philos. Trans. R. Soc. London*, vol. 142, pp. 463–562, Jan. 1852.
- [5] G. G. Stokes, "On the Change of Refrangibility of Light. No. II," *Philos. Trans. R. Soc. London*, vol. 143, pp. 385–396, 1853.
- [6] N. Bohr, "I. On the constitution of atoms and molecules," *London, Edinburgh, Dublin Philos.*, vol. 26, no. 151, pp. 1–25, 1913.
- [7] P. T. So, C. Y. Dong, B. R. Masters, and K. M. Berland, "Two-photon excitation fluorescence microscopy," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 399–429, 2000.
- [8] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.*, vol. 19, no. 11, p. 780, Jun. 1994.
- [9] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, no. v. 1. Springer, 2006.
- [10] M. Kasha, "Phosphorescence and the Role of the Triplet State in the Electronic Excitation of Complex Molecules.," *Chem. Rev.*, vol. 41, no. 2, pp. 401–419, Oct. 1947.
- [11] M. Kasha, "Characterization of Electronic Transitions in Complex Molecules," *Discuss. Faraday Soc.*, no. 9, pp. 14–19, 1950.
- [12] "Franck-Condon principle." [Online]. Available: [https://en.wikipedia.org/wiki/Franck?Condon\\_principle](https://en.wikipedia.org/wiki/Franck%3FCondon_principle). [Accessed: 07-Aug-2015].
- [13] T. Förster, "Intermolecular energy transference and fluorescence," *Ann Phys.*, vol. 2, pp. 55–75, 1948.
- [14] R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule FRET," *Nat Methods*, vol. 5, no. 6, pp. 507–516, 2008.
- [15] E. Shaw, P. St-Pierre, K. McCluskey, D. A. Lafontaine, and J. C. Penedo, "Using sm-FRET and Denaturants to Reveal Folding Landscapes.," *Methods Enzymol.*, vol. 549, pp. 313–41, Jan. 2014.
- [16] A. Demchenko, *Introduction to fluorescence sensing*, 1st ed. Springer Netherlands, 2009.
- [17] J. N. Forkey, M. E. Quinlan, and Y. E. Goldman, "Protein structural dynamics by single-molecule fluorescence polarization.," *Prog. Biophys. Mol. Biol.*, vol. 74, no. 1–2, pp. 1–35, Jan. 2000.
- [18] S. Blouin, T. D. Craggs, D. A. Lafontaine, and J. C. Penedo, "Functional studies of DNA-protein interactions using FRET techniques.," *Methods Mol. Biol.*, vol. 543, pp. 475–502, Jan. 2009.



- [19] A. N. Kapanidis and S. Weiss, "Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules," *J. Chem. Phys.*, vol. 117, no. 24, p. 10953, Dec. 2002.
- [20] H. Sahoo, "Förster resonance energy transfer – A spectroscopic nanoruler: Principle and applications," *J. Photochem. Photobiol. C Photochem. Rev.*, vol. 12, no. 1, pp. 20–30, Mar. 2011.
- [21] F. Yuan, L. Griffin, L. Phelps, V. Buschmann, K. Weston, and N. L. Greenbaum, "Use of a novel Förster resonance energy transfer method to identify locations of site-bound metal ions in the U2-U6 snRNA complex.," *Nucleic Acids Res.*, vol. 35, no. 9, pp. 2833–45, Jan. 2007.
- [22] A. K. Wozniak, G. F. Schröder, H. Grubmüller, C. A. M. Seidel, and F. Oesterhelt, "Single-molecule FRET measures bends and kinks in DNA.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 47, pp. 18337–42, Nov. 2008.
- [23] H. Edelhoch, L. Brand, and M. Wilchek, "Fluorescence Studies with Tryptophyl Peptides \*," *Biochemistry*, vol. 6, no. 2, pp. 547–559, Feb. 1967.
- [24] B. Rotman, "Measurement of activity of single molecules of beta-D-galactosidase.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 47, no. 12, pp. 1981–91, Dec. 1961.
- [25] T. Hirschfeld, "Optical microscopic observation of single small molecules.," *Appl. Opt.*, vol. 15, no. 12, pp. 2965–6, Dec. 1976.
- [26] W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," *Phys. Rev. Lett.*, vol. 62, no. 21, pp. 2535–2538, May 1989.
- [27] M. Orrit and J. Bernard, "Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal," *Phys. Rev. Lett.*, vol. 65, no. 21, pp. 2716–2719, Nov. 1990.
- [28] L. Möckl, D. C. Lamb, and C. Bräuchle, "Super-resolved fluorescence microscopy: Nobel Prize in Chemistry 2014 for Eric Betzig, Stefan Hell, and William E. Moerner.," *Angew. Chem. Int. Ed. Engl.*, vol. 53, no. 51, pp. 13972–7, Dec. 2014.
- [29] J. K. Trautman, J. J. Macklin, L. E. Brus, and E. Betzig, "Near-field spectroscopy of single molecules at room temperature," *Nature*, vol. 369, no. 6475, pp. 40–42, May 1994.
- [30] N. G. Walter, C. Y. Huang, A. J. Manzo, and M. A. Sobhy, "Do-it-yourself guide: how to use the modern single-molecule toolkit," *Nat Methods*, vol. 5, no. 6, pp. 475–489, 2008.
- [31] R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes.," *Biophys. J.*, vol. 82, no. 5, pp. 2775–2783, 2002.
- [32] S. W. Hell, "Toward fluorescence nanoscopy," *Nat Biotech*, vol. 21, no. 11, pp. 1347–1355, Nov. 2003.
- [33] S. Inoué, "Foundations of confocal scanned imaging in light microscopy," in *Handbook of Biological Confocal Microscopy: Third Edition*, 2006, pp. 1–19.
- [34] E. Toprak, C. Kural, and P. R. Selvin, *Super-Accuracy and Super-Resolution: Getting Around the Diffraction Limit*, vol. 475, no. C. 2010.
- [35] M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning*, vol. 10, no. 4, pp. 128–138, Aug. 1988.
- [36] C. Cremer and T. Cremer, "Considerations on a laser-scanning-microscope with high resolution and depth of field," *Microsc. Acta*, vol. 81, no. 1, pp. 31–44, 1978.
- [37] D. Axelrod, "Cell-substrate contacts illuminated by total internal reflection fluorescence," *J Cell Biol*, vol. 89, no. 1, pp. 141–145, 1981.

- [38] D. Li, Ed., *Encyclopedia of Microfluidics and Nanofluidics*. Boston, MA: Springer US, 2008.
- [39] A. L. Stout and D. Axelrod, "Evanescent Field Excitation of Fluorescence by Epi-Illumination Microscopy," *Appl. Opt.*, vol. 28, no. 24, pp. 5237–5242, 1989.
- [40] S. Weiss, "Fluorescence Spectroscopy of Single Biomolecules," *Science (80-. )*, vol. 283, no. 5408, pp. 1676–1683, Mar. 1999.
- [41] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss, "Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor," *Proc Natl Acad Sci U S A*, vol. 93, no. 13, pp. 6264–6268, 1996.
- [42] T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, and T. Yanagida, "Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution.," *Nature*, vol. 374, no. 6522, pp. 555–9, Apr. 1995.
- [43] Y. Ishii, T. Yoshida, T. Funatsu, T. Wazawa, and T. Yanagida, "Fluorescence resonance energy transfer between single fluorophores attached to a coiled-coil protein in aqueous solution," *Chem. Phys.*, vol. 247, no. 1, pp. 163–173, Aug. 1999.
- [44] G. J. Schütz, W. Trapesinger, and T. Schmidt, "Direct observation of ligand colocalization on individual receptor molecules.," *Biophys. J.*, vol. 74, no. 5, pp. 2223–6, May 1998.
- [45] T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss, "Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism," *Proc. Natl. Acad. Sci.*, vol. 96, no. 3, pp. 893–898, Feb. 1999.
- [46] R. Kühnemuth and C. A. M. Seidel, "Principles of Single Molecule Multiparameter Fluorescence Spectroscopy," *Single Mol.*, vol. 2, no. 4, pp. 251–254, 2001.
- [47] M. Prummer, B. Sick, A. Renn, and U. P. Wild, "Multiparameter Microscopy and Spectroscopy for Single-Molecule Analytics," *Anal. Chem.*, vol. 76, pp. 1633–1640, 2004.
- [48] J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken, and C. A. M. Seidel, "Single-molecule detection and identification of multiple species by multiparameter fluorescence detection.," *Anal. Chem.*, vol. 78, no. 6, pp. 2039–50, Mar. 2006.
- [49] A. Einstein, "Investigations on the Theory of the Brownian Movement," *Ann. Phys.*, vol. 17, p. 549, 1905.
- [50] A. Lehninger, D. Nelson, and M. Cox, *Principles of Biochemistry*. 2008.
- [51] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*. W. H. Freeman, 2008.
- [52] F. H. Crick, "On protein synthesis.," *Symp. Soc. Exp. Biol.*, vol. 12, pp. 138–63, Jan. 1958.
- [53] B. Maddox, "The double helix and the 'wronged heroine'.," *Nature*, vol. 421, no. 6921, pp. 407–8, Jan. 2003.
- [54] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.," *Nature*, vol. 171, no. 4356, pp. 737–8, Apr. 1953.
- [55] R. E. Franklin and R. G. Gosling, "Molecular Configuration in Sodium Thymonucleate," *Nature*, vol. 171, no. 4356, pp. 740–741, Apr. 1953.
- [56] M. Purushothaman, A. Varghese, P. K. Mandal, and N. Gautham, "Structure of d(CCCCGGTACCGGG)2 at 1.65 Å resolution.," *Acta Crystallogr. Sect. F, Struct. Biol. Commun.*, vol. 70, no. Pt 7, pp. 860–5, Jul. 2014.

- [57] M. W. Szulik, P. S. Pallan, B. Nocek, M. Voehler, S. Banerjee, S. Brooks, A. Joachimiak, M. Egli, B. F. Eichman, and M. P. Stone, "Differential stabilities and sequence-dependent base pair opening dynamics of Watson-Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine.," *Biochemistry*, vol. 54, no. 5, pp. 1294–305, Feb. 2015.
- [58] S. Atwell, E. Meggers, G. Spraggon, and P. G. Schultz, "Structure of a copper-mediated base pair in DNA.," *J. Am. Chem. Soc.*, vol. 123, no. 49, pp. 12364–7, Dec. 2001.
- [59] R. E. Dickerson, H. R. Drew, B. N. Conner, M. L. Kopka, and P. E. Pjura, "Helix Geometry and Hydration in A-DNA, B-DNA, and Z-DNA," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 47, no. 0, pp. 13–24, Jan. 1983.
- [60] J. N. Barr and R. Fearn, "How RNA viruses maintain their genome integrity," *J. Gen. Virol.*, vol. 91, no. 6, pp. 1373–1387, 2010.
- [61] T. R. Sosnick and T. Pan, "RNA folding: models and perspectives," *Curr. Opin. Struct. Biol.*, vol. 13, no. 3, pp. 309–316, 2003.
- [62] D. Thirumalai and C. Hyeon, "Theory of RNA Folding: From Hairpins to Ribozymes," in *NonProtein Coding RNAs*, vol. 13, no. D, 2009, pp. 27–47.
- [63] J. Pan, D. Thirumalai, and S. A. Woodson, "Folding of RNA involves parallel pathways.," *J. Mol. Biol.*, vol. 273, no. 1, pp. 7–13, Oct. 1997.
- [64] R. Behrouzi, J. H. Roh, D. Kilburn, R. M. Briber, and S. A. Woodson, "Cooperative tertiary interaction network guides RNA folding.," *Cell*, vol. 149, no. 2, pp. 348–57, Apr. 2012.
- [65] S. E. Butcher and A. M. Pyle, "The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks," *Acc. Chem. Res.*, vol. 44, no. 12, pp. 1302–1311, 2011.
- [66] D. K. Treiber, "Kinetic Intermediates Trapped by Native Interactions in RNA Folding," *Science (80-. )*, vol. 279, no. 5358, pp. 1943–1946, Mar. 1998.
- [67] X. Zhuang and M. Rief, "Single-molecule folding," *Curr. Opin. Struct. Biol.*, vol. 13, no. 1, pp. 88–97, Feb. 2003.
- [68] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna, "Crystal structure of a group I ribozyme domain: principles of RNA packing.," *Science*, vol. 273, no. 5282, pp. 1678–85, Sep. 1996.
- [69] V. Lehnert, L. Jaeger, F. Michele, and E. Westhof, "New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme," *Chem. Biol.*, vol. 3, no. 12, pp. 993–1009, Dec. 1996.
- [70] W. Gilbert, "Origin of life: The RNA world," *Nature*, vol. 319, no. 6055, pp. 618–618, Feb. 1986.
- [71] A. Serganov and D. J. Patel, "Ribozymes, riboswitches and beyond: regulation of gene expression without proteins.," *Nat. Rev. Genet.*, vol. 8, no. 10, pp. 776–90, Oct. 2007.
- [72] M. B. Hoagland, M. L. Stephenson, J. F. Scott, L. I. Hecht, and P. C. Zamecnik, "A soluble ribonucleic acid intermediate in protein synthesis.," *J. Biol. Chem.*, vol. 231, no. 1, pp. 241–57, Mar. 1958.
- [73] R. C. Lee, R. L. Feinbaum, and V. Ambros, "The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*.," *Cell*, vol. 75, no. 5, pp. 843–54, Dec. 1993.
- [74] L. He and G. J. Hannon, "MicroRNAs: small RNAs with a big role in gene

- regulation.," *Nat. Rev. Genet.*, vol. 5, no. 7, pp. 522–531, 2004.
- [75] M. J. Fedor and J. R. Williamson, "The catalytic diversity of RNAs.," *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 5, pp. 399–412, May 2005.
- [76] W. G. Scott, "RNA catalysis," *Curr. Opin. Struct. Biol.*, vol. 8, no. 6, pp. 720–726, Dec. 1998.
- [77] C. Carola and F. Eckstein, "Nucleic acid enzymes.," *Curr. Opin. Chem. Biol.*, vol. 3, no. 3, pp. 274–83, Jun. 1999.
- [78] R. Fiammengo and A. Jäschke, "Nucleic acid enzymes.," *Curr. Opin. Biotechnol.*, vol. 16, no. 6, pp. 614–21, Dec. 2005.
- [79] A. M. Pyle, "Ribozymes: a distinct class of metalloenzymes.," *Science*, vol. 261, no. 5122, pp. 709–714, 1993.
- [80] T. A. Steitz and P. B. Moore, "RNA, the first macromolecular catalyst: The ribosome is a ribozyme," *Trends in Biochemical Sciences*, vol. 28, no. 8, pp. 411–418, 2003.
- [81] A. Roth, Z. Weinberg, A. G. Y. Chen, P. B. Kim, T. D. Ames, and R. R. Breaker, "A widespread self-cleaving ribozyme class is revealed by bioinformatics.," *Nat. Chem. Biol.*, vol. 10, no. 1, pp. 56–60, Jan. 2014.
- [82] T. Lönnerberg and H. Lönnerberg, "Chemical models for ribozyme action," *Current Opinion in Chemical Biology*, vol. 9, no. 6, pp. 665–673, 2005.
- [83] R. H. Symons, "Small catalytic RNAs.," *Annu. Rev. Biochem.*, vol. 61, pp. 641–71, Jan. 1992.
- [84] R. Tremblay, J. Mulhbachter, S. Blouin, J. C. Penedo, and D. Lafontaine, "Natural Functional Nucleic Acids: Ribozymes and Riboswitches," in *Functional Nucleic Acids for Analytical Applications SE - 2*, L. Yingfu and L. Yi, Eds. Springer New York, 2009, pp. 11–46.
- [85] D. M. J. Lilley, "The origins of RNA catalysis in ribozymes.," *Trends Biochem. Sci.*, vol. 28, no. 9, pp. 495–501, Oct. 2003.
- [86] N. G. Walter, K. J. Hampel, K. M. Brown, and J. M. Burke, "Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer.," *EMBO J.*, vol. 17, no. 8, pp. 2378–91, May 1998.
- [87] A. I. H. Murchie, J. B. Thomson, F. Walter, and D. M. J. Lilley, "Folding of the Hairpin Ribozyme in Its Natural Conformation Achieves Close Physical Proximity of the Loops," *Mol. Cell*, vol. 1, no. 6, pp. 873–881, May 1998.
- [88] N. G. Walter, J. M. Burke, and D. P. Millar, "Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction.," *Nat. Struct. Biol.*, vol. 6, no. 6, pp. 544–549, 1999.
- [89] X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu, "Correlating structural dynamics and function in single ribozyme molecules.," *Science*, vol. 296, no. 5572, pp. 1473–6, May 2002.
- [90] E. Tan, T. J. Wilson, M. K. Nahas, R. M. Clegg, D. M. J. Lilley, and T. Ha, "A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 16, pp. 9308–13, Aug. 2003.
- [91] M. J. B. Pereira, E. N. Nikolova, S. L. Hiley, D. Jaikaran, R. A. Collins, and N. G. Walter, "Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis.," *J. Mol. Biol.*, vol. 382, no. 2, pp. 496–509, Oct. 2008.
- [92] S. E. McDowell, J. M. Jun, and N. G. Walter, "Long-range tertiary interactions in single hammerhead ribozymes bias motional sampling toward catalytically active conformations.," *RNA*, vol. 16, no. 12, pp. 2414–2426, 2010.

- [93] X. Zhuang, "Single-molecule RNA science," *Annu Rev Biophys Biomol Struct*, vol. 34, pp. 399–414, 2005.
- [94] S. Weiss, "Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy," *Nat. Struct. Biol.*, vol. 7, no. 9, pp. 724–729, 2000.
- [95] J. R. Widom, S. Dhakal, L. A. Heinicke, and N. G. Walter, "Single-molecule tools for enzymology, structural biology, systems biology and nanotechnology: an update.," *Arch. Toxicol.*, vol. 88, no. 11, pp. 1965–85, Nov. 2014.
- [96] G. Bokinsky, D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter, and X. Zhuang, "Single-molecule transition-state analysis of RNA folding.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 16, pp. 9302–7, Aug. 2003.
- [97] J. B. Murray, A. A. Seyhan, N. G. Walter, J. M. Burke, and W. G. Scott, "The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone.," *Chem. Biol.*, vol. 5, no. 10, pp. 587–95, Oct. 1998.
- [98] F. Walter, A. I. Murchie, J. B. Thomson, and D. M. Lilley, "Structure and activity of the hairpin ribozyme in its natural junction conformation: effect of metal ions.," *Biochemistry*, vol. 37, no. 40, pp. 14195–203, Oct. 1998.
- [99] N. G. Walter and J. M. Burke, "The hairpin ribozyme: structure, assembly and catalysis.," *Curr. Opin. Chem. Biol.*, vol. 2, no. 1, pp. 24–30, 1998.
- [100] D. M. J. Lilley, "Folding and catalysis by the hairpin ribozyme," *FEBS Letters*, vol. 452, no. 1–2, pp. 26–30, 1999.
- [101] D. M. J. Lilley, "Structure, folding and mechanisms of ribozymes.," *Curr. Opin. Struct. Biol.*, vol. 15, no. 3, pp. 313–23, Jun. 2005.
- [102] T. J. Wilson, M. Nahas, L. Araki, S. Harusawa, T. Ha, and D. M. J. Lilley, "RNA folding and the origins of catalytic activity in the hairpin ribozyme.," *Blood Cells. Mol. Dis.*, vol. 38, no. 1, pp. 8–14, Jan. 2007.
- [103] D. Rueda, G. Bokinsky, M. M. Rhodes, M. J. Rust, X. Zhuang, and N. G. Walter, "Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 27, pp. 10066–71, Jul. 2004.
- [104] B. Okumus, T. J. Wilson, D. M. J. Lilley, and T. Ha, "Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic.," *Biophys. J.*, vol. 87, no. 4, pp. 2798–806, Oct. 2004.
- [105] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, and R. R. Breaker, "Genetic control by a metabolite binding mRNA," *Chem Biol*, vol. 9, no. 9, p. 1043, 2002.
- [106] R. R. Breaker, "Prospects for riboswitch discovery and analysis.," *Mol. Cell*, vol. 43, no. 6, pp. 867–79, Sep. 2011.
- [107] N. Sudarsan, J. E. Barrick, and R. R. Breaker, "Metabolite-binding RNA domains are present in the genes of eukaryotes.," *RNA*, vol. 9, no. 6, pp. 644–7, Jun. 2003.
- [108] S. Blouin, J. Mulhbach, J. C. Penedo, and D. A. Lafontaine, "Riboswitches: ancient and promising genetic regulators," *Chembiochem*, vol. 10, no. 3, pp. 400–416, 2009.
- [109] M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, and R. R. Breaker, "A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression," *Science (80-. )*, vol. 306, no. 5694, pp. 275–279, 2004.
- [110] C. A. Kellenberger and M. C. Hammond, "In vitro analysis of riboswitch-Spinach aptamer fusions as metabolite-sensing fluorescent biosensors.," *Methods Enzymol.*, vol. 550, pp. 147–72, Jan. 2015.
- [111] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, and R. R. Breaker, "An mRNA

- structure that controls gene expression by binding S-adenosylmethionine.," *Nat. Struct. Biol.*, vol. 10, no. 9, pp. 701–7, Sep. 2003.
- [112] S. Eschbach, P. St-Pierre, J. C. Penedo, and D. A. Lafontaine, "Folding of the SAM-I riboswitch: A tale with a twist," *RNA Biol*, vol. 9, no. 4, 2012.
- [113] N. Sudarsan, J. K. Wickiser, S. Nakamura, M. S. Ebert, and R. R. Breaker, "An mRNA structure in bacteria that controls gene expression by binding lysine.," *Genes Dev.*, vol. 17, no. 21, pp. 2688–97, Nov. 2003.
- [114] A. Serganov, L. Huang, and D. J. Patel, "Structural insights into amino acid binding and gene control by a lysine riboswitch.," *Nature*, vol. 455, no. 7217, pp. 1263–1267, 2008.
- [115] B. Heppell, J. Mulhbacher, J. C. Penedo, and D. A. Lafontaine, "Application of fluorescent measurements for characterization of riboswitch-ligand interactions.," *Methods Mol. Biol.*, vol. 540, pp. 25–37, 2009.
- [116] R. T. Batey, "Structure and mechanism of purine-binding riboswitches," *Quarterly Reviews of Biophysics*, vol. 45, no. 03. pp. 345–381, 2012.
- [117] W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, and R. R. Breaker, "Control of gene expression by a natural metabolite-responsive ribozyme.," *Nature*, vol. 428, no. 6980, pp. 281–286, 2004.
- [118] J. A. Collins, I. Irnov, S. Baker, and W. C. Winkler, "Mechanism of mRNA destabilization by the glmS ribozyme," *Genes Dev.*, vol. 21, no. 24, pp. 3356–3368, 2007.
- [119] P. Y. Watson and M. J. Fedor, "The glmS riboswitch integrates signals from activating and inhibitory metabolites in vivo.," *Nat. Struct. Mol. Biol.*, vol. 18, no. 3, pp. 359–363, 2011.
- [120] E. Loh, O. Dussurget, J. Gripenland, K. Vaitkevicius, T. Tiensuu, P. Mandin, F. Repoila, C. Buchrieser, P. Cossart, and J. Johansson, "A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*.," *Cell*, vol. 139, no. 4, pp. 770–9, Nov. 2009.
- [121] E. Jabri, "Non-coding RNA: Small, but in control," *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 5, pp. 361–361, May 2005.
- [122] S. A. Woodson, "Metal ions and RNA folding: a highly charged topic with a dynamic future," *Curr Opin Chem Biol*, vol. 9, no. 2, pp. 104–109, 2005.
- [123] D. Thirumalai, N. Lee, S. A. Woodson, and D. K. Klimov, "Early Events in RNA Folding," *Annu. Rev. Phys. Chem.*, vol. 52, no. 1, pp. 751–762, 2001.
- [124] G. A. Perdrizet, I. Artsimovitch, R. Furman, T. R. Sosnick, and T. Pan, "Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch," *Proceedings of the National Academy of Sciences*, vol. 109, no. 9. pp. 3323–3328, 2012.
- [125] K. C. Suddala, A. J. Rinaldi, J. Feng, A. M. Mustoe, C. D. Eichhorn, J. A. Liberman, J. E. Wedekind, H. M. Al-Hashimi, C. L. Brooks, and N. G. Walter, "Single transcriptional and translational preQ1 riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure.," *Nucleic Acids Res.*, vol. 41, no. 22, pp. 10462–75, Dec. 2013.
- [126] G. G. Hammes, Y.-C. Chang, and T. G. Oas, "Conformational selection or induced fit: a flux description of reaction mechanism.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 33, pp. 13737–41, Aug. 2009.
- [127] K. C. Suddala and N. G. Walter, "Riboswitch structure and dynamics by smFRET microscopy.," *Methods Enzymol.*, vol. 549, pp. 343–73, Jan. 2014.

- [128] P. A. Dalgarno, J. Bordello, R. Morris, P. St-Pierre, A. Dubé, I. D. W. Samuel, D. A. Lafontaine, and J. C. Penedo, "Single-molecule chemical denaturation of riboswitches," *Nucleic Acids Res*, 2013.
- [129] K. G. Daniels, N. K. Tonthat, D. R. McClure, Y.-C. Chang, X. Liu, M. A. Schumacher, C. A. Fierke, S. C. Schmidler, and T. G. Oas, "Ligand concentration regulates the pathways of coupled protein folding and binding.," *J. Am. Chem. Soc.*, vol. 136, no. 3, pp. 822–5, Jan. 2014.
- [130] E. Kim, S. Lee, A. Jeon, J. M. Choi, H.-S. Lee, S. Hohng, and H.-S. Kim, "A single-molecule dissection of ligand binding to a protein with intrinsic dynamics.," *Nat. Chem. Biol.*, vol. 9, no. 5, pp. 313–8, May 2013.
- [131] I. R. Price, A. Gaballa, F. Ding, J. D. Helmann, and A. Ke, "Mn(2+)-sensing mechanisms of yybP-ykoY orphan riboswitches.," *Mol. Cell*, vol. 57, no. 6, pp. 1110–23, Mar. 2015.
- [132] A. Wachter, "Gene regulation by structured mRNA elements," *Trends in Genetics*, vol. 30, no. 5, pp. 172–181, 2014.
- [133] W. Winkler, A. Nahvi, and R. R. Breaker, "Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.," *Nature*, vol. 419, no. 6910, pp. 952–956, 2002.
- [134] A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov, and E. Nudler, "Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria," *Cell*, vol. 111, no. 5, pp. 747–756, 2002.
- [135] B. A. M. McDaniel, F. J. Grundy, I. Artsimovitch, and T. M. Henkin, "Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 6, pp. 3083–3088, 2003.
- [136] V. Epshtein, A. S. Mironov, and E. Nudler, "The riboswitch-mediated control of sulfur metabolism in bacteria.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 9, pp. 5052–5056, 2003.
- [137] W. C. Winkler, S. Cohen-Chalamish, and R. R. Breaker, "An mRNA structure that controls gene expression by binding FMN.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 25, pp. 15908–13, Dec. 2002.
- [138] E. E. Regulski, R. H. Moy, Z. Weinberg, J. E. Barrick, Z. Yao, W. L. Ruzzo, and R. R. Breaker, "A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism.," *Mol. Microbiol.*, vol. 68, no. 4, pp. 918–32, May 2008.
- [139] T. D. Ames, D. A. Rodionov, Z. Weinberg, and R. R. Breaker, "A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate.," *Chem. Biol.*, vol. 17, no. 7, pp. 681–5, Jul. 2010.
- [140] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker, "Riboswitches Control Fundamental Biochemical Pathways in *Bacillus subtilis* and Other Bacteria," *Cell*, vol. 113, no. 5, pp. 577–586, May 2003.
- [141] M. Mandal and R. R. Breaker, "Adenine riboswitches and gene activation by disruption of a transcription terminator," *Nat Struct Mol Biol*, vol. 11, no. 1, pp. 29–35, 2004.
- [142] J. N. Kim, A. Roth, and R. R. Breaker, "Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 41, pp. 16092–7, Oct. 2007.
- [143] A. Roth, W. C. Winkler, E. E. Regulski, B. W. K. Lee, J. Lim, I. Jona, J. E. Barrick, A. Ritwik, J. N. Kim, R. Welz, D. Iwata-Reuyl, and R. R. Breaker, "A riboswitch

- selective for the queuosine precursor preQ1 contains an unusually small aptamer domain.," *Nat. Struct. Mol. Biol.*, vol. 14, no. 4, pp. 308–17, Apr. 2007.
- [144] N. Sudarsan, E. R. Lee, Z. Weinberg, R. H. Moy, J. N. Kim, K. H. Link, and R. R. Breaker, "Riboswitches in eubacteria sense the second messenger cyclic di-GMP.," *Science*, vol. 321, no. 5887, pp. 411–3, Jul. 2008.
- [145] J. W. Nelson, N. Sudarsan, K. Furukawa, Z. Weinberg, J. X. Wang, and R. R. Breaker, "Riboswitches in eubacteria sense the second messenger c-di-AMP.," *Nat. Chem. Biol.*, vol. 9, no. 12, pp. 834–9, Dec. 2013.
- [146] F. J. Grundy, S. C. Lehman, and T. M. Henkin, "The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 21, pp. 12057–62, Oct. 2003.
- [147] T. D. Ames and R. R. Breaker, "Bacterial aptamers that selectively bind glutamine.," *RNA Biol.*, vol. 8, no. 1, pp. 82–89, Oct. 2014.
- [148] M. J. Cromie, Y. Shi, T. Latifi, and E. A. Groisman, "An RNA sensor for intracellular Mg(2+).," *Cell*, vol. 125, no. 1, pp. 71–84, Apr. 2006.
- [149] C. E. Dann, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, and W. C. Winkler, "Structure and mechanism of a metal-sensing regulatory RNA.," *Cell*, vol. 130, no. 5, pp. 878–92, Sep. 2007.
- [150] J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge, and R. R. Breaker, "Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride," *Science (80-. )*, vol. 335, no. 6065, pp. 233–235, Dec. 2011.
- [151] J. X. Wang, E. R. Lee, D. R. Morales, J. Lim, and R. R. Breaker, "Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling.," *Mol. Cell*, vol. 29, no. 6, pp. 691–702, Mar. 2008.
- [152] X. Jia, J. Zhang, W. Sun, W. He, H. Jiang, D. Chen, and A. I. H. Murchie, "Riboswitch control of aminoglycoside antibiotic resistance.," *Cell*, vol. 152, no. 1–2, pp. 68–81, Jan. 2013.
- [153] P. St-Pierre, K. McCluskey, E. Shaw, C. J. Penedo, and D. A. Lafontaine, "Fluorescence tools to investigate riboswitch structural dynamics," *Biochim. Biophys. Acta - Gene Regul. Mech.*, 2014.
- [154] K. McCluskey, E. Shaw, D. A. Lafontaine, and J. C. Penedo, "Single-molecule fluorescence of nucleic acids.," *Methods Mol. Biol.*, vol. 1076, pp. 759–91, Jan. 2014.
- [155] J. F. Lemay, J. C. Penedo, R. Tremblay, D. M. Lilley, and D. A. Lafontaine, "Folding of the adenine riboswitch," *Chem Biol*, vol. 13, no. 8, pp. 857–868, 2006.
- [156] L. Chen, E. Cressina, N. Dixon, K. Erixon, K. Agyei-Owusu, J. Micklefield, A. G. Smith, C. Abell, and F. J. Leeper, "Probing riboswitch-ligand interactions using thiamine pyrophosphate analogues.," *Org. Biomol. Chem.*, vol. 10, no. 30, pp. 5924–31, Aug. 2012.
- [157] J. Mulhbacher, E. Brouillette, M. Allard, L. C. Fortier, F. Malouin, and D. A. Lafontaine, "Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways," *PLoS Pathog*, vol. 6, no. 4, p. e1000865, 2010.
- [158] R. Zhao and D. Rueda, "RNA folding dynamics by single-molecule fluorescence resonance energy transfer," *Methods*, vol. 49, no. 2, pp. 112–117, 2009.
- [159] B. Heppell, S. Blouin, A. M. Dussault, J. Mulhbacher, E. Ennifar, J. C. Penedo, and D. A. Lafontaine, "Molecular insights into the ligand-controlled organization of the SAM-I riboswitch," *Nat. Chem. Biol.*, vol. 7, no. 6, pp. 384–392, 2011.
- [160] A. Haller, U. Rieder, M. Aigner, S. C. Blanchard, and R. Micura, "Conformational capture of the SAM-II riboswitch," *Nat. Chem. Biol.*, vol. 7, no. 6, pp. 393–400, May



- 2011.
- [161] L. R. Fiegland, A. D. Garst, R. T. Batey, and D. J. Nesbitt, "Single-molecule studies of the lysine riboswitch reveal effector-dependent conformational dynamics of the aptamer domain.," *Biochemistry*, vol. 51, no. 45, pp. 9223–33, Nov. 2012.
- [162] M. D. Brenner, M. S. Scanlan, M. K. Nahas, T. Ha, and S. K. Silverman, "Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine.," *Biochemistry*, vol. 49, no. 8, pp. 1596–605, Mar. 2010.
- [163] A. Haller, R. B. Altman, M. F. Soulière, S. C. Blanchard, and R. Micura, "Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 11, pp. 4188–93, Mar. 2013.
- [164] C. E. Aitken, R. A. Marshall, and J. D. Puglisi, "An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments," *Biophys J*, vol. 94, no. 5, pp. 1826–1835, 2008.
- [165] S. Preus, "DecayFit - Fluorescence Decay Analysis Software 1.3." FluorTools.
- [166] S. A. McKinney, C. Joo, and T. Ha, "Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling," *Biophys J*, vol. 91, no. 5, pp. 1941–1951, 2006.
- [167] S. R. Eddy, "What is a hidden Markov model?," *Nat. Biotechnol.*, vol. 22, no. 10, pp. 1315–6, Oct. 2004.
- [168] F. Qin and L. Li, "Model-Based Fitting of Single-Channel Dwell-Time Distributions," *Biophys J*, vol. 87, no. 3, pp. 1657–1671, 2004.
- [169] M. Blanco and N. G. Walter, "Analysis of complex single-molecule FRET time trajectories," *Methods Enzymol.*, vol. 472, pp. 153–178, 2010.
- [170] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez Jr, and C. H. Wiggins, "Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data," *Biophys J*, vol. 97, no. 12, pp. 3196–3205, 2009.
- [171] J. O. Rawlings and D. a Dickey, *Applied Regression Analysis : A Research Tool , Second Edition Springer Texts in Statistics*, vol. 41, no. 1. 1998.
- [172] J. C. Penedo, T. J. Wilson, S. D. Jayasena, A. Khvorova, and D. M. J. Lilley, "Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements," *RNA*, vol. 10, no. 5, pp. 880–888, May 2004.
- [173] R. Tremblay, J. F. Lemay, S. Blouin, J. Mulhbachter, E. Bonneau, P. Legault, P. Dupont, J. C. Penedo, and D. A. Lafontaine, "Constitutive regulatory activity of an evolutionarily excluded riboswitch variant," *J Biol Chem*, vol. 286, no. 31, pp. 27406–27415, 2011.
- [174] D. K. Wilson, F. B. Rudolph, and F. A. Quioco, "Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations.," *Science*, vol. 252, no. 5010, pp. 1278–84, May 1991.
- [175] D. Leipply, D. Lambert, and D. E. Draper, "Ion-RNA interactions thermodynamic analysis of the effects of mono- and divalent ions on RNA conformational equilibria.," *Methods Enzymol.*, vol. 469, pp. 433–63, Jan. 2009.
- [176] E. D. Holmstrom, J. L. Fiore, and D. J. Nesbitt, "Thermodynamic origins of monovalent facilitated RNA folding.," *Biochemistry*, vol. 51, no. 18, pp. 3732–43, May 2012.
- [177] J. Noeske, H. Schwalbe, and J. Wöhnert, "Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain," *Nucleic Acids Res*, vol. 35, no. 15, pp. 5262–5273, 2007.

- [178] D. Lambert, D. Leipply, R. Shiman, and D. E. Draper, "The influence of monovalent cation size on the stability of RNA tertiary structures.," *J. Mol. Biol.*, vol. 390, no. 4, pp. 791–804, Jul. 2009.
- [179] A. Savinov, C. F. Perez, and S. M. Block, "Single-molecule studies of riboswitch folding.," *Biochim. Biophys. Acta*, vol. 1839, no. 10, pp. 1030–1045, Oct. 2014.
- [180] A. Borgia, P. M. Williams, and J. Clarke, "Single-molecule studies of protein folding.," *Annu. Rev. Biochem.*, vol. 77, pp. 101–25, Jan. 2008.
- [181] X. Michalet, S. Weiss, and M. Jäger, "Single-molecule fluorescence studies of protein folding and conformational dynamics.," *Chem. Rev.*, vol. 106, no. 5, pp. 1785–813, May 2006.
- [182] B. Schuler and W. A. Eaton, "Protein folding studied by single-molecule FRET.," *Curr. Opin. Struct. Biol.*, vol. 18, no. 1, pp. 16–26, Feb. 2008.
- [183] C. M. Dobson, "Protein folding and misfolding.," *Nature*, vol. 426, no. 6968, pp. 884–90, Dec. 2003.
- [184] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy.," *Nat. Methods*, vol. 5, no. 6, pp. 491–505, Jun. 2008.
- [185] T. Pan and T. R. Sosnick, "Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity," *Nat. Struct. Biol.*, vol. 4, no. 11, pp. 931–938, Nov. 1997.
- [186] W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, and S. M. Block, "Direct observation of hierarchical folding in single riboswitch aptamers.," *Science*, vol. 319, no. 5863, pp. 630–3, Feb. 2008.
- [187] P. Venkatesu, M.-J. Lee, and H.-M. Lin, "Thermodynamic characterization of the osmolyte effect on protein stability and the effect of GdnHCl on the protein denatured state.," *J. Phys. Chem. B*, vol. 111, no. 30, pp. 9045–56, Aug. 2007.
- [188] W. Reisner, N. B. Larsen, A. Silahatoglu, A. Kristensen, N. Tommerup, J. O. Tegenfeldt, and H. Flyvbjerg, "Single-molecule denaturation mapping of DNA in nanofluidic channels.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 30, pp. 13294–9, Jul. 2010.
- [189] U. D. Priyakumar, C. Hyeon, D. Thirumalai, and A. D. MacKerell, "Urea destabilizes RNA by forming stacking interactions and multiple hydrogen bonds with nucleic acid bases," *J. Am. Chem. Soc.*, vol. 131, no. 49, pp. 17759–17761, 2009.
- [190] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech, "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.," *Cell*, vol. 31, no. 1, pp. 147–157, 1982.
- [191] C. R. Burke and A. Lupták, "Catalytic RNA," in *eLS*, John Wiley & Sons, Ltd, 2001.
- [192] G. Bokinsky, D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter, and X. Zhuang, "Single-molecule transition-state analysis of RNA folding.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 16, pp. 9302–9307, 2003.
- [193] S. Schubert and J. Kurreck, "Ribozyme- and deoxyribozyme-strategies for medical applications.," *Curr. Drug Targets*, vol. 5, no. 8, pp. 667–81, Nov. 2004.
- [194] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, and G. Bruening, "Autolytic processing of dimeric plant virus satellite RNA.," *Science*, vol. 231, no. 4745, pp. 1577–80, Mar. 1986.
- [195] C. J. Hutchins, P. D. Rathjen, A. C. Forster, and R. H. Symons, "Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid," *Nucleic Acids Res.*, vol. 14, no. 9, pp. 3627–3640, May 1986.

- [196] T. O. Diener, "Circular RNAs: relics of precellular evolution?," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 23, pp. 9370–4, Dec. 1989.
- [197] J. Han and J. M. Burke, "Model for general acid-base catalysis by the hammerhead ribozyme: PH-activity relationships of G8 and G12 variants at the putative active site," *Biochemistry*, vol. 44, no. 21, pp. 7864–7870, 2005.
- [198] C. Hammann, A. Luptak, J. Perreault, and M. de la Pena, "The ubiquitous hammerhead ribozyme," *RNA*, vol. 18, no. 5, pp. 871–885, 2012.
- [199] W. G. Scott, J. T. Finch, and A. Klug, "The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage," *Cell*, vol. 81, no. 7, pp. 991–1002, 1995.
- [200] H. W. Pley, K. M. Flaherty, and D. B. McKay, "Three-dimensional structure of a hammerhead ribozyme.," *Nature*, vol. 372, no. 6501, pp. 68–74, 1994.
- [201] S. R. Morrissey, T. E. Horton, and V. J. DeRose, "Mn<sup>2+</sup> Sites in the Hammerhead Ribozyme Investigated by EPR and Continuous-Wave Q-band ENDOR Spectroscopies," *J. Am. Chem. Soc.*, vol. 122, no. 14, pp. 3473–3481, 2000.
- [202] O. Schiemann, J. Fritscher, N. Kisseleva, S. T. Sigurdsson, and T. F. Prisner, "Structural Investigation of a High-Affinity MnII Binding Site in the Hammerhead Ribozyme by EPR Spectroscopy and DFT Calculations. Effects of Neomycin B on Metal-Ion Binding," *ChemBiochem*, vol. 4, no. 10, pp. 1057–1065, 2003.
- [203] N. Kisseleva, A. Khvorova, E. Westhof, and O. Schiemann, "Binding of manganese (II) to a tertiary stabilized hammerhead ribozyme as studied by electron paramagnetic resonance spectroscopy," *RNA*, vol. 11, no. 1, pp. 1–6, 2005.
- [204] T. E. Edwards and S. T. Sigurdsson, "EPR spectroscopic analysis of U7 hammerhead ribozyme dynamics during metal ion induced folding," *Biochemistry*, vol. 44, no. 38, pp. 12870–12878, 2005.
- [205] O. Schiemann, R. Carmieli, and D. Goldfarb, "W-band<sup>31</sup>P-ENDOR on the high-affinity Mn<sup>2+</sup> binding site in the minimal and tertiary stabilized hammerhead ribozymes," *Appl. Magn. Reson.*, vol. 31, no. 3–4, pp. 543–552, 2007.
- [206] G. S. Bassi, A. I. Murchie, F. Walter, R. M. Clegg, and D. M. Lilley, "Ion-induced folding of the hammerhead ribozyme: a fluorescence resonance energy transfer study.," *EMBO J.*, vol. 16, no. 24, pp. 7481–7489, 1997.
- [207] G. S. Bassi, N. E. Møllegaard, A. I. H. Murchie, and D. M. J. Lilley, "RNA folding and misfolding of the hammerhead ribozyme," *Biochemistry*, vol. 38, no. 11, pp. 3345–3354, 1999.
- [208] G. S. Bassi, A. I. Murchie, and D. M. Lilley, "The ion-induced folding of the hammerhead ribozyme: core sequence changes that perturb folding into the active conformation.," *RNA*, vol. 2, no. 8, pp. 756–68, Aug. 1996.
- [209] M. De la Peña, S. Gago, and R. Flores, "Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity.," *EMBO J.*, vol. 22, no. 20, pp. 5561–70, Oct. 2003.
- [210] A. Khvorova, A. Lescoute, E. Westhof, and S. D. Jayasena, "Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity.," *Nat. Struct. Biol.*, vol. 10, no. 9, pp. 708–712, 2003.
- [211] M. Martick and W. G. Scott, "Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis," *Cell*, vol. 126, no. 2, pp. 309–320, 2006.
- [212] D. E. Ruffner and O. C. Uhlenbeck, "Thiophosphate interference experiments locate phosphates important for the hammerhead RNA self-cleavage reaction.," *Nucleic Acids Res.*, vol. 18, no. 20, pp. 6025–6029, 1990.

- [213] W. G. Scott, J. B. Murray, J. R. Arnold, B. L. Stoddard, and A. Klug, "Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme.," *Science*, vol. 274, no. 5295, pp. 2065–2069, 1996.
- [214] C. Cremer and T. Cremer, "Considerations on a laser-scanning-microscope with high resolution and depth of field.," *Microsc. Acta*, vol. 81, no. 1, pp. 31–44, Sep. 1978.
- [215] P. Schwille, F. J. Meyer-Almes, and R. Rigler, "Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution.," *Biophys. J.*, vol. 72, no. 4, pp. 1878–1886, 1997.
- [216] J. Ries and P. Schwille, "Fluorescence correlation spectroscopy," *BioEssays*, vol. 34, no. 5, pp. 361–368, 2012.
- [217] O. Krichevsky, "Fluorescence correlation spectroscopy: the technique and its applications," *Reports Prog. Phys.*, vol. 65, pp. 251–297, 2002.
- [218] M. D. Eisaman, J. Fan, A. Migdall, and S. V Polyakov, "Invited review article: Single-photon sources and detectors.," *Rev. Sci. Instrum.*, vol. 82, no. 7, p. 071101, Jul. 2011.
- [219] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, vol. 81, no. ISBN-10 3-540-26047-1. 2005.
- [220] W. Becker, *The bh TCSPC Handbook*, 5th ed. Berlin: Becker & Hickl GmbH, 2012.
- [221] D. Magde, E. Elson, and W. W. Webb, "Thermodynamic fluctuations in a reacting system measurement by fluorescence correlation spectroscopy," *Phys. Rev. Lett.*, vol. 29, no. 11, pp. 705–708, 1972.
- [222] R. Rigler, Ü. Mets, J. Widengren, and P. Kask, "Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion," *Eur. Biophys. J.*, vol. 22, no. 3, Aug. 1993.
- [223] E. Sezgin and P. Schwille, "Fluorescence techniques to study lipid dynamics," *Cold Spring Harb. Perspect. Biol.*, vol. 3, 2011.
- [224] E. Haustein and P. Schwille, "Fluorescence correlation spectroscopy: novel variations of an established technique.," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 36, pp. 151–169, 2007.
- [225] M. Eigen and R. Rigler, "Sorting single molecules: application to diagnostics and evolutionary biotechnology.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 13, pp. 5740–5747, 1994.
- [226] P. O. Gendron, F. Avaltroni, and K. J. Wilkinson, "Diffusion coefficients of several rhodamine derivatives as determined by pulsed field gradient nuclear magnetic resonance and fluorescence correlation spectroscopy," *J. Fluoresc.*, vol. 18, no. 6, pp. 1093–1101, 2008.
- [227] J. W. Krieger and J. Langowski, "QuickFit 3.0 (status: beta, compiled: 12/12/2014, SVN: 3157): A data evaluation application for biophysics." 2014.
- [228] C. Rouillon, M. Zhou, J. Zhang, A. Politis, V. Beilsten-Edmands, G. Cannone, S. Graham, C. V. Robinson, L. Spagnolo, and M. F. White, "Structure of the CRISPR interference complex CSM reveals key similarities with cascade," *Mol. Cell*, vol. 52, no. 1, pp. 124–134, 2013.
- [229] M. B. Borgia, A. Borgia, R. B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke, "Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins.," *Nature*, vol. 474, no. 7353, pp. 662–665, 2011.
- [230] G. Cevc and H. Richardsen, "Lipid vesicles and membrane fusion," *Adv. Drug Deliv. Rev.*, vol. 38, no. 3, pp. 207–232, Aug. 1999.
- [231] B. Valeur and M. N. Berberan-Santos, *Molecular fluorescence: principles and*

*applications*. John Wiley & Sons, 2012.

- [232] B. Chon, K. Briggman, and J. Hwang, "Single molecule confocal fluorescence lifetime correlation spectroscopy for accurate nanoparticle size determination.," *Phys. Chem. Chem. Phys.*, vol. 16, no. 26, pp. 13418–25, Jul. 2014.
- [233] K. T. Schroeder, P. Daldrop, and D. M. J. Lilley, "RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn.," *Structure*, vol. 19, no. 9, pp. 1233–40, Sep. 2011.
- [234] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera—a visualization system for exploratory research and analysis.," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–12, Oct. 2004.
- [235] S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke, and C. A. M. Seidel, "A toolkit and benchmark study for FRET-restrained high-precision structural modeling.," *Nat. Methods*, vol. 9, no. 12, pp. 1218–25, Dec. 2012.
- [236] E. A. Alemán, C. de Silva, E. M. Patrick, K. Musier-Forsyth, and D. Rueda, "Single-Molecule Fluorescence Using Nucleotide Analogs: A Proof-of-Principle.," *J. Phys. Chem. Lett.*, vol. 5, no. 5, pp. 777–781, Mar. 2014.
- [237] J. R. Taylor, *An Introduction to Error Analysis*, vol. 101. 1997.