FRET Below and Beyond: New tools for single-molecule fluorescence studies of DNA and Proteins

Benjamin Ambrose

A Thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy in the Faculty of Science

Department of Chemistry

April 2021

Supervisor: Dr T. D. Craggs

Abstract

Single-molecule FRET (smFRET) is a family of techniques within single-molecule fluorescence spectroscopy, which involve measuring the FRET efficiency between a pair of dyes within or between single molecules. As FRET results in a signal that is dependent on the interdye distance, measuring FRET on single molecules allows the experimenter to monitor conformations of biomolecules in a manner free from ensemble averaging. In recent years, advances in smFRET techniques have given access to absolute intramolecular distances for resolving structures, and a wealth of insight into biomolecular kinetics at equilibrium that would not be feasible with ensemble techniques. However, a lack of standardization in technique, and high entry barriers to acquiring instrumentation have made smFRET techniques inaccessible to all but the most specialist of single-molecule fluorescence spectroscopy researchers. Here I present work demonstrating the reproducibility of the accurate FRET technique, alongside many other smFRET labs, on a cost effective and open-source confocal smFRET microscope, as well as a new contact-quenching based method for measuring short distances, which can work in parallel with smFRET experiments, adding to the tool kit of single-molecule fluorescence spectroscopists. Finally, I will present work using these methods to investigate various biomolecular systems, exploring the conformational landscapes of DNA binding proteins, and the flexibility of the damaged substrates they recognise.

Declaration of Authorship

I, the author, confirm that this Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not been previously been presented for an award at this, or any other, University.

As with all collaborative work, much of the data here would not be possible without the contributions of others. Wherever possible at the beginning of chapters and in the text I have endeavoured to make clear the contributions of other researchers.

Work presented in this thesis has formed part of the following publications, as indicated at the beginning of each chapter;

- Bennet, I. A. et al. Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1. *Nucleic Acids Res*. **46**, 5618–5633 (2018).

- Hellenkamp, B. et al. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat. Methods* 15, 669–676 (2018).

- Ambrose, B. et al. The smfBox is an open-source platform for single-molecule FRET. *Nat. Commun.* 11, 5641 (2020).

Acknowledgements

I would primarily like to thank all past and present members of the Craggs Lab, including Tim himself, for what has been an enjoyable and fruitful three years. None of this work would have been possible without the heroic efforts of John Cully and James Baxter in designing and building (and rebuilding) the original smfBox. The masters students who have worked under my guidance have made genuine contributions to my own scientific efforts as evidenced in this thesis, and I hope they learned as much from my supervision as I learned supervising them. James Piggin was a brave ally in my initial battles with the smfBox back in the early days when we had no way to actually analyse any of the data we were recording, and his keen IT skills were invaluable to the efforts of making the microscope operable with a single computer instead of two. I also thank both Matthew Willmott and Callum Johnston for battling through the preliminary quenching experiments, especially during the long and intermittent periods of getting questionable data, which we now know was largely down to the lasers not being plugged in correctly. I would like to expressly thank both of the Erasmus students Marleen Aaldering and Anna van den Boom for building the TCSPC spectrometer I used in this thesis, as well as offering an apology to Anna for advising her to clean a cuvette with a solvent that would eventually melt it. I would like to thank my fellow PhD students (and post-doc), Dylan George, Alice Rhind-Tutt, Victoria Hill, and Mahmoud Abelhamid for fruitful scientific discussions, both in and outside of the pub. I would like to especially thank Alice for surviving the comings and goings of E81 with me and suffering through my endless jokes about physicists, as well as Mahmoud for revolutionising the lab with his 3D-printing prowess.

I must also thank the seemingly endless list of collaborators who I have worked with along the way. I thank Ian Bennet, Dave Finger, and Jane Grasby for their work with FEN1, and Dave in particular for showing me how just about everything in the department works, sometimes more than once. I thank Benji Bateman for his work on the LABVIEW acquisition software, and Elliot Steele for turning it into smOTTER, even though he probably now regrets ever suggesting that he could. I thank Olivia Hill for embarking on the TIRF-FRET experiments with me, which we regretfully did not get far enough with to include in this thesis. I thank Ali Twelvetrees and Evie Smith for their invaluable work as test users of the smfBox, and for suffering through my rambling descriptions of how it works. I thank David Williams, Esther Allen, and everyone else in the Williams lab for their help working on ATL1. I thank Jamieson Howard, Steven Quinn, Mark Leake, Agnes Noy, and Nettie Alevropoulos-Borrill for their collaboration working on RepA. I would like to thank Grant Hill and Rob Shaw for teaching me molecular dynamics, and Tristan-Johnston Wood for running the simulations I have shown in this work. I thank Jenny Clark, Eitan Lerner, and Shimon Weiss for valuable discussions on qqFRET. I also have to thank Anders Barth for all of his help working with PAM, as well as offering his expertise on single-molecule techniques in general. In that spirit, I also thank the entire smFRET community for the many fruitful interactions I have had with them to various extents, in particular to Claus Seidel, for filling my head with a hundred ideas about quenching on a long flight from San Diego to London, some of which were helpful ideas. I also thank everyone else who I have probably forgotten to thank, it was not on purpose.

Lastly I would like to once again thank Tim Craggs for getting me through these past 3 years and doing an absolutely stellar job at his first try as a PhD supervisor. If my tenure in the lab can be called a success, then I hope there are many more successes to come. Despite the many gaffes alluded to above, I would say that the Craggs lab has a bright future ahead of it. And a lot of Tequila, probably.

Table of Contents

Glossary of Terms and Abbreviations

Å	Ångstrom
A	Adenine/Adenosine
$A_{ex}A_{em}$ / AA	Acceptor emitted photons under acceptor excitation
ALEX	Alternating laser excitation
aALEX	Asymmetric alternating laser excitation
APBS	All photons burst search
APD	Avalanche photodiode
AFM	Atomic force microscopy
ATL1	Alkyltransferase-like protein 1
AV	Accessible volume
BRET	Bioluminescent resonance energy transfer
BSA	Bovine serum albumin
BVA	Burst variance analysis
C	Cytosine/Cytidine
CCD	Charge coupled device
CPU	Central processing unit
Cryo-EM	Cryo-electron microscopy
C6-dT	deoxyribo-thymine modified with a 6 carbon linker
C2-dT	deoxyribo-thymine modified with a 2 carbon linker
Dir. Ex	Direct excitation
DCBS	Dual channel burst search
$D_{ex}D_{em}$ / DD	Donor emitted photons under donor excitation
$D_{ex}A_{em}$ / DA	Acceptor emitted photons under donor excitation
dPDA	dynamic probability/photon distribution analysis
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
E*	Uncorrected FRET efficiency
ES	Efficiency-stoichiometry
FCS	Fluorescence correlation spectroscopy
FCCS	Fluorescence cross correlation spectroscopy
FEN1	Flap endonuclease-1
FPS	FRET Positioning and screening
FRET	Förster/Fluorescence resonance energy transfer
G	Guanine/Guanosine
GFP	Green fluorescent protein
HDF5 / H5	Hierarchical data format 5
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HIV	Human immunodeficiency virus
HMM	Hidden markov model
LED	Light emitting diode
Lk	Leakage
MalE	Maltose binding protein
MD	Molecular dynamics
MFD	Multi-parameter fluorescence detection
ms-ALEX	millisecond alternating laser excitation, ie. TIRF-ALEX
NA	Numerical aperture
NHS	N-hydroxysuccinimide
NIR	Near infrared
NMDA	N-methyl-d-aspartate
NMR	Nuclear magnetic resonance
ns	Nanosecond

ns-ALEX	Nanosecond alternating laser excitation, *also PIE*
NSOM	Near field scanning optical microscopy
OGG1	Oxoguanine glycosylase
O6mG	O^6-methylguanine
PAM	PIE analysis with MATLAB
PAX	Periodic acceptor excitation
PDA	Probability/photon distribution analysis
PDB	Protein data bank
PIE	Pulsed interleaved excitation, *also ns-ALEX*
qqFRET	Quantitative quenching FRET
QR	Quenching Ratio
RepA	Replicative Helicase
RET	Resonance energy transfer
R_0	Förster radius, or critical distance, giving 50% FRET efficiency
R6G	Rhodamine-6G
S*	Uncorrected stoichiometry
SAXS	Small angle x-ray scattering
smFRET	Single-molecule FRET
SPAD	Single-photon avalanche detector
SPIM	Selective-plane illumination microscopy
spFRET	Single-pair FRET
T	Thymine/Thymidine
TCSPC	Time correlated single photon counting
TD-DFT	Time dependent density functional theory
TIRF	Total internal reflection fluorescence
TIRFM	Total internal reflection fluorescence microscopy
Tris	Tris-aminomethane
XRD	X-ray diffraction
μs-ALEX	Microsecond alternating laser excitation
UvrD	DNA Helicase II
α	Accurate FRET correction leakage term
β	Accurate FRET correction excitation efficiency term
δ	Accurate FRET correction direct excitation term
γ	Accurate FRET correction detection efficiency term
2CDE	2 Channel kernel density estimator
8-oxo-G	8-oxo-Guanine

Chapter 1. Introduction

1.1. Chapter Overview

In the following work I present an open-source confocal smFRET microscope capable of various single-molecule fluorescence spectroscopy techniques, including accurate FRET for absolute distance determination. In addition to validating the capabilities of this microscope to perform various single-molecule fluorescence spectroscopy techniques, I use this microscope to participate in a worldwide benchmarking study validating the accurate FRET technique, develop a new method for measuring short distances, and apply these techniques to a number of DNA and protein systems. First, I will introduce fluorescence, the theory behind FRET as a phenomenon and an experimental technique, and the various forms of smFRET techniques, some of which will be employed in the following experimental chapters.

1.2 Luminescence

Luminescence occurs when a material in an electronically excited state relaxes to a lower energy state via the emission of a photon. These excited states are typically reached first by absorption of a photon, moving from the ground state S_0 to S_1. S_0 and S_1 are *single* states, meaning that the electrons are paired, ie, the electron in the higher orbital has the opposite spin as the electron in the lower orbital. As shown in Fig. 1.2.1, after excitation, energy is rapidly lost (to the local environment) via relaxation to the lowest vibrational state of S_1. From here, further relaxation can occur via a number of routes.

In Fluorescence, the electron returns to S_0 by emitting a photon, and then further relaxes to the lowest vibrational state of $S_{0.}$ As electrons are excited from the lowest vibrational state of S_0 to higher vibrational states of $S_{1,}$ and from the lowest vibrational state of S_1 to higher vibrational states of $S_{0,}$ the emitted photon will usually be of lower energy than that of the absorbed photon, with the wavelength spectrum of absorbed photons being dependent on the vibrational levels of the excited state, and the emission spectrum dependent on those of the ground state.

Alternatively, the electron can undergo "intersystem crossing", moving to a triplet state. In a triplet state, the electron has an opposite spin, and as such the movement from a singlet state to a triplet state (or back) is "forbidden", as the electron must change spin. Relaxation from a triplet state by emission of a photon is called phosphorescence, and occurs much more slowly than fluorescence as the electron must make another forbidden transition from T_1 to $S_{0.}$

Not all relaxation processes involve the emission of a photon however. In so called "radiationless relaxation", the electron moves from a low energy vibrational state of S_1 to a significantly higher vibrational state of S_0 via internal conversion, and then continues to relax without emitting a photon.

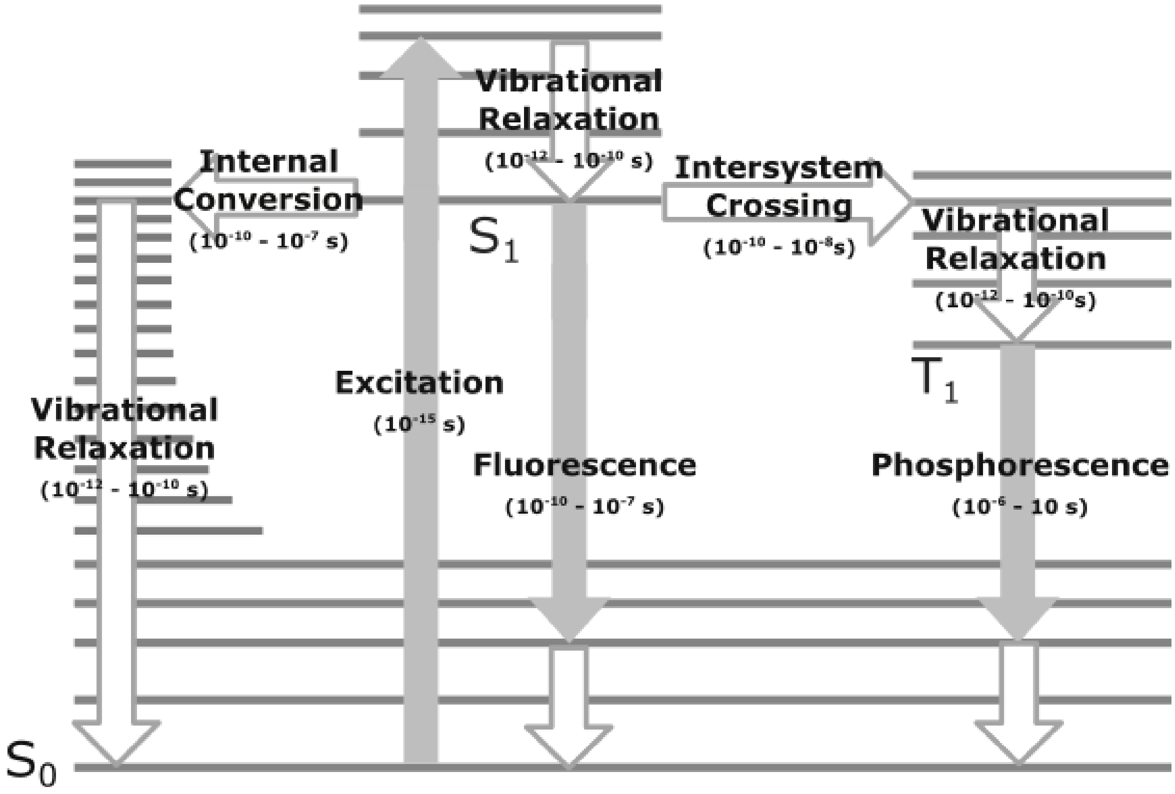


Fig. 1.2.1: Jablonski diagram of fluorescence and related processes with time scales. After excitation from singlet state S_0 to S_1 and vibrational relaxation, a fluorophore can further relax through a number of processes. Internal conversion to a higher vibrational state of S_0 can result in radiationless relaxation. A photon can be emitted by fluorescence, returning to S_0, or after intersystem crossing to a triplet state, a photon can be emitted by phosphorescence.

Whenever a fluorescent molecule is exposed to conditions which reduce its fluorescence emission (by providing alternative, non-radiative relaxation paths), its fluorescence is said to be “quenched”. Quenching mechanisms are typically divided into two forms; static quenching and dynamic quenching. In static quenching, the interaction occurs in the ground state, preventing the molecule from ever reaching an excited state. In dynamic quenching however, the molecule can reach the excited state, but loses the energy non-radiatively. Many dynamic quenching mechanisms involve *energy transfer*, to another molecule. Whilst FRET, one of the more well-known mechanisms of energy transfer, is the subject of much of this thesis, one other form of energy transfer is dexter electron transfer, where the excited electron itself is transferred from one molecule to another via a wavefunction overlap between the two. A third form of dynamic quenching is exciplex formation, where the molecule forms an *excited state complex* with another molecule[1].

1.3. FRET Theory

FRET, which stands for either Förster resonance energy transfer, or Fluorescence resonance energy transfer, is a mechanism by which excited state energy can transfer non-radiatively between two chromophores. The phenomenon is often used in techniques where one or more of the chromophores involved are fluorescent, and it is the fluorescence of either (or both) molecules which is measured, hence *fluorescence* resonance energy transfer. However, the phenomenon refers strictly to the energy transfer, and not the fluorescence emission which may occur before or after.

FRET has a strong distance dependence over length scales relevant to biomolecules (nanometres and Ångstroms), and as such it has been employed to great effect in biochemistry and biophysics for determining structural information, probing interactions which occur over distances too small to monitor by traditional microscopy techniques, and under conditions which are not amenable to crystallisation for x-ray diffraction experiments[2,3]. However, long before FRET could be used in a biological setting, the mathematical and photophysical details of the technique were elucidated by physicists in the first half of the 20th century.

An experiment involving fluorescence of a thallium-mercury vapor mix[4] is believed to be the first observation of FRET at the time called "sensitized fluorescence". Later, Perrin described energy transfer between two identical fluorophores (now called Homo-FRET), which included a distance dependence[5], and would be built on by Förster to form a complete quantum mechanical description of FRET.

Förster's first publication on FRET was in 1946 in which he explored Perrins work in the context of photosynthesis and several other experiments involving dye mixtures[6], and then in 1948 gave a full quantum mechanical description of the process[7], detailing many of the equations and concepts used by FRET researchers today. As such, it is Förster for whom the "F" in the FRET acronym is sometimes named. The exact choice of words to be used in the acronym is at time of writing a matter of some debate within the research community; whilst *fluorescence* can misleadingly suggest an incorrect description of the phenomenon, there is some reluctance to honor Förster due to his membership of the National Socialist German Workers' Party. Whilst there are many alternatives that have been used such as just "RET", the word *FRET* has attained some status of name recognition in of itself within the literature, and so one suggested alternative is to simply not expand the acronym at all. Therefore, for the remainder of this work the phenomenon will be referred to just as "FRET".

Förster's description of "intermolecular energy migration" showed that radiationless transfer of energy from one molecule to another other occurs via a dipole-dipole interaction in which the excited state of the first molecule, termed the *donor,* couples to the *acceptor* (Fig.

1.3.1). The transfer rate is dependent on the rate of coupling, which in turn is inversely proportional to the sixth power of the distance between the two molecules.

In Förster's theory, transfer of photons from the donor to the acceptor will proceed with a rate characteristic for any two dyes given the distance between them and the relative orientation of their dipoles. The rate of this transfer can be compared to the rates of other decay process, including donor fluorescence, to give an equation for the efficiency of transfer:

$$E = \frac{k_{ET}}{k_{ET} + k_f + k_i}$$

Eq.1.3.1

Where k_{ET} is the rate of energy transfer, k_f is the rate of donor fluorescence and k_i is the rate of non-fluorescence decay (in the absence of an acceptor).

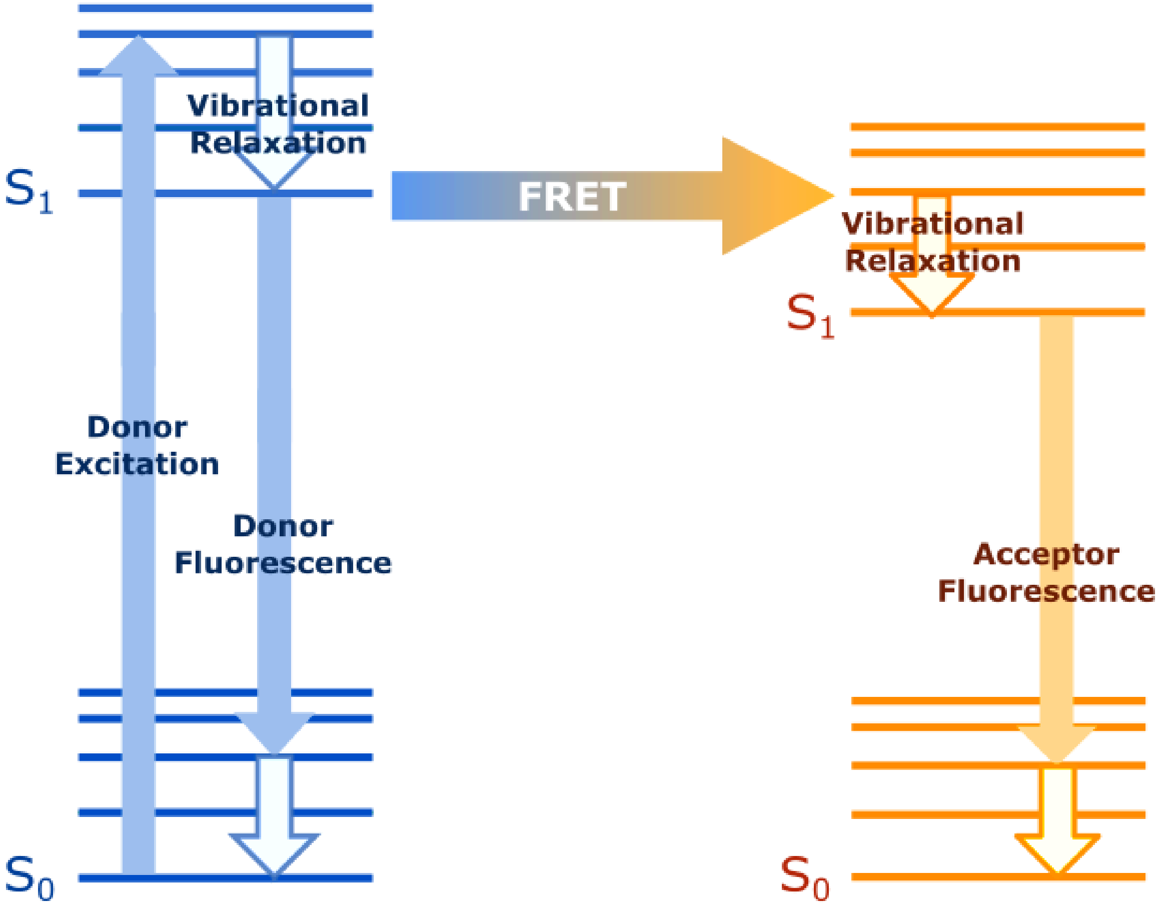


Fig. 1.3.1 Jablonski diagram of FRET. Upon excitation and vibrational relaxation, the donor excited state can decay either through fluorescence or through resonance energy transfer to the acceptor. The acceptor then fluoresces, but only after also undergoing vibrational relaxation, which further reduces the energy of the emitted photon.

It is notable that Förster not only gave a full quantum mechanical description of energy transfer, but that he expressed the rate of energy transfer in experimentally measurable values, such that the efficiency of transfer relative to relaxation of the donor can be expressed as a critical distance, R_0 where the transfer efficiency is 50%, due to the rate of transfer being equal to competing de-excitation processes. This critical distance, today known as the Förster radius, or simply R_0, can be determined for any two dyes, using the following equation:

$$\frac{R_0}{\text{Å}} = 0.2108 \sqrt[6]{\frac{\Phi_{F,D}\kappa^2}{n_{im}^4} \; \frac{J}{M^{-1}cm^{-1}nm^4}}$$

Eq. 1.3.2

Where ϕ_D is the donor quantum yield, κ is the orientation factor, N_A is Avogadro's constant, n is the refractive index of the inter-dye medium, J is the spectral overlap integral (Eq. 1.3.4). The units for R_0 are Ångstroms when the overlap integral is expressed as $M^{-1}cm^{-1}nm^4$.

Thus, when all other parameters can be assumed to be constant and the R_0 for any two dyes is known, a simple relationship between transfer efficiency and distance emerges and is given by:

$$E = \frac{1}{1 + \left(\frac{\mathrm{r}}{R_0}\right)^6}$$

Eq. 1.3.3

Where r is the distance between the two fluorophores expressed in the same units as the R_0.

The overlap integral of a dye pair is effectively a measure of the degree to which the energies of the donor excited state correspond to the energies over which the acceptor can absorb photons. It is given as the product of acceptor excitation coefficient (ε_A) the donor emission intensity, and wavelength (λ) to the fourth power, integrated across all wavelengths, where the donor emission spectrum is normalised to 1. (Fig.1.3.2, and Eq.1.3.4). Therefore, when choosing an appropriate dye pair, dyes with a high degree of overlap between donor emission spectrum and acceptor excitation spectrum are typically desired for their consequently large R_0, and vice versa for a smaller R_0.

The equation for the J integral is;

$$J = \int_0^\infty \bar{F}_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda$$

$$\text{Where; } \int_0^\infty \bar{F}_D(\lambda)d\lambda = 1$$

Eq. 1.3.4

Where ε_A is the extinction coefficient of the acceptor in $M^{-1}cm^{-1}$, and λ is wavelength in nm, giving J units of $M^{-1}cm^{-1}nm^4$.

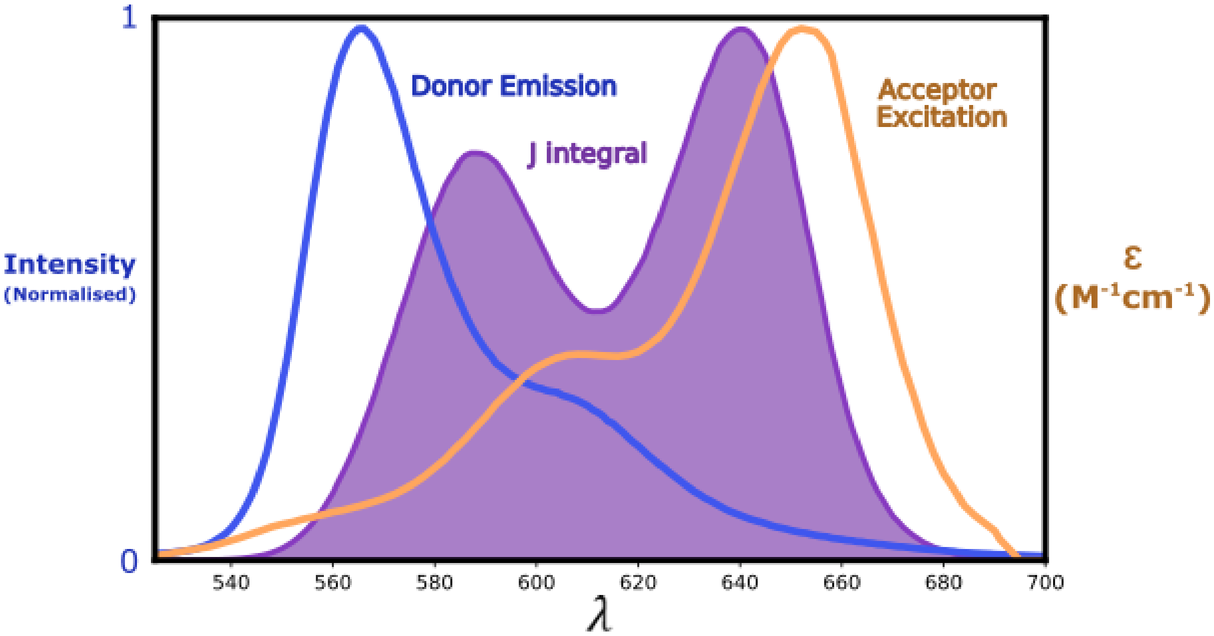


Fig1.3.2 Showing the spectral overlap density. The emission spectrum of the donor is in blue, the absorption spectrum of the acceptor is in the orange, and the J integral is in purple.

The spectral overlap dependence can often lead to an incorrect interpretation of the FRET mechanism, as at first glance it would appear to describe the proportion of photons which can be emitted by the donor and subsequently absorbed by the acceptor, however FRET is a non-radiative process, not a reabsorption process, and the energy transmitted is not in the form of a real emitted photon. The overlap integral is in fact a measure of resonance between the two fluorophores, as wavelengths within the overlap correspond to energy transitions over which the donor may relax and the acceptor may excite, weighted by the likelihood that those states are populated. It is useful that this parameter is expressed in a form which can be derived from two experimentally measurable spectra.

Another important variable in the determination of R_0 is kappa squared; the orientation factor. It is a measure of alignment of the acceptor excitation dipole with the *electric field* of the donor emission dipole. The orientation factor is given by the following equation;

$$\kappa^2 = (cos\theta_{DA} - 3cos\theta_D cos\theta_A)^2$$

Eq. 1.3.5

Where θ_{AD} is the angle of the unit vector between the donor and acceptor dipole moments, and θ_D and θ_A are the angles of the emission dipole moment of the donor and the absorption dipole moment of the acceptor with respect to this unit vector (see Fig. 1.3.3).

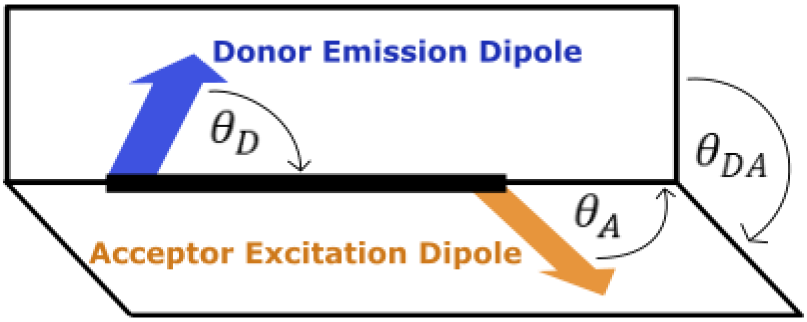


Fig. 1.3.3 The angles used in calculation of the kappa squared.

Eq 1.3.5 above gives kappa squared as a function of both the alignment between the two dipoles, and the angle between them, which may lead to the common misconception that the key factor in kappa squared is that the dipoles themselves are aligned, ie. "when the dipoles are parallel kappa squared will be high, and when they are perpendicular kappa squared will be zero". However, the equation simply uses the orientation of the donor dipole to refer to the electric field around it, as seen in Fig. 1.3.4, kappa squared can be as low as zero when the dipoles are parallel, if the acceptor dipole is perpendicular to the donor electric field, and as high as 2 when the dipoles are perpendicular, if the acceptor is parallel to the donor electric field.

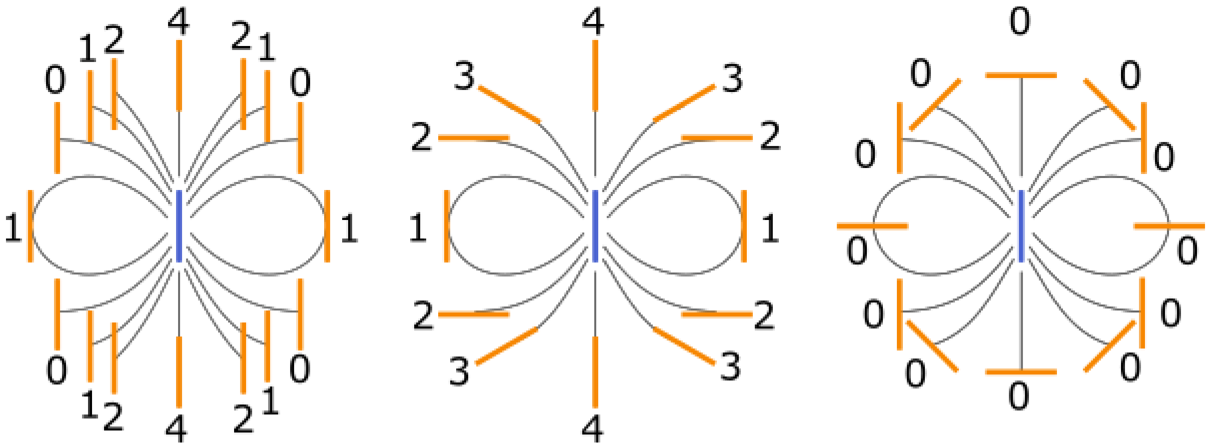


Fig. 1.3.4 Some examples of possible kappa squared, with the donor dipole in blue, it's electric field as grey lines, and the acceptor dipole in orange. On the left, the acceptor dipole is parallel to the donor dipole, and kappa squared will be anywhere from 0 to 4. In the middle, the acceptor dipole is parallel to the electric field of the donor dipole, and can be anywhere from 1 to 4. On the right, the acceptor dipole is perpendicular to the electric field of the donor, in which case kappa squared is always 0.

As kappa squared can assume a range of quantities between 0 and 4, its value is of great importance to the final relationship between distance and FRET efficiency. Since the vast majority of FRET experiments are concerned with distance rather than orientation, it is often desired that the value of κ^2 remains constant and known. To this end, conditions are typically sought such that one or both fluorophores may freely rotate and translate with respect to each other such that the dynamic average of kappa squared is ⅔. The assumption of kappa squared as ⅔ is an important factor in inferring distances from FRET efficiencies, and much thought has been given to whether it is valid, and under what conditions it remains so[8–12].

1.4. FRET as a Spectroscopic Ruler

Almost 20 years after Förster's description of the physical mechanism, the first experiment was performed to demonstrate that FRET can be used as a measure of distance[13]. Stryer and Haugland demonstrated that, given FRET has a strong distance dependence, measuring the efficiency of the process by exciting the donor molecule and measuring the relative fluorescence of the donor and the acceptor could give insight into the distance between them. Stryer and Haugland synthesised poly-L-proline peptides of varying length, with an α-naphthyl group at one end as a donor and a dansyl group at the other as an acceptor (Fig. 1.3.1).

They then measured the excitation spectrum of these peptides at the emission wavelength of the acceptor, and found that for molecules with short poly-proline spacers, fluorescence emission of the acceptor could be detected when excited at the shorter excitation wavelength of the donor (Fig. 1.4.1). For longer spacers, this transfer efficiency was reduced.

By analysing the relationship between the inter-dye distances and the measured FRET efficiency, Stryer and Haugland experimentally confirmed the 6th power dependence theorised by Förster, and suggested that FRET could be used as a *spectroscopic ruler* for distances between 10-60 Å, given dye pairs with R_0 of 15-45 Å. Half a century since, FRET has been used to great effect as a spectroscopic ruler, and Stryer and Haugland's prediction has been wrong only in the sense that fluorophore pairs with slightly higher R_0 of 30-70 Ångstroms have been more favoured. Today, similar experiments are conducted using DNA as a spectroscopic ruler[14–16], however the polyproline ruler originally used to test FRET in the ensemble has been revisited using single-molecule FRET[17].

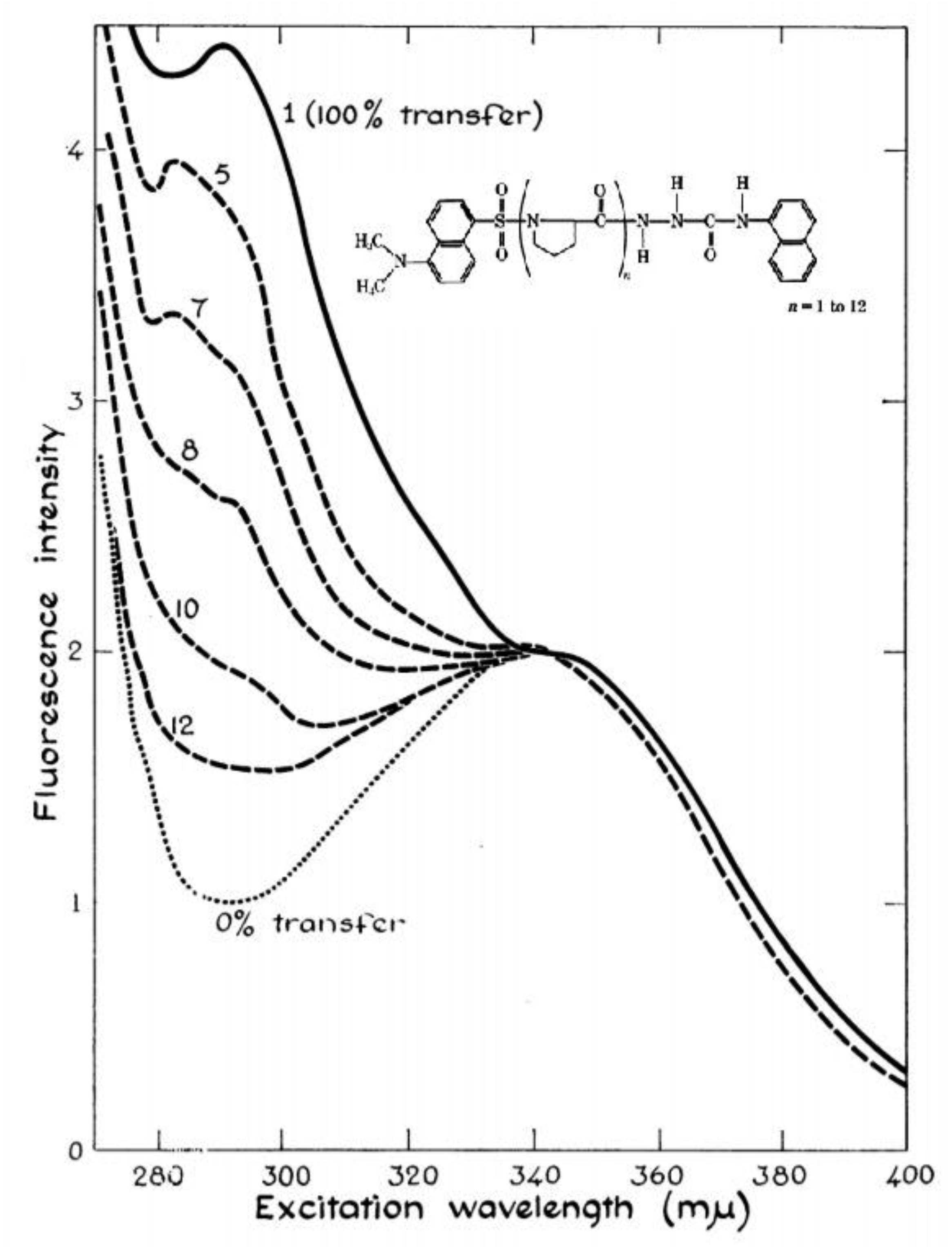


Fig. 1.4.1 Excitation spectrum of dansyl-L-prolyl-hydrazide (acceptor only) and dansyl-(L-prolyl)$_n$-α-naphythyl (FRET molecules) in ethanol. Adapted from the original publication by Stryer and Haugland[13]. The structure of the ruler is shown in the top right.

Today FRET is used widely to probe such distances on the order of nanometres. Whilst measuring fluorescence of whole solutions in the manner of Stryers' original experiment (hereby referred to as *ensemble-FRET*) can be used to reliably measure relative distance changes across R_0 in a qualitative binary fashion, single-molecule FRET techniques (hereby referred to as smFRET) can more accurately measure absolute distances at or close to the R_0. In either scenario, it is the R_0 itself which dictates the range over which a dye pair is sensitive. A variety of ranges over which FRET is sensitive have been quoted in the literature, 1-10 nm[18], 1-7.5 nm[19], 3-10 nm[20], and 2-10 nm[21]. The true range of sensitivity is largely dependent on the FRET pair used in question, and the sensitivity reduces further from the Förster radius for any given dye pair (Fig. 1.4.2). R_0's for dyes used in smFRET experiments can vary from the low 2.8 nm of the Cy2-Cy7 pair[22] to the high 7.3 nm of the Cy5-Cy5.5 pair[23].

A generalised relationship between the R_0 and FRET efficiency can be shown by expressing intermolecular distance as a fraction of R_0 (Fig. 1.4.2 a). It is worth noting that the sigmoidal shape of the graph imparts two key characteristics; first, that the changes in FRET efficiency resulting in distance changes fall off rapidly above 75% and below 25% FRET, resulting in an approximately linear region in which it is sensitive, and secondly, that the linear region is steep, meaning that small changes in distance result in large changes in FRET efficiency (Fig1.4.2 b). As a consequence of the steepness of the linear region, so long as fluorophores can be placed at a distance sufficiently close to the Förster radius, the FRET signal will be sensitive to changes in distance on the order of single Ångstroms, despite the distance itself being on the order of several nanometers. This sensitivity range aligns excellently with the dimensions which structural molecular biology concerns itself with; as proteins and DNA are several nanometers across, but their functions are dependent on Ångstrom scale movements and interactions.

Thus, dye placement in FRET experimental design often falls into one of two strategies. In one strategy more common to ensemble-FRET experiments, fluorophores are placed such that the two states of interest produce transfer efficiencies either side of the linear region, resulting in low precision distance information but a high signal change which can be correlated against time or concentration of an effector ligand. The second experimental design strategy more common (but not exclusive) to smFRET experiments, involves attaching the fluorophores such that most, if not all observed states occur in the linear region, for the purposes of either detecting subtle conformational changes within the molecule of interest or determining absolute distance information for resolving structures.

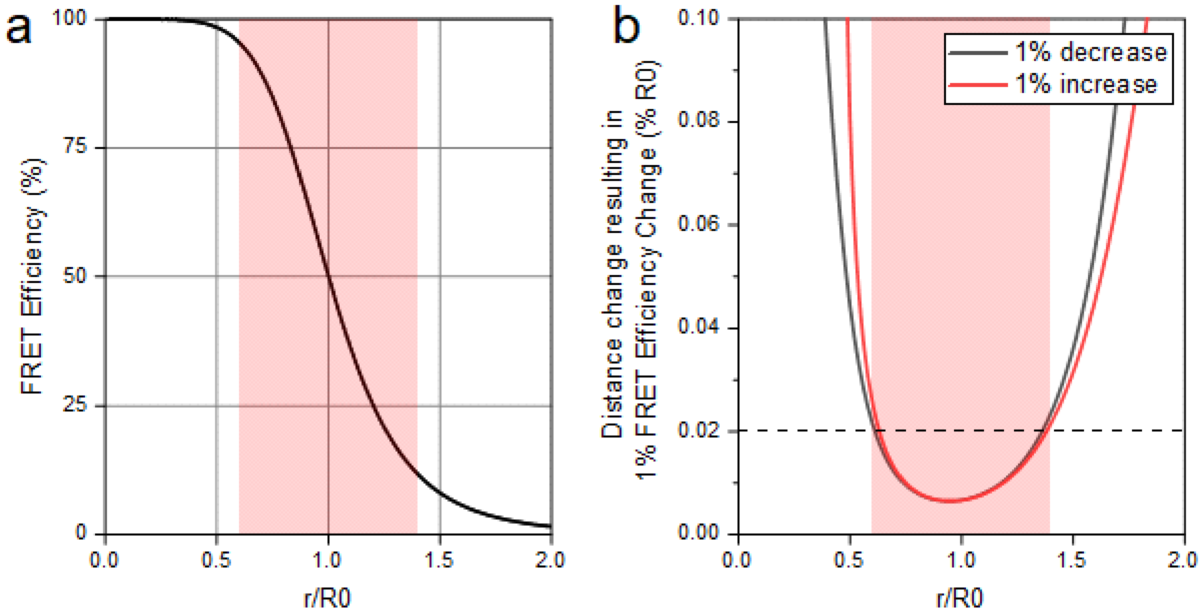


Fig. 1.4.2 a. FRET Efficiency against distance between dyes expressed in multiples of R_0. Shaded in red is the "linear region" **b.** Distance changes resulting in 1% FRET efficiency change against distance (either up or down). Between approximately 0.6 and 1.4 R_0 (shaded red), a single percentage point of FRET efficiency corresponds to a distance change of less than 2% of the R_0. The experimental implication of this is that if one can accurately detect a 1% change in FRET efficiency, then the experiment will be sensitive to changes in structure on the order of single Ångstroms (2% of a typical R_0 of 5 nm is 1 Å). However, outside of this region, even very large changes in R_0 will produce indistinguishably small changes in FRET efficiency.

The first strategy has been employed to great effect throughout structural and mechanistic biology to give dynamic proximity information[24–26]. FRET is a staple tool for investigating protein-protein interactions, and to this end proteins are often genetically conjugated to fluorescent proteins such as green fluorescent protein (GFP) or GFP derivatives for *in vivo* study[27]. FRET has even seen uses in biotechnological assays, as a reporter for glutamate transition in iGluSnFR[28] and as a HIV-1 protease activity reporter[29].

The second strategy of determining quantitative information on inter-dye distance has been equally widespread in scope. Extracting intramolecular distance information from FRET has been employed to investigate protein folding[30], DNA polymerase[31], RNA polymerase[32], Hsp90[33], G-quadruplexes[34], nucleosomes[35], and neuroreceptors[36] to name just a few.

1.5. Single-molecule FRET

Perhaps the most notable advancement in FRET measurement techniques to date is single-molecule-FRET (smFRET); a family of techniques which measure the FRET efficiency of either a single doubly labelled molecule, or of a single pair of donor and acceptor labelled molecules. As smFRET techniques sometimes involve measuring two (or more) distinct molecules which each have a label, they are sometimes instead referred to as single-pair FRET[37] or spFRET.

Shortly after the first detection of single fluorescent molecules[38] and usage in a biochemical experiment measuring single ATP turnovers of mysosin[39], the technique was combined with FRET[40]. Using DNA as a spacer labelled with tetramethylrhodamine (TMR) at one end and Texas Red at the other, the FRET efficiency between a single pair of the two dyes was measured by near-field scanning optical microscopy (NSOM). By measuring the FRET efficiency between a single pair of fluorophores, FRET can be used as a single-molecule technique, rather than an ensemble one.

Single-molecule techniques have many notable advantages over ensemble ones, crucially being the measurement of dynamic events, whose signal is lost in the ensemble due to averaging. In the case of FRET, by measuring molecules independently, heterogeneous subpopulations can be investigated, and multiple inter-dye distances can be determined, rather than simply the average of all distances. As signals corresponding to subpopulations can be extracted one molecule at a time, kinetics of both forward and reverse processes can be measured even when the whole solution is at equilibrium. This removes the need for synchronisation to measure rates, as is necessary in ensemble biophysical techniques (ie. stopped-flow fluorescence). Furthermore, the ability to investigate each molecule on a case by case basis allows mislabelled or partially bleached molecules to be removed from final statistics, reporting on only signals from molecules displaying the intended combination of dyes[41].

Since 1996, improvements in smFRET have come from new single molecule isolation techniques, with diffraction limited confocal excitation[42] allowing observation of freely diffusing molecules, and total internal reflection microscopy (TIRFM) for surface attachment experiments favoured over the original NSOM[19]. Additional excitation schemes such as alternating laser excitation[18] (ALEX), and pulsed interleaved excitation (PIE)[43] along with combining FRET with a multitude of other metrics[44] have given access to additional information beyond simply FRET efficiency. Lastly, access to improved detectors and brighter, more stable commercially available dyes have since provided incremental but notable improvements in signal to noise ratio.

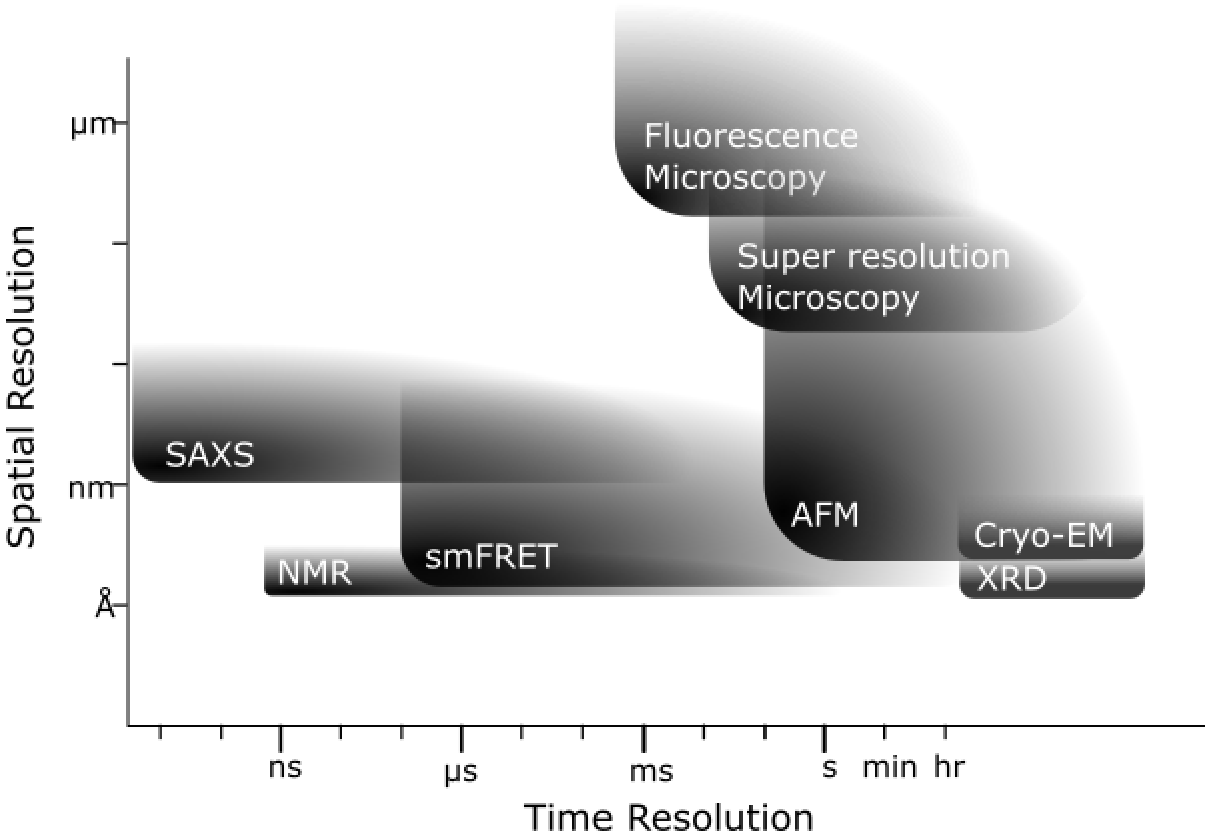


Fig. 1.5.1: Spatial and temporal resolutions of various biophysical techniques. Positions shown here are approximate and non-standard variations of these techniques may have different values.

When compared to other structural biology techniques (Fig. 1.5.1), smFRET occupies a unique niche of spatial range and resolution with fast temporal resolution, even aside from sensitivity to sample heterogeneity. FRET is sensitive to Ångstrom scale changes in structures over several nanometres and has been used to measure interconversion rates faster than microseconds[45]. X-ray diffraction and Cryo-electron microscopy give excellent spatial resolution in structures but lack capacity to measure the time scales of dynamics due to the need to crystallise or freeze samples. NMR is capable of measuring fast dynamics over short distances in ensemble, but lacks the long distance range of FRET. Small angle x-ray scattering can interrogate subpopulations with negligible time averaging due to the femtosecond time scale of scattering, but gives ambiguous spatial information.

1.6. smFRET Techniques

The key restrictive experimental parameter in the design of single-molecule fluorescence techniques is that of signal to noise ratio. The quality of information that can be derived from a single emitter (or in the case of FRET, two emitters) is highly dependent both on the raw magnitude of photons detected (due to high shot-noise at low photon counts) and the ratio of signal photons compared to photons originating from all other obfuscating sources. Thus, in order to reduce incoming light from background sources against the fluorophore signal, the defining strategy common to single-molecule techniques is to reduce the detection volume, such that the limited background from a decreased volume of solvent in the detection zone can be sufficiently overwhelmed by the otherwise low fluorescent intensity of a single emitter.

The two prevalent strategies to achieve this reduction in detection volume in smFRET experiments are total internal reflection fluorescence microscopy (TIRFM), in which the detection volume is reduced to a ~100 nm deep plane above the surface of the sample coverslip, and confocal microscopy, in which the detection volume is reduced to a spheroidal diffraction limited spot of approximately 1 femtoliter (or 1 μm^3).

In TIRFM-FRET, the excitation laser is totally internally reflected by the glass-solvent interface, and fluorescent molecules are excited by the resulting evanescent wave (Fig. 1.6.1), the resultant emitted light is then focused onto highly sensitive CCD cameras[46]. As the penetration depth of the evanescent wave is limited by the wavelength of light (Eq. 1.6.1), only molecules very close to the glass-solvent interface will be excited, and therefore detected. This technique deliberately limits detection in depth (z) in order to decrease background, but has effectively no limitation in detection in x and y, allowing many molecules to be individually detected simultaneously. The penetration depth, typically on the order of 100 nm is found by the following equation:

$$d = \lambda/4\pi(\eta_1^2 sin^2 - n_1^2)^{-1/2}$$

Eq. 1.6.1

Where λ is the wavelength of the light, n_1 and n_2 are the refractive indices of the media.

Due to the reduced depth, freely diffusing molecules will be detected only very transiently, and the photon detection efficiencies in TIRFM are insufficient to precisely assign FRET efficiencies to molecules observed over such short time scales. To circumvent this, molecules of interest are attached to glass slides either directly via surface chemistry[47], or indirectly via encapsulation in vesicles which are themselves attached via surface chemistry[48].

Surface attachment can introduce an additional level of steric inhibition to the molecule of interest, however vesicle encapsulation is an effective way to circumvent this problem[49].

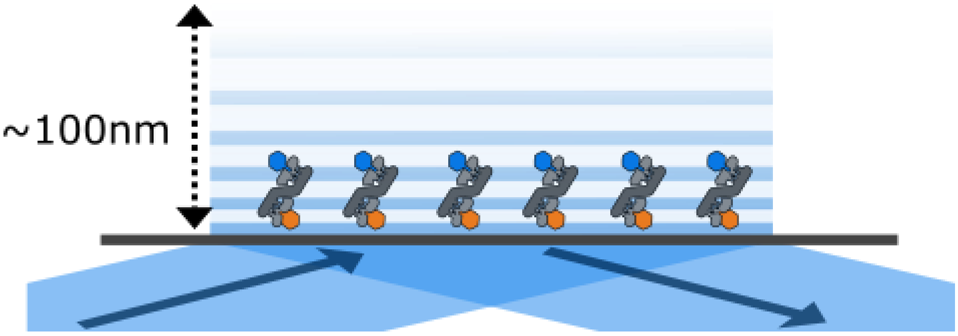


Fig. 1.6.1 Schematic of TIRFM, an evanescent field created by total internal reflection of a laser beam off of a glass coverslip, which decays rapidly at approximately 100 nm from the surface. Molecules of interest are tethered to the coverslip surface, usually using streptavidin and biotin, which form a strong but non-covalent bond.

In confocal smFRET, the excitation laser is focused through a high numerical aperture (NA) lens into a spot at some depth into the solution (Fig. 1.6.1), and emitted light is focused back through the same lens, and directed through a pinhole (removing out of focus light), onto single-photon avalanche-diode (SPAD) which can detect individual photons with quantum efficiencies >50% and timestamp them via a transistor-transistor logic (TTL) pulse on the order of nanoseconds[50]. The combination of focusing excitation light into a tight diffraction limited spot, and optically sectioning emitted light in a limited focal plane (via a pinhole) results in a combined detection zone approximately 1 cubic micrometer in volume. This technique is typically used to interrogate freely diffusing molecules as they pass through this volume, at sufficiently low concentrations (10's of pM) such that there is one or less molecule in the volume at a time, remaining for times typically on the order of milliseconds. The lack of a need to produce an image to single out molecules in space, as in TIRFM, means the entire detector can be used for each molecule, allowing for the collection of many more photons per second, and so accurate FRET efficiencies can still be assigned with typically 10-100 photons per burst despite short integration times.

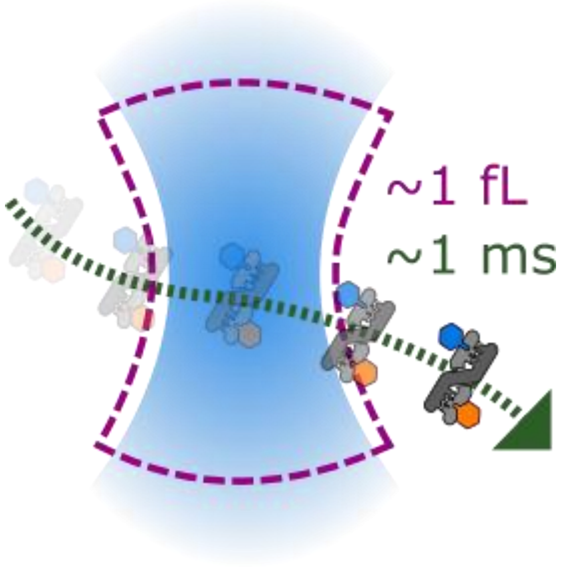


Fig. 1.6.2 Schematic of confocal type smFRET. A detection volume of approximately 1 femtoliter is formed by focussing a laser through a high numerical-aperture objective lens, and using a lens-pinhole-lens system in the path of the collected emission to reject light originating from out of the focal plane. Molecules are detected as they freely diffuse through the detection volume, with a typical residence time on the order of 1 millisecond, depending on the diffusion coefficient of the molecule.

These two single-molecule schemes each have strengths and weaknesses which make them appropriate for different kinds of experiment (Table 1.6.1). Whilst both techniques are appropriate for assigning static accurate FRET distances, the differing time scales of detection separate them into unique niches for investigating dynamic changes and subpopulations. TIRF-smFRET can allow for observation of individual molecules on the order of 10's of seconds to minutes, and can measure many individual molecules in parallel, so is useful for investigating structural dynamics over longer time scales. However, the need for a longer integration time significantly reduces precision of FRET efficiencies over short time scales and can complicate the separation of subpopulations with fast interconversion rates. Furthermore, the need for imaging within ~100 nm of the surface demands the use of attachment chemistry which can increase the setup complexity (typically requiring the use of flow-cells for sequential addition of surface attachment components), and the proximity to the surface may alter the behaviour of the biomolecule under study[51] (however such "surface effects" can be avoided using vesicle immobilization[49]).

One advantage of confocal smFRET, is the high photon rate (on the order of 100 kHZ), which allows for more precise assignment of FRET efficiencies with just millisecond snapshots of data for each molecule, and so faster subpopulation dynamics can be observed for freely diffusing molecules. Whilst the short diffusion time limits direct observation of changes over longer than 10 ms, various statistical techniques are available to infer interconversion rates at the subpopulation level on time scales up to whole seconds[52].

Table 1.6.1: Comparison of TIRF and Confocal

Experimental parameter	**TIRF**	**Confocal**
Minimum averaging time	~20ms	~0.5ms
Maximum observation time	>Up to 30 minutes (limited by photobleaching)	~10ms
Number of simultaneously detected molecules	>100	<1
Typical number of modifications to molecule of interest	3 (fluorophores + biotin)	2 (fluorophores)
Freely diffusing?	Only with vesicle encapsulation	Yes

It is worth noting that whilst the majority of smFRET experiments fall succinctly into one of these two categories, there are alternative schemes which combine advantages of both. For example, surface confocal experiments use a confocal detection scheme focused on a surface immobilised molecule, giving the high time resolution of confocal combined with the long measurement times typically afforded by imaging techniques[53]. Similarly, an anti-brownian electrokinetic trap (ABEL) can be used to immobilise a charged molecule in solution via voltages applied to the sample[54], giving long observation times with no chemical tether.

Single-molecule fluorescence techniques are not limited to simply measuring donor and acceptor emission upon donor excitation, and smFRET is frequently combined (or "multiplexed") with other detection parameters in order to gain additional information about the molecule of interest, or to enhance the ability to accurately determine FRET efficiency.

The most common of these additional parameters to measure is acceptor emission under direct excitation (by a longer wavelength laser). By measuring the fluorescence emission of the molecule under direct acceptor excitation in addition to donor excitation, the experimenter can easily distinguish between low FRET and donor only molecules, and to determine if the acceptor has entered a dark state during detection. Furthermore, spectral cross talk such as donor emission into the acceptor channel (leakage), and acceptor excitation by the donor laser (direct excitation), can be more easily quantified, allowing for more accurate determination of FRET efficiency.

$$E^{*} = \frac{D_{ex}A_{em}}{D_{ex}D + D_{ex}A_{em}}$$

Eq. 1.6.2

$$S^{*} = \frac{D_{ex}D_{em} + D_{ex}A_{em}}{D_{ex}D + D_{ex}A_{em} + A_{ex}A_{em}}$$

Eq. 1.6.3

This can be done by alternating laser excitation (ALEX) between donor and acceptor wavelengths[41] at a frequency such that emission can be detected under both donor and acceptor excitation multiple times for each molecule. The result is such that the experimenter gains access to not only FRET efficiency of any molecule (Eq. 1.6.2), but stoichiometry (Eq. 1.6.3), which are often plotted together on an "ES plot" (Fig. 1.6.3). In TIRF experiments this typically involves alternating on the order of 1-10 times per second (ms-ALEX), however due to the lower residence times (~1 ms) in confocal experiments alternation rates on the order of 10 KHz are required. Alternatively, if pulsed lasers are used, the pulses from two colours can be delayed from one another on the order of 10's of nanoseconds, resulting in a technique known as pulsed interleaved excitation (PIE)[43] or nanosecond-ALEX (ns-ALEX)[41].

However, precisely modulating two lasers can be complex and expensive, so an alternative method with similar results, involves periodic acceptor excitation (PAX), where the donor laser is not modulated[55]. This reduces the complexity and cost of a confocal setup, however it can increase the complexity of analysis, as the donor remains excited during acceptor excitation, so donor emission leakage, and FRET emission must be subtracted from the data gathered under acceptor excitation.

Finding the fluorescence stoichiometry of the two fluorophores like this can allow for identification of donor only and acceptor only labelled molecules[18]. Not only is this useful in order to ensure the final analysed population is strictly doubly labelled, but by looking at the measured FRET efficiency and stoichiometry of singly labelled species, (either arising from a doubly labelled sample or intentionally singly labelled), one can quantify spectral crosstalk by determining the leakage of donor emission into the acceptor channel, and the amount of direct excitation of the acceptor by the donor laser. Furthermore; the relationship between FRET efficiency and stoichiometry of molecules can be used to quantify the inequality between detected emission from each dye (the product of detection efficiency and quantum yield). The quantification and subsequent correction of these three factors can aid in the accurate determination of "true" FRET efficiency[14,21,22], as will be discussed in Chapter 2.

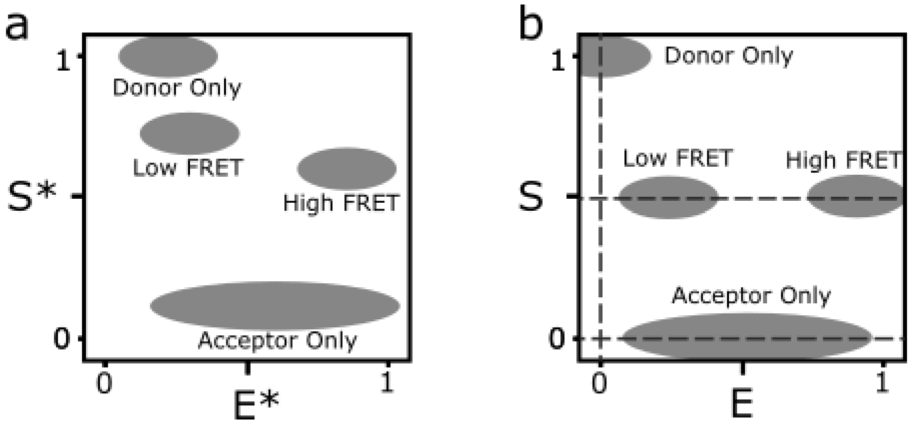


Fig1.6.3 Schematics of typical 2D Efficiency-Stoichiometry (ES) histograms obtained from an smFRET experiment, where the FRET efficiency and stoichiometry of each burst is individually plotted. **a.** Uncorrected stoichiometry (S*) against uncorrected FRET efficiency (E*), also known as the proximity ratio. Donor only molecules appear with a stoichiometry of 1, acceptor only molecules appear with a low stoichiometry, and doubly labelled molecules appear with an intermediate stoichiometry. Molecules with a shorter inter-dye distance will appear with a higher E*. **b.** A corrected ES plot. After removing spectral crosstalk, donor only molecules exhibit a FRET efficiency of zero and acceptor only molecules have a stoichiometry of 0. After correcting for detection and excitation efficiencies, all doubly labelled molecules appear with a stoichiometry of 0.5.

In addition to filtering molecules with incorrect labelling stoichiometry, and access to accurate FRET correction parameters, direct acceptor excitation schemes like ALEX and PIE can also inform on changes in local environment of either dye which have altered their relative brightness. A technique named protein induced fluorescence enhancement (PIFE)[56–58] involves the use cyanine dyes, which increase in brightness when a protein binds nearby, as the protein sterically restricts the *cis-trans* photo-isomerisation of the dye that would otherwise compete with fluorescence[59]. This increase in brightness can be exploited at the single molecule level to monitor individual binding events of proteins. Whilst PIFE as a single-molecule technique is powerful on its own, it can be combined with smFRET to measure both intramolecular distances *and* protein binding in parallel[60]. With PIFE-FRET, molecules where a protein is bound in proximity to the cyanine donor will appear with a higher apparent stoichiometry. The same principle of cyanine brightness enhancement can be used with nucleic acids, and is called NAIFE (nucleic acid-induced fluorescence enhancement). A related technique, called stacking-induced fluorescence increase (SIFI)[61], exploits the increase in fluorescence of cyanine dyes when stacked against nucleotide bases, rather than being restricted by a bound protein. As such, SIFI can be used to measure base stacking at the single-molecule level and investigate through-backbone DNA allostery and DNA hairpin dynamics[61,62].

The idea of multiplexing the measurement of different fluorescent parameters in confocal smFRET was first introduced as multi-parameter fluorescence detection (MFD)[63]. Whilst the later use of an additional longer wavelength laser for ALEX or PIE can be considered a form of MFD, the original MFD involved the simultaneous measurement of fluorescence lifetime by time-correlated single-photon counting (TCSPC), and anisotropy using polarisation filters, in addition to intensity and emission colour of each dye. The measurement of these parameters gives additional information on the chemical environment and orientation of the dyes, allowing for confirmation that parameters such as quantum yield and kappa squared required for understanding FRET remain constant[44].

In PIE and MFD, rather than continuous wave excitation, a picosecond pulse laser is repeated (Fig. 1.6.4), and TCSPC is used to measure the "microtime" of the emitted photons with respect to the laser pulse, in addition to the "macrotime" with respect to the beginning of the experiment as in continuous excitation schemes. Photon arrival microtimes are summed across pulses to produce a fluorescence lifetime decay curve from each dye.

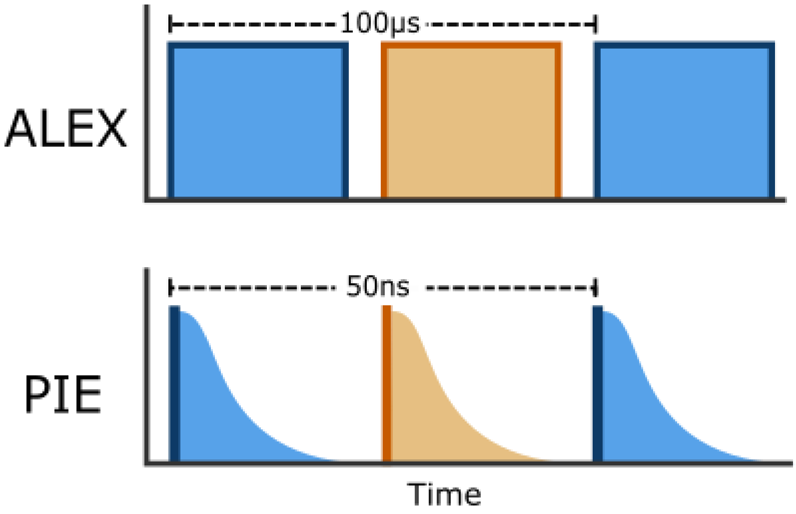


Fig1.6.4 ALEX and PIE excitation schemes. Laser pulses are in dark outline, and recorded photons are in lighter shade. Note that in PIE, typically only one emitted photon is recorded per pulse, and are summed across pulses to produce the decay curves shown in the shaded area.

As discussed earlier, one drawback of freely diffusing confocal detection is the low throughput imposed by the need to effectively measure less than one molecule at a time. One way around this problem is to use a liquid crystal on silicon (LCOS) light modulator to focus the excitation beam onto multiple spots in the focal plane, and subsequently focus the emission pathway onto an array of single photon avalanche detectors (SPADs). Using this technique, the confocal scheme has been parallelised to 8-spots[64] and later to 48-spots[65], rapidly increasing the speed at which data can be obtained.

1.7 DNA and Labelling

Deoxyribonucleic acid, (DNA), has historically been overwhelmingly the molecule of choice for smFRET method development and validation[14,40,41,54,66] due to its relative inflexibility and high availability for modification and design to place fluorophores, whose relative positions can then be accurately predicted owing to the wealth of existing research on DNA structure.

DNA is composed of nucleotide monomers, comprised of a phosphate, 2-deoxyribose, and a variable nucleobase. The phosphate of one nucleotide, linked to the 5’ carbon of its deoxyribose sugar, is reacted with the 3’ hydroxyl of the next nucleotide, to form a continuous alternating ribose-phosphate backbone. Two complementary strands of DNA may anneal into an antiparallel right handed double helix via hydrogen bonding between nucleobases[67]. In DNA, bases hydrogen bond to one another via strict Watson-Crick base pairing (Fig. 1.7.1), where thymine (T) forms two hydrogen bonds with adenine (A), and cytosine (C) forms three hydrogen bonds with guanine (G). It is also worth noting that the larger purines A and G bond to the smaller pyrimidines T and C, rather than with each other, such that each base pair is approximately the same size.

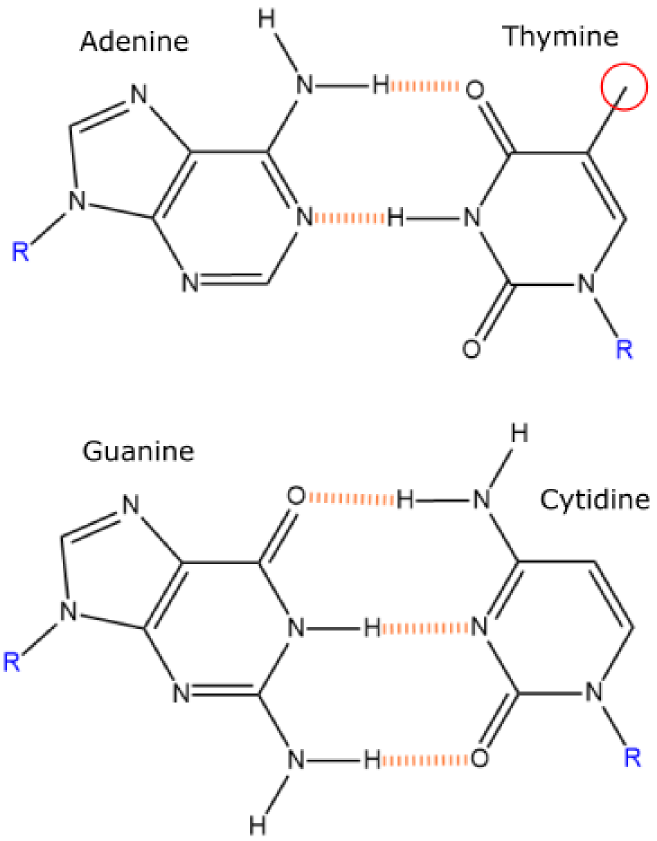


Fig1.7.1 Base pairing in DNA. Dashed orange lines are hydrogen bonds, the “R” in blue indicate the deoxyribose sugar. Circled in red on Thymine is the carbon atom in the major groove which is extended to a flexible linker for labelling with fluorophores (See Fig. 1.6.2)

When labelling DNA with fluorophores, amine reactive probes charged with N-hydroxysuccinimide (NHS) ester are used, due to the absence of primary amines in natural DNA. Amine modified bases are incorporated into the strand via normal phosphoramidite synthesis, which can then be reacted with dye NHS esters. Primary amines can be added to 3’ or 5’ ends of DNA, however experiments with fluorophores conjugated like this have shown that the dyes can stack like bases against the terminal base pairs[68]. In order to preserve the assumption of $\kappa^2 = 2/3$ which is crucial to accurate FRET determination, fluorophores must be able to rotate freely, therefore a more common approach is to conjugate amine reactive probes to an internal thymine modified with a flexible linker protruding out of the major groove (Fig. 1.7.2)

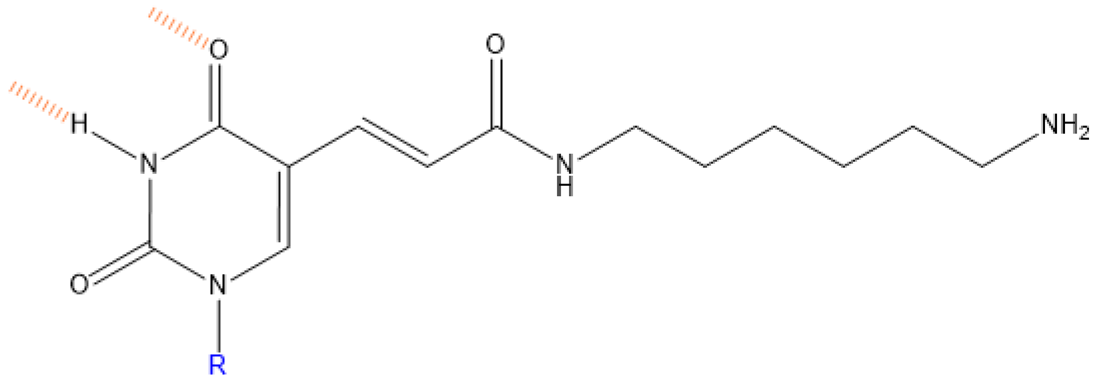


Fig. 1.7.2 Structure of C6-dT. Orange dashed lines indicate hydrogen bonds to Adenine, the R in blue indicates the deoxyribose sugar.

Any FRET experiment requires, at a minimum, the attachment of two fluorophores to a molecule, or molecules of interest. Many FRET experiments in molecular biology for determining simple on/off proximity information typically involve the genetic fusion of fluorescent proteins, (or in the case of BRET, bioluminescent proteins[69]), to the protein of interest. However, GFP and its many derivatives are regarded as inappropriate for single molecule FRET due to their large size (Fig. 1.6.3), instability, and low brightness compared to small organic fluorophores. As such, the majority of smFRET experiments involve in some way the chemical conjugation of small organic dyes to proteins or DNA.

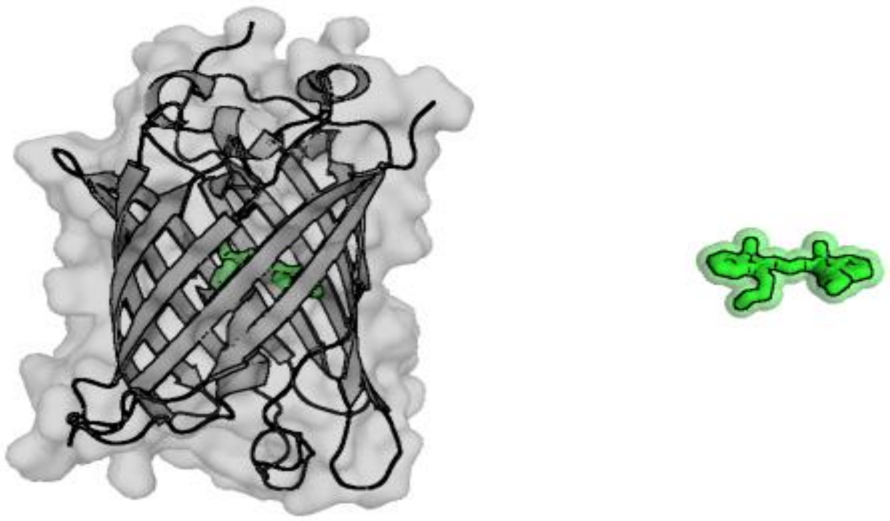


Fig1.7.3 Structures of GFP and Cyanine-3 (Cy3). Note that despite the chromophore inside the GFP β-barrel being approximately the same size as Cy3, the whole protein itself is many times larger.

Fluorophores for smFRET typically come from the Cyanine (eg. Cy3, Cy3B, Cy5, Alexa Fluor 647), or Rhodamine (eg. TAMRA, Atto 550, Atto 647N, Alexa Fluor 546) families of dyes, due to their high extinction coefficients, high quantum yields, photostability, and commercial availability for simple conjugation. Whilst DNA is typically labelled using an NHS ester reaction to a synthetically incorporated primary amine, proteins are usually labelled via maleimide reaction to cysteine residues, although NHS esters can also be used to label lysines or amino termini.

Maleimide labelling usually involves mutation of the protein of interest to include cysteines at desired labelling sites, and to remove off-target cysteines from the surface of the protein. Cysteine is not a common amino acid, so the total number of modifications is typically minimal, however it is imperative that activity assays are done on the final labelled protein to rule out abrogation of biologically relevant characteristics either by mutation or by conjugation to the label. Cysteine is often mutated to or from amino acids with a similar steric profile, such as serine and alanine, to prevent such problems.

These conjugation reactions to proteins are typically done *in vitro* on purified products, due to the high availability of off target thiols and primary amines in living cells, however *in vivo* labelling techniques have also recently emerged. Genetic fusion to reactive proteins such as SNAP tags[70], or using synthetic biological techniques to incorporate unnatural amino acids such as p-azidophenylalanine can result in recombinant proteins with high specificity to click conjugated fluorophores[71]. The labelling reactions can then occur either upon uptake of label by the cell, or directly with cell lysate.

More than two dyes can be attached to a single molecule, and past experiments have used three dyes to give three FRET coordinates in one experiment[23]. The complexity of information from labelling with three fluorophores simultaneously is far greater than that of doing three separate experiments with different two fluorophore labelling schemes, as changes in one FRET efficiency can be correlated against changes in another. Analysis of dynamics with three colour systems can allow for investigations into coordinated molecular motions[72] establishing the relationship between conformations in one part of molecule to conformations in another, which cannot be easily established using distinct experiments with just one FRET pair each.

Four dye experiments are also possible, and have been realised in a number of cases, giving 6 possible inter-dye distances[73]. However with the addition of each dye experiments become increasingly complex due to crowding in spectral space and increasing complexity of corrections required to recover each FRET efficiency should accurate distances need to be recovered.

1.8 smFRET analysis

Due to the wide variation in techniques for smFRET, there are a number of different software packages available for analysis; for TIRFM, examples include Twotone[74] from the Kapanidis lab, SMART[75] from the Herschlag lab, and iSMS[76] from the Birkedal lab. For confocal type smFRET, there is the FRETBursts[20] python package from the Weiss lab and PAM[77], a MATLAB based package from the Lamb lab, in addition to many other custom written scripts used by individual labs and not made widely available.

Analysis of TIRF-smFRET data requires mapping the two colour channels together, identification of single-molecules in the image, and correction for local differences in detection efficiency and noise. Conversely confocal smFRET data has no spatial resolution, but to find molecules diffusing through the confocal volume, “bursts” of photons are identified along the time axis.

Burst analysis in freely diffusing methods must first determine which regions in the photon stream correspond to a single-molecule, by identifying periods of increased photon flux significantly greater than that of the background level. The most common burst search methods are some variation on the sliding window algorithm[63,66,78]. In the sliding window burst search, photons are marked as belonging to a burst if at least L successive photons have at least M photons arriving before or after it within a time window of length T. The values M and T effectively form a photon rate threshold, above which photons are considered to be higher than expected from just background. An alternative form of the sliding window instead looks for L successive photons where the surrounding M photons have a rate F times higher than the background rate[20], which allows for more reliable comparison between experiments with differing background fluorescence.

The sliding window algorithm when applied to all channels together is often referred to as an all photon burst search (APBS), however the algorithm can be modified for ALEX experiments to search two channels simultaneously, and find the intersection between photon bursts found in both channels[66] (Fig. 1.8.1). This *dual channel* burst search (DCBS), when applied to the donor excitation channel and acceptor excitation channels will efficiently return only molecules which respond to both lasers (ie. have both a donor and acceptor), as well as cutting off the end of a burst should one of the dyes bleach.

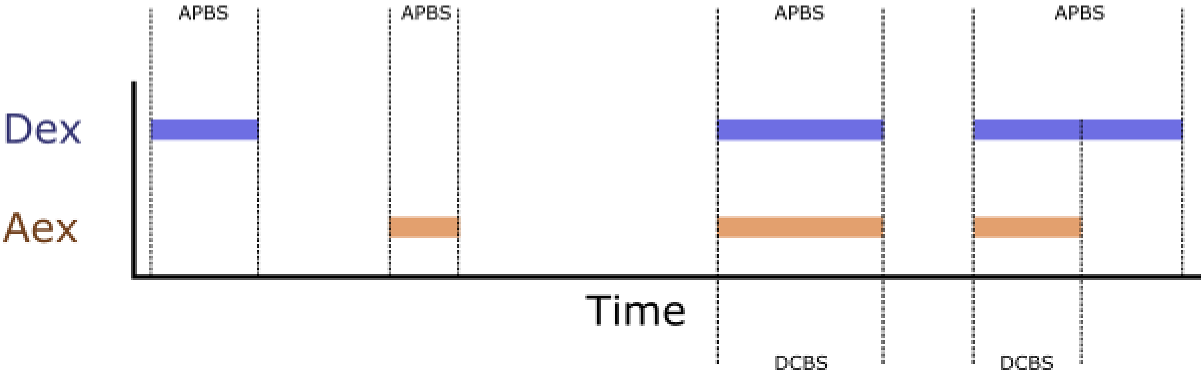


Fig. 1.8.1 Schematic of how the all photons burst search and dual channel burst search work. The APBS returns a burst wherever there is a signal regardless of the channel, whereas the DCBS only returns the intersection of signals between the two channels. The first two bursts here appear to be donor only and acceptor only, so are not returned by the DCBS. The third appears to be doubly labelled. The second burst appears to be doubly labelled, but the acceptor bleaches mid-burst. APBS returns the entire burst, however the DCBS returns only the part where there are two dyes.

After identification of molecules in either image based smFRET or freely diffusing type, accurate FRET analysis takes largely the same form, subtracting background, correcting for spectral cross talk, and detection efficiency of either dye to recover accurate FRET efficiency, as will be discussed in greater detail later.

However simply determining true FRET efficiency and having knowledge of the R_0 is insufficient to find distance for anything but a completely static dye pair. As discussed above, dyes are typically attached to the molecule of interest by flexible linkers, and so random motions of either fluorophore around their attachment point create a distribution of possible FRET efficiencies which may be sampled at any given moment in time. This distribution of FRET efficiencies, averaged across the measurement integration time, produces an average FRET efficiency not necessarily equal to a static treatment of the distance between the attachment points. In order to better inform labelling strategy, and to aid in conversion between FRET efficiency and intramolecular distance, the motion of the fluorophore around the molecule must be accounted for.

One coarse grained approach to modelling this effect involves taking a known model of the molecule of interest, and finding an accessible volume, or AV (Fig. 1.8.2) around the attachment point of each fluorophore. This can be done using the FRET Positioning and Screening software (FPS)[79]. AV's are calculated using dimensions of the dye and flexible linker, and determining all locations in 3-dimensional space which the dye can access without clashing with the molecule it is tethered to, or its own linker. Distances between points in the two AV's can then be sampled, averaged, and converted to FRET efficiencies to find the value of $\langle E \rangle$ for the two attachment points. These AV's can then be used for preliminary screens of a molecule of interest to determine the most informative labelling positions, and for interpreting FRET efficiencies obtained from experiments

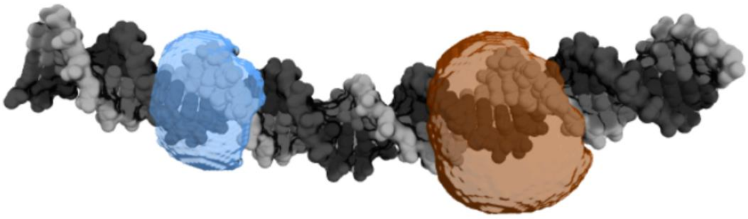


Fig 1.8.2 Two AV clouds modelled onto a bent DNA structure using FPS.

smFRET is not just limited to determining single distances however, and considerable thought has been given to the idea of extracting information on kinetics and dynamic heterogeneity from smFRET data. In surface immobilized experiments, the experimenter has access to relatively long molecular histories for each molecule, which are often analysed using hidden Markov models (HMMs) and dwell time analysis[80]. For freely diffusing confocal data however, the observation times are low and there is often insufficient data for dwell time analysis on any given molecule, and so alternative approaches are required.

When analysing the distribution of FRET efficiencies measured in an experiment, it is common to fit this distribution with a Gaussian curve, or sums of Gaussian curves. Whilst this can give a satisfactory fit and is sometimes appropriate (less so at extremes of the E*), the standard deviation of the Gaussian curve (or full width at half maximum) does not directly correspond to real properties of the observed population. The standard deviation of the Gaussian will always be a composite of various factors contributing to the width of the distribution, many of which correspond to experimental parameters such as *shot noise* and background and may not be directly correlated to actual variation in the sample.

Shot noise arises from the fact that a finite number of photons are measured of each molecule, and so the ratio of donor and acceptor photons in each burst will only be an approximation of the true FRET efficiency, stochastically distributed in a binomial fashion (Fig. 1.8.3). Background adds additional width, as over the short time scales of each burst the exact contribution to overall signal will similarly vary stochastically and can only be estimated. The mean background can be subtracted, but the resulting value will still vary as much as the underlying background. In fact, standard deviation in this instance cannot even give a reliable measure of confidence in population FRET efficiency, as some proportion of the underlying width will always correspond to a real distribution of FRET efficiencies[81], excepting the theoretical but experimentally irrelevant case of rigid and unmoving fluorophores.

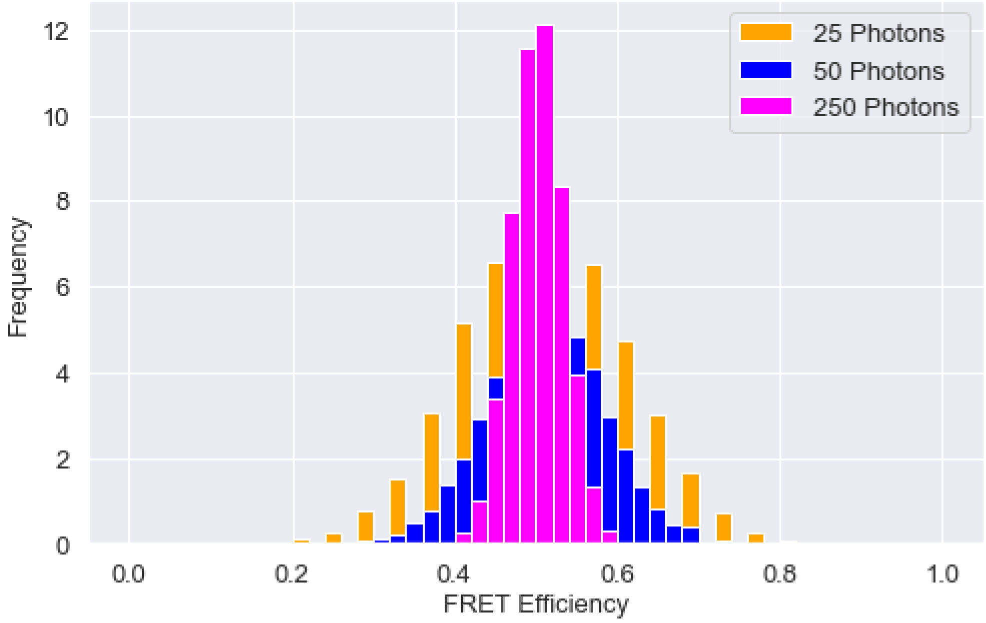


Fig. 1.8.3 Simulated data generated using the PDA algorithm[66] demonstrating shot noise broadening as a function of burst size (number of photons). In magenta is a histogram of bursts generated with an underlying FRET efficiency of 0.5 and 250 photons. With only 50 photons (blue), the histogram is much wider, and with 25 photons per burst (orange) the histogram is not only wider but has gaps where there are no possible fractions with a denominator of 25.

For confocal experiments, many of the approaches to understand the nature of the experimentally observed distribution of proximity ratios follows from the works of Antonik et al.[82] and the shortly followed work of Nir et al.[66]. Antonik gives a mathematical description of the distribution of proximity ratio values starting with a known distribution of burst width. Briefly; the derivation begins by assuming that the distribution of proximity ratios for a fixed FRET efficiency and fixed burst size will follow a binomial distribution, as the likelihood of any photon being emitted by the donor or acceptor will be fixed each time. Thus, the probability of observing any given proximity ratio can be described by summing this binomial distribution across all values in the burst size distribution of the experiment.

By mathematically describing the distribution in terms of a combination of distributions which incorporate the expected variation, one can thus derive the quantity of width which arises due to real fluctuations in the observed molecules, in a manner referred to as *probability distribution analysis*, or PDA (alternatively, *photon* distribution analysis).

Nir et al., expanded on this, and demonstrated that whilst the distribution described by Antonik is valid, it is highly CPU intensive to compute for large burst size distributions, and in fact the necessary number of calculations scales by a power law with the size of the largest burst. This problem is compounded by the fact that to fit experimental data to this distribution, the minimisation step would require the distribution to be recalculated multiple times. Nir et al., presented an alternative method whereby for each burst in the distribution, a proximity ratio is randomly generated according to a binomial distribution with size S and probability E (where S is the number of photons in the burst, and E is the expected proximity ratio). This randomisation is repeated N times for each burst in the distribution. The resulting distribution of randomly generated ratios can then be divided by N to give a good approximation of the ideal experimental distribution. Background can additionally be accounted for by subtracting a random value generated by a Poisson distribution where the mean is the product of burst length and background rate for each burst sampled.

The distributions generated by this type of analysis give a much better fit than a simple Gaussian distribution, and form the basis for more complex analysis which can handle E* distributions with multiple populations, which may dynamically interconvert. Dynamic PDA[83,84] (dPDA) expands upon the work of Antonik et al. and Nir et al., including a kinetic model for two interconverting states with different distances and forward and backward rates, which is then used to populate an underlying distance distribution from which to generate an expected E* distribution given that molecules may interconvert between states during the measurement. Whilst the amplitudes of the two static peaks in the E* distribution can be used to infer the relative forward and backward rates, the amplitude of the intermediate population can inform the absolute rates, as faster interconversion will increase the number of intermediate molecules detected.

In this way, it is possible to determine kinetic rates of interconversion for biomolecules exhibiting distinct conformational states, even when the solution is at equilibrium, and where no molecular history is available (as in surface immobilized experiments). However, dPDA is effectively “agnostic” to whether or not the intermediate population is truly dynamic, and so burst variance analysis[85] (BVA) and the FRET 2 channel kernel density estimator[86] (2CDE) can be used as a hypothesis test to determine which, if any populations on a proximity ratio histogram are dynamic. Both tests operate on photon arrival times to determine if the proximity ratio has changed mid burst, however they are different in their mathematical construction.

BVA looks for deviations from an expected sub-burst E* variance given the E* of the whole burst, whereas 2CDE is formulated such that excessive fluctuations in E* will give the same deviation in score (>10) irrespective of the E* of the whole burst.

Chapter 2. Open-source smFRET for accurate distances and dynamics

Chapter Overview

In this chapter I expand upon the theory behind performing accurate FRET corrections for absolute distance determination, as briefly outlined in the Chapter 1, and discuss the growing field of open-source microscopy projects, which until now lacks a platform for performing confocal smFRET measurements. I then present a cost-effective, open-source confocal microscope platform, the *smfBox*, for single-molecule FRET, with operational software and analysis software solutions. I use this microscope to participate in a benchmarking study validating the accurate FRET technique, and explore the other capabilities of the smfBox for measuring dynamics and diffusion.

Contributions to work in this chapter

Accurate FRET data collected by myself appears here alongside data from numerous other labs as part of the work in Hellenkamp et al., 2018 (see publications below). The DNA standards themselves were provided by the Hugel lab.

The smfBox was initially designed and built by James M. Baxter, John Cully, and Johathan Shewring. 3D renders of the smfBox shown here were made by Marleen Aaldering. The LABVIEW acquisition software was written by John Cully and Benji C. Bateman. The smOTTER acquisition software was written by Elliot M. Steele. The smFRET data on DNA hairpins at various salt concentrations was collected by Matthew Willmott.

Publications from work in this chapter

- Hellenkamp, B. et al. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat. Methods* 15, 669–676 (2018).

- Ambrose, B. et al. The smfBox is an open-source platform for single-molecule FRET. *Nat. Commun.* 11, 5641 (2020).

2.1 Introduction

Outline

Here I present data validating and optimising the *smfBox,* an open-source confocal smFRET platform, and use it to participate in a multi-lab, worldwide, blind study validating a standardised accurate FRET correction procedure. As well as validating it's capabilities for other techniques such as dynamics analysis with dPDA[83,84] and FRET-2CDE[86], and fluorescence correlation spectroscopy (FCS). The aim of this work is to not only help validate, standardise, and test the precision of accurate FRET correction, but also to validate the smfBox as a platform to enable other biophysicists to access this method.

Accurate FRET correction

The original publication of alternating laser excitation (ALEX)[41] briefly mentioned the idea of using information gained from stoichiometry to correct proximity ratio to absolute FRET efficiency for conversion to distance. The exact procedure to do so was subsequently outlined in a later publication[18]**.** The procedure is as follows;

First, we define the three photon channels:

$D_{ex}D_{em}$: donor channel photons collected under donor excitation
$D_{ex}A_{em}$: acceptor channel photons collected under donor excitation
$A_{ex}A_{em}$: acceptor channel photons collected under acceptor excitation

A fourth channel also exists:

$A_{ex}D_{em}$: donor channel photons collected under acceptor excitation

However these $A_{ex}D_{em}$ photons are not relevant to analysis, as they do not arise from FRET or any other fluorescence process, and the channel is largely dark counts from the detector. In confocal smFRET, the raw data typically exists first as two channels, representing photons from each of the two detectors. Information about the laser duty cycles is then used to divide the photons in each channel based on which laser was on when the photon reached the detector.

Next, backgrounds are subtracted from each channel. This can be done a number of ways as discussed in Chapter 1. Background subtraction is critical, especially for channels with a high background, as this can add photons and alter the FRET efficiency calculated. In this work, backgrounds are calculated using data from within the sample using the method in FRETBursts[20], rather than by taking a separate background sample. In this method, inter-photon delays are calculated, and delays below a certain threshold are discarded (assumed to originate from bursts). The background rate is then calculated by subtracting the threshold time from the remaining delays, and taking the inverse of the mean. The validity of the background calculated from this threshold can then be visually checked by plotting a “fit” line with Eq.2.1.1 over a histogram of inter-photon delays to see if the tail of the distribution has been correctly fit. (SeeFig 2.3.9.c). The background is calculated in this way from each channel and then subtracted from bursts (according to the length of the burst in time).

$$F = ke^{-bt}$$

Eq. 2.1.1

Where F is the frequency of interphoton delays at time t, b is the background rate, and k is a scaling factor to account for the number of photons used in the calculation.

With backgrounds subtracted, E* and S*, the uncorrected FRET efficiency (proximity ratio) and uncorrected stoichiometries can be calculated as such:

$$E^{*} = \frac{D_{ex}A_{em}}{D_{ex}D + D_{ex}A_{em}}$$

Eq. 2.1.2

$$S^{*} = \frac{D_{ex}A_{em} + D_{ex}A_{em}}{D_{ex}D + D_{ex}A_{em} + A_{ex}A_{em}}$$

Eq. 2.1.3

There are a number of confounding variables which cause E* and S* to differ from the “true” FRET efficiency and dye stoichiometry. There are two kinds of spectral crosstalk, the first being leakage of the donor emission into the acceptor channel (due to the long wavelength tails typically exhibited by fluorophore emission spectra), and direct excitation of the acceptor by the donor laser (due to the short wavelength tails typically exhibited by fluorophore excitation spectra). In addition to spectral crosstalk, the donor and acceptor may not be equally detected, due to different quantum yields, different filter bandwidths for each channel, and wavelength dependent quantum efficiency of the detectors. Take for example, a dye pair where 100% of donor photons are detected but only 50% of acceptor photons are detected, in this case a “true” FRET efficiency of 0.5 would instead present as 0.33 if calculated using Eq 2.1.2. For similar reasons, unequal excitation efficiencies (from extinction coefficients and laser powers) can result in an incorrect determination of dye stoichiometry with Eq. 2.1.3.

Thus, to determine absolute FRET efficiencies and stoichiometries, the data must be corrected for leakage, direct excitation, detection efficiencies, and excitation efficiencies. We begin by determining the spectral cross talk parameters from donor only and acceptor only molecules:

$$\alpha = \frac{E_{Donly}}{1 - E_{Donly}}$$

Eq. 2.1.4

The leakage of photons from the donor into the acceptor channel can cause donor only molecules to appear with a FRET efficiency of more than zero, depending on the exact amount of leakage. As such, from the proximity ratio of the donor in absence of acceptor, we determine α, the leakage parameter, using Eq. 2.1.4. This can be done either with a sample deliberately labelled with only a donor, or by using the high stoichiometry (typically 1) of the donor only species to extract it from a data set also containing acceptor labelled molecules.

$$\delta = \frac{S_{Aonly}}{1 - S_{Aonly}}$$

Eq. 2.1.5

Similarly, the direct excitation of the acceptor by the donor laser can cause acceptor only molecules to present with a dye stoichiometry of more than zero. We can use Eq. 2.1.5 to determine δ, the direct excitation correction factor, from acceptor molecules in the absence of a donor. As with leakage and donor only molecules, this can be done either with a deliberately acceptor only labelled sample, or by extracting acceptor only molecules from the data set using their lower stoichiometry.

The $D_{ex}A_{em}$ channel can now have spectral crosstalk photons subtracted, leaving only photons originating from FRET, which we will call $D_{ex}F_{em}$:

$$D_{ex}F_{em} = D_{ex}A_{em} - \alpha D_{ex}A_{em} - \delta A_{ex}A_{em}$$

Eq. 2.1.6

Recalculating E* and S* using $D_{ex}F_{em}$ instead of $D_{ex}A_{em}$, the apparent FRET efficiency of the donor only species becomes 0, and the apparent stoichiometry of the acceptor only species becomes 0 (See Fig. 2.1.1). The background and crosstalk corrected data can now be used to determine the detection efficiency and excitation efficiency correction factors.

As can be seen in Fig. 2.1.1, uncorrected doubly labelled molecules can fall on a gradient, with S* being dependent on E*. The slope of this gradient is dependent on detection efficiencies whereas the height (or S* intercept) is dependent on excitation efficiencies. For example, if the acceptor is less bright (or less well detected) than the donor, then at higher FRET efficiencies the doubly labelled molecules will have lower stoichiometry, as less of the photons absorbed under donor excitation will be detected. Likewise, if the acceptor is less well excited by the acceptor laser than the donor is by the donor laser, then the stoichiometry of all doubly labelled populations will appear to be higher.

The ratios of detection efficiencies for each dye (γ), and excitation efficiencies for each dye (β), can be determined from the ES plot by fitting with an equation:

$$S^{*} = (1 + \gamma\beta + (1 - \gamma)\beta E^{*})^{-1}$$

Eq. 2.1.7

Eq. 2.1.6 can be fit through all doubly labelled bursts in a data set exhibiting multiple FRET efficiencies, or by determining the E* and S* centers of multiple FRET species and fitting through those. In either case, the γ and β parameters determined from the fit can then be used in the final correction process, with α and δ, to give fully corrected E and S:

$$E = \frac{D_{ex}A_{em} - \alpha D_{ex}D_{em} - \delta A_{ex}A_{em}}{\gamma D_{ex}D_{em} + D_{ex}A_{em} - \alpha D_{ex}D_{em} - \delta A_{ex}A_{em}}$$

Eq. 2.1.8

$$S = \frac{\gamma D_{ex}D_{em} + D_{ex}A_{em} - \alpha D_{ex}D_{em} - \delta A_{ex}A_{em}}{\gamma D_{ex}D_{em} + D_{ex}A_{em} - \alpha D_{ex}D_{em} - \delta A_{ex}A_{em} + A_{ex}A_{em}/\beta}$$

Eq. 2.1.9

With γ and β correction factors applied, all doubly labelled populations now appear with a dye stoichiometry of 0.5 (See Fig. 2.1.1), and have correct FRET efficiency which can now be converted to distances.

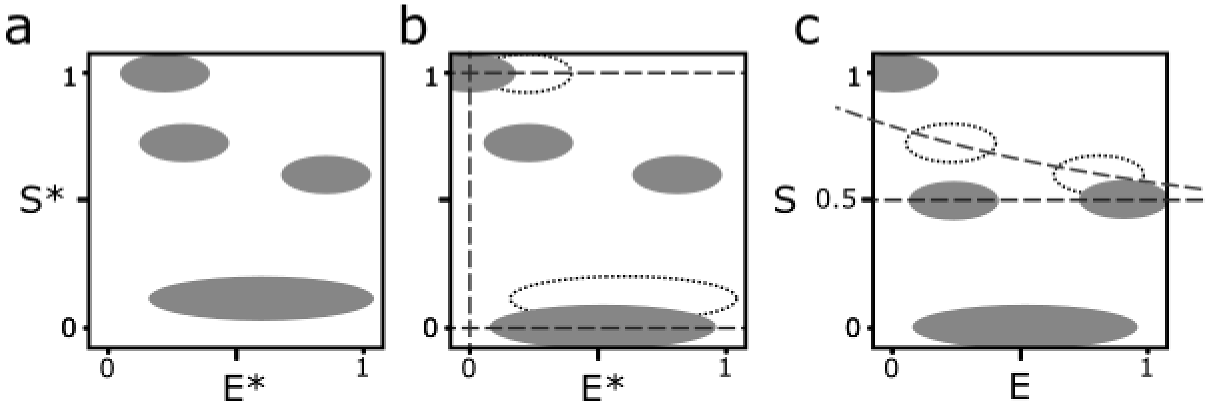


Fig. 2.1.1 Schematics of ES histograms throughout the accurate FRET correction process. **a.** Uncorrected data. The donor only population at S* = 1 has a non-zero E*, and the acceptor only population appear with a low but non-zero S*. The doubly labelled species may not appear with S* = 0.5, and populations with different E* may have different S*, forming a gradient. **b.** After spectral crosstalk correction, the donor only population goes to E = 0 and the acceptor only population goes to S = 0. **c.** Fully corrected ES plot, with both FRET populations going to S = 0.5, and showing accurate FRET efficiencies.

The accurate FRET correction process, yielding absolute FRET efficiencies, has been a very powerful tool for biophysicists, enabling the determination of absolute distances within biomolecules as part of an integrative approach to structure resolution[2]. This capacity of smFRET to yield accurate intramolecular measurements has been employed to restrain structures of biomolecules in a range of systems, in conjunction with other structural techniques.

In DNA polymerases, accurate FRET distances have been used with molecular dynamics (MD) to determine that DNA polymerase recognises a pre-bent, partially melted DNA conformation for replication[87]. In an archaeal RNA polymerase (from *Methanocaldococcus jannaschii*), accurate FRET was used with the Bayesian nanopositioning system (NPS[88]) to reveal how archeal polymerases interact with transcription factors to melt DNA without a dedicated translocase or helicase[32]. In heat shock protein 90 (Hsp90), accurate FRET distances were used to corroborate existing x-ray structures of the closed state, as well as determining an ensemble of conformations in the open state[89]. In a similar case, accurate FRET distances were used to validate two conformational structures of N-methyl-d-aspartate (NMDA) receptors seen in crystal structures, and confirm the existence of a previously unseen but theorised state thought to be responsible for the activation of the receptor upon NMDA binding[36]. Distances determined by FRET were used to elucidate the folding mechanism of DNA G-Quadruplexes as a complex multi-step process[34]. Accurate FRET, as well as a host of other multi-modal fluorescence spectroscopy techniques, have been employed to investigate multi-histone chromatin units, revealing short lived states key to maintaining heterochromatin suprastructure while allowing access to gene regulation machinery[35].

However, despite the wide applicability and demonstrable power of accurate FRET distance determination in structural and mechanistic biology, there has not yet been a wide uptake of the method within the wider field biophysics. One possible reason for this is that the process of correcting FRET efficiencies to be properly accurate has not been standardised, and the reproducibility and precision of such methods is unclear due to a lack of generalised protocols. Furthermore, the instrumentation required to carry out smFRET experiments with ALEX can represent a barrier to entry itself, as the microscopes used to make such measurements must either be bought as part of an expensive imaging system, or self-built and operated with homemade software.

Open-source microscopy

A number of open-source microscopy projects have been developed in recent years, both as software packages and microscope designs themselves. By publishing custom microscopy platforms, other biophysicists can utilise methods that would otherwise require buying specialist equipment or software packages which may be prohibitively expensive, and difficult to modify for newer and extended forms of techniques.

In the space of open-source microscopy hardware, a large number of projects have provided templates for other labs to construct microscopes capable of new and powerful techniques. To name just a few; the miCube[90] is capable of super-resolution microscopy, total internal fluorescence (TIRF), and LED brightfield microscopy. The openSPIM[91] is a light-sheet microscope for selective-plane illumination microscopy (SPIM). Another project details a cost-effective microscope for scanning two-photon microscopy[92].

In addition to whole microscopes, some projects detail open-source plans for common microscope components. NanoJ-Fluidics[93] is a microscope fluidics system that can be constructed from LEGO. The LaserEngine[94] is an open-source design for collimating multiple affordable lasers with speckle and noise reduction. The OpenFlexure[95] is a 3D-printable microscope stage for sample translation and focusing.

In terms of software; ThunderSTORM[96] is an ImageJ based package for single-molecule localization data. Nano-J-SQUIRREL[97], a similar package provides quantitative assessment of super-resolution microscopy image quality for identifying methodological defects. fairSIM[98] is a software package for analysis of structured illumination microscopy (SIM) data.

Specifically in the field of single-molecule fluorescence spectroscopy, and smFRET in particular, there are few freely available packages published for data analysis, and none for data acquisition. PIE analysis with MATLAB (PAM[77]) is capable of handling data from multiple types of single-molecule spectroscopy experiments, and the python package FRETBursts[20] can perform burst analysis from both ALEX and PIE experiments. The photon-HDF5[99] file standard is a single-molecule specific branch of the HDF5 (hierarchical data format 5) file type which keeps experiment specific metadata with photon data for context specific interpretation of the file.

Whilst the miCube[90], for example, can be used for TIRF based smFRET, no open-source project until now has detailed the construction of a confocal microscope capable of smFRET.

2.2 Methods and Materials

DNA Constructs

For the DNA hairpins in this study, oligonucleotides containing amino-modified thymine nucleotides (on C6 linkers) were purchased from LGC Biosearch, and labelled according to the protocol outlined below. Complementary labelled oligos (Table 2.2.21) were then mixed at equimolar concentration in annealing buffer, heated to 95 °C, and allowed to cool overnight to form the duplex hairpins (Table 2.2.2). For the straight duplex DNA standards, labelled and HPLC purified oligonucleotides with C2 linkers were ordered from IBA GmbH (Göttingen) by the Hugel group before being distributed to participating researchers.

Table 2.2.1 DNA oligonucleotides used in this chapter to make the DNA constructs in Table 2.2.2

Oligo Name	Sequence	Label
MJW-001	5’**-T(C6 amino)-**GG ATT AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA TCC AAA GGA TGT ATG GTA ATG GGA CGA AGA ATG AGG-3’	Cy3B on C6 linker
MJW-002	5’-CCT CAT TCT TCG TCC CAT TAC CA**-T(C6 amino)-** ACA TCC-3’	Atto647N on C6 linker
MJW-003	5’-CCT CAT TCT TCG TCC CAT **-T(C6 amino)-** AC CAT ACA TCC 3’	Atto647N on C6 linker
MJW-004	5’-CCTCATTCTTCG **-T(C6 amino)-** CCCATTACCATACATCC-3’	Atto647N on C6 linker
T 31(C2), A strand	biotin-5’-CCA GAC AAA CAC TCA AAC AAA CTC GAC ACT **-T(C6 amino)-**TC AGC TC -3’	Atto647N on C2 linker
T 31(C2), D strand	5’- GAG CTG AAA GTG TCG AGT TTG TTT GAG TGT **-T(C6 amino)-**TG TCT GG - 3’	Atto550/Atto 647N on C2 linker
T 23(C2), D strand	5’- GAG CTG AAA GTG TCG AGT TTG T**-T(C6 amino)**-T GAG TGT TTG TCT GG - 3’	Atto550/Atto 647N on C2 linker
T 10(C2), D strand	5’- GAG CTG AAA GTG TCG AGT **-T(C6 amino)-**TG TTT GAG TGT TTG TCT GG - 3’	Atto550/Atto 647N on C2 linker

Table 2.2.2: DNA constructs used in this chapter.

Construct Name	Donor strand	Acceptor strand
1a	T 31(C2), D strand	T 31(C2), A strand
1b	T 23(C2), D strand	T 31(C2), A strand
1c	T 10(C2), D strand	T 31(C2), A strand
High FRET hairpin	MJW-001	MJW-002
Mid FRET hairpin	MJW-001	MJW-003
Low FRET hairpin	MJW-001	MJW004

DNA labelling and purification

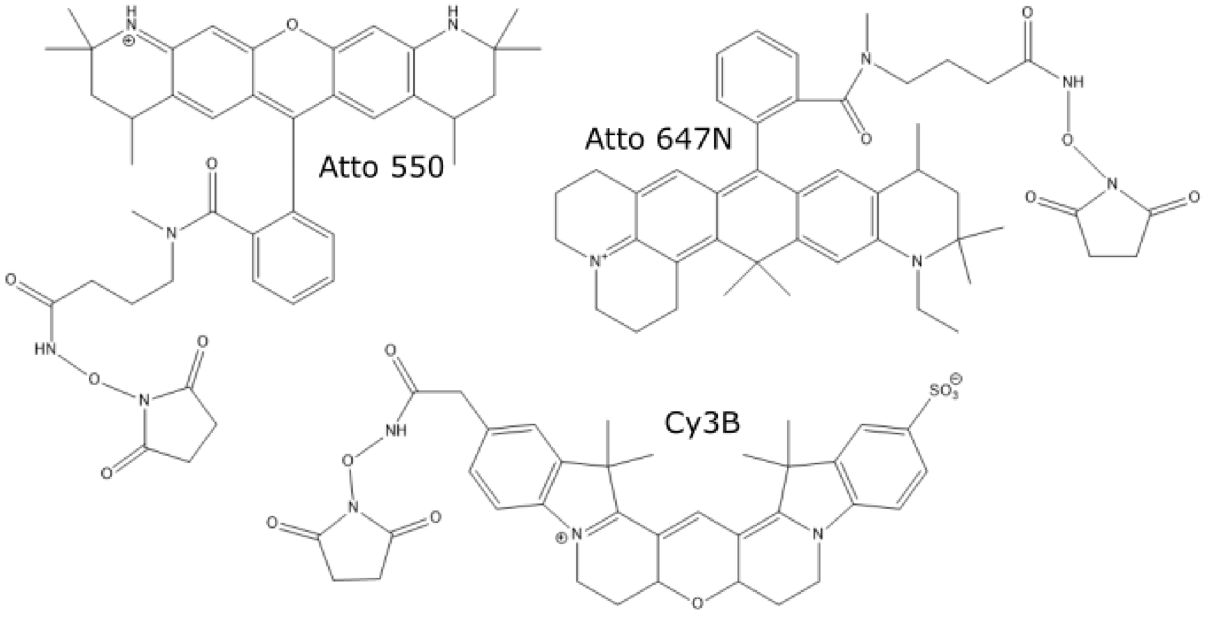


Fig. 2.2.1: Dye NHS esters used in this chapter.

Fluorescent dye N-hydroxysuccinimide (NHS) esters (Fig. 2.2.1) were obtained lyophilised (GE Healthcare, Atto-tec) in 1 mg aliquots. Dyes were dissolved in dry dimethyl sulfoxide (DMSO) and split into 50 nmol aliquots. Aliquots were re-lyophilised in a vacuum concentrator and stored at -80 °C. DNA oligonucleotides with internal amino groups were suspended in labelling buffer (100 mM Sodium Tetraborate, pH 8.3) to 200 µM, and dye aliquots for labelling were resuspended in 5 µl of DMSO. A borate buffer is used for labelling as opposed to an amine based buffer (ie. Tris) so as to not interfere with the amine specific reaction of the NHS ester. 25 ul of DNA were then mixed with one 5 ul aliquot containing 50 nmoles of dye NHS ester, and left in a light-tight thermoshaker at 50 °C and 600 rpm for 1 hour. After 1 hour, a second aliquot is added, the temperature is lowered to 37 °C and left overnight at 600 rpm.

To initially remove the free dye from solution, an ethanol precipitation is performed as follows: 3.5 µl of 3M sodium acetate pH 5.2 and 100 µl of ice-cold 100% ethanol are added, the solution is vortexed and incubated at -20 °C for at least an hour. The solution is then spun in a centrifuge at 16,000 G and 4 °C for 15 minutes to pellet the precipitated DNA. The supernatant is removed and discarded with care not to disturb the pellet, and spun again with 200 µl of ice-cold 70% ethanol at 16,000 G, 4 °C for 5 minutes to wash the pellet. The supernatant is again discarded and the pellet is air dried in the dark.

To separate the unlabelled from the labelled oligonucleotides, denaturing polyacrylamide gel electrophoresis (PAGE) purification is performed as follows: a polyacrylamide gel with urea and TBE buffer (130 mM Tris pH 7.6, 45 mM boric acid, 2.5 mM EDTA) is cast with 20 ml of 40% polyacrylamide, 4 ml of 10x TBE, 14.4 g urea, made up to 40 ml with H_2O and polymerised with addition of 200 µl of 10% ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) before pouring. The gel is pre-run in 1x TBE at a constant power of 30-40 W until it is at approximately 55 °C. Labelled DNA oligonucleotides are resuspended in 50% formamide and loaded into the gel. The gel is run at constant power, in the dark, until the visibly coloured band has reached approximately ⅔ of the way down the gel. The gel is imaged by UV-shadow against a fluorescent TLC plate, to confirm separation between the labelled and unlabelled bands. The labelled bands are then excised from the gel, crushed, and soaked in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) overnight.

The solutions are then spun down (low speed tabletop centrifuge) to pellet the polyacrylamide, and the oligonucleotide containing supernatant is purified using Bio-Spin® 6 columns (Bio-Rad) to remove the urea and borate content of the solution. The columns are first equilibrated with TE buffer according to product instructions before application of the sample.

Purified oligos are then checked for concentration and labelling efficiency on a NanoDrop spectrophotometer (Thermofisher) by first subtracting the 260 nm absorbance contribution from the dye.

The smfBox

The smfBox presented in this chapter is an open-source, cost effective, confocal single-molecule FRET microscope, with operational software which produces photon-HDF5 files for analysis in either PAM[77] or FRETBursts[99] (see below). For full build specifications and parts lists, see the results section.

Operational Software

The smfBox can be operated using either smOTTER, an open-source standalone implementation written in C++, or a set of LabVIEW VI's which can be more easily customised. The LabVIEW software records raw photon streams as .txt files which can be converted to photon HDF5 using Jupyter Notebooks with the phconvert python package. smOTTER saves data directly as photon HDF5 files for analysis.

Analysis Software

HDF5 data files generated by the smfBox can be analysed using either FRETBurst or the MATLAB based PAM[77]. Where specified, data shown here was analysed either with Jupyter Notebooks running FRETBursts 0.7 in a custom Anaconda 5.3.0 environment (see smfBox GitHub[100] for install file), or using PAM.

Accurate FRET

Duplex DNA constructs (see Table 2.2.2), were provided by the Hugel Lab as part of the blind, multi-lab FRET study[14]. DNAs were diluted to approximately 100 pM in observation buffer (20 mM $MgCl_2$, 5 mM NaCl, 5 mM Tris, pH 7.5), and ~50 µl placed on a coverslip passivated with 1 mg/ml BSA, and data were acquired by the smfBox. Analysis was done with Jupyter Notebooks using FRETBursts[20]. Background in each channel were estimated by means of an exponential fit of inter-photon delays. Bursts were identified using an all photon sliding window[66] algorithm as described in Chapter 1, with $L = 10$ and $F = 45$ for both channels, and background was subtracted. Spectral cross talk factors were found by combining data from all standards and extracting bursts with a stoichiometry >0.95 as the donor only population and <0.175 for the acceptor only population to calculate α and δ, respectively. A dual channel burst search (DCBS) was then used to extract doubly-labelled bursts from each 30-min acquisition, and used to find E and S with single Gaussian fits. Combined data from all three oligos were then plotted together and fitted with Eq. 2.1.6 to obtain γ and β. Corrected FRET efficiencies of all doubly-labelled bursts were then obtained using all four correction parameters as described above in this chapter.

Hairpin dynamics

DNA Hairpins (see Table 2.2.2) were diluted to approximately 100 pM in Hairpin buffer (Tris 50 mM pH 7.5, BSA 0.1 mg/ml, EDTA 1 mM, Glycerol 5%, DTT 1 mM) with additional NaCl where specified, placed into a chamber made of two coverslips and a silicone gasket (to enable >2 h acquisitions with no sample evaporation, therefore maintaining a constant salt concentration), and data were acquired by the smfBox. Data were analysed using the MATLAB software package PAM[77]. Bursts were selected using a sliding window dual channel burst search, with a 50 photon threshold and a 500 µs window size, with FRET-2CDE[86] values calculated. Doubly-labelled bursts were selected between 0.2 and 0.85 S, and bursts were cut into 0.5, 1, and 1.5 ms lengths. To access the precision of the kinetic parameters, acquisitions were split into subsets of 2000 molecules before further analysis. Dynamic PDA[83] was then used to fit a two-state model to the data using the histogram library method implemented in PAM.

Fluorescence Correlation Spectroscopy

Samples were diluted to a concentration higher than would be normally used for single-molecule experiments (~ 10 nM), in water in the case of Rhodamine 6G, and in buffer (5 mM Tris, pH 7.5, 20 mM $MgCl_2$ and 5 mM NaCl) in the case of DNA. Data were acquired for 3 minutes under continuous wave excitation with the 515 nm laser, at a reduced laser power (100 uW before the objective). Correlation curves were generated and fitted using PAM[77]. A simple 3D diffusion model (Eq. FCS) was used, with the diffusion coefficient for Rhodamine 6G being fixed at 414 $\mu m^2\ s^{-1}$ as determined by a previous study[101]. Confocal volume is calculated using Eq. Vol

$$G(\tau) = 1 + \frac{1}{N_p} \cdot \frac{1}{\left(1 + \frac{\tau}{\tau_D}\right)} \cdot \frac{1}{\sqrt{1 + (\frac{w_0}{z_0})^2 \cdot \frac{\tau}{\tau_D}}}$$

Eq. 2.2.1

$$Vol = \left(\frac{\pi}{2}\right)^{3/2} w_0^2 z_0$$

Eq. 2.2.2

2.3 Results

The smfBox

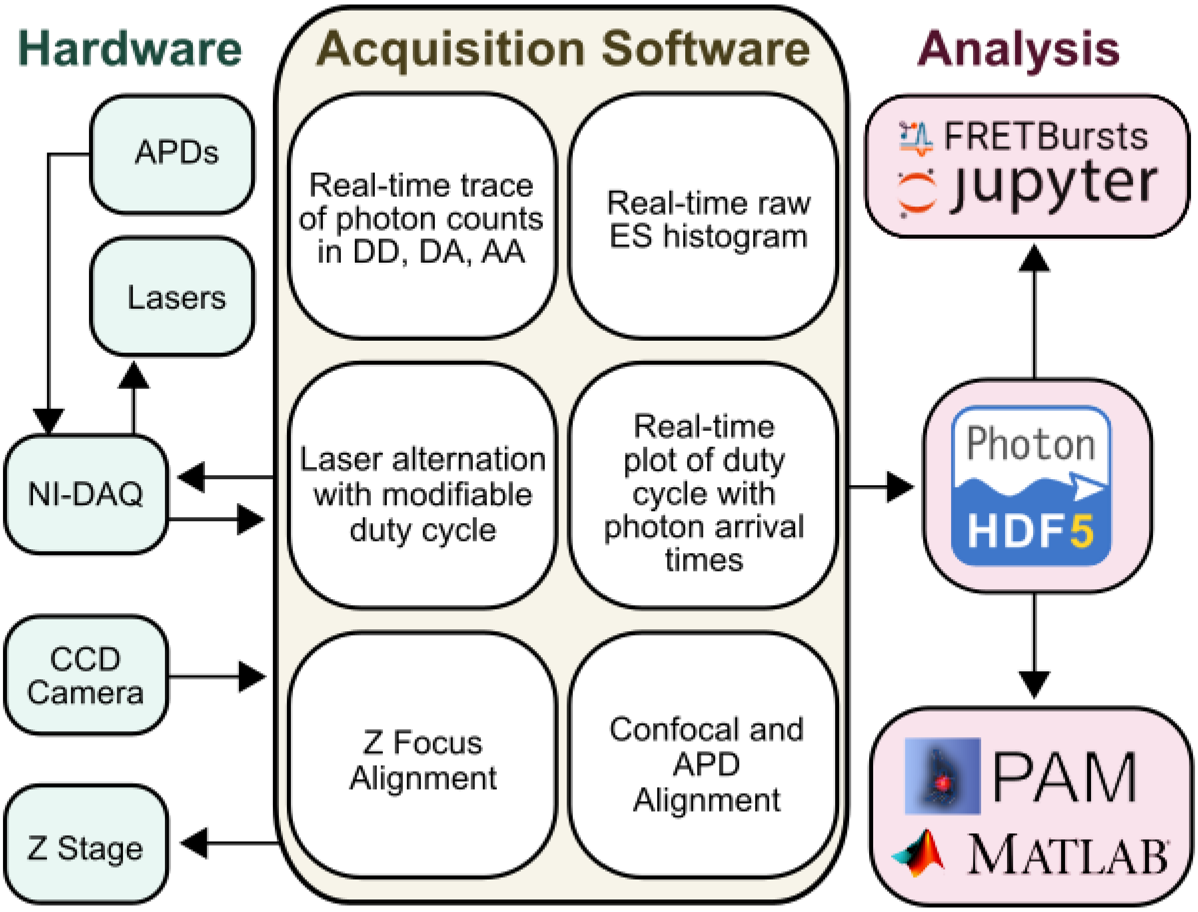


Fig. 2.3.1 Flowchart of the smfBox platform. The acquisition software controls both lasers and avalanche photodetectors (APD's) via the NI-DAQ board. The user can control the Z-stage to focus, using the image from the CCD camera to inform the confocal focus position. The acquisition software shows real-time trace of photon counts which can be plotted on a raw ES histogram to inspect the sample before acquiring data. Laser duty cycle can be modified and inspected against photon counts in each channel. The alignment tab shows a time trace with longer time binning and y-axis for aligning the APD's and confocal system with a concentrated sample. Data is exported as photon-HDF5 files for analysis in either FRETBursts or PAM.

The smfBox platform is a home-built confocal microscopy capable of single-molecule FRET with alternating laser excitation (ALEX). Operational software; smOTTER, interacts with all hardware to focus the z-stage stage, control lasers, and record photon data (Fig. 2.3.1). All build instructions, with shopping list, operational software, analysis pipeline, and sample data are available online[100].

The smfBox alternates two lasers (515 nm at ~220 µW, and 635 nm at ~70 µW, Omicron LuxX plus lasers, powers measured immediately before the excitation dichroic) by TTL-controlled modulation of electronic shutters. Following the laser path in Fig. 2.3.2; the beams are coupled into a single-mode fibre before being collimated, (L1) and cropped by an iris to 5 mm, stabilised by two mirrors (M1, M2), and into the microscope body, first passing through a beamsplitter BS1. Light reflected off of the beamsplitter reaches a nanosecond rise time photodiode (Thorlabs, DET10A/M) for monitoring of laser power. Light passing through the beamsplitter enters the microscope body. See Fig. 2.3.3 for an exploded view of the excitation path.

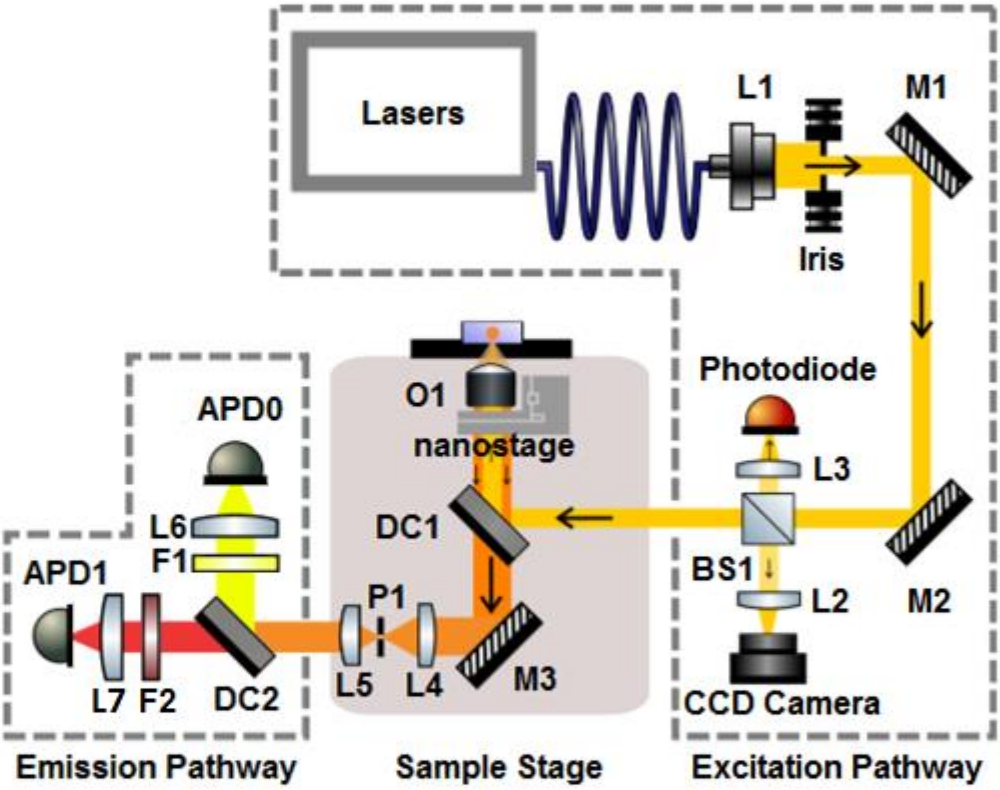


Fig. 2.3.2 2D Schematic of the smfBox showing the Excitation pathway, box, and emission pathway, with all optical components labelled as per the figures below.

Inside the microscope body, an excitation dichroic mirror DC1 (Chroma ZT532/640 rpc 3 mm) directs the beam into an objective O1 (Olympus UPLSAPO ×60 NA = 1.35 oil immersion), mounted on a nanostage for focussing. The same objective collects the emission, which passes through the excitation dichroic DC1, reflected off of another mirror M3, and focused through a 20 μm pinhole P1 by two lenses, L4 and L5. See Fig. 2.3.4 for an exploded view of the optics inside the microscope body.

Light reflected off of the sample and collected by the objective, being the same wavelength as the incident laser light, is reflected by DC2 rather than transmitted, where it leaves the microscopy body back up the excitation path towards beamsplitter BS1. On the way into the box the beamsplitter directs light to the nanosecond photodetector, but on the way out, scattered light is directed to a lens L3, and a CCD camera (Thorlabs DCC1545M). Imaging the scattered light allows the experimenter to determine whether the confocal volume is focused above or below the glass-solvent interface of the sample.

In the emission path, after the lens-pinhole-lens system, the emission light is split into two colour channels by a longpass dichroic mirror DC2 (Chroma NC395323—T640lpxr). Each of the two colour channels is cleaned up by filters F1 and F2, before being focussed by lenses L6 and L7 onto the two avalanche photodetectors, APD1 and APD2 (SPCM-AQRH-14 and SPCM-NIR-14, Excelitas). See Fig. 2.3.5 for an exploded view of the emission path. Photon arrival times and detector channel are recorded by a national instruments card (PCIe-6353), see Fig. 2.3.7.

The microscope body (Fig. 2.3.6) is made of anodised aluminium parts with a light-tight lid, and shutter mechanism which blocks the laser when the lid is open. This effectively makes the smfBox a Class 1 laser product as no laser light will escape under normal use.

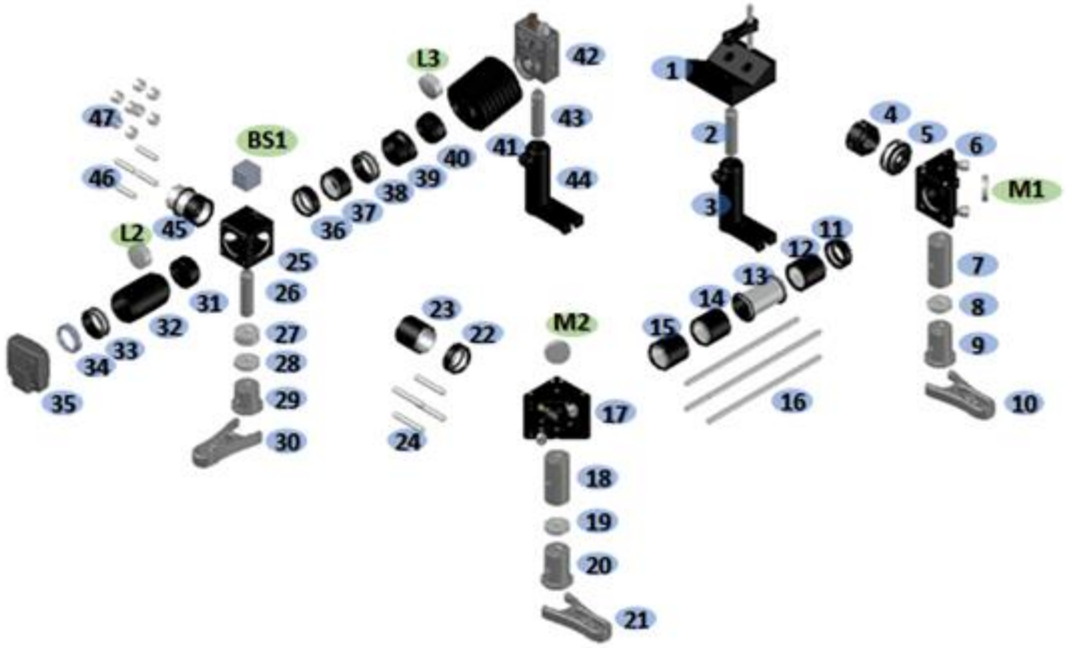


Fig. 2.3.3: Exploded view of the excitation path with all parts labelled according to Table 2.3.1.

Table 2.3.1 Excitation path components

Part Nr	Product Code	Use	Part Nr	Product Code	Use
1	VC3C/M	V-shaped collimating lens holder	27	RS10/M	Post spacer BS1 (L=10mm)
2	TR50/M	Collimating lens post	28	RS5/M	Post spacer (L=5mm)
3	UPH50/M	Post holder	29	RS1P8E	Pedestal Pillar Post BS1 (L=1")
4	SM1P1	Attachment lens to iris	30	CF125C/M	Clamping fork for BS1 post
5	SM1D12C	Iris	31	SM1L03	Tube 1 BS1 to DCC
6	KCB1C/M	Kinematic Mirror Mount for M1	32	SM1L20C	Slotted lens tube for L2
7	RS50/M	Pillar post (L=50 mm)	33	SM1T2	Tube coupler BS1 to DCC
8	RS5/M	Post spacer (L=5mm)	34	SM1A9	Adapter SM1/C-mount BS1 to DCC
9	RS1.5P/M	Pedestal pillar post (L=38mm)	35	DCC1545M	CCD Camera
10	CF125C/M	Clamping fork for M1 post	36	SM1L03	Tube 1 BS1 to photodetector
11	SM1L03	Tube 1 M1 to M2	37	SM1L05	Tube 2 BS1 to photodetector
12	SM1L10	Tube 2 M1 to M2	38	SM1T2	Tube coupler BS1 to photodetector
13	SM1T20	Tube coupler M1 to M2	39	SM1L03T	Angled tube BS1 to photodetector
14	SM1L10	Tube 3 M1 to M2	40	SM1L03	Tube with L3
15	SM1L10	Tube 4 M1 to M2	41	SM1B3	Tube Bellow BS1 to photodetector
16	ER6-P4	4x rods M1 to M2 (6")	42	DET10A/M	Photodetector
17	KCB1C/M	Kinematic Mirror Mount for M2	43	UPH50/M	Post holder
18	RS50/M	Pillar post (L=50 mm)	44	TR50/M	Post for photodetector
19	RS5/M	Post spacer (L=5mm)	45	SM1V10	Tube BS1 to box
20	RS1.5P/M	Pedestal pillar post (L=38mm)	46	ER1-P4	4x Cage Rods BS1 to box
21	CF125C/M	Clamping fork for M1 post	47	ERSCA-P4	2x4 Rod adapter BS1 to box
22	SM1L03	Tube 1 M2 to BS1	M1	BB1-E02	First Mirror
23	SM1S10	Tube 2 M2 to BS1	M2	BB1-E02	Second Mirror
24	ER1.5-P4	1x4 rods M2 to BS1 (1.5")	BS1	21012	10/90 Beamsplitter
25	CM1-DCH/M	Cage Cube for BS1	L2	AC254-030-AML	Lens for Camera
26	TR50/M	Post for BS1(L=50mm)	L3	49793	Lens for Photodetector

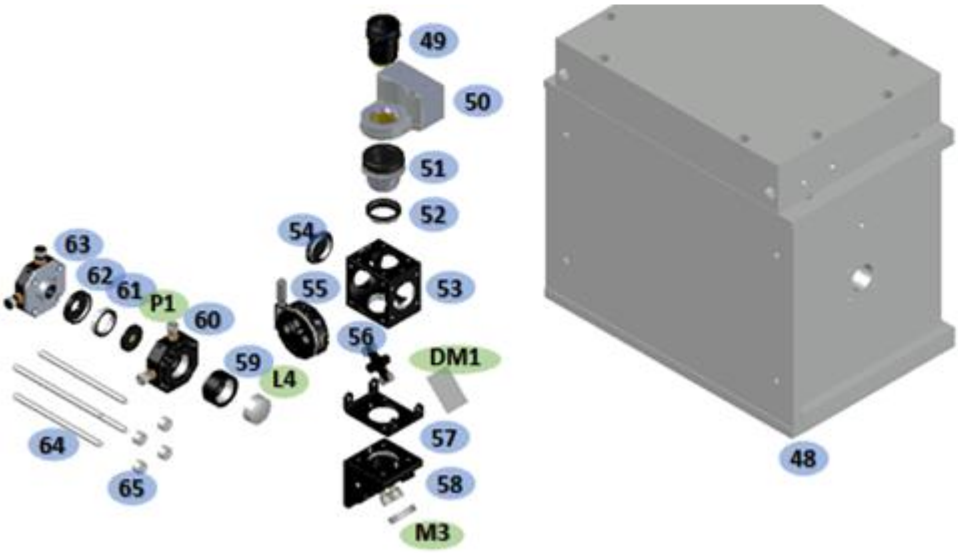


Fig. 2.3.4 An exploded view of the optical components inside the microscope body, with parts labelled according to Table 2.3.2

Table 2.3.2 Part numbered microscope body components.

Part	Product	Use	Part Nr	Product Code	Use
48	N/A	Machined Aluminium Box	59	SM1L05	Lens Tube L4 (0.5")
49	N1480700	Objective	60	CXY1	Translating Lens Mount for L4
50	FOC.300	Single axis nanostage	61	SM1A3	Adapter (Ext. SM1 Int. RMS) for P1
51	N/A	Piezoconcept bushing (with	62	SM1A1	Adapter (Ext. SM05 Int. SM1) for P1
52	SM1A12	FOC bushing adapter (Ext. SM1 Int.	63	CXYZ05/M	Translating Mount for P1
53	C4W	Cage Cube DM1	64	ER4-P4	1x4 4" Rods
54	SM1CP2	Cap for spare DM1 cage cube ports	65	ERSCA-P4	1x4 Rod Adapter
55	B4CRP/M	Precision mount for DM1	DM1	ZT532/640rpc	Excitation Dichroic
56	FFM1	Clamp for DM1	M3	MM3-311-t6-1	Hard Mirror between DM1 and L4
57	C4W-CC	Cage Cube Connector M3 to DM1	L4	Edmund - 49792	Lens before pinhole (50 mm FL)
58	KCB1/M	Right angle KM mount for M3	P1	PNH-20	Pinhole (20 µm)

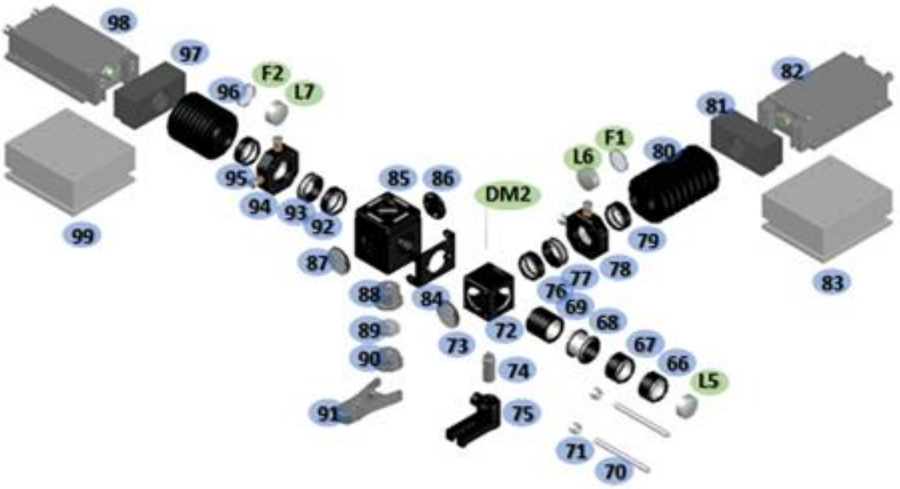


Fig. 2.3.5 An exploded view of the emission path, with parts labelled according to Table 2.3.3

Table 2.3.3 Part numbered emission path components.

Part	Product	Use	Part Nr	Product Code	Use
66	SM1L05	Tube 1 box to DM2 with L5	86	SM1CP2	Caps for spare ports of filter
67	SM1L05	Tube 2 box to DM2	87	SM1CP2	cubes
68	SM1T10	Tube coupler box to DM2	88	RS075P/M	Pedestal pillar post (L=19mm)
69	SM1L10	Tube 3 box to DM2	89	RS5/M	Spacer (L=5mm)
70	ER3	2x Rods box to DM2 3"	90	RS05P	Pedestal pillar post (L=0.5")
71	ERSCA	2x Rod adapters box to DM2	91	CF125C/M	Clamping fork filter cube
72	CM1-DCH/M	Cage cube for DM2	92	SM1L03	Tube 1 DM2 to APD0
73	SM1CP2	Cap for spare port of DM2 cage cube	93	SM1T2	Tube adapter DC2 to APD1
74	TR30/M	Post DM2 (L=30 mm)	94	CXY1	Translating lens mount
75	UPH30/M	Post Holder DM2 (L=30mm)	95	SM1L03	Tube for F2 DC2 to APD1
76	SM1L03	Tube 1 DM2 to APD0	96	SM1B3	Lens tube bellow DC1 to APD1
77	SM1T2	Tube adapter DM2 to APD0	97	Fabricated Al	Connector tube to APD1
78	CXY1	Translating lens mount	98	SPCM-AWRH-14	APD1 (Excelitas)
79	SM1L03	Tube for F1 DM2 to APD0	99	Fabricated Al	Post APD1
80	SM1B3	Lens tube bellow DM1 to APD0	L5	Edmund - 49793	Lens after Pinhole (63.5 mm FL)
81	Fabricated Al	Connector tube to APD0	L7	AC254-75-B-ML	APD1 Lens
82	SPCM-AWRH-	APD0 (Excelitas)	L6	AC254-100-A-ML	APD0 Lens
83	Fabricated Al	Post APD0	F2	FF01-679/41-25	Acceptor clean-up filter
84	CM1-CC	connector DM2 to filter cube	F1	FF01-571/72-25	Donor clean-up filter
85	DFM1/M	Fluorescence filter cube	DM2	NC395323-T640lpxr	Emission Dichroic

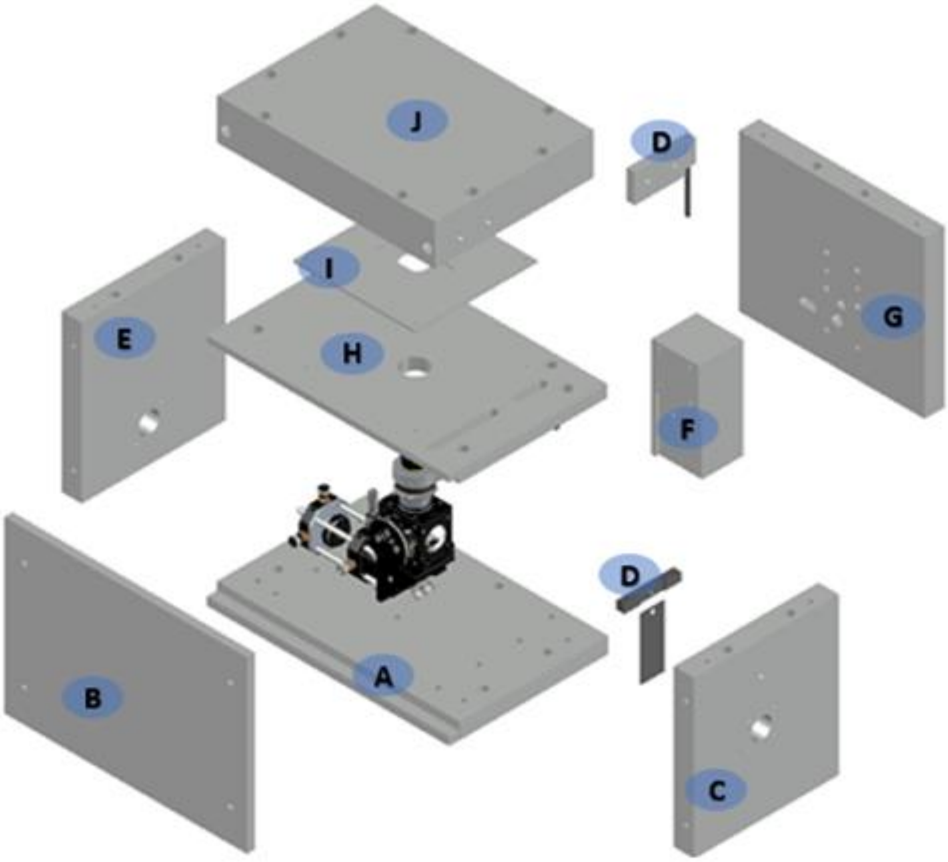


Fig. 2.3.6: An exploded view of the aluminium box with components labelled according to Table 2.3.4.

Table 2.3.4 Microscope body components

Part Nr	Use	Part Nr	Use
A	Base Plate	F	Block
B	Front Plate	G	Back Plate
C	Laser-In Plate	H	Top Plate
D	Shutter	I	Stage
E	Laser-Out Plate	J	Lid

Full build instructions for the excitation path (Fig. 2.3.3, Table 2.3.1), emission path (Fig. 2.3.5, Table 2.3.3), microscope body (Fig. 2.3.6, Table 2.3.4), and internal microscope components (Fig. 2.3.4, Table 2.3.2) are available online with scrollable animations showing how each piece fits together[100].

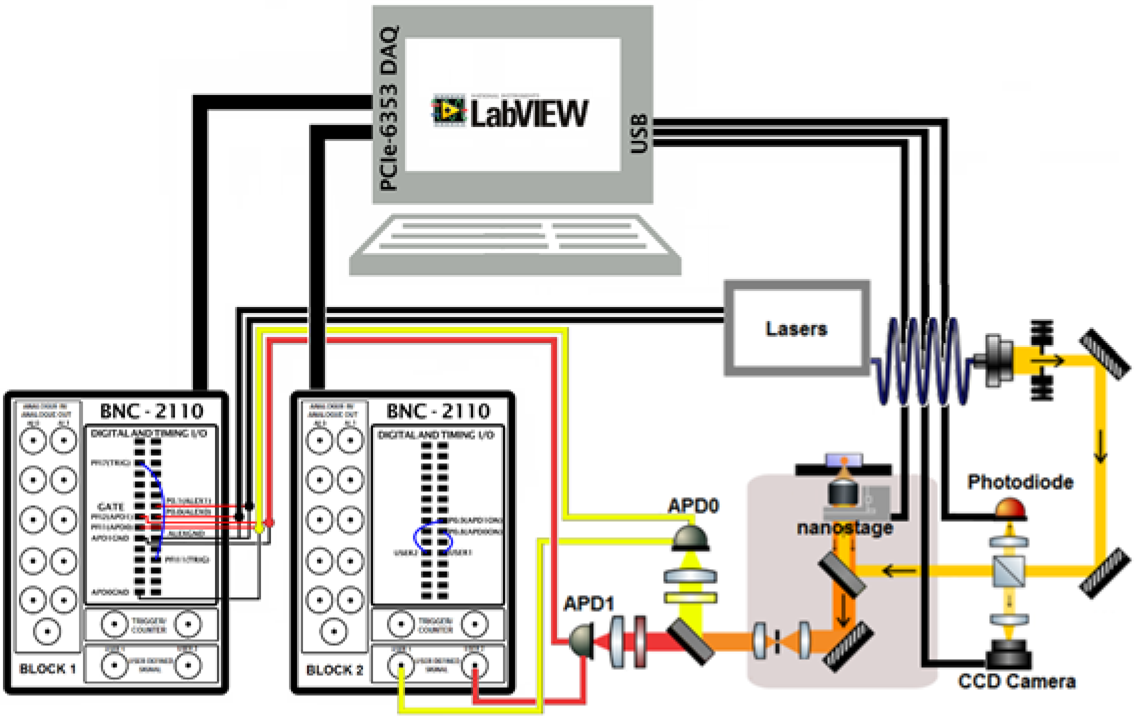


Fig. 2.3.7: Wiring diagram of electronics in the smfBox. Lasers are controlled, and APD's are gated and monitored via BNC adaptors connected to the NI-DAQ board in the PC. The CCD camera and nanostage for Z focussing, as well as the photodiode for laser power monitoring are connected directly via USB.

The wavelength dependent optics consist of; an excitation polychroic mirror, emission dichroic, and cleanup filters designed to excite dyes at 515 nm and 638 nm, and detect them at 530-620 nm and 660-710 nm (see Fig. 2.3.8). This allows for the use of many popular green donor fluorophores such as Cy3, Cy3B, Atto 532, Atto 550, and Alexa Fluor 546, and far red fluorophores such as Cy5, Alexa Fluor 647, and Atto 647N. However, the smfBox could theoretically be constructed for the use of dyes in any spectral range, and the use of an additional laser and an extended emission path could easily allow for 3 colour experiments.

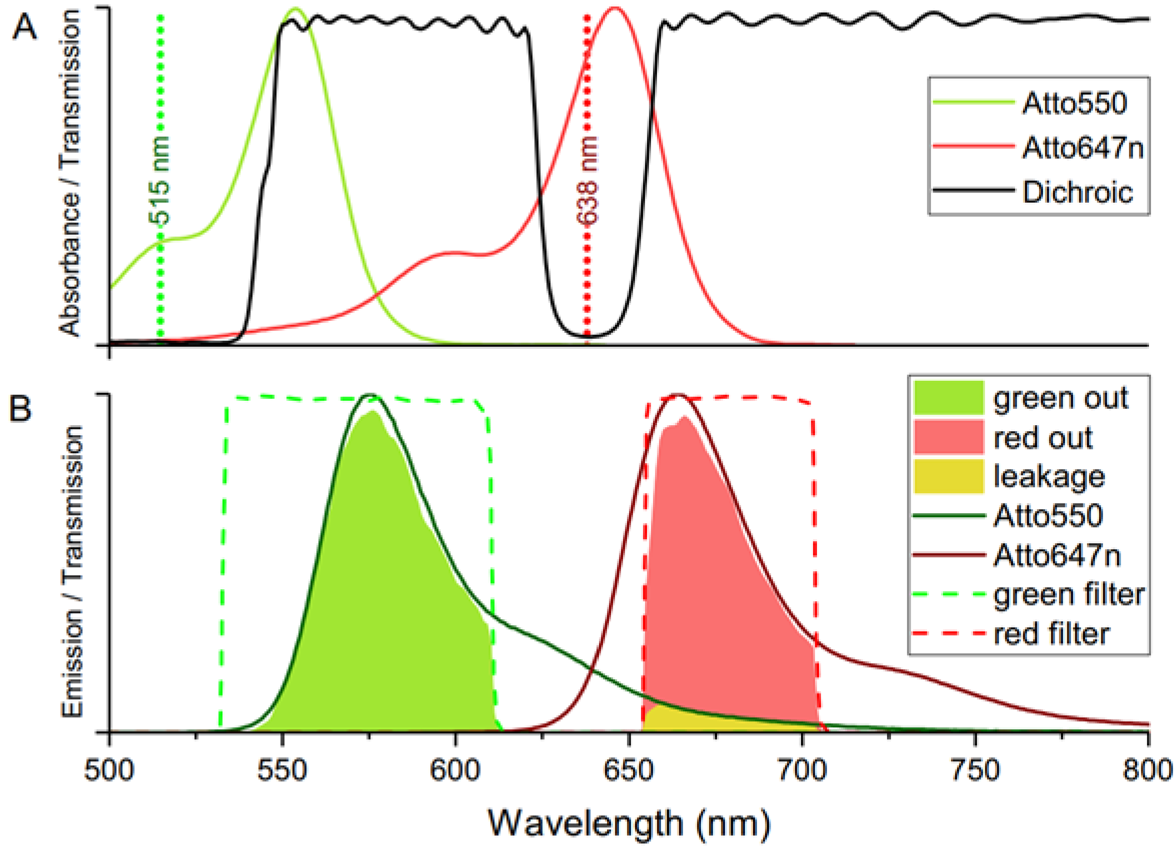


Fig. 2.3.8: Spectra of optical components in the smfBox. **a:** The excitation dichroic in black with laser lines and typical dye absorption spectra in green and red. **b:** The two emission channels shown in dashed lines, with typical emission spectra. Block colours show the spectra of emission which ultimately reaches the detectors.

FRET Analysis workflow

In order to make smFRET analysis accessible and reproducible, self-explanatory Jupyter Notebooks were made for the smfBox platform which use the FRETBursts[20] python package (Fig. 2.3.9). The Jupyter notebooks described here guide the user through the process of importing photon-HDF5 files[99], and checking the data for problems such as incorrect duty cycles (as specified within the metadata created by the acquisition software), calculating background (via Eq. 2.1.1) or fluctuating background levels (Fig. 2.3.10) which can be indicative of evaporation or sample loss to the coverslip.

Uncorrected data can be worked with directly, or the correction notebooks can be used to determine correction parameters for accurate FRET, starting with determining alpha and delta, then gamma and beta (see Fig. 2.3.11).

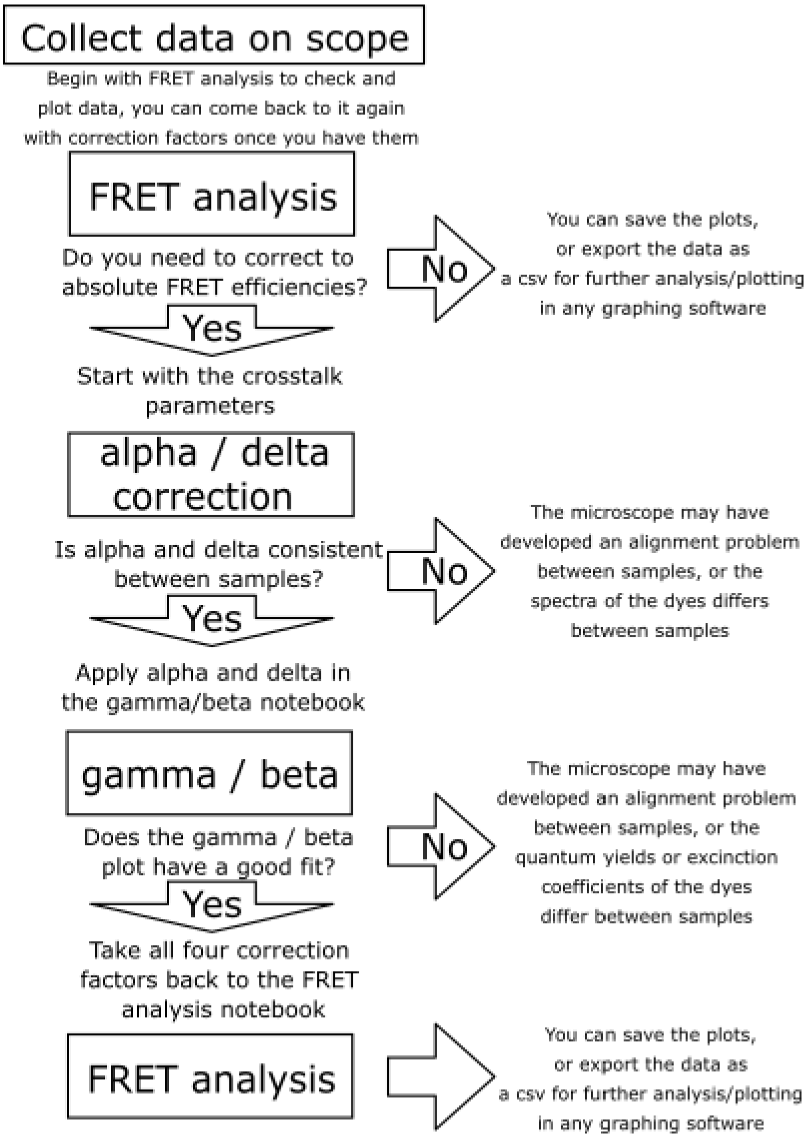


Fig. 2.3.9 The work flow for analysing smFRET data with Jupyter Notebooks. The FRET analysis notebook can be used for an initial inspection of the data. If accurate FRET corrections are not required the data can be plotted within this notebook or the burst matrices can be exported as csv files for further work in another program. If Accurate FRET correction is required, the alpha/delta correction notebook is first used to acquire leakage and direct excitation parameters from the donor only and acceptor only populations. These are then applied in the gamma/beta notebook, where gamma and beta are determined using the spectral crosstalk corrected apparent FRET efficiencies and stoichiometries of distinct populations. All four correction factors can then be applied in the FRET analysis notebook for analysis of absolute FRET efficiencies.

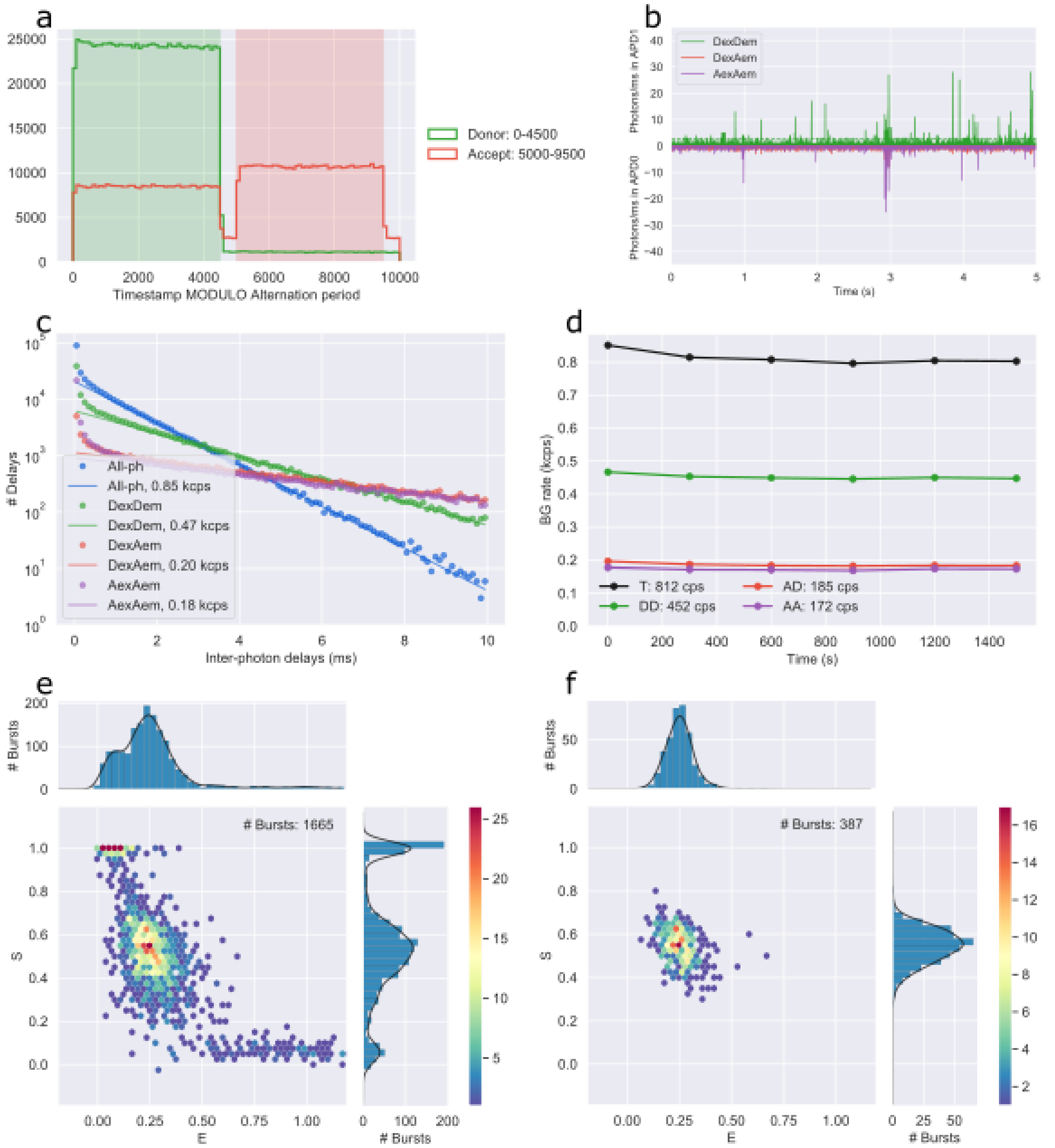


Fig 2.3.10 Sample graphs obtained from the FRET analysis notebook. **a:** Laser duty cycle plot. The green and red lines show the photons arriving within that part of the laser duty cycle across the entire experiment. The shaded areas show the time periods that each laser is on. If the laser duty cycle parameters supplied are correct then the shaded areas will line up with the detected photons. **b:** Time trace for the first 5 seconds of the experiment with photons the DexDem channel shown in green, DexAem in red, and AexAem shown in purple. **c**: Background fits for each of the three channels and all four combined (blue). A histogram of interphoton delays is fitted with an exponential distribution, by ignoring delays shorter than a given threshold (1.5 ms here) which originate from fluorescent bursts, and taking the sample mean (minus the threshold). The fit to the remaining photons after the threshold gives the background photon rate (see Eq. 2.1.1 for fit equation). **d**: A plot of background in each channel over time, this can be useful to see if any dramatic changes have occurred throughout the experiment. **e**: ES histogram of all burst populations, from an all photons burst search. **f:** ES histogram of the doubly labelled population, from a dual channel burst search.

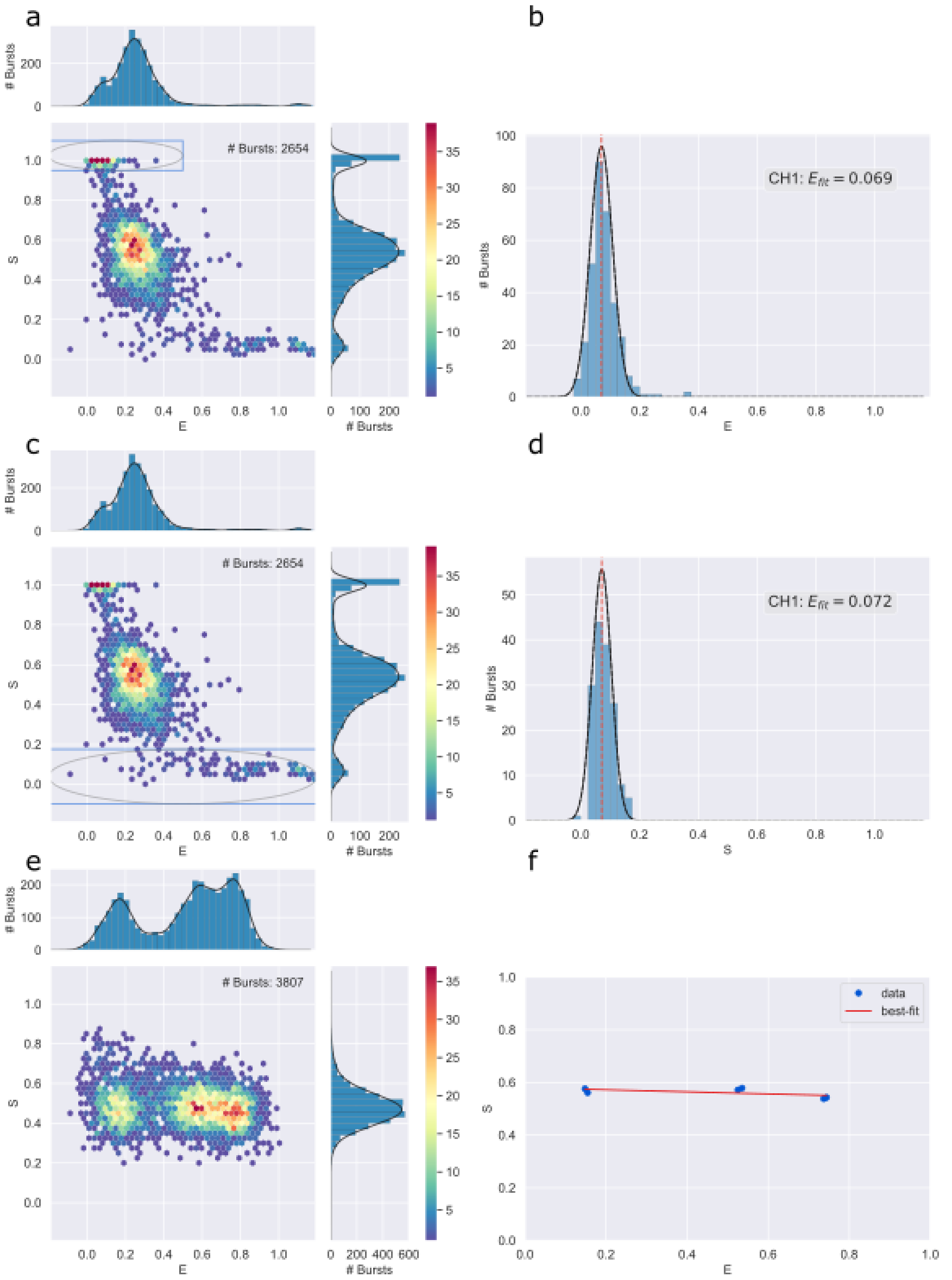


Fig 2.3.11 The plots from the correction factor notebooks. **a:** ES histogram of an all photons burst search with the donor only population selected. **b:** 1D E histogram of the donor only population fitted with a Gaussian distribution to determine the leakage factor. **c:** ES histogram of an all photons burst search with the acceptor only population selected. **d:** 1D S histogram of the acceptor only population fitted with a Gaussian distribution to determine the direct excitation factor. **e**: ES histogram of a dual channel burst search of multiple FRET populations. **f**: The centers of those populations plotted S against E and fitted with Eq. 6 to determine gamma and beta.

Microscope optimisation: Raman Scatter

The original microscope optics chosen for the smfBox used a 582/75 filter (BrightLine HC, Semrock) for the donor emission cleanup. The background intensity in the $D_{ex}D_{em}$ channel was noticeably high compared to the other channels (Fig. 2.3.12), even with ultrapure water which should not exhibit any fluorescence background.

A high background level of photon counts can be detrimental to smFRET experiments, as additional photons are added to the channels used to calculate the proximity ratio. Background levels can be determined and subtracted such that the centre point of FRET efficiency distributions are not changed, however as the number of background counts within the length of a burst (~ 1 ms) is random and Poisson distributed; the width of the distribution is increased. When determining absolute FRET efficiencies this can increase the number of bursts required to determine a FRET efficiency within a given error range. When measuring multiple FRET populations increased width can make it harder to disentangle populations of similar FRET efficiencies, which is especially important for determining dynamics.

One possible source of this increased background level in the channel is Raman scatter of the 515 nm laser from water in the confocal volume. As seen in Fig. 2.3.13, approximately half of the Raman scatter peak of water from 515 nm excitation should appear in the light transmitted through the 582/75 filter. However, swapping to a 575/72 filter should reject almost all of this peak whilst only slightly reducing the detected emission from donor fluorophores. If the background in the $D_{ex}D_{em}$ channel is primarily from Raman scatter then the significantly larger reduction in transmission of the Raman peak compared to the Donor emission should result in an increase in signal to noise.

After installing a 575/72 filter, the background from a water sample is reduced significantly (Fig. 2.3.14), from 2.27 to 0.53 KHz. This large decrease in noise represents a significant improvement for the smfBox, and demonstrates a need to carefully check emission filters against expected Raman peaks for the lasers in the microscope in future, if the smfBox is to be modified for a different set of dyes.

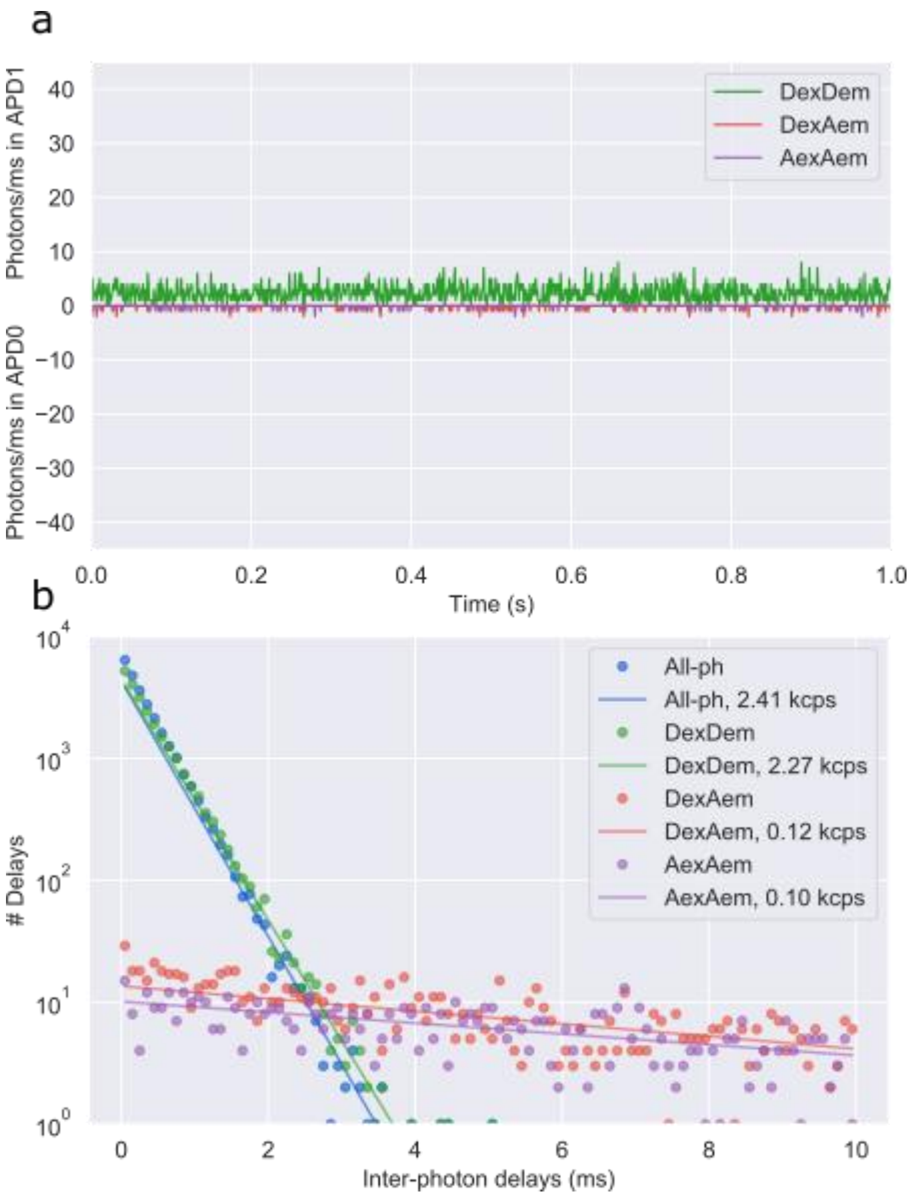


Fig 2.3.12 Background from water (photons per millisecond) using the original donor emission filter (582/75) on the smfBox. **a.** Time trace showing a much higher background in $D_{ex}D_{em}$ than in the other two channels. **b:** Background fits (Eq. 2.1.1) for each channel, the $D_{ex}D_{em}$ background rate is 2.27 kHz, whereas the background in $D_{ex}D_{em}$ and AexAem is 0.12 and 0.10 KHz.

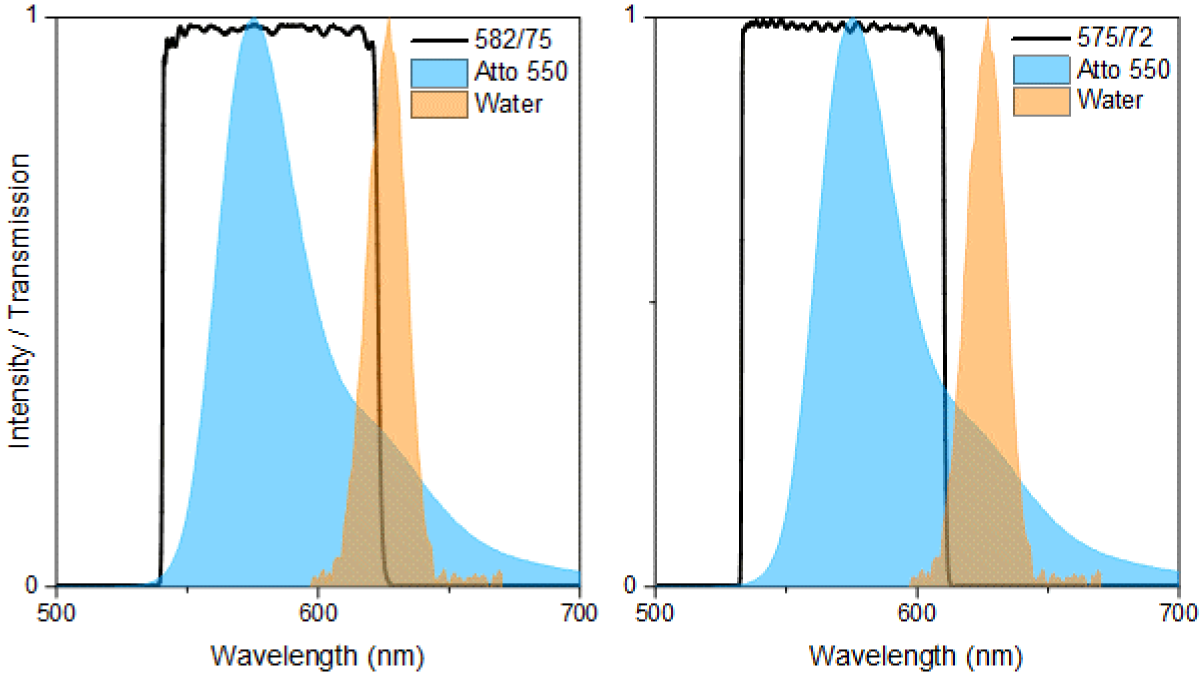


Fig. 2.3.13 Raman scatter spectrum of water (at 515 nm excitation, measured in a Fluorimeter) plotted with the emission spectrum of Atto 550, and the original donor emission filter for the smfBox, 582/75 (left). On the right is the same plot with a prospective replacement filter which would reduce the Raman scatter transmitted to the detector whilst only slightly reducing the detected emission of the donor.

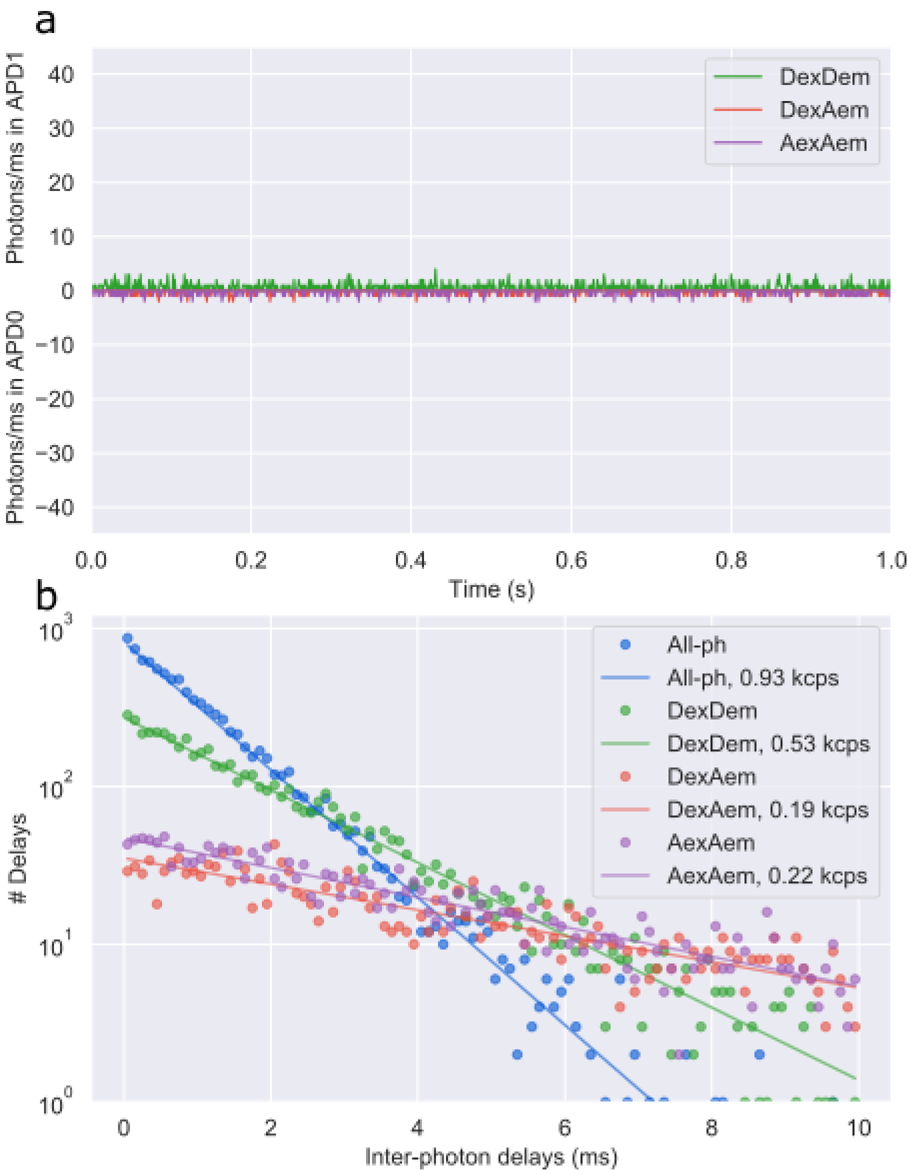


Fig. 2.3.14 Background from water using the new donor emission filter (575/72) on the smfBox. **a.** Time trace showing a much lower background in $D_{ex}D_{em}$ than before. **b:** Background fits for each channel, the $D_{ex}D_{em}$ background rate is now 0.52 KHz, less than a quarter than the 2.27 KHz with the original filter.

Laser Power and Focus optimisation

In order to get high quality and consistent data for both smFRET and FCS, it is desired that the confocal volume remains a consistent size and shape. For smFRET, the confocal volume size will affect signal to noise ratio and burst duration, whereas for FCS this is doubly important, as a consistent confocal volume is assumed when sharing parameters between the fits of different acquisition.

One way that the confocal volume can be warped is by using excessive laser power[102]. With increasingly higher irradiance of laser light, the dyes can become "saturated" spending a significant proportion of time in the excited state, which causes the relationship between excitation intensity and emission intensity to become non-linear, as there are fewer molecules left in the ground state to be excited. This effect will be dependent on the relaxation rate of the dye and it's extinction coefficient at the laser wavelength. By exciting concentrated solutions of dyes with varying laser power and recording the apparent brightness (Fig. 2.3.15), we can see that the relationship between the laser power and brightness is non-linear. However, the curve is approximately linear at low laser powers, where the incident laser light is not saturating.

Performing experiments beyond the linear region of these power response curves can increase the apparent confocal volume size, as molecules within different regions of the confocal will have their emission intensity disproportionately increased. For example; if the laser power is doubled, then molecules will be twice as bright at the edges of the confocal volume, where the flux of laser light is low enough to be non-saturating, however molecules at the centre where laser power is higher will have their brightness increased by a factor of less than two. This has the effect of "squashing" the Gaussian profile of the detection volume, and increasing its overall size.

If we fit these data at different laser powers with FCS correlation curves (Fig. 2.3.16), using the known diffusion coefficient of Rhodamine 6G, we see that the confocal volume begins to sharply increase beyond 100 μW, the same laser intensity at which the response curve in Fig. 2.3.14 becomes non-linear.

Therefore, for FCS experiments, powers for both the green and red lasers should be kept at or below 100 μW in order to prevent warping of the confocal volume. Note that this 100 μW corresponds to the saturating laser irradiance for the beam width used in these experiments (5 mm), if the beam width is changed then the confocal volume waist will be different, and so the relationship between absolute laser power and irradiance at the confocal volume will result in a different saturating laser power. For smFRET burst analysis experiments, laser intensities beyond the saturation point will increase background intensity faster than signal intensity, which will have a negative impact on signal to noise ratio. However, a not insignificant amount of noise in bursts comes from shot-noise, which arises from low signal intensity. Therefore, laser powers beyond the linear region may be desired for smFRET experiments, as the reduction in shot-noise from increasing the absolute signal will offset the noise from increased signal to background ratio.

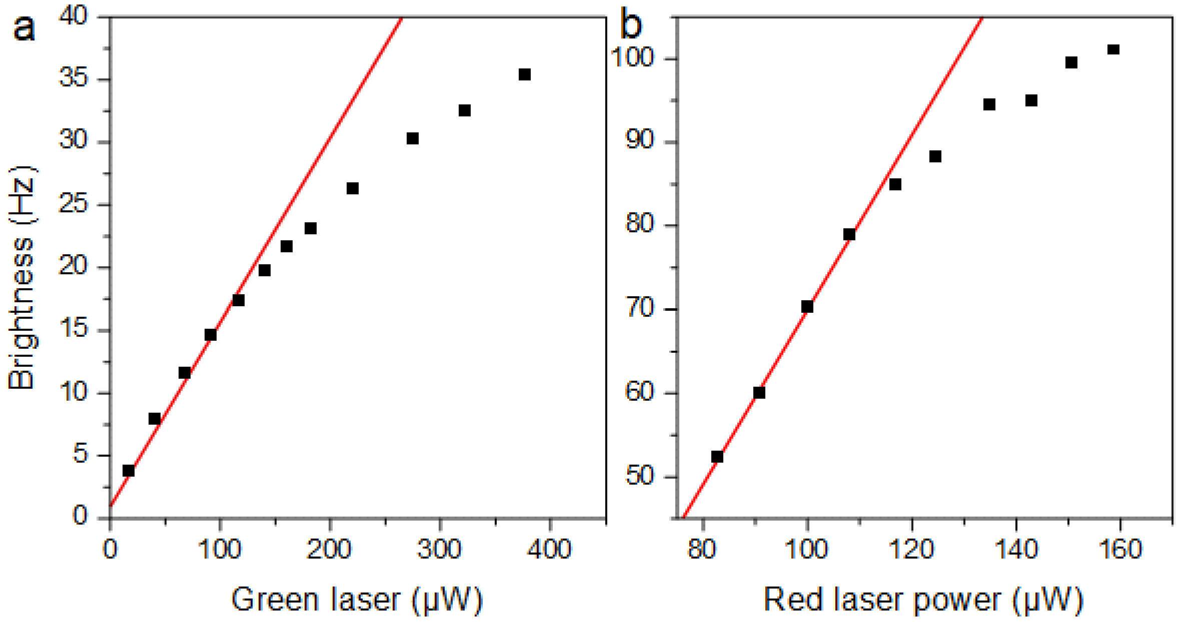


Fig. 2.3.15 Sample brightness at different laser powers. **a.** The brightness measured from a concentrated (10 nM) sample of rhodamine 6G, plotted against green laser power (measured before the objective). The response to increasing laser power becomes non-linear beyond ~100 µW. **b.** The brightness measured from a concentrated sample of Atto 647N (10 nM) plotted against red laser power (measured before the objective). The response here also becomes non-linear after ~100 µW. Red lines are eye guides and not fits to the data.

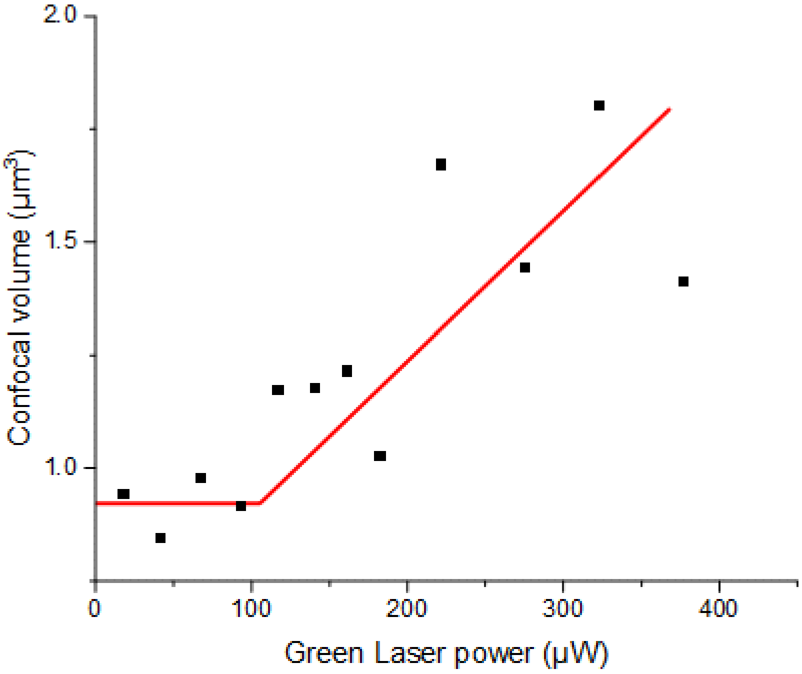


Fig. 2.3.16 Confocal volume determined from FCS on a sample of known diffusion coefficient (rhodamine 6G) plotted against laser power. As a consequence of the non-linear response shown in Fig. 2.3.15, the apparent confocal volume size increases beyond ~100 µW. Red lines are eye guides and not fits to the data.

Another source of deviation in confocal volume size and shape can come from focus depth. For freely diffusing confocal experiments, the detection volume is focused just above the glass coverslip, tens of microns into the sample solution. For the smfBox, this is done by moving the z-control of the nanostage until the back reflection on the glass-water interface shrinks to a spot, indicating that the focus is on the glass surface, then moving the z-control up above it.

It is important to keep the confocal focus at a consistent and optimised distance above the coverslip surface, as the refractive index mismatch between the solution and the glass can cause the confocal volume to stretch. By performing FCS experiments on a dye of known diffusion coefficient (Rhodamine 6G), and varying the focus depth (Fig. 2.3.17), we can see that the confocal volume remains relatively constant between 10-30 μm, but then stretches dramatically beyond this. Thus, to keep the confocal volume small, and of consistent shape, focus depth should be kept at or around 20 μm, with extra care taken for this in FCS experiments in particular.

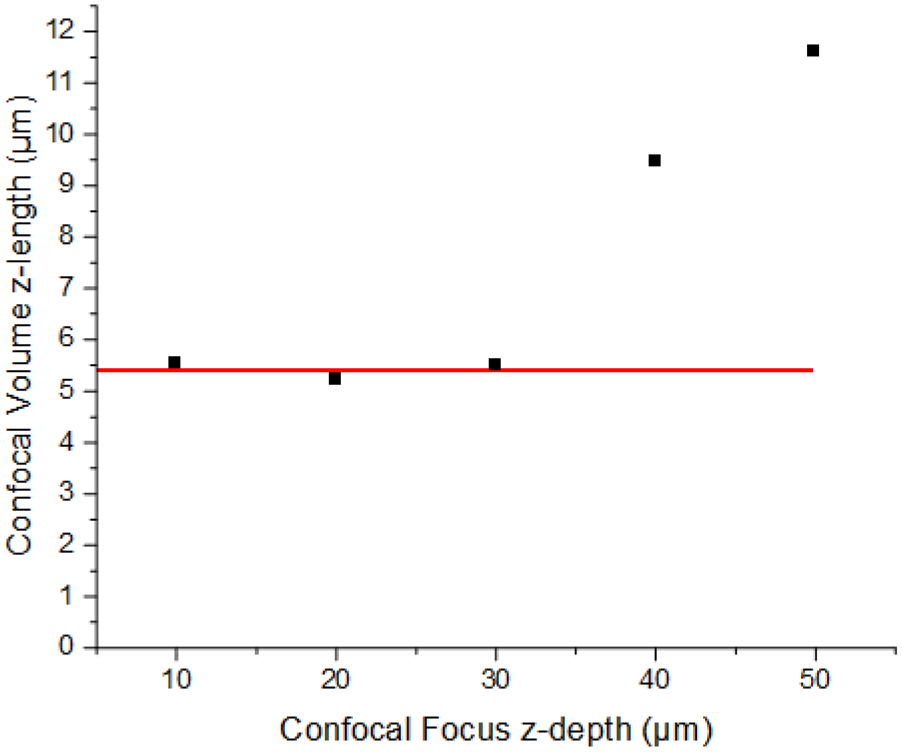


Fig. 2.3.17 Confocal volume length along the z-axis determined from FCS on a sample of known diffusion coefficient (Rhodamine 6G) at different focus depths into the sample. The confocal volume remains an approximately constant length from 10-30 μm, however beyond 30 μm the confocal volume becomes stretched due to the refractive index mismatch between the glass coverslip and the sample. The Red line is an eye guide and not a fit to the data.

Asymmetric ALEX

The rate and duty cycle of the donor and acceptor lasers of the smfBox can be completely customised using either of the provided acquisition software. It is even possible to leave one laser on all the time (a technique known as periodic acceptor excitation or PAX[55]) or to leave one laser permanently off if ALEX is not required. For the experiments in this work we chose a duty cycle length of 100 µs, as this is fast enough to allow several observations with either laser per molecule (observation time of ~1 ms). If there are too few alternations then there is a reduced ability of algorithms like the dual channel burst search (DCBS) to discriminate between molecules which have bleached during observation. However, the lasers have a rise/fall time of several microseconds (See Fig. 2.3.18), so repeating too fast would decrease the total time the lasers are on at full power. Furthermore, in either the case of converting to HDF5 from the LABVIEW software, or in directly saving to HDF5 in smOTTER, the details of the ALEX cycle are contained within the metadata, which is then read out by FRETBursts[16], so analysis is not further complicated by changing any of these parameters.

Here I demonstrate the capabilities and effects of altering the ALEX cycle with a number of asymmetric excitation schemes, in which the amount of time either laser is on for is not equal. Typically ALEX schemes are symmetrical, however there is not strictly any reason why the acceptor laser need be on for the same amount of time as the donor laser. In fact, since all the information for the FRET efficiency comes under donor excitation, it would intuitively make sense to maximise the time this laser is on if greater precision in FRET efficiency is desired. The four datasets shown (Fig. 2.3.18) were taken using duplex 1b as with the accurate FRET measurements in this chapter. A dual channel burst search (F = 15, m = 10) was used for analysis, and a burst selection of 50 photons under donor excitation and 50 photons under acceptor excitation.

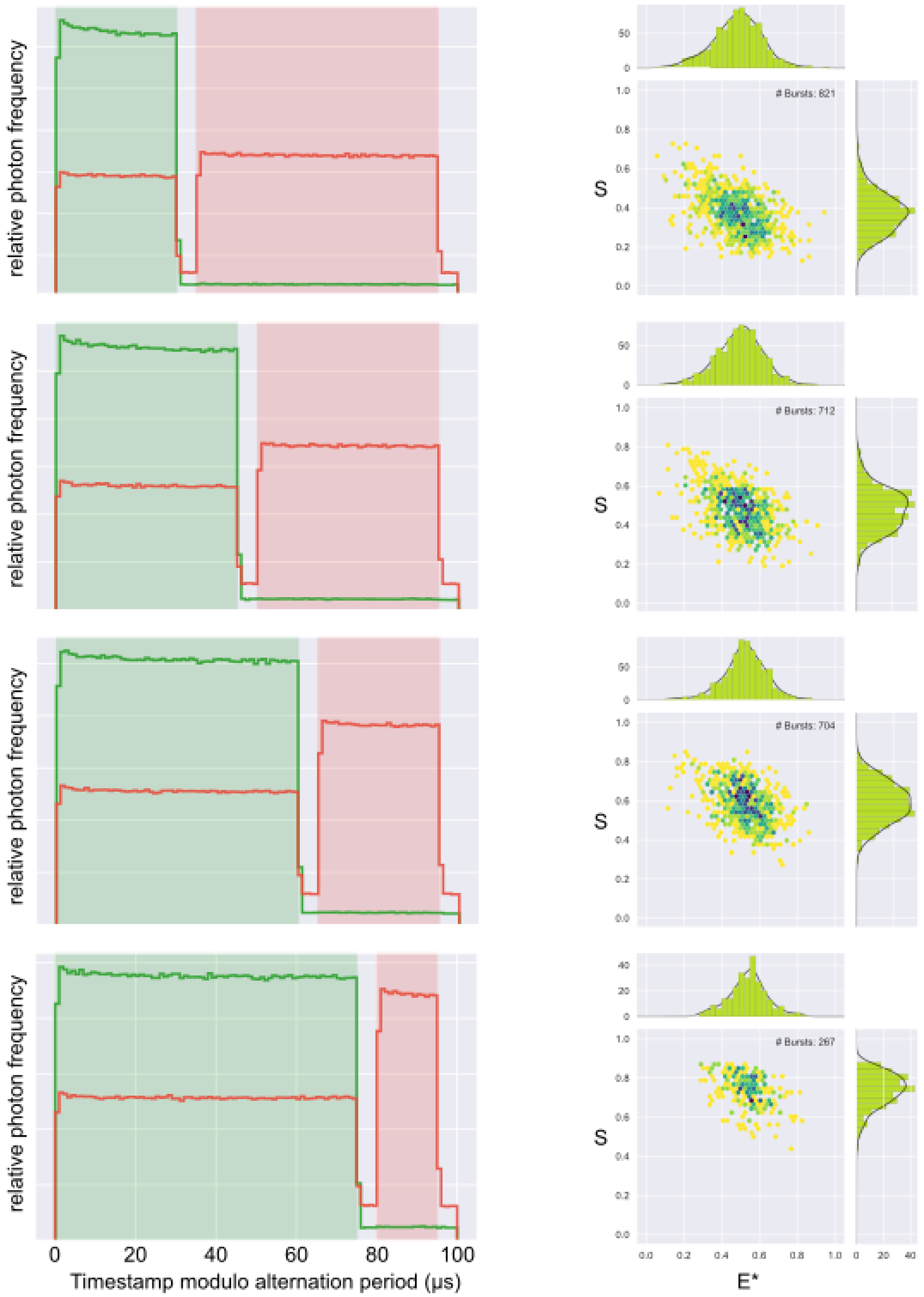


Fig. 2.3.18 Four different ALEX schemes. On the left are photon arrival time plots with detected photons as green and red lines and shaded boxes representing which laser was on. On the right are uncorrected ES Histograms of DNA standard 1b measured with these ALEX schemes. From top to bottom the green/red excitation times are 30/60 µs, 45/45 µs, 60/30 µs, and 75/15 µs. In all cases a 5 µs wait time was left between each laser.

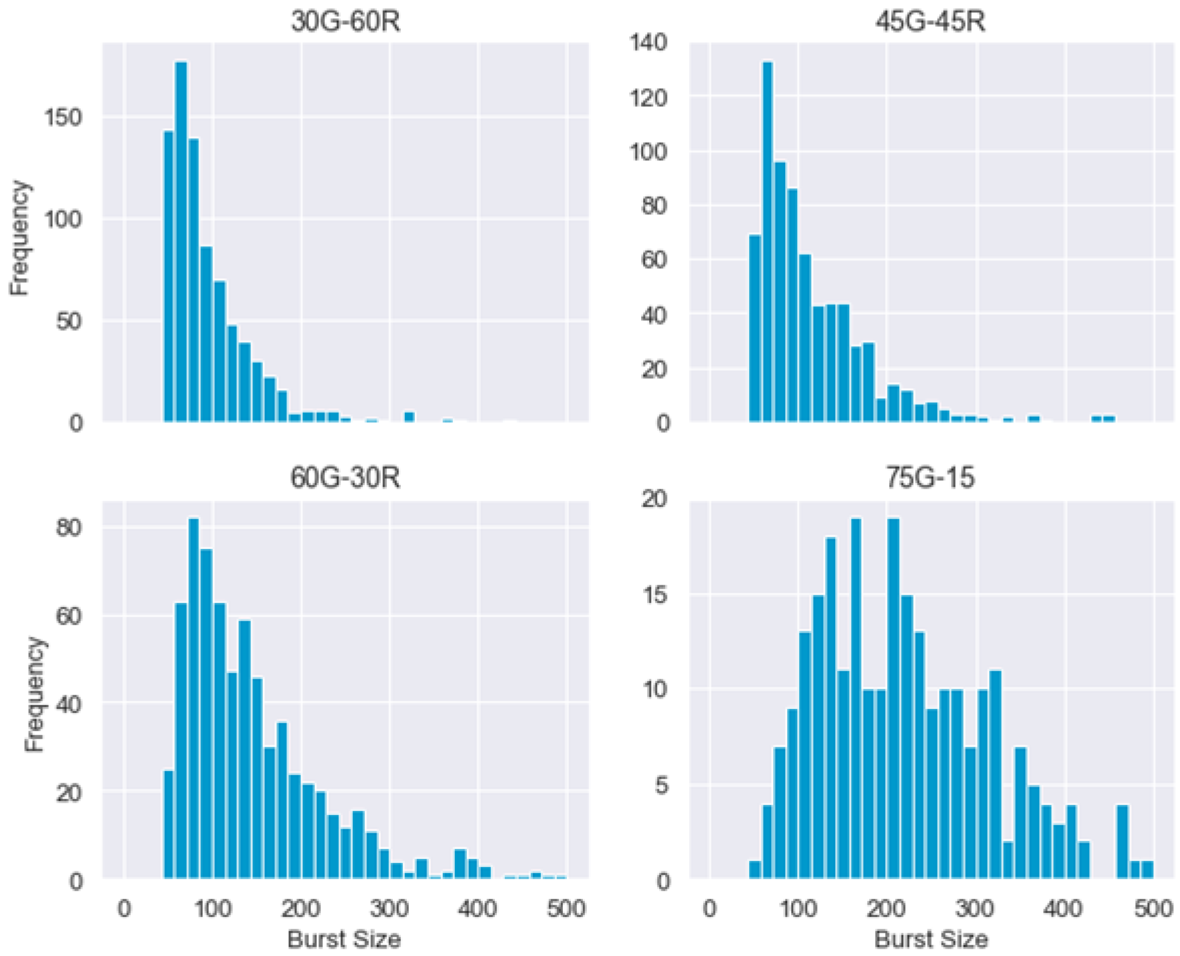


Fig. 2.3.19 Burst size histograms for each excitation scheme. Plotted are the number of photons under green excitation after the selection of >50 photons, which is why there are no bursts <50 photons shown.

Changing the relative amounts of excitation by either laser shows a predictable trend (Fig. 2.3.18). The S* position in each ES plot increases with the length of donor excitation, due to the increased number of photons detected under donor excitation. However, two other effects become apparent. Firstly, the width of the distribution in E* decreases due to the decreased contribution of shot-noise from the increased number of photons acquired. Secondly, the number of doubly labelled bursts which are found by the burst search algorithm decreases due to the reduced information under acceptor excitation, which is required to confidently assert that any burst is doubly labelled. From these two effects it becomes apparent there may be an optimum excitation scheme, which may not necessarily be symmetrical, since the complexity of information inferred under donor excitation is greater than that obtained under acceptor excitation (two colours versus one). As expected, we can see that the burst size histogram shifts to an increasing number of photons under donor excitation when the donor excitation period is increased (Fig. 2.3.19), consistent with the reduction in width in FRET efficiency histogram due to decreased shot noise.

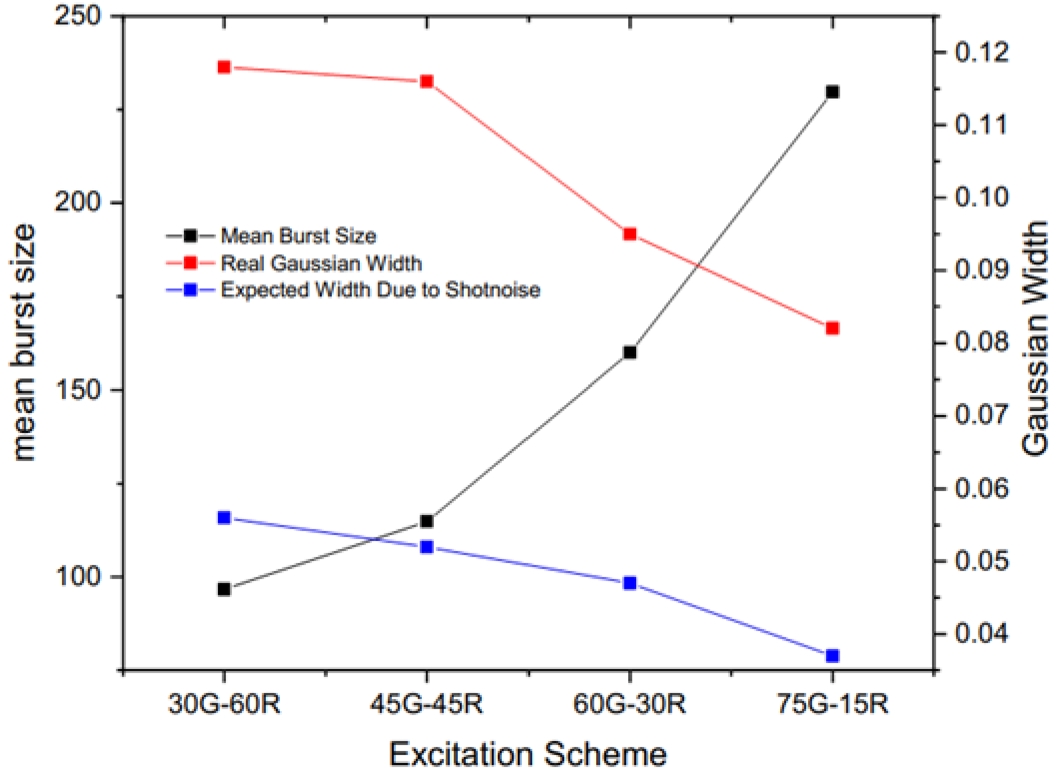


Fig. 2.3.20 Burst statistics under each excitation scheme, showing: the mean number of photons under green excitation (black); the width of the Gaussian fitted to the 1D FRET efficiency distribution (red); and the expected Gaussian width computed from the burst size distribution (blue) according to a previously published algorithm[66]. N=1

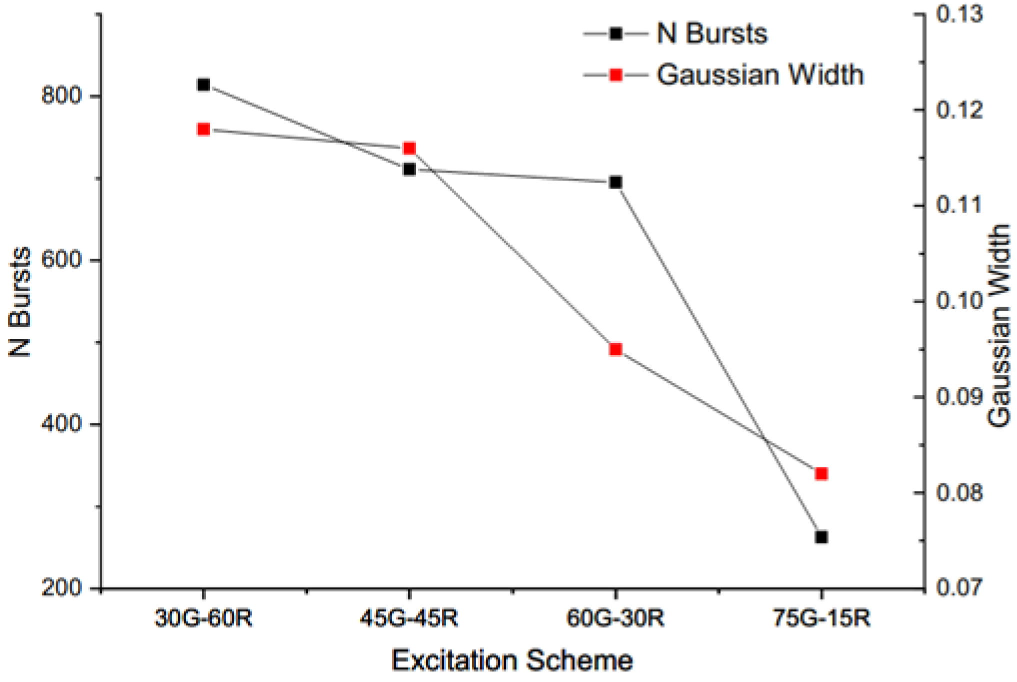


Fig 2.3.21 Four different excitation schemes with the number of bursts found by a DCBS and the width of the Gaussian distributions fitted to the FRET efficiencies. N=1

To further emphasise this point, we see that Gaussian width decreases with increasing burst size, as can be predicted by a binomial computation of shot noise from the burst size distribution[66] (Fig. 2.3.20). Note that the width of the Gaussian distribution fitted to the experimental data is larger than predicted from shot noise alone, as shot noise is not the only form of noise present, but represents a considerable proportion nonetheless. Other forms of noise which make up for the discrepancy are background contribution, and flexing of the DNA molecule, which has been previously discussed in detail[82,83,83]. Finally in Fig. 2.3.21 we can see the effects of each excitation scheme on both Gaussian width and number of detected bursts. Whilst more testing may be necessary to fully explore this relationship, it can be seen phenomenologically that increasing donor excitation to twice that of the acceptor can markedly decrease the width of the obtained FRET distribution without having to resort to increasing laser power (and hence increasing photobleaching), whilst having little to no effect on the ability to confidently assert that a burst is doubly labelled. From this plot we would recommend as an immediate improvement over symmetric ALEX a 'sweet-spot' of 60G-30R for the duty cycle, to both maximise the number of detected bursts and minimise the shot noise. The exact sweet spot may vary depending on application however; if trying to detect a relatively small shift in FRET efficiency then tighter histograms at the expense of a few less bursts as shown may be desirable, however if quantifying two populations which are already well separated in FRET efficiency, then tighter histograms are less of an advantage.

Accurate FRET Validation

In order to validate the accurate FRET technique, DNA standards labelled with dyes at varying differences were distributed to smFRET laboratories worldwide. The labelling positions of each standard were not known to experiment participants prior to the measurements, and so FRET efficiencies were to be independently measured and reported for comparison. Here I present accurate FRET data on three of the standards in the study, 1a, 1b, and 1c, labelled with Atto 550 and Atto 647N to give low, mid, and high FRET efficiencies.

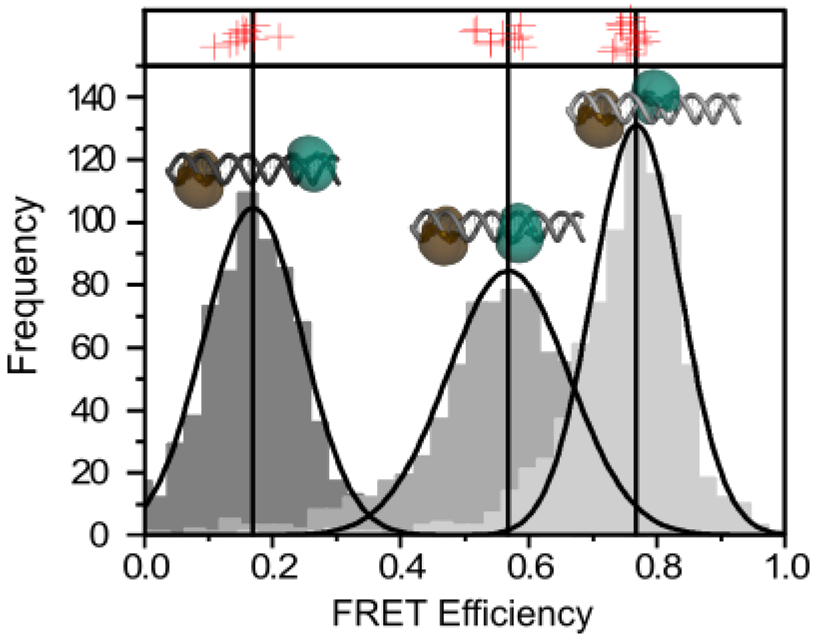


Fig. 2.3.22 Fully corrected FRET efficiency histograms of three doubly labelled DNA standards (1a, 1b and 1c, cartoons with dye accessible volumes; for sequences see Table 2.2.1) measured using the smfBox (grey). Vertical black lines and curves show Gaussian fits of our data, E = 0.17 ± 0.07, E = 0.57 ± 0.1, E = 0.77 ± 0.07 (mean ± sd), compared to the results from 20 other labs as part of a multi-lab benchmarking study

Each standard was measured on the smfBox, and data were analysed using FRETBursts Jupyter Notebooks as described in this chapter. The accurate FRET efficiencies determined fall well within the cluster of measurements reported by other labs (Fig. 2.3.22, Table 2.3.5). This has the dual purpose of confirming the validity of the accurate FRET method as part of the multi-lab study, and also validating the capability of the smfBox to recover accurate FRET efficiencies at least as well as commercially available microscope platforms in used by other smFRET groups.

In terms of the results of all participants in the study, the standard deviations for all smFRET standards were between 2-5% , and the mean FRET efficiency deviated by 0.1-4% from expected values predicted from a static DNA model. This would imply an overall accuracy of somewhere between 0.4 and 2.4 Å for intramolecular distances determined from accurate FRET.

Table 2.3.5 Summary statistics for the accurate FRET validation. The means and standard deviations of FRET efficiencies from all participants in the benchmarking study[5] are shown, alongside the FRET efficiency mean (E) and width (**σ**) of a Gaussian fit to data from the smfBox, and the standard error of the mean (σ/√n).

	Benchmarking study	**smfBox Data**			
Sample	**E**	**E**	**σ**	**N bursts**	$\boldsymbol{\sigma/\sqrt{n}}$
1a	**0.15 ±0.02**	**0.17**	**0.07**	**694**	**0.003**
1b	**0.56 ±0.03**	**0.57**	**0.10**	**717**	**0.004**
1c	**0.76 ±0.015**	**0.77**	**0.07**	**799**	**0.002**

Measuring Dynamics

To validate capabilities of the smfBox to investigate dynamics information, as well as distance measurements, a number of smFRET experiments were carried out using a DNA hairpin. A hairpin is a DNA construct which anneals to itself, with a long flexible region linking the two complementary sequences. DNA hairpins convert between two states, a closed state where the complementary regions on the hairpin are annealed to one another, and an open state where they are not. The rates of interconversion between these two states are dependent on the length of the complementary sequence, as well as temperature, pH, and cation content of the solvent. As such, hairpin structures can be tuned by sequence to change on and off rates, and also by solvent properties, making them a useful test system for measuring rates of conformational change.

The DNA hairpin used here is a replica of the one used in a previous smFRET study on DNA hairpin dynamics[103] using the same sequence and dyes (Cy3B, Atto 647N) and labelling positions, as well as two other labelling positions. In all three constructs, the dyes are distant when the hairpin is open, resulting in a low FRET efficiency, and when the hairpin is closed, the dyes will be close, giving a higher FRET efficiency. The three constructs differ by the position of the acceptor on the immobile "stem" of the hairpin, giving different FRET efficiencies in the closed state.

The hairpin data exhibits two clear FRET efficiency peaks, with a smaller intermediate population bridging the two (Fig. 2.3.23, Fig. 2.3.24). Dynamic probability distribution analysis (dPDA[83]) can be used to fit these three populations, and extract forward and backward kinetic rates. dPDA works on the assumption that the intermediate FRET population arises from molecules interconverting between the other two states during the observation period, but does not check if this assumption is true in of itself, so it is prudent to first test if this assumption is valid. In this specific case, it can be safely assumed that the intermediate FRET population is a dynamic species, as it is known that this hairpin interconverts on the millisecond timescale of our measurements. However, as we are demonstrating the capabilities of the smfBox rather than performing novel research on hairpins, the initial test for dynamics will be carried out anyway.

The FRET-2CDE[86] filter can be used as a hypothesis test to check for the presence of dynamics in the smFRET data. The 2CDE filter runs and compares kernel density estimators (KDE) at the photon timestamp level in both channels to score bursts for the likelihood that the FRET efficiency of the molecule has changed during the observation. A "static" molecule which remains the same FRET efficiency throughout the burst will be expected to have a 2CDE score of 10, and so bursts with a 2CDE score significantly higher than 10 can be flagged as "dynamic".

First, a mixed sample containing two *different, static* dsDNA standards is used as a negative control for the 2CDE test (Fig. 2CDE). As the heterogeneity in this sample does not arise from transitions between states, none of the bursts exhibit large enough fluctuations in FRET for there to be any significant 2CDE scores above 10. For the hairpin however, the intermediate population causes a very obvious "upside down smile" shape, giving higher 2CDE values in the middle, and lower 2CDE values at the edges where bursts do not transition during the measurement.

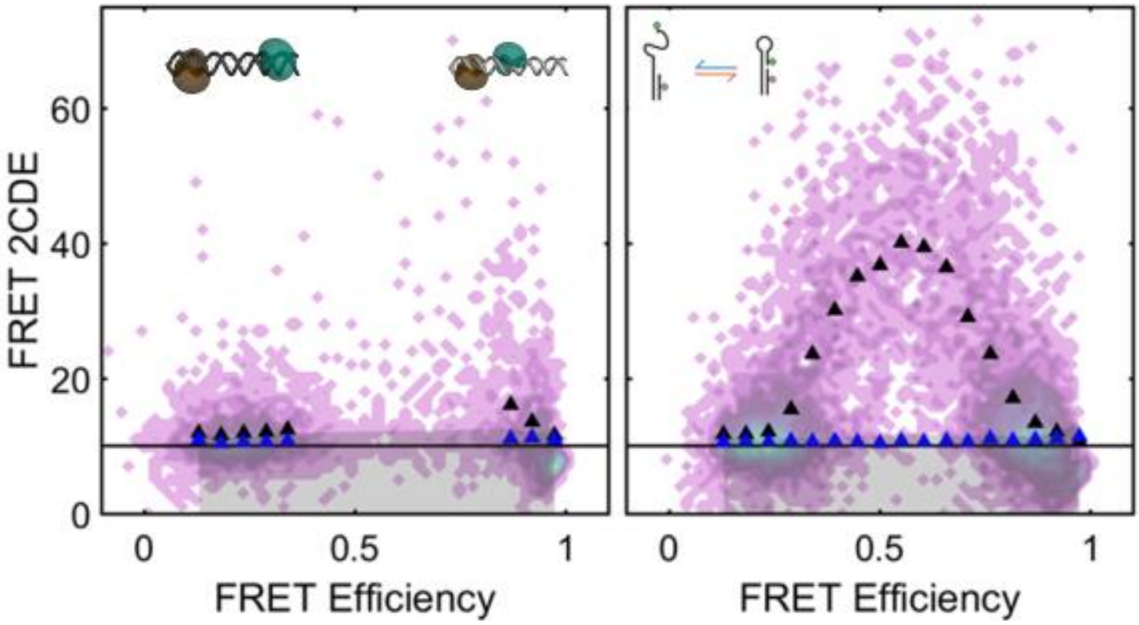


Fig. 2.3.23 2CDE plots of a static dsDNA mixture (left) and a dynamic hairpin (right). FRET-2CDE analysis[28] can be done in PAM on data from the smfBox in order to verify the presence of dynamics in the sample. A high 2CDE value (average shown in black triangles) suggests that bursts are dynamic, ie. Interconverting between FRET efficiencies during observation, whereas low 2CDE bursts along a straight line (blue triangles) suggest a static FRET efficiency.

Next, having confirmed that the intermediate population in the hairpin smFRET data is dynamic in nature, the three FRET populations can be fit via dPDA to yield forward and backward rates of transition. This is first done for a range of salt concentrations (Fig. 2.3.24). Cations, like Na^+, can screen the electrostatic repulsion of phosphate groups in DNA, increasing flexibility and stabilising duplex structure. Thus, higher sodium chloride concentrations will alter both the forward and reverse rates for the hairpin, by lowering the energy barrier to bend the linker region into a pin, and increasing the energy barrier to melting of the annealed region.

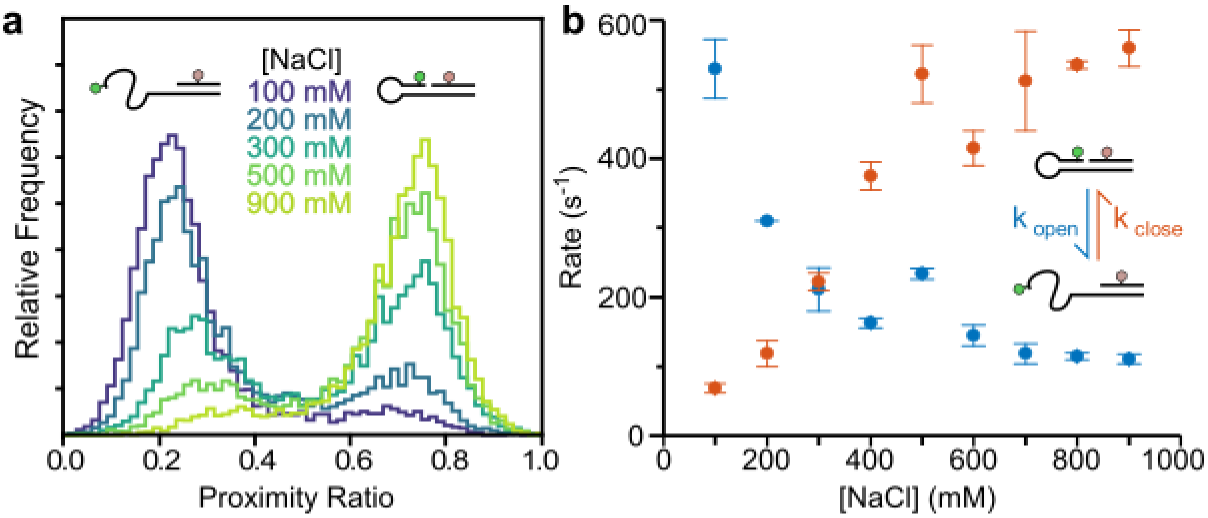


Fig. 2.3.24 dPDA data of hairpins at different salt concentrations. **a:** Proximity ratio (uncorrected FRET efficiency) histograms of a DNA hairpin at indicated salt concentrations **b:** Salt dependent rates for hairpin opening (k_{open}) and closing (k_{close}) determined by dynamic photon distribution analysis (dPDA) (mean ± SD, $n = 2$ with >1000 molecules per technical repeat at each [NaCl])

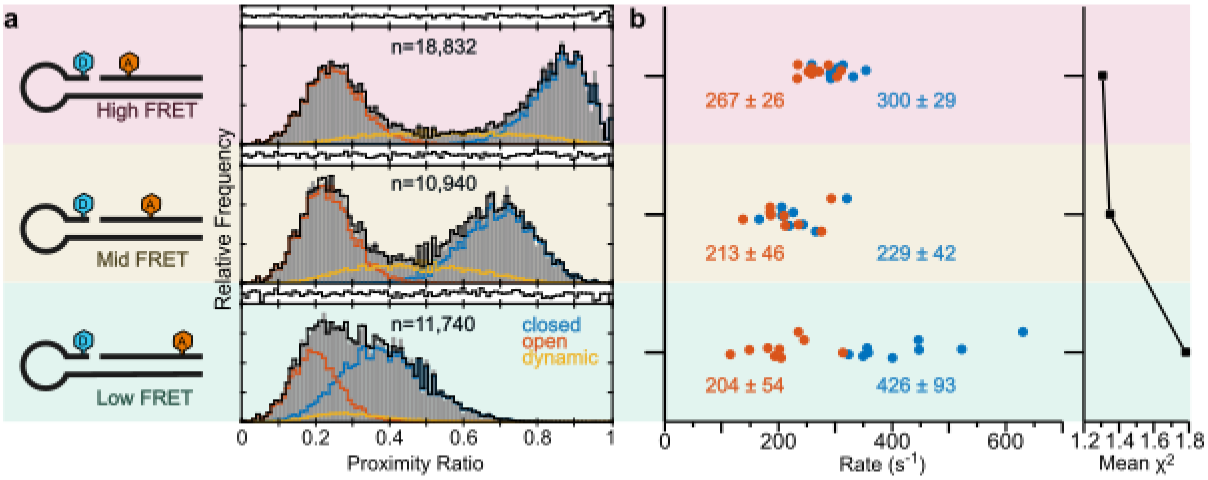


Fig. 2.3.25 dPDA data of hairpins with different FRET efficiency changes. **a:** Proximity ratio histograms of High-, Mid- and Low-FRET hairpins (at 300 mM NaCl). Data (grey) were fit using dPDA (black) to a two-state model, comprising a closed population (blue), open population (orange) and interconverting dynamic population (yellow). **c:** Plot of rates determined from dPDA of nine data sets for each hairpin, each containing 2000 molecules, quoting the mean and standard deviation across the data sets, with the mean chi-squared of the fits plotted to the right

With increased NaCl concentration we can see that the low FRET population falls and the high FRET population rises (Fig. 2.3.24), caused by the DNA hairpin shifting to the closed state. By using dPDA to fit the data we find that the forward rate into the closed hairpin state increases, and the corresponding backward rate decreases. This trend matches observations made of the same hairpin in a previous study, using TIRFM based smFRET[103]. The exact kinetic rates obtained here do not perfectly match those determined in the original study; here the opening and closing rates plateau at approximately 550 and 100 s^{-1} respectively, whereas the original study found rates of 200 and 50 s^{-1} at higher salt concentrations. This could be for any number of reasons, for example, the slow frame rate used in TIRFM may bias the measurements to low rates of transition. The overall shape of the curves is consistent however, and plateau at the same salt concentration (~600 mM)

In order to test the precision and limits of the dPDA method further, the same hairpin sequence was labelled at two additional positions, yielding a higher and lower FRET efficiency in the closed state. As dPDA works by fitting to the intermediate FRET population between the two states, one would expect that the closer those two FRET efficiencies are, the harder it will be for dPDA to recover kinetic rates. Likewise, with a larger change in FRET efficiency, the rates may be recovered with greater precision, as the dynamic intermediate population would be less obscured by the breadth of the two static populations.

As evidenced in Fig. 2.3.25 the rates recovered from the high FRET hairpin have a much lower standard deviation (26 and 29 s^{-1}) than those from the original mid FRET hairpin (46 and 42 s^{-1}). The low FRET hairpin on the other hand, gives a much larger spread of rates (54 and 93 s^{-1}), and the mean rate of closing is significantly higher, suggesting a possible reduction in accuracy as well as precision. The slight discrepancy in mean rates recovered at different FRET efficiencies could be due to the dye itself altering the dynamics of the hairpin, although it is not clear how. Alternatively, this may be caused by the dPDA method itself having slight biases depending on the degree of overlap between the two states.

In either case, this data represents a good rule of thumb going forward for experimental design, confirming the hypothesis about the relationship between FRET change and precision in rates. When selecting dye positions for dynamics analysis, the two states should ideally have very low and very high FRET efficiencies. This design constraint is less compatible with the methodology of an accurate FRET experiment, where FRET efficiencies can be converted to distances with most precision when close to the R_0, yielding 50% FRET efficiency.

Fluorescence Correlation spectroscopy

Whilst not originally designed to do so, the smfBox as presented here is capable of fluorescence correlation spectroscopy (FCS) and fluorescence cross correlation spectroscopy (FCCS) with no hardware modifications. Rather than identifying individual molecules and performing calculation on their photon counts (as with smFRET), FCS works by measuring correlations and fluctuations in signal originating from dyes diffusing in and out of the confocal volume[104]. As such, FCS can work at much higher concentrations than smFRET, and logistically speaking the sample concentration is the only key difference between an smFRET measurement and an FCS measurement.

To validate the capability of the smfBox to perform FCS, example data is presented (Fig 2.3.25) using a solution of free dye of known diffusion coefficient (Rhodamine 6G) to determine confocal volume parameters, and a 37bp doubly labelled duplex DNA which is determined to have a diffusion coefficient of 88 $\mu m^2 s^{-1}$.

The diffusion coefficient of 88 $\mu m^2 s^{-1}$ is a good match to what would be expected of a DNA oligomer of this length[105].

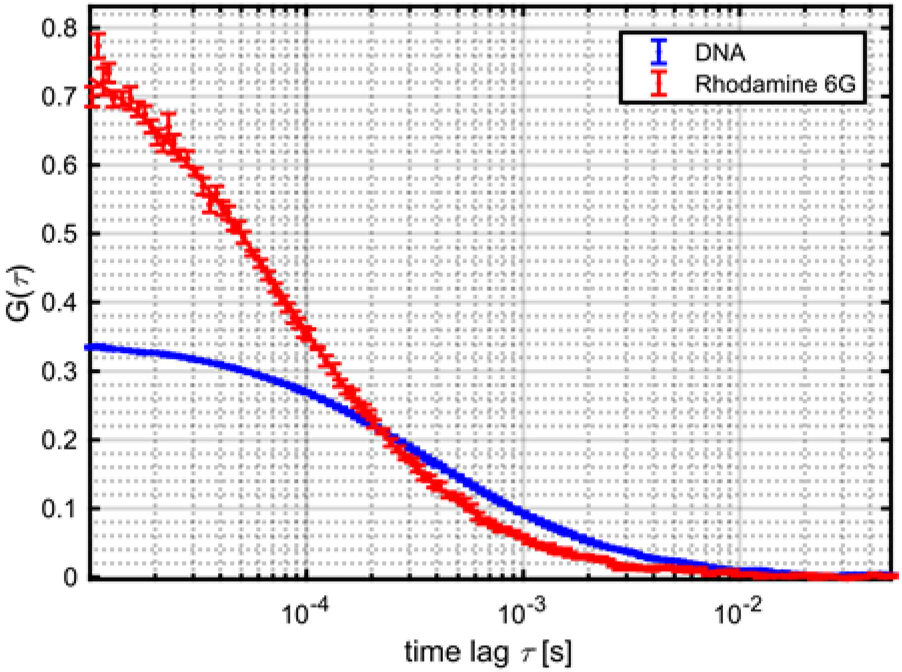


Fig. 2.3.26 Correlation curves for Rhodamine 6G (red) and a 37 bp doubly labelled DNA (blue). See Table 2.3.6 for fit parameters.

Table 2.3.6 FCS curve fit parameters. N: average number of molecules in the confocal volume D: diffusion coefficient (fixed for Rhodamine 6G, measured for DNA) W_0, Z_0: lateral and axial dimensions of the confocal volume, respectively, calculated by fitting the R6G data assuming a 3D Gaussian profile.

Sample	**Counts (KHz)**	**Brightness (KHz)**	**N**	**D($\mu m^2 s^{-1}$)**	**W_0(μm)**	**Z_0(μm)**
R6G	**58**	**135**	**0.43**	**414**	**0.36**	**4.31**
DNA	**205**	**199**	**1.03**	**88**	**0.36**	**4.31**

2.4 Discussion

As part of a wider collaboration with other smFRET laboratories[14] I have demonstrated the replicability of a standardised accurate FRET technique, capable of recovering absolute FRET efficiencies corresponding to an intramolecular distance accuracy of 0.4-2.5 Å. This work has not only validated the accurate FRET method itself, but demonstrated that the smfBox, an open-source and cost effective confocal microscope, can recover such FRET efficiencies just as well as commercially available and custom built microscopes alike used in other labs. Whilst this work only pertains to intramolecular measurements made on relatively static DNA molecules, further work is now underway to repeat the validation of this technique in a protein system; the maltose binding protein MalE, which interconverts between an open and closed state[106]. The dynamic nature of this protein system will allow for determining the lab-to-lab reproducibility of kinetics measurements as well as testing the accurate FRET method outside of nucleic acids.

In addition to accurate FRET, the smfBox is capable of dynamics analysis, both identifying dynamic species and recovering kinetic rates of conformational transition on the order of 1000 s^{-1}. With no hardware modifications required, the smfBox can also perform FCS and other correlation spectroscopy techniques, recovering diffusion parameters and molecular brightnesses. Furthermore, I have demonstrated the possible advantages of an asymmetric alternating excitation scheme (aALEX), acquiring more signal photons for precise FRET efficiency calculation without significant loss of discriminatory power for filtering molecules by dye stoichiometry.

Further extensions to the smfBox may entail the addition of pulsed lasers, allowing pulsed interleaved excitation (PIE[43]) experiments, as an alternative to ALEX. PIE would allow for measurement of fluorescence lifetimes at the single-molecule level, and the addition of two more detectors with polarising beamspitters would enable multiparameter fluorescence detection (MFD[107]), accessing anisotropic decays as well. Lifetimes and anisotropies would allow for more rigorous analysis of collected data, controlling against spectroscopic changes not arising from intramolecular distance changes, and providing extra ways to detect dynamics.

Leaving aside further additions to the hardware of the smfBox, even in its current state it has a strong potential for investigating biophysical systems and developing new fluorescence tools for structural analysis, as will be demonstrated in the following chapters.

Chapter 3. A contact quenching method for measuring interactions below 3 nanometres

Chapter Overview

In this chapter I outline the problems involved with attempting to measure shorter intramolecular distances with smFRET, and some of the possible approaches to do so. I then present work developing one of these approaches, which exploits a contact induced quenching phenomenon seen with the dyes Cy3B and Atto 647N. I use confocal smFRET with the smfBox, accessible volume (AV) modelling, ensemble fluorescence spectroscopy, and time-correlated single-photon counting (TCSPC) to elucidate the origins of the contact quenching effect and explore ways to measure it, with the aim of developing an additional technique to add to the toolkit of single-molecule fluorescence spectroscopy.

Contributions to work in this chapter

Under my supervision, Matthew Willmott and Callum Johnston each helped synthesize, and collect preliminary data, on the DNA quenching constructs in this chapter. This work formed the basis of the smFRET experiments shown here.

The fluorescence spectroscopy data in this chapter was collected with Callum Johnston using a Horiba Duetta, which was on temporary loan from Horiba Scientific.

The TCSPC data shown here was collected using a custom spectrometer built by Anna van den Boom, who also aided in the data analysis.

The preliminary molecular dynamics simulations shown at the end of this chapter were run by Tristan-Johnston Wood and Robert A. Shaw

3.1 Introduction

Measuring Shorter Distances

As shown in previous chapters, smFRET can be used to accurately and reliably measure intramolecular distance changes within protein and nucleotide systems under a variety of conditions, acting as a powerful tool for investigating biological mechanisms especially those that involve molecular motion. A range of upper and lower limits for the sensitivity of these methods are often quoted, 1-10 nm [18], 2-8 nm[19], 3-10 nm[20], and 2-10 nm[21], however general consensus is that sensitivity falls off rapidly as you deviate from the 3-8 nm range where the R_0 of most FRET dye pairs falls. This is because the sigmoidal relationship between inter-dye distance and FRET efficiency (Fig. 3.1.1) causes large changes in distance to correspond to ever smaller changes in FRET efficiency as you go above or below the R_0, which are thus more difficult to measure with sufficient precision to distinguish.

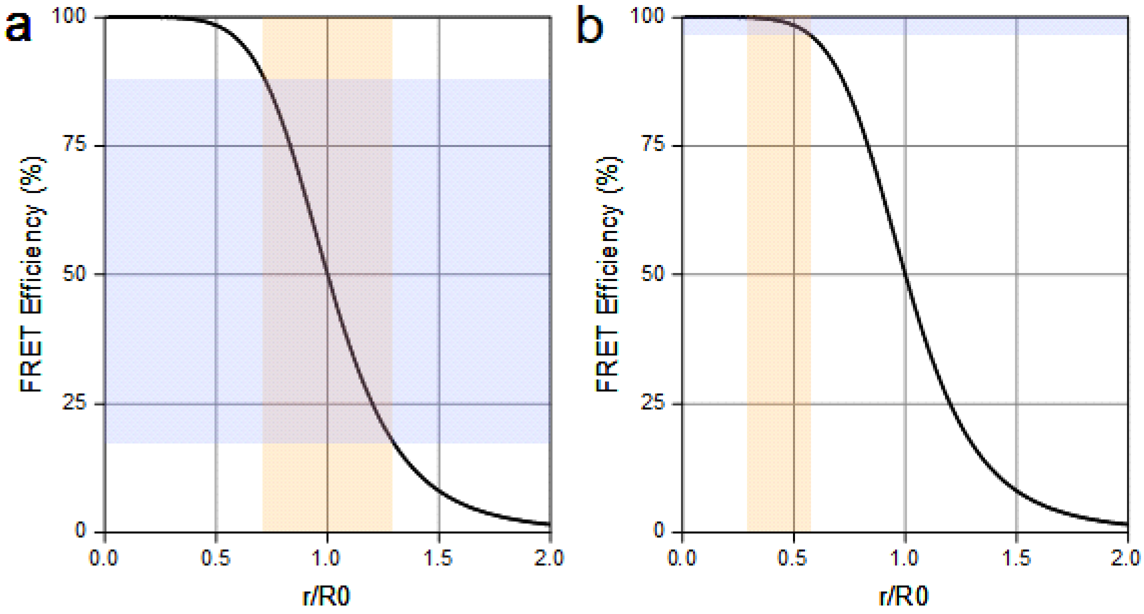


Fig. 3.1.1 FRET efficiency vs inter-dye distance as a function of the R_0. **a**: A relatively small region of distances around the R_0 (orange) corresponds to a large range of FRET efficiencies (blue). In this "linear" region, FRET is most sensitive. **b:** Much below the R_0, large changes in distance result in indistinguishably high FRET efficiency.

Many biomolecular structural rearrangements occur between subunits much closer than this 3 nm limit, and a common strategy for measuring such changes is to place dyes on more distant parts of the molecule so as to keep them within the sensitive range of the dye pair. However, many important molecules of interest, such as enzyme drug targets, are themselves too small for even this strategy to work. For example, HIV-1 protease, a drug target which has seen much interest in development of HIV therapies, has flaps (Fig. 3.1.2) which lift up and over a drug binding site[108,109]. These flaps move only a few Ångstroms apart, and the entire enzyme is only 3 nanometres across along the vector of flap movement, so the potential of standard smFRET approaches to measure drug binding in this system would be of very limited sensitivity. In such systems, an alternative approach which goes below typical limits of FRET measurements would be of great use.

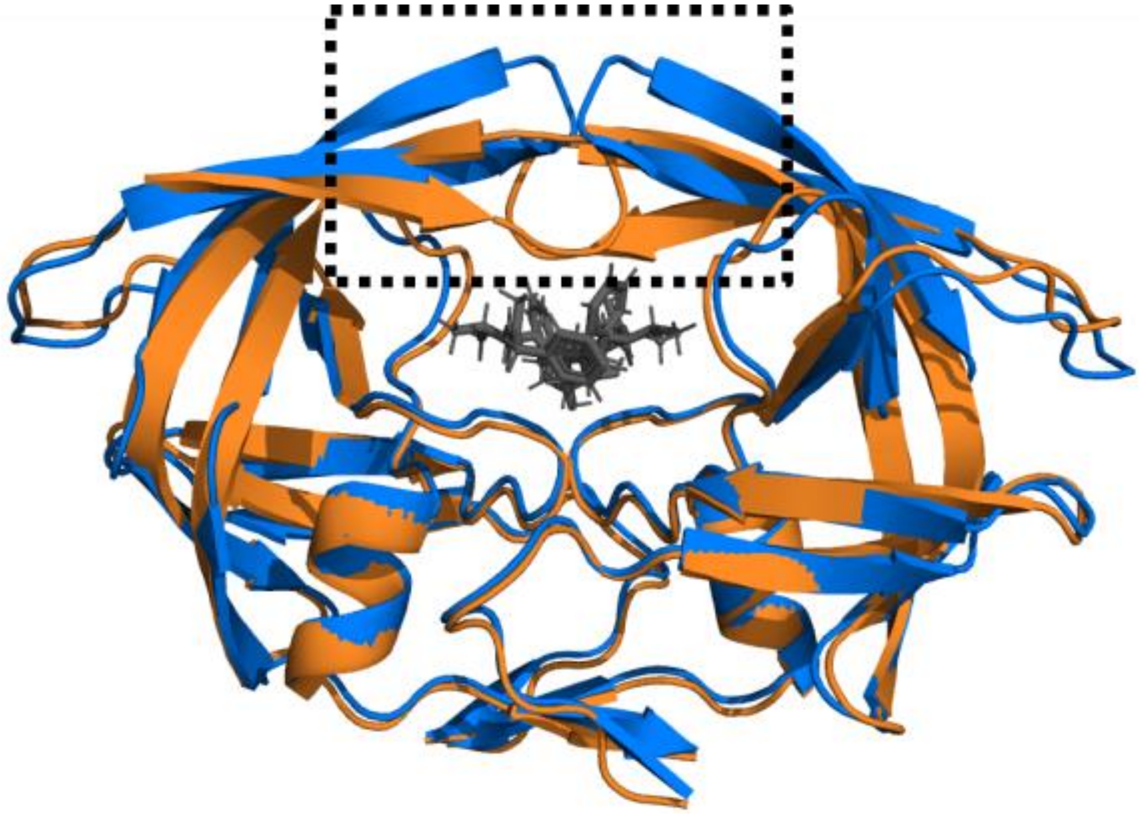


Fig. 3.1.2 HIV-1 protease in the open conformation[108] (PDB: 1HHP) shown in blue, and the closed conformation[109] (PDB: 1HVI) shown in orange, locked in place by a synthetic inhibitor (Abbott 77003), shown in black. Note the flap movement (dashed lines) over the inhibitor upon binding.

Furthermore, the recent use of smFRET in structural biology of using networks of accurately measured intramolecular distances to restrain structural models[32,89,110–113], could benefit from pushing the limits of which distances can be precisely measured, as this would reduce the amount of prior knowledge of structure required when choosing dye placement. A means by which to make sense of extremely high FRET efficiencies would potentially allow more structural information to be gained from the same number of labelling coordinates.

As effective distance range is intrinsically linked to R_0, one strategy to measure higher or lower distances would be to alter the R_0, however there are “soft” limits on R_0’s that are feasible for FRET measurements. The available variables to alter in order to change R_0 are; donor quantum yield, acceptor extinction coefficient, spectral overlap, kappa squared and inter-dye refractive index (Eq. 3.1.1). Increasing refractive index will decrease R_0, whereas R_0 will increase with donor quantum yield, acceptor extinction coefficient. Kappa squared and refractive index are typically not deliberately altered, as free dye rotation allows the useful assumption of kappa squared = ⅔, and refractive index is difficult to change without seriously altering the molecule of interest, as it would involve extreme changes in solvent properties.

$$\frac{R_0}{\mathring{A}} = 0.2108\sqrt[6]{\frac{\Phi_{F,D}\kappa^2}{\eta_{im}^4}\ \frac{J}{M^{-1}cm^{-1}nm^4}}$$

Eq. 3.1.1

Where ϕ_D is the donor quantum yield, κ is the orientation factor, N_A is Avogadro's constant, n is the refractive index of the inter-dye medium, J is the spectral overlap integral (see Chapter 1 for more detail).

This leaves just the parameters of the dyes themselves. These parameters tend to have upper limits due to photophysics and lower limits imposed by the need for the dyes to be easily detectable (see Table 3.1.1). To exacerbate this problem; shorter wavelength dyes tend to have higher quantum yields, and longer wavelength dyes tend to have higher extinction coefficients (Fig. 3.1.3), so increasing spectral separation by choosing a bluer donor and redder acceptor may come at the expense of increased quantum yield and extinction coefficient, reducing the intended impact on R_0. In addition to this, R_0 only increases with the sixth root of these parameters, so large changes in them produce only small changes in the range of sensitivity. It is for these reasons that it is so difficult to increase the sensitivity range of FRET at shorter and longer distances just by seeking dyes with different spectral properties.

Table 3.1.1 limits imposed on feasible R_0 by dye parameters.

Parameter	**Meaning**	**Upper Limit and cause**	**Lower limit and cause**
$\Phi_{F,D}$	Fluorescence quantum yield of Donor	1.0, but typically no higher than 0.9	0, but typically no lower than 0.1 due to constraints on detectability
ε_A	Extinction coefficient of Acceptor	Typically no higher than 250,000	Theoretically very low, however limited by detectability if the method requires excitation of acceptor (ALEX, PIE etc.)
J	Spectral overlap	Limited by high crosstalk caused by spectrally similar dyes	Low detectability of very low and high wavelength dyes

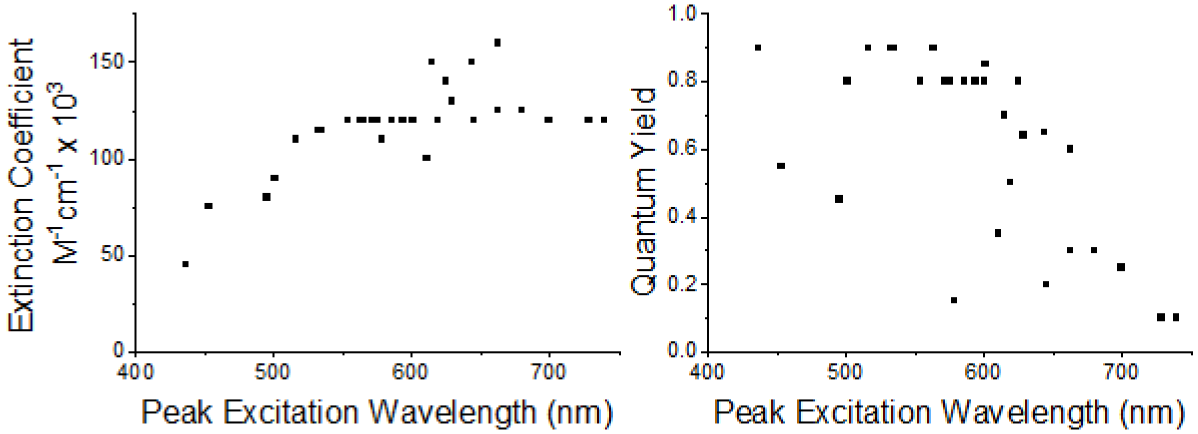
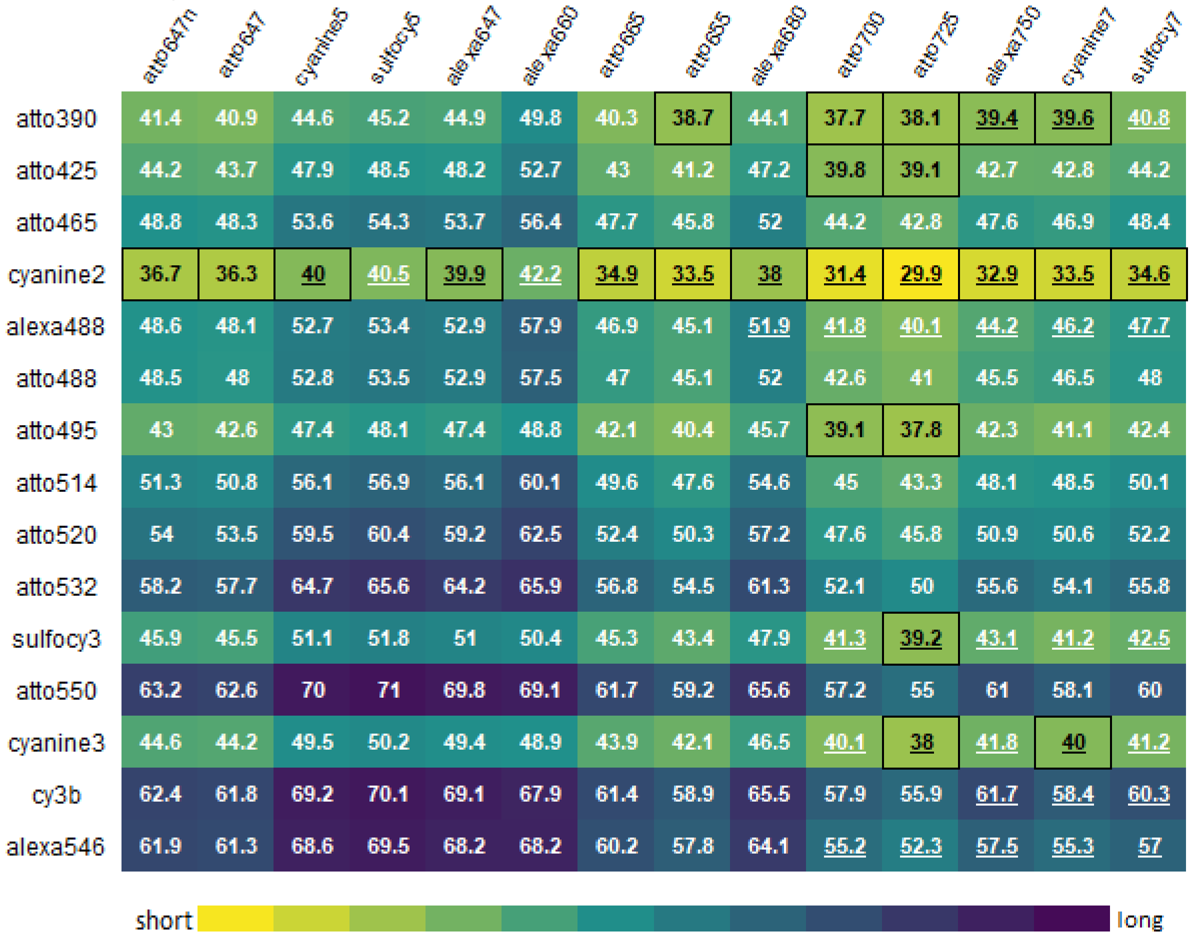


Fig. 3.1.3 Extinction coefficients and fluorescence quantum yields of Atto-tec fluorophores plotted against their peak excitation wavelengths, demonstrating that using more blue donor dyes and more red acceptor dyes to decrease spectral overlap will tend to increase the other two parameters which go into R_0, limiting the potential to decrease it. Quantum yields and extinction coefficients were obtained from the Atto-tec catalogue[114].



	atto647n	atto647	cyanine5	sulfocy5	alexa647	alexa660	atto665	atto655	alexa680	atto700	atto725	alexa750	cyanine7	sulfocy7
atto390	41.4	40.9	44.6	45.2	44.9	49.8	40.3	38.7	44.1	37.7	38.1	39.4	39.6	40.8
atto425	44.2	43.7	47.9	48.5	48.2	52.7	43	41.2	47.2	39.8	39.1	42.7	42.8	44.2
atto465	48.8	48.3	53.6	54.3	53.7	56.4	47.7	45.8	52	44.2	42.8	47.6	46.9	48.4
cyanine2	36.7	36.3	40	40.5	39.9	42.2	34.9	33.5	38	31.4	29.9	32.9	33.5	34.6
alexa488	48.6	48.1	52.7	53.4	52.9	57.9	46.9	45.1	51.9	41.8	40.1	44.2	46.2	47.7
atto488	48.5	48	52.8	53.5	52.9	57.5	47	45.1	52	42.6	41	45.5	46.5	48
atto495	43	42.6	47.4	48.1	47.4	48.8	42.1	40.4	45.7	39.1	37.8	42.3	41.1	42.4
atto514	51.3	50.8	56.1	56.9	56.1	60.1	49.6	47.6	54.6	45	43.3	48.1	48.5	50.1
atto520	54	53.5	59.5	60.4	59.2	62.5	52.4	50.3	57.2	47.6	45.8	50.9	50.6	52.2
atto532	58.2	57.7	64.7	65.6	64.2	65.9	56.8	54.5	61.3	52.1	50	55.6	54.1	55.8
sulfocy3	45.9	45.5	51.1	51.8	51	50.4	45.3	43.4	47.9	41.3	39.2	43.1	41.2	42.5
atto550	63.2	62.6	70	71	69.8	69.1	61.7	59.2	65.6	57.2	55	61	58.1	60
cyanine3	44.6	44.2	49.5	50.2	49.4	48.9	43.9	42.1	46.5	40.1	38	41.8	40	41.2
cy3b	62.4	61.8	69.2	70.1	69.1	67.9	61.4	58.9	65.5	57.9	55.9	61.7	58.4	60.3
alexa546	61.9	61.3	68.6	69.5	68.2	68.2	60.2	57.8	64.1	55.2	52.3	57.5	55.3	57

Fig. 3.1.4 Förster radii (in Å) for some common organic fluorophores used in fluorescence microscopy, with prospective donors on the left and acceptors on the top. Radii were calculated using quantum yields, extinction coefficients, and spectra provided by commercial suppliers. Kappa squared was set to ⅔, and the inter-dye refractive index was set to 1.4, an intermediate value between 1.33 of pure water, and 1.45 typical of proteins and DNA[115]. Radii below 40 Å are outlined in black. For underlined values, the provided donor emission spectrum did not extend beyond the peak of the acceptor absorption spectrum, so these numbers are likely below the true Förster radii.

The difficulties in increasing the sensitive range of FRET measurements at shorter distances are not limited to just R_0. It has been shown that at shorter distances (~2 nm) the ideal dipole approximation (IDE) upon which Försters equations depend becomes no longer applicable[8]. Furthermore, at close distances the accessible volumes (AV) of the dyes can overlap, allowing some degree of dye-dye contact. Not only can these contacts cause kappa squared to deviate from the assumed ⅔, but a number of photophysical effects caused by the dye contact can further obfuscate analysis[53,116].

PET Quenching

One other way of measuring conformational changes over short distances is to use photoinduced electron transfer (PET)[117]. Traditionally, this technique involves placing a fluorophore in proximity to an amino acid (tryptophan or tyrosine) which can quench it via electron transfer, however non amino acid quenchers can also be used[118]. Whilst this technique can be used to probe distances below 3 nm, it requires either the fortuitous presence of a quenching amino acid in the region of interest, or the attachment of a nitroxide radical group rather than a second fluorophore. A quenching technique which operates in a similar way, but between two fluorophores typically used for FRET would have the advantage of, allowing the experimenter to measure both long and short distances with the same toolkit.

Collisional Contact Quenching

Dye-dye contact effects in smFRET systems have been previously investigated and exploited in the common donor-acceptor pair Cy3B/Atto647N[110,116,119–121], and briefly in other dye pairs[53] (TMR, Cy3, and Cy5). In 2007 Brune et al, measured dye-dye contact quenching by linking together Cy3B and Atto647N via the amino acid cysteine[116]. The authors used UV-Vis spectroscopy to measure the blue-shifting of Cy3B and the red-shifting of the Atto647N peak absorbances, and likened this to the formation of H-dimers, suggesting that the proximity of the dyes was causing them to dimerise due to hydrophobic interactions. They further demonstrated that these spectral shifts are abrogated in Ethanol, supporting the conclusion that they are caused by hydrophobically driven dimerisation. Additionally, they used ALEX-smFRET to show that in ethanol, the doubly labelled cysteine showed a distinct population of intermediate stoichiometry, which disappears in aqueous solution. However, it is not apparent whether this loss of doubly labelled population originates from quenching of the donor, or acceptor, or both. The authors of this study intended for this work to be a warning against placing dyes too close together in smFRET experiments, however the change in signal detectable by ALEX would go on to form the basis of its own assay for resolving extremely small distance changes.

In 2010, Cordes et al., used a Cy3B/Atto647N pair labelled on either strand of a DNA substrate to measure the formation of a transcription bubble by RNA polymerase (RNAP); in the absence of RNAP the duplex structure of DNA allowed the dyes to contact each other and hence quench, whereas in the presence of RNAP, the separation of the strands reduced the quenching effect[119]. The authors of this paper also noted a blue-shift in the absorption spectrum of Cy3B (from a 560 nm peak to 548 nm). This same labelling scheme was later used to measure transcription bubble dynamics in conjunction with traditional FRET measurements[120]. This quenching assay has also been employed to investigate influenza virus promoter RNA[121] and conformations, and DNA polymerase recognition of gapped DNA substrates[110].

Evidently, exploiting the effects of contact quenching seen in the Cy3B/Atto647N pair would be of great use in extending the smFRET toolkit and recovering proximity information below the limit imposed by typical methods. So far, this quenching system has been used to monitor binary changes, in systems that are either open or closed (on or off), much like ensemble FRET assays or FRET based biosensors which report either one state or the other (see Table 3.1.2 for comparison). In this chapter present a series of experiments aiming to better understand the contact quenching mechanism in Cy3B/Atto647N with the goal of extending its applicability to measuring precise, but relative distance changes (as is possible in smFRET) below the 3 Å limit, which will then pave the way toward making absolute, accurate conformational measurements (as in accurate FRET).

Table 3.1.2 Current and future possible applications of contact quenching compared to various technology levels of FRET techniques.

State of Art	FRET Analogy	Contact Quenching	Feasibility
Binary	Ensemble FRET assays and biosensors, in on/off states	Measuring no quenching to high quenching	Currently in use, employed in multiple publications [110,116,119–121]
Quantitative; Relative	smFRET measurements of different proximity ratios	Measuring small changes that elicit varying degrees of quenching	May require further understanding. The aim of this work is to make this possible.
Quantitative; Absolute	Accurate FRET corrections, recovering absolute FRET efficiencies	Using quenching measurement to inform predictions of absolute conformational state	Future work, discussed later in this chapter

Here I present a combination of smFRET, fluorescence spectroscopy and time-correlated single-photon counting (TCSPC), to investigate the contact quenching phenomenon between Cy3B and Atto647N on a DNA duplex scaffold. The aim of this work is to investigate both the kinetics and photophysics of the contact quenching effect so as to better understand the relationship between dye proximity and measurable quenching parameters, and develop the applicability of the technique for experiments in biophysical systems.

3.2. Materials and Methods

DNA Constructs

Oligonucleotides containing amino-modified thymine residues (on either C2 or C6 linkers) were purchased from Eurogentec, and labelled with Cy3B (GE Healthcare) and Atto647N (Atto-tec) N-hydroxysuccinimide (NHS) esters (Fig. 3.2.2) in as described in Chapter 2. Purified oligos were checked for concentration and labelling efficiency on a NanoDrop spectrophotometer (Thermofisher) by first subtracting the 260 nm absorbance contribution from the dye. Complementary labelled oligos (Table 3.2.1) were then mixed at equimolar concentration in annealing buffer, heated to 95 °C, and allowed to cool overnight to form duplex DNA (Table 3.2.2). Note that whilst the acceptor oligos were purified with a low apparent labelling efficiency (possibly due to extinction coefficient change upon conjugation), the donor strands each have high labelling efficiencies and so the majority of duplex molecules with an acceptor will also have a donor.

Table 3.2.1: Names, sequences, and labelling of DNA oligonucleotides used in this chapter

Oligo Name	Sequence	Label	Labelling Efficiency (%)
BMA21	5’-CCGGGAC-T(C6-Amino)-GATACAGAGCGGAGGG-3’	Cy3B on C6 linker	100.4
BMA22	5’-CCGGGACTGA-T(C6-Amino)-ACAGAGCGGAGGG-3’	Cy3B on C6 linker	101.0
BMA23	5’-CCCTCCGCTCTG-T(C6-Amino)-ATCAGTCCCGG-3’	Atto647N on C6 linker	68.8
BMA24	5’-CCCTCCGCTCTGTA-T(C6-Amino)-CAGTCCCGG-3’	Atto647N on C6 linker	69.8
BMA25	5’-CCCTCCGCTC-T(C6-Amino)-GTATCAGTCCCGG-3’	Atto647N on C6 linker	64.5
BMA26	5’-CCCTCCGC-T(C6-Amino)-CTGTATCAGTCCCGG-3’	Atto647N on C6 linker	90

Table 3.2.2 DNA Duplex constructs used in this chapter, the two labelled oligo strands used to make them (see Table 3.2.1), and the base pair separation between the labelling positions. In the case of BMA2224, the labelling positions are 5’ of one another, which is denoted as a separation of -1 bp.

Duplex Name	Donor strand	Acceptor strand	Seperation (bp)
BMA2123	BMA21-Cy3B	BMA23-Atto647N	4
BMA2124	BMA21-Cy3B	BMA24-Atto647N	2
BMA2125	BMA21-Cy3B	BMA25-Atto647N	6
BMA2126	BMA21-Cy3B	BMA26-Atto647N	8
BMA2223	BMA22-Cy3B	BMA23-Atto647N	1
BMA2224	BMA22-Cy3B	BMA24-Atto647N	-1
BMA2225	BMA22-Cy3B	BMA25-Atto647N	3
BMA2226	BMA22-Cy3B	BMA26-Atto647N	5
BMAC3B	BMA21-Cy3B	BMA26 - No Label	-
BMA647N	BMA21 - No Label	BMA26-Atto647N	-

Fig. 3.2.1: Schematic of DNA duplex scaffold and labelling positions used here. Donor positions are in blue, acceptor positions are in orange.

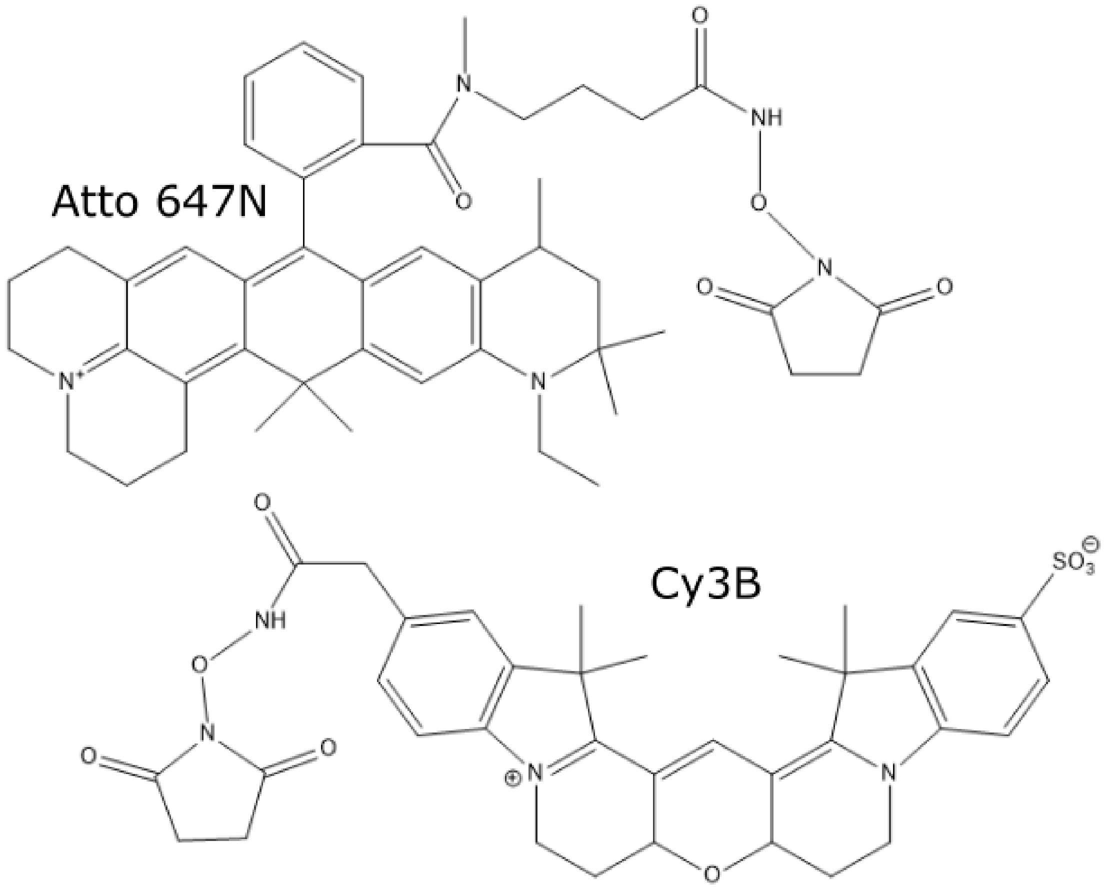


Fig. 3.2.2: Dye NHS esters used in this study

smFRET data acquisition

Single-molecule quenching measurements were conducted using the smfBox as described in earlier chapters. Laser powers were set to 65% green, 15% red, and the ALEX duty cycle was 45-5-45-5, and acquisition was controlled by the smOTTER software. Samples were diluted in quenching buffer (50 mM Tris 8.0, 5 mM NaCl, 20 mM MgCl$_2$, 0.1 mg/ml BSA, 1 mM DTT), to sub nanomolar concentrations to give approximately 1 to 5 bursts per second checked by the live display in smOTTER. Appropriately diluted samples were placed in airtight chambers made of a silicone gasket between two glass coverslips and data were acquired for 1 hour and saved as photon HDF5 files.

smFRET data analysis

Single-molecule quenching data were analysed via FRETBursts[20] Jupyter Notebooks (Chapter 2). Backgrounds are subtracted from each channel, and a single channel sliding window burst search (L=10, m=10, F=15) was executed on the $A_{ex}A_{em}$ channel only. Bursts with at least 30 photons in the $A_{ex}A_{em}$ channel were then selected, and bursts longer than 5 ms were rejected. A double Gaussian function was fitted to the stoichiometry distribution and the amplitudes of low and mid S taken as the population sizes. The quenching ratio (QR) was then calculated as the ratio of bursts in the acceptor only population (see results section for detail).

Compound Overlap calculation

Accessible volumes (AV's) were generated using the Olga software from the Seidel lab[122] with a fixed grid increment of 0.2 Å (unless specified otherwise) and AV parameters shown in Table 3.2.3.

Overlap is then calculated by finding the number of points in one cloud which are inside of the other using an algorithm implemented in python. The algorithm first "snaps" one AV onto the same 3D co-ordinate grid as the other by shifting all points by the difference in each dimension. The two coordinate arrays are then converted to sets and compared to find the intersection. Pictures of AV models are rendered using PyMOL[123].

Table 3.2.3 Dye and linker dimensions used in AV calculations

Dye	Linker Length (Å)	Linker Width (Å)	Dye Radius (Å)
Cy3B	19	4.5	3.5
Atto 647N	22	4.5	3.5

Spectra

Fluorescence spectra were taken using a Horiba Duetta. Samples were diluted to 100 nM in quenching buffer and measured in a 1 cm pathlength quartz cuvette (cleaned with ethanol and distilled water between measurements), resulting in 0.012 and 0.015 absorbance units at the peaks of Cy3B and Atto 647N, which is low enough to make the inner filter effect negligible. Samples were excited from 350-750 nm in steps of 3 nm, and emission was collected from 500-800 nm to produce an excitation emission map (EEM). Collected EEM's were reduced to averaged 1D spectra and manipulated in python. For excitation spectra, emission was summed from 700-750 nm to reduce noise. For emission spectra, intensities were summed across 5 excitation wavelengths (15 nm bandwidth total).

TCSPC

Time-correlated single-photon counting was performed using a home-made spectrophotometer built from 3D printed parts, a 4 W supercontinuum laser fixed at a 40 MHz repetition rate (Fiannium Whitelase SC-400-4, now obsolete, closest replacement is *NKT-FIU-15*), and a Becker & Hickl HPM-100-07 detector. Samples were diluted to 50 nM in quenching buffer in a 1 cm pathlength cuvette, excitation light was filtered to 610±10 nm and emission collected through a

665 nm longpass filter. Instrument response function (IRF) was measured with quenching buffer and no emission filters. Measured curves were fitted (with IRF) in FLIMfit 5.1.1 with either a single or double exponential decay (where specified).

3.3. Results

Compound Overlap

One way to link measured quenching to structural information is to use the spatial overlaps between the accessible volumes (AV's) of each dye. If the dyes exhibit quenched behaviour when they touch, then the degree of quenching should be closely related to the quantity of AV space they can both explore. Dyes positioned such that their AV's overlap only slightly should rarely contact one another and thus exhibit little quenching. Conversely, a high AV overlap presents more opportunities for the dyes to contact and thus should exhibit greater quenching. If AV overlap correlates well to measured quenching, then it may act as a useful intermediate variable linking quenching and intramolecular conformation.

To explore this idea further, we first define the *compound overlap* (O_C), which is given as the product of the fraction of the donor AV in the acceptor AV, and fraction of the acceptor AV in the donor. If the AVs are defined as a set of points (as simulated AV's often are), then this equation can be written as Eq. 3.3.1, and so long as these points are spaced by an equal increment in both AVs, it can be simplified to Eq. 3.3.2, given that the number of points in the donor cloud which are in the acceptor cloud will be equal to the number of points in the acceptor cloud which are in the donor cloud.

$$O_C = \frac{n(D \cap A)}{n(D)} \cdot \frac{n(A \cap D)}{n(A)}$$

Eq. 3.3.1

$$O_C = \frac{n(D \cap A)^2}{n(D) \cdot n(A)}$$

Eq. 3.3.2

Expressing spatial overlap like this is useful, as it accounts for differences in size in the two AV clouds, and also represents the probability that both dyes are occupying the overlapping region at the same time, given a uniform exploration of their respective AV's, hence the name *compound overlap*; the compound probability that both the donor is in the acceptor AV and that the acceptor is in the donor AV. The assumption of uniform AV exploration is likely incorrect to some degree, as it has been shown that dyes do not occupy all points in their AV's with equal probability[124], spending more time in some regions than others. However, as an initial model going forward we will assume that a uniform AV is a good enough approximation for estimating the likelihood that dyes will contact.

Construct Design

In order to design DNA constructs for this study, an initial calculation of compound overlap was done across different base pair separations on either strand of a duplex DNA model (Fig. 3.3.1). As this was just an initial scan, AV's were generated with a larger 0.4 Å grid increment and snapping one AV to the other in only one direction (as explored later on) to increase calculation speed. The calculation was done on a DNA model sequence of one thymine residue to be labelled with the donor, and a stretch of adenines in both directions (to label the thymines on the opposite strands with the acceptor). To differentiate between the different directions on the DNA, base pair distances are reported as positive when the dyes are 3' of each other (ie. the bases between the dyes on the donor strand are 3' of the base the donor is conjugated to), and negative when they are 5' of each other.

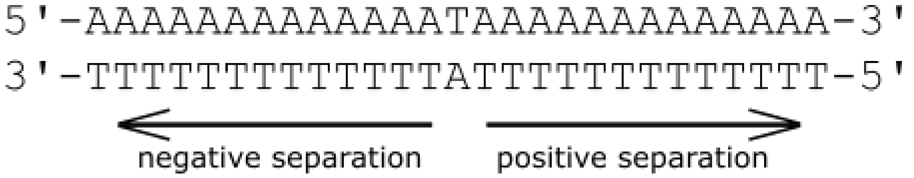


Fig. 3.3.1: Poly-T model DNA used to initially test AV overlaps in silico. The donor is attached to the central T on the top strand, with the other acceptor placed on the bottom strand. Pairs where the dyes are 3' of each other (along their own strand) are referred to here as positive separation, whereas pairs 5' of each other are referred to as negative separation

As seen in Fig. 3.3.2, the overlap does not peak around zero base pair separation (not included here, as bases are conjugated to thymines, and two thymines will not pair under Watson-Crick rules). The distribution instead peaks around 2-3 bp 3', as the AV's extend further in the 3' direction along the major groove than they do 5'. As a consequence of this, dyes separated by 1bp have very different overlaps depending on the orientation of their separation.

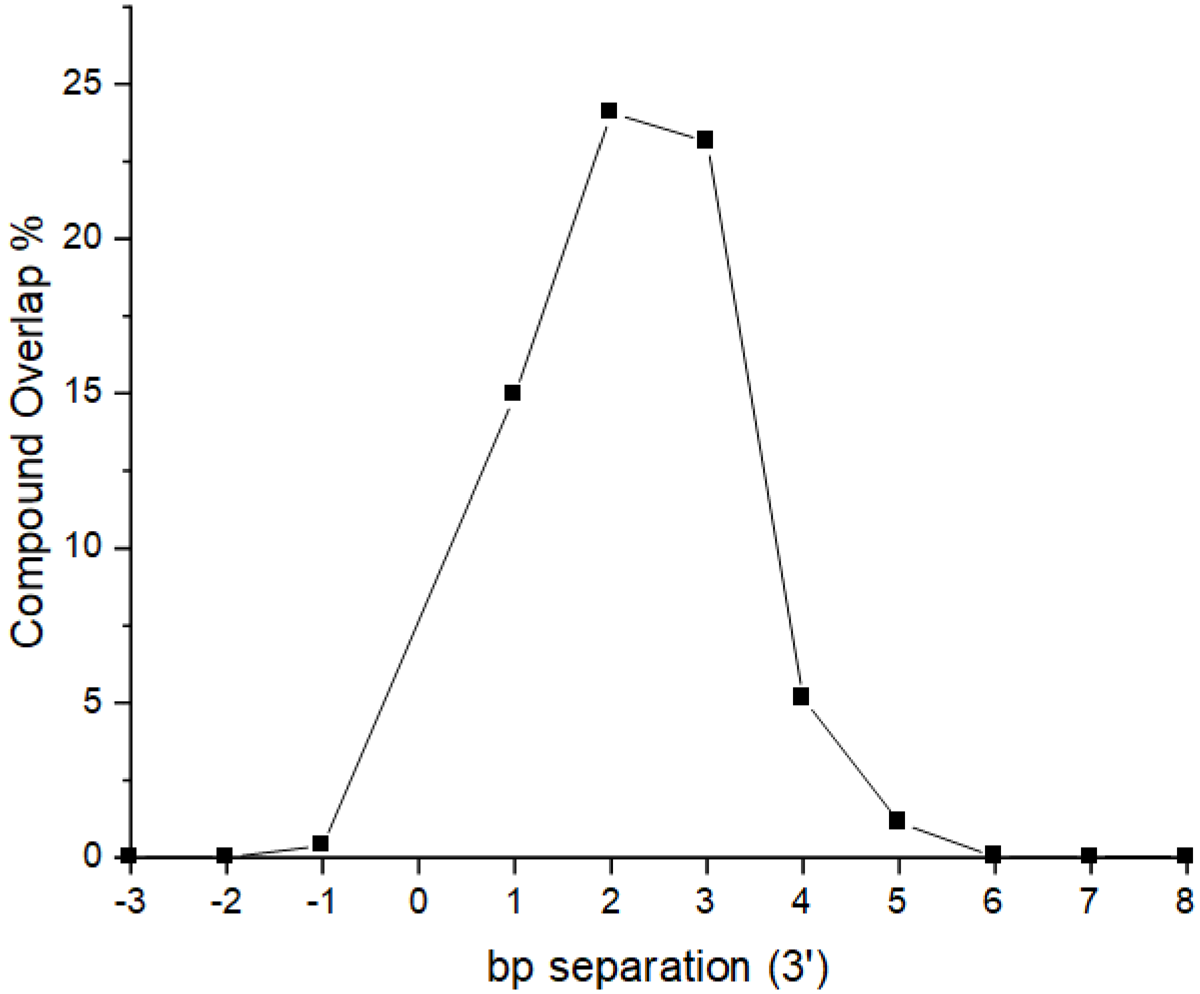


Fig. 3.3.2 Initial scan of compound overlaps against base pair separation between dyes in the 3' direction. Note that the peak is at 2-3 bp, not 0 bp.

To access a range of compound overlaps, DNA constructs were designed made of strands labelled at 2 donor positions and 3 acceptor positions (Fig. 3.2.1) for a total of 8 constructs with separations of -1, 1, 2, 3, 4, 5, 6, and 8 bp. AV clouds (Fig. 3.3.3) were generated using the exact sequence to be used in experiments rather than on a poly-T model as described above.

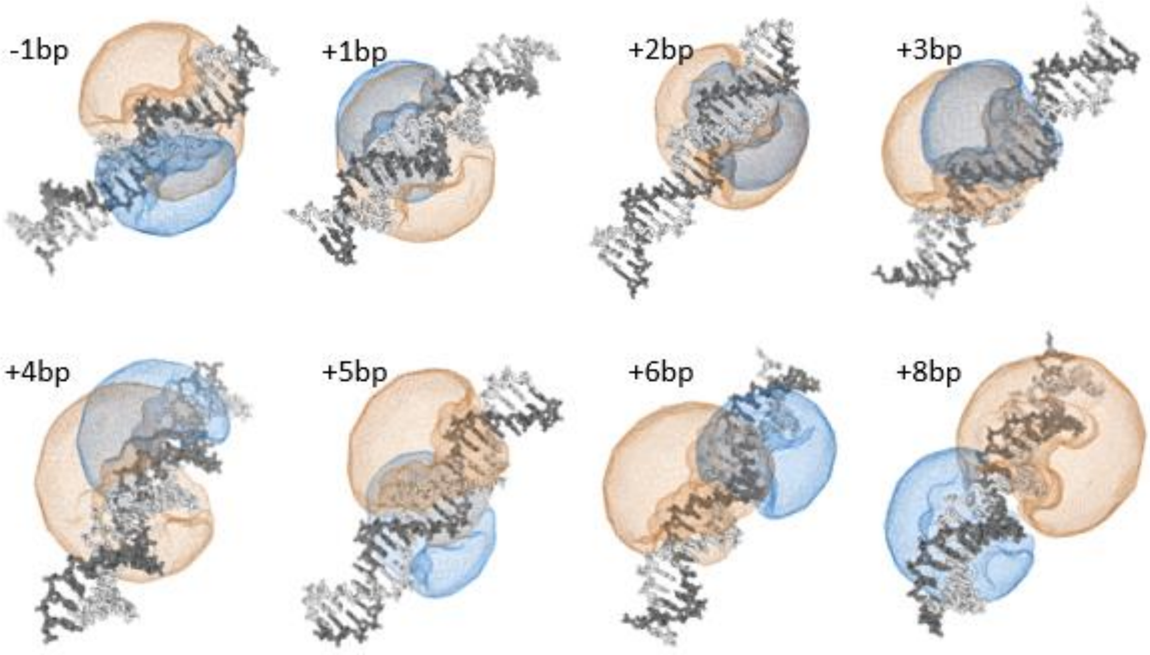
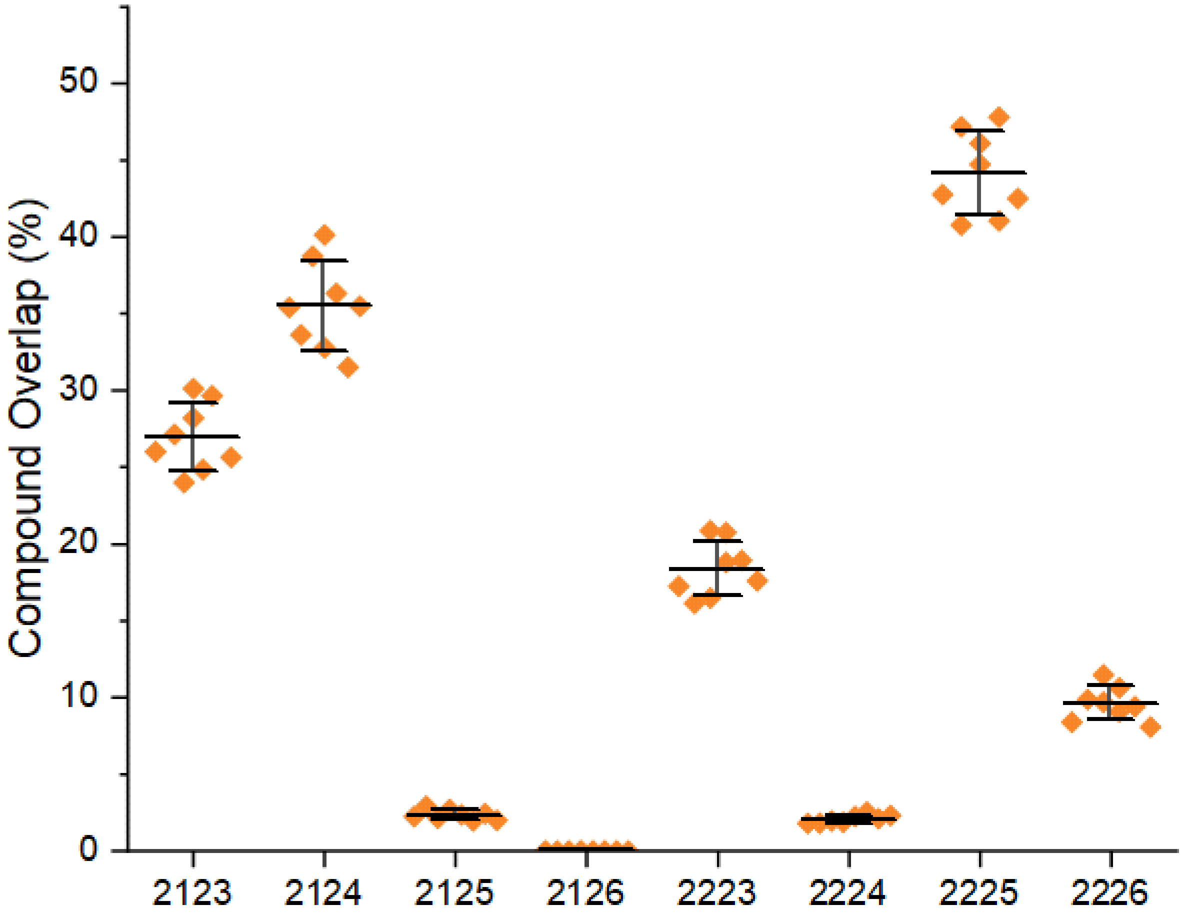


Fig. 3.3.3 The 8 DNA quenching constructs to be used for experiments, with AV clouds shown as meshes. Donor clouds are in blue, acceptor clouds are in orange. The donor clouds are smaller due to Cy3B having a shorter linker than Atto647N. Note the extreme difference in overlap between -1bp and +1bp.

As the overlap calculation used involves “snapping” one AV to the grid of the other, this may introduce some error, as the AV’s are moved relative to each other by a small distance between 0 and 0.2 Ångstroms in each of the 3 grid dimensions. In order to quantify this error, all 8 possible “snap” directions (combinations of positive and negative shifts in x, y, z) were implemented for 8 of the AV’s used in this study, and compared (Fig. 3.3.4). The same overlap data is shown again in Fig. 3.3.5, against base pair separation. As with Fig. 3.3.2, the peak overlap is not at 0 bp but at 3 bp. As the standard deviations are small, being at most 2.7 percentage points, the error introduced by snapping on a 0.2 Ångstrom grid will be minimal, and the average of these 8 overlaps will be used going forward.



	2123	2124	2125	2126	2223	2224	2225	2226
mean	26.9	35.5	2.35	0.0163	18.4	2.09	44.1	9.59
st. dev	1.96	2.57	0.293	0.0065	1.59	0.235	2.41	0.983

Fig. 3.3.4 Compound overlaps for all 8 possible snap directions for the 8 quenching constructs. Means and standard deviations are shown as horizontal lines on the chart and reported below.

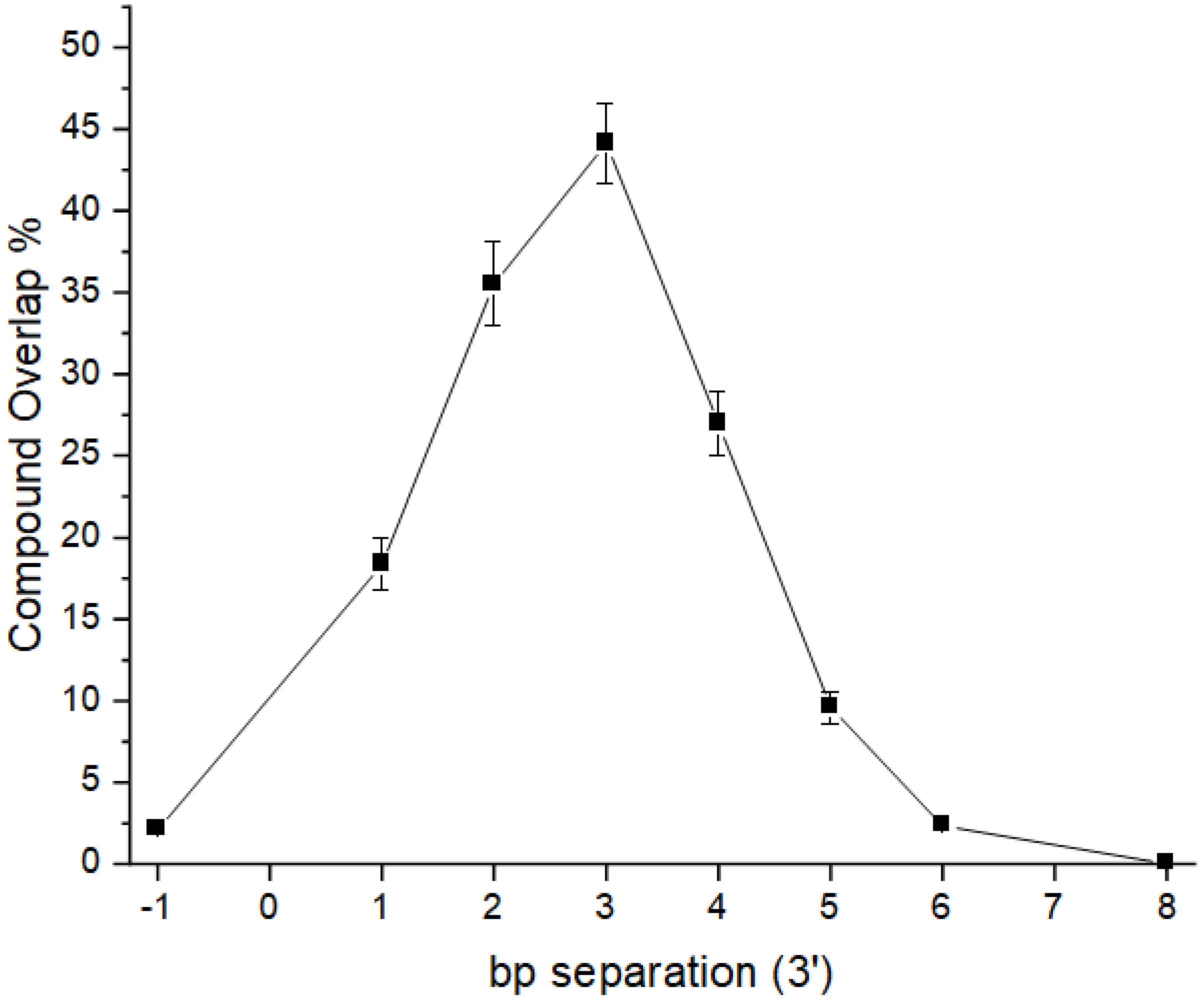


Fig. 3.3.5 Compound overlap against bp separation for the 8 quenching constructs, with error bars showing standard deviation of 8 calculations as in Fig. 3.3.4. Note that the distribution is not centred on 0 bp separation.

smFRET ALEX measurements of DNA constructs

To test the relationship between AV overlap and contact quenching, 8 DNA constructs were designed, made from 2 donor strands and 4 acceptor strands. These positions were chosen to provide a range of compound overlaps, with base pair separations from 1bp 3' to 8bp 5' (see Table 3.2.2). These constructs were measured using smFRET ALEX, as it has been shown that the quenching effect causes the number of detected "doubly labelled" (or rather; molecules which emit under excitation of both lasers) to decrease relative to the acceptor only population (See Fig. 3.3.6). Whilst it is not clear whether both dyes quench, or only one dye quenches, it is clear that the doubly labelled $S_{0.5}$ population disappears with respect to the apparently singly labelled S_0. Thus by measuring the ratio of the two, we can quantify the amount of quenching of the dyes.

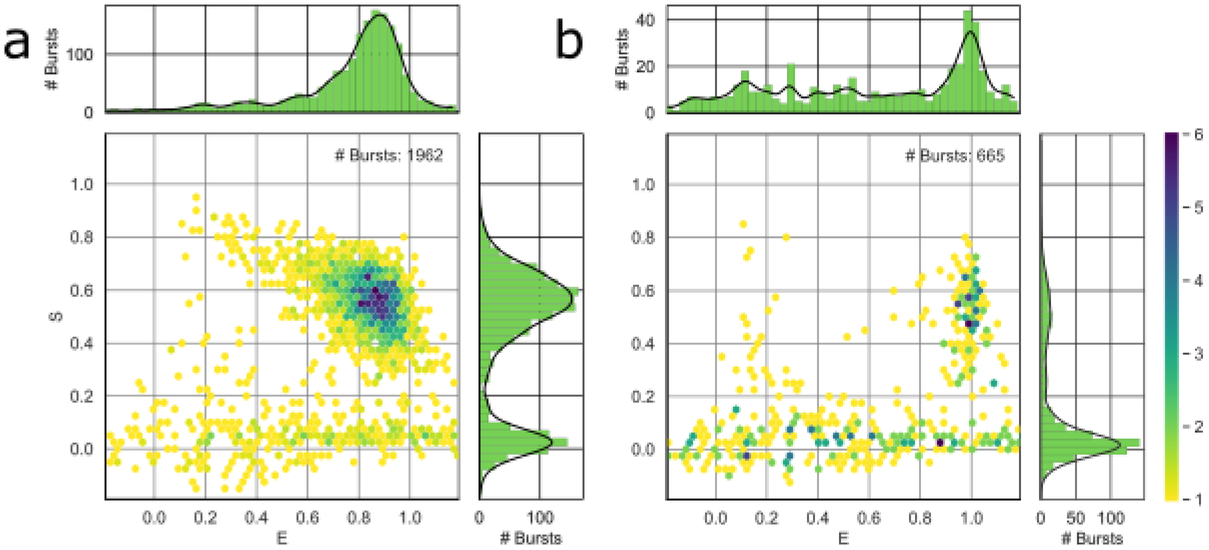


Fig. 3.3.6: Example ES plots of two quenching constructs, obtained from a burst search focussed solely on the acceptor excitation channel **a.** BMA2126 which has the lowest AV overlap and **b.** BMA2124 which has the highest AV overlap. Notice that with higher overlap, the population at intermediate stoichiometry decreases relative to the acceptor only population.

Thus, quenching can be reported from ALEX using the number of molecules detected under excitation of *only* the acceptor laser (appearing with low stoichiometry), as a fraction of *all* molecules responding to the acceptor laser (Eq. 3.3.3). In other words; the low S population, divided by the sum of the mid and low S populations.

$$QR = \frac{S_0}{S_{0.5} + S_0}$$

Eq. 3.3.3

From Fig. 3.3.7 we can see that as the dyes are brought closer together, the quenching ratio as measured by smFRET increases, due to the increased overlap between their AV's. The majority of the detectable change happens between 10 and 25 Å, which is well below the 30 Å limit of normal FRET measurements. However, the distance between the AV's themselves does not in of itself say anything useful about the likelihood of the dyes making contact, and so to build a kinetic model with which to fit this data, it will instead be plotted against the compound overlap metric discussed above.

As can be seen in Fig. 3.3.8, the QR also increases with compound overlap, but rises quickly before plateauing rather than forming a sigmoidal shape as with Fig. 3.3.7. The plateau occurs at approximately 20%, leaving 30% of molecules still detectable as doubly labelled. This suggests that even with maximal overlap, the dyes are not quenched all of the time.

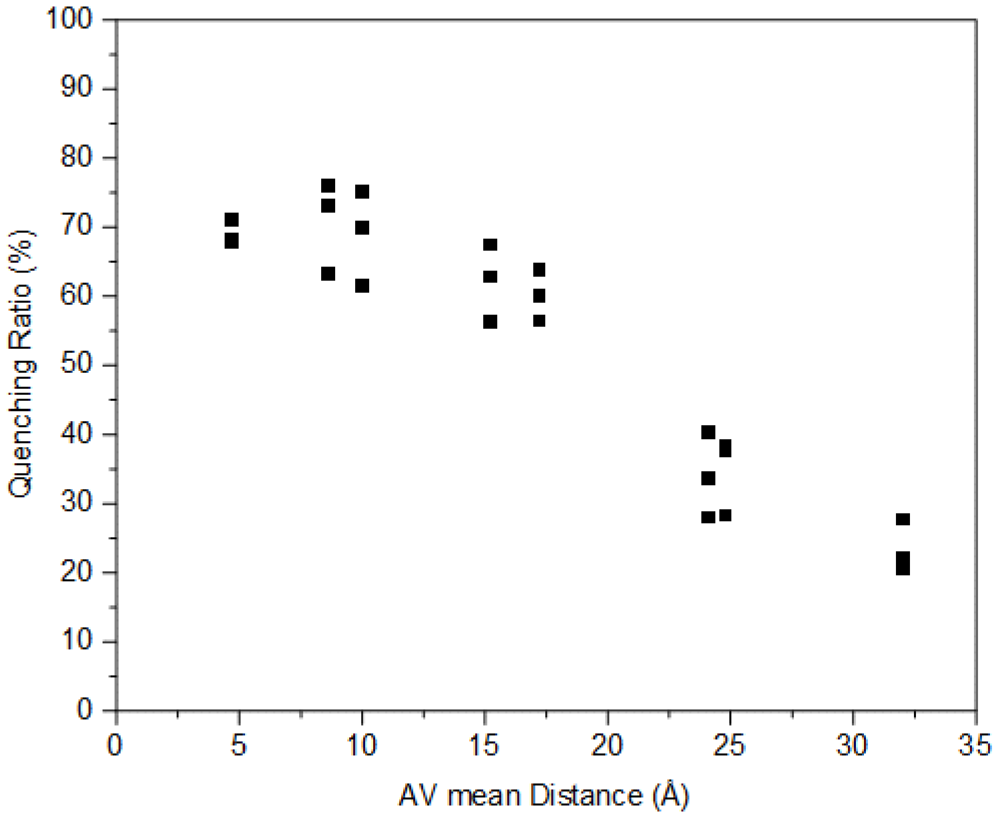


Fig. 3.3.7 Quenching ratio against distance (between centre of AV's), with three repeat measurements for each of the 8 constructs.

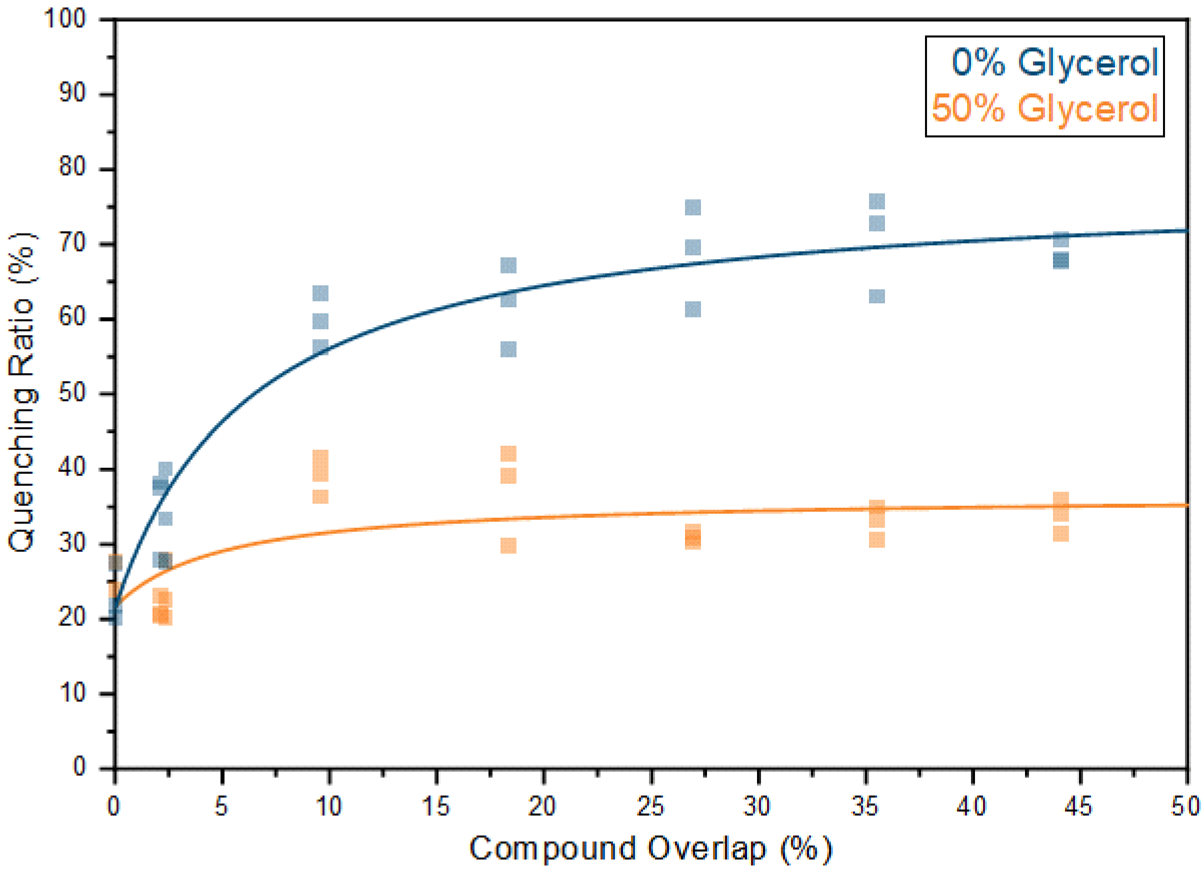


Fig. 3.3.8 Quenching Ratio against overlap for 8 DNA constructs with a range of compound overlaps, with three repeats for each construct. The curves shown here are the result of fitting with equations described below. The parameters of the fit are shown in Table 3.3.1

From here we can build a model, starting with compound overlap, to describe the expected degree of quenching, and then use these equations to fit the data and either validate or improve the model.

If we assume that the rate at which dyes come together will be proportional to the compound overlap, then we can describe the proportion of time that dyes spend bound together so long as we can also describe the rate that they come apart. Here we will make another assumption; that the rate of dissociation between the dyes is constant irrespective of where they are tethered. If this is true, then the fraction of time in the bound state can be related to compound overlap by equation 3.3.4:

$$Bound\ Fraction = \left(\frac{k_{off}}{k_{diff} O_c} + 1\right)^{-1}$$

Eq. 3.3.4

Where k_{off} is the rate of dissociation between bound dyes, and k_{diff} is the constant of proportionality between the on rate and compound overlap (O_c). Assuming that bound molecules dye pairs go on to quench and present as low stoichiometry molecules, then from here we can begin to develop an equation for the quenching ratio (QR); If bound molecules are not quenched all of the time, then the fraction of bound molecules will need to be scaled by Q_e, the fraction of time spent in the quenched state once bound. This simple model can be seen as a schematic in Fig. 3.3.9

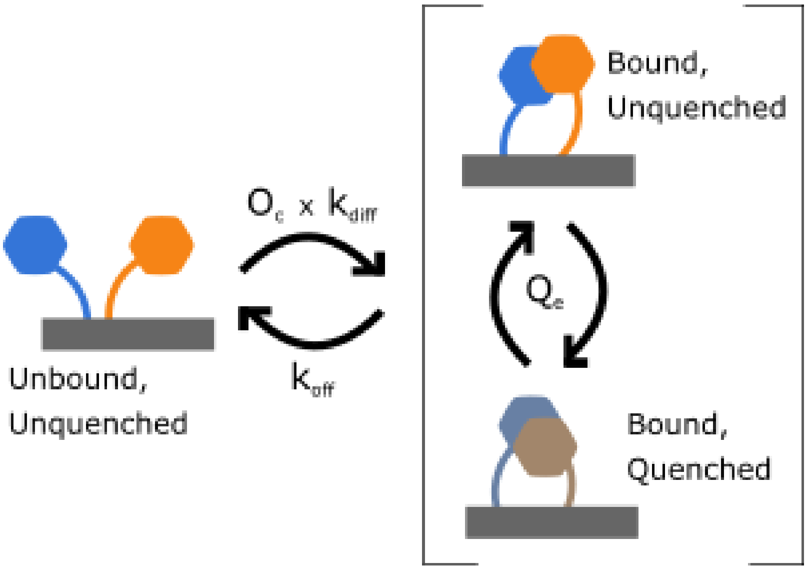


Fig. 3.3.9 A schematic of the kinetic model for quenching. The rate into the bound state is dependent on compound overlap and diffusion. The rate out of the bound state, k_{off}, is independent of overlap and assumed as constant for all dye attachment positions. In the bound state, some proportion of molecules will be quenched, the proportion of which is represented by the variable Q_e.

Additionally, if k_{off} (the rate dyes come apart), and k_{diff} (the rate that dyes come together given perfect AV overlap), are assumed characteristic for a given system, they can be simplified to a single constant k. This gives Eq. 3.3.5

$$Quenched\ Fraction\ = Q_e\left(\frac{\mathrm{k}}{O_c}+1\right)^{-1}$$

Eq. 3.3.5

However, this equation gives the fraction of quenched molecules, rather than the quenching ratio, which is calculated relative to the remaining acceptor only molecules. We can find the predicted QR with respect to the number of doubly labelled molecules (DA) and singly labelled molecules (detected as S_0) using the quenched (QF) above, as quenched fraction describes the number of molecules lost from the doubly labelled population:

$$QR = \frac{S_0}{S_0 + DA(1-QF)}$$

Eq. 3.3.6

Now we can express DA in terms of the baseline acceptor only fraction, A_{only}, which is simply the QR when there is no quenching:

When QF = 0, $S_{0.5} = DA$, therefore;

$$A_{only} = \frac{S_0}{DA + S_0}$$

Eq. 3.3.7

And then rearraging to;

$$DA = S_0(\frac{1}{A_{only}} - 1)$$

Eq. 3.3.8

Now substituting Eq. 3.3.8 and Eq. 3.3.5 into Eq. 3.3.6;

$$QR = \frac{S_0}{S_0 + S_0(\frac{1}{A_{only}} - 1)\left(1 - Q_e(\frac{k}{O_c} + 1\right)}$$

Eq. 3.3.9

Cancelling S_0, we now have a final equation for QR;

$$QR = \frac{1}{1 + \left(\frac{1}{A_{only}} - 1\right)\left(1 - Q_e(\frac{k}{O_c} + 1\right)}$$

Eq. 3.3.10

This equation can be used to find the kinetic parameters k and Q_e using the experimentally observed relationship between QR and O_c, so long as the value A_{only} (the fraction of acceptor only species which are truly acceptor only) remains constant between samples.

In addition to a standard measurement of the quenching constructs by smFRET in quenching buffer, the same measurement was made in a 50% glycerol version of the same buffer. By increasing the viscosity of the solution and thereby reducing the diffusion of the dyes, the model would predict a decrease in quenching, so long as there is no concomitant decrease in the off rate. In the 50% glycerol buffer we see a much lower plateau as expected, levelling off at 30% QR rather than 70%. Reassuringly, both curves begin at the same QR, which indicates that the high glycerol concentration has no effect on the measurement outside of the capacity for the dyes to contact.

Table. 3.3.1 Parameters derived from fitting the data in Fig. 3.3.8 with Eq. 3.3.10

Condition	**Q_e**	**k**	**A_{only}**
0% Glycerol	0.929±0.016	1.76±0.63	0.215±0.03
50% Glycerol	0.521+0.095	3.01±3.52	0.217±0.04

When the smFRET ALEX data are fit with the Eq. 3.3.10, we see an A_{only} value of 0.215 (Table. 3.3.1), as to be expected, as this is simply the amount of acceptor only species with zero quenching. The parameter Q_e, which represents the amount of species in the bound state which go on to appear as quenched, comes to 0.521, suggesting that even when the dyes are in contact some 10% will still appear unquenched. Finally the value k comes to 1.76, this value is difficult to interpret on its own as it represents the ratio of the off rate and some constant of proportionality describing dye diffusion.

When the same treatment is given to the 50% glycerol data, a very similar A_{only} value of 0.217 is returned. As discussed above, this confirms that the high glycerol concentration has no effect on the zero overlap measurement. The k value increases to 3.01, suggesting a decrease in the diffusion rate relative to the off rate, as expected, however the very high error on this value of 3.52, makes it difficult to say whether k has actually changed significantly. The Q_e value on the other hand decreases to 0.521 with an error of only 0.095, suggesting a significant decrease in the number of molecules in the bound state which then quench.

Whilst the increase in k fits with the prediction from the kinetic model, the large change in Q_e however is more difficult to explain, as it is hard to imagine a reason why glycerol would decrease the rate of transition into the quenched state for bound molecules, or increase the rate of unquenching. It is possible that the model need be refined to better explain this, or that glycerol has additional effects on the system beyond just the reduction in diffusion as was initially hypothesised. For example, the glycerol could be forming hydrogen bonds with the dyes which sterically prevent a transition into the quenched state, rather than just slowing their diffusion. The refractive index change from the glycerol is likely not the cause, as this would only alter the FRET efficiency, rather than the stoichiometry of bursts.

Fluorescence spectra

Excitation and emission maps (EEMS) were gathered for several quenching constructs as well as a construct with a single Cy3B (BMAC3B) and one with a single Atto 647N (BMA647N). These data are rich with information on both the excitation and emission of the dyes across the visible spectrum.

First; the excitation spectrum of a high quenching construct (BMA2123) is compared to the spectra of Cy3B and Atto 647N only (Fig. 3.3.10). Here we can clearly see a blue shift in the excitation spectrum of Cy3B and a slight redshift in the spectrum of Atto 647N. Whilst this is indicative of a change in the spectrum of the quenched state compared to the unquenched state, the dyes in each construct are not in the “quenched” state 100% of the time, so the spectrum here is a composite of the spectra of both the dyes in their unquenched and quenched states, rather than a direct spectrum of only the quenched state.

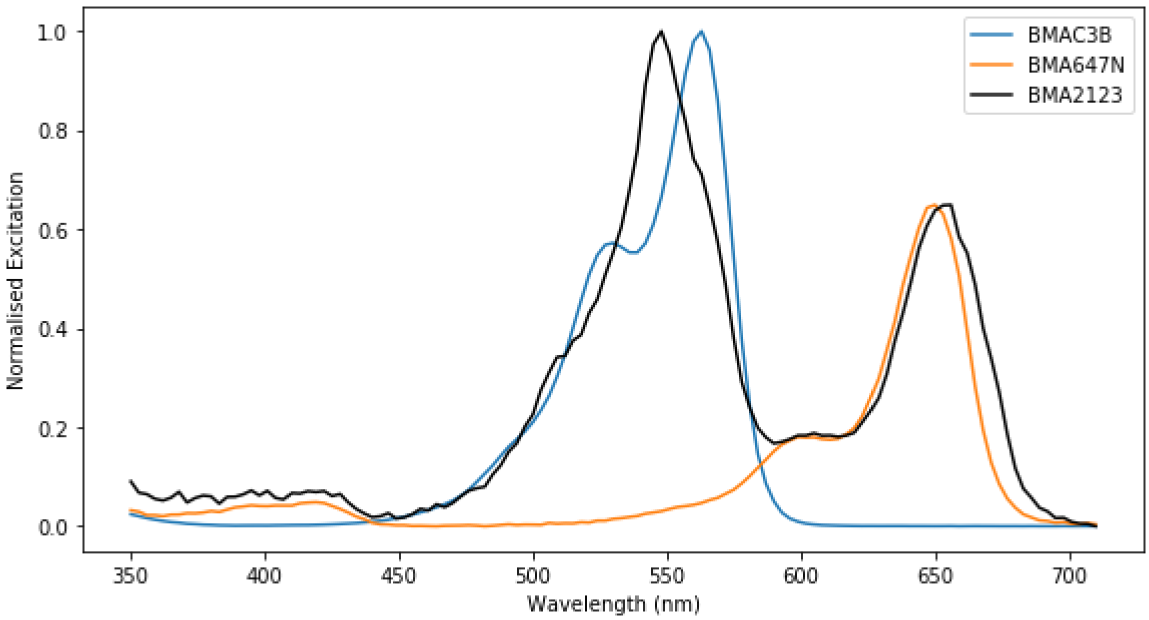


Fig. 3.3.10 Excitation spectra of Cy3B (blue) and Atto647N (orange) tethered to DNA alone, and BMA2123 (black), one of the quenching constructs with a high compound overlap. On the quenching construct, Cy3B exhibits a hypsochromic (blue) shift from a peak of 563 to 548 nm, whereas Atto647N exhibits a bathochromic (red) shift from 641 to 647 nm.

These spectral shifts for both dyes can be examined for multiple constructs of different overlap (Fig. 3.3.11). For Cy3B, the spectral shift appears clearly as a decrease in the amplitude of one peak at 563 nm and a corresponding increase in the amplitude of another peak at 548 nm. For Atto 647N the rise and fall of the two peaks is less apparent, appearing instead as a gradual shift from a peak of 641 to 647 nm. This is likely due to the much smaller peak shift with respect to the width of the spectrum, causing the sum of the two curves to appear as one single curve.

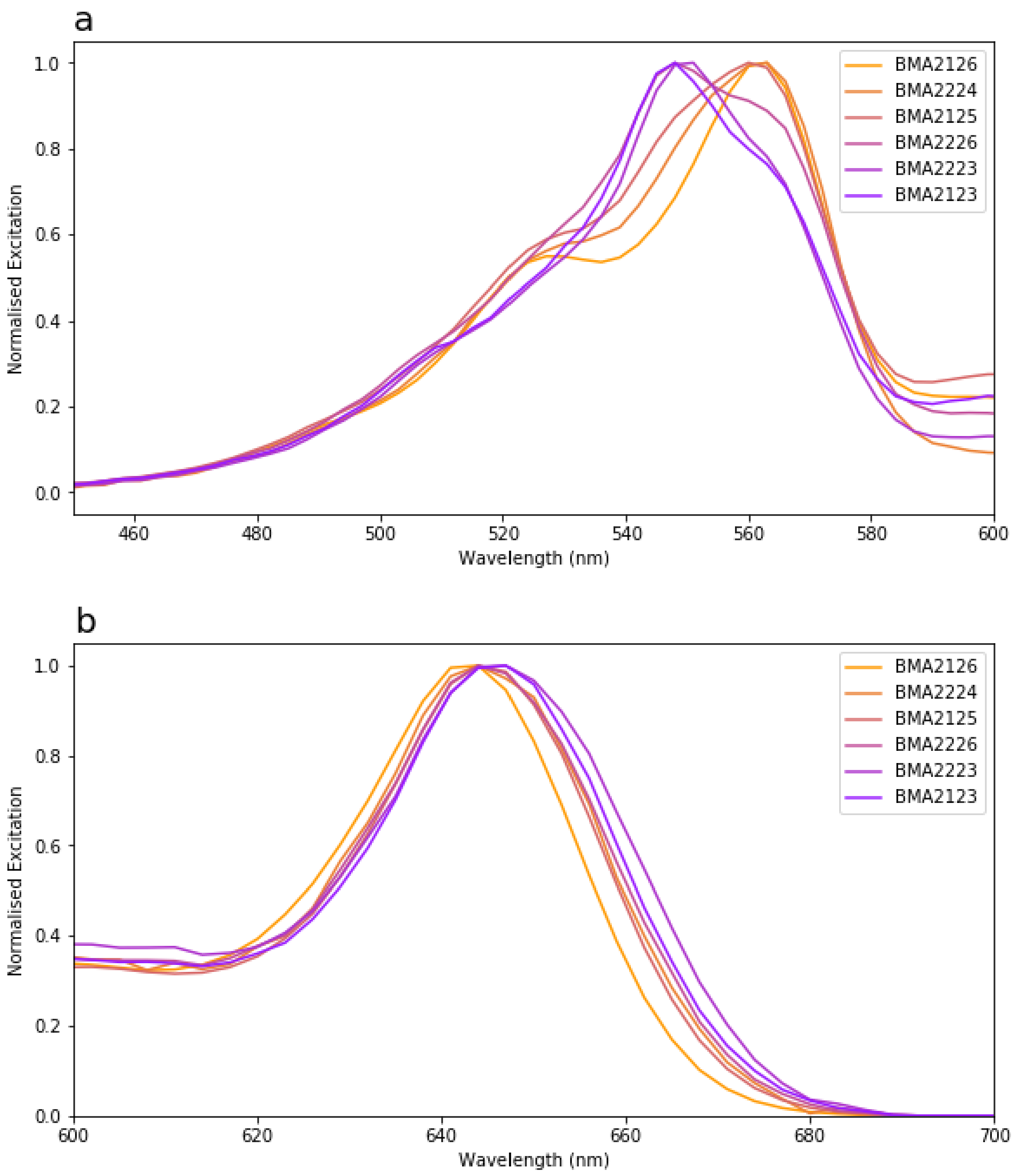


Fig. 3.3.11 Excitation spectra of 6 quenching constructs of varying compound overlap (shown here increasing from orange to blue). **a.** Spectra normalised to the donor peak. A peak at 563 nm can be seen to decrease in amplitude relative to peak at 548 nm which increases in amplitude. **b.** Spectra normalised to the acceptor peak. Here the wavelength shift is too small to observe changes in amplitudes between two distinct peaks.

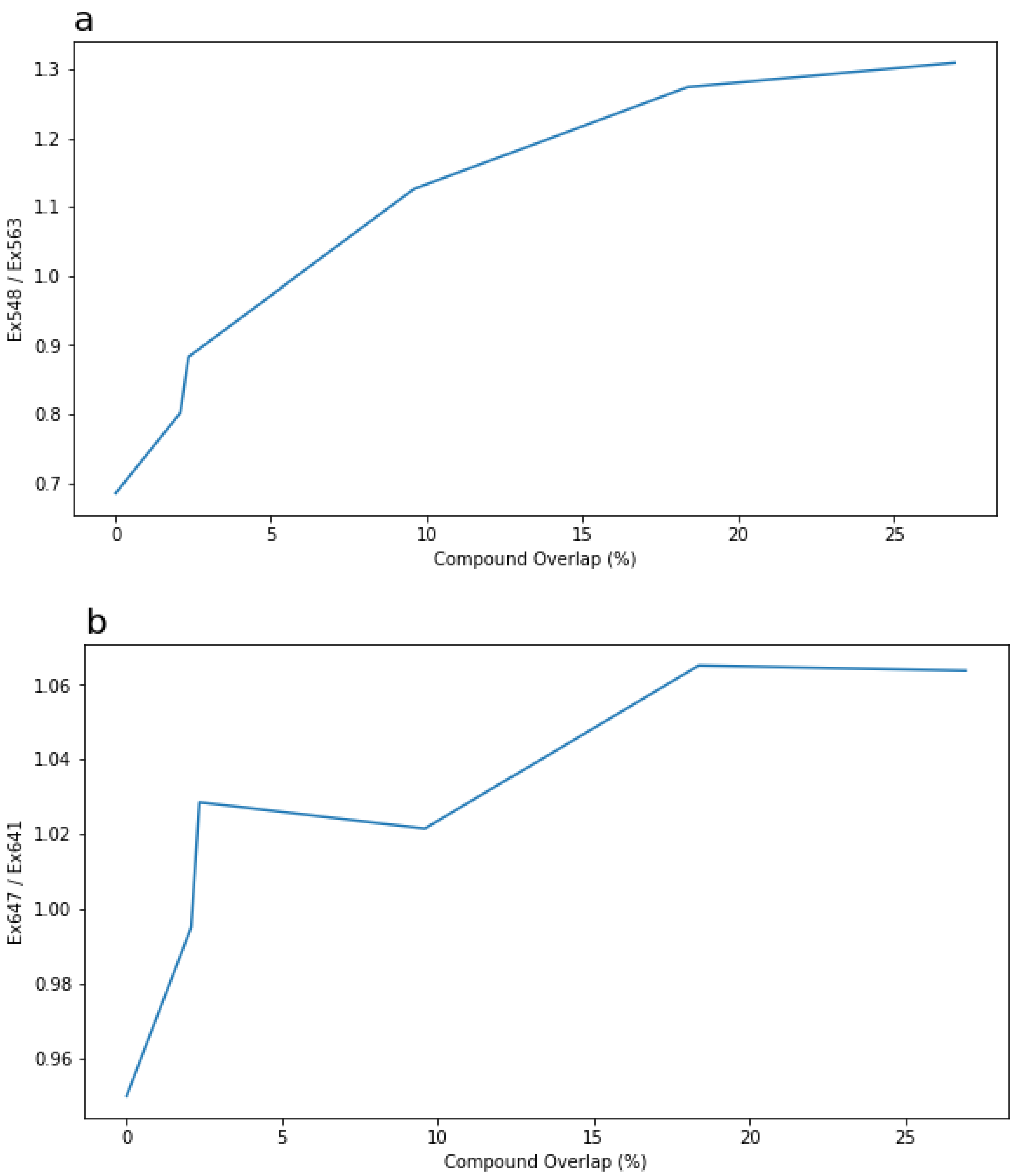


Fig 3.3.12 Intensities at the quenched peak as a fraction of intensity at the unquenched peak, plotted against compound overlap. **a.** The 548/563 ratio increases smoothly with compound overlap, plateauing at approximately double that of the value for unquenched Cy3B. **b.** The 647/641 ratio also increases with compound overlap, but is less clear, likely owing to the much smaller shift.

By dividing the height of one peak by the other and plotting against compound overlap, the spectral shifts of both dyes with respect to AV overlap (and thus dye-dye contact) can be monitored (Fig. 3.3.12). The 548/563 nm ratio for Cy3B increases smoothly with respect to compound overlap, plateauing at higher overlaps in much the same way that the smFRET data does. The 647/641 nm ratio for Atto 647N also increases and plateaus but less obviously, although this is likely a problem with the precision of determining the peak shift, due to the small size of the shift, as detailed above.

Next, the emission spectrum of the high overlap construct BMA2123 is checked against the original emission spectrum of each dye, both at green and red excitation wavelengths. At the green, donor excitation wavelengths of 505-531 nm (Fig. 3.3.13), the emission of Cy3B does not shift by any noticeable amount in the quenched state, whereas the peak emission of Atto 647N does shift when compared to that of Atto 647N alone.

When excited at the red, acceptor excitation wavelengths of 629-644 nm however (Fig. 3.3.14), the emission spectrum of Atto 647N does not exhibit this shift, despite the shift in its excitation spectrum under the same conditions (Fig. 3.3.11). Nevertheless, the shift in the acceptor emission spectrum under green excitation can similarly be tracked, increasing with AV overlap before plateauing (Fig 3.3.15).

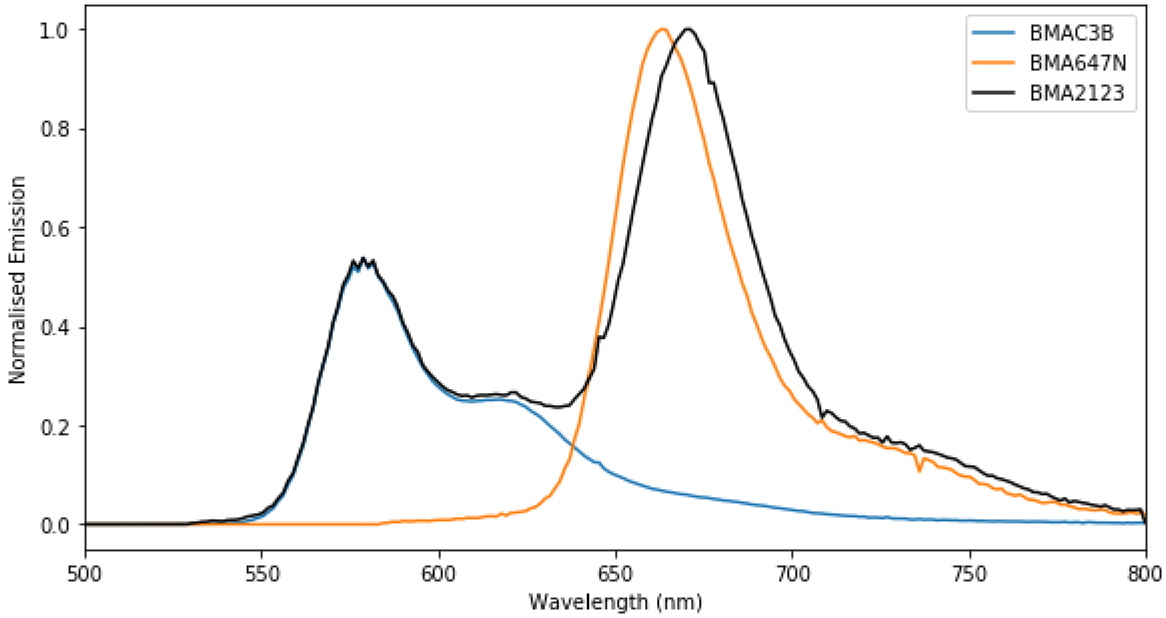


Fig. 3.3.13 Emission spectra of Cy3B (blue) and Atto647N (orange) tethered to DNA alone, and BMA2123 (black), one of the quenching constructs with a high compound overlap. There is little to no shift in the spectrum of Cy3B in the quenched construct, however when excited at a shorter wavelength Atto647N exhibits a notable bathochromic shift.

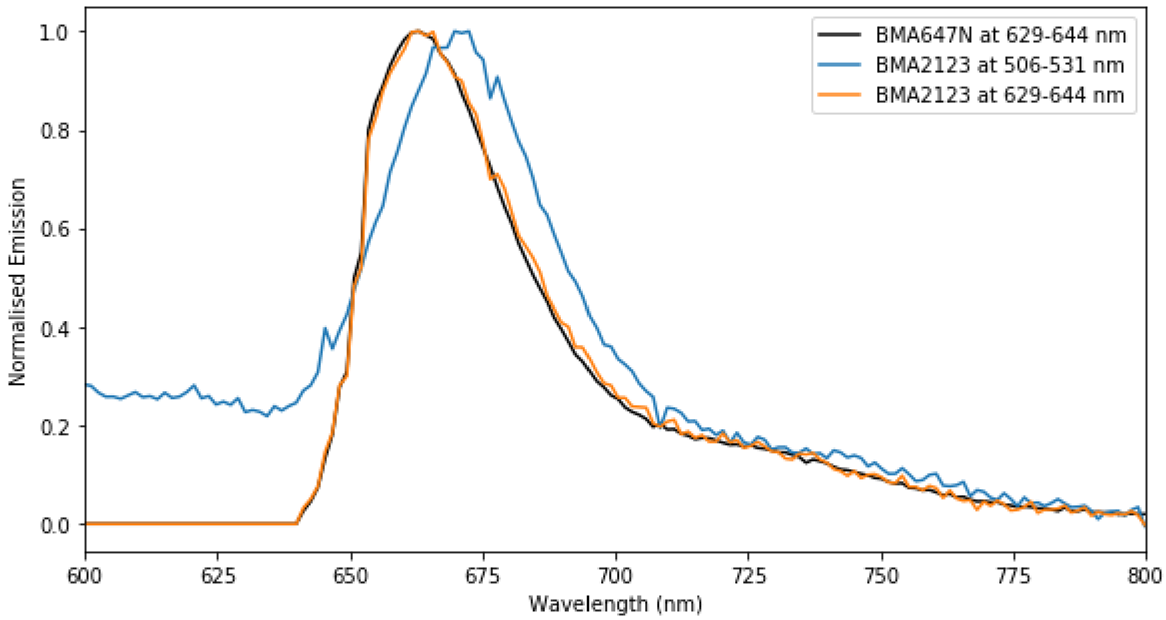


Fig. 3.3.14 Atto647N Emission spectra of BMA2123 excited at 506-531 nm (blue) and at 629-644 nm (orange), with the spectrum of Atto647N in the absence of a donor (black) for comparison. The bathochromic shift in Atto647N is only noticeable when excited in the absorption range of Cy3B.

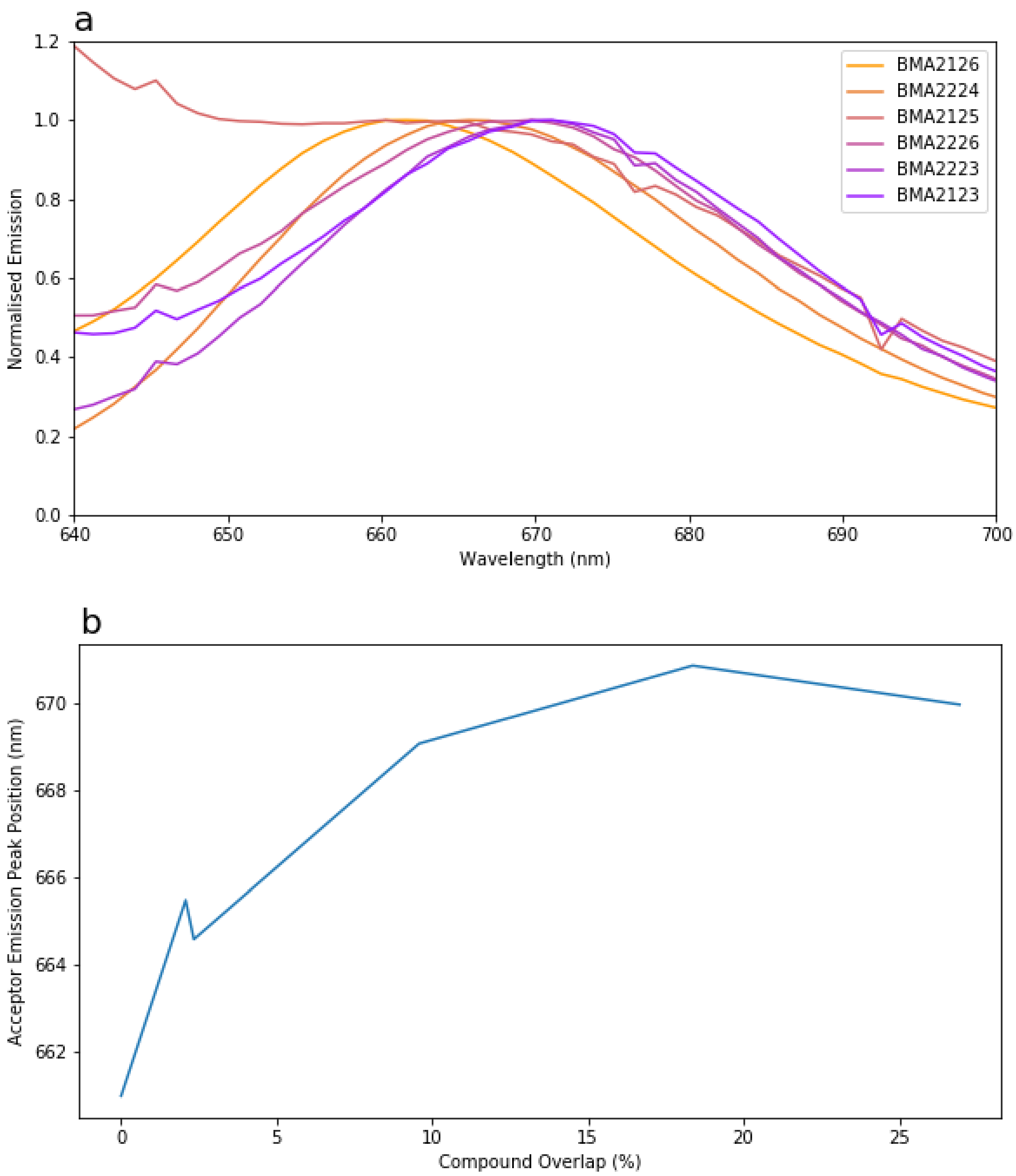


Fig. 3.3.15 Tracking the shift in emission spectrum of Atto647N. **a.** Emission spectra of each construct, with increasing overlaps shown from orange to blue. **b.** Peak wavelength against compound overlap.

TCSPC

Time correlated single photon counting (TCSPC) can be used to determine FRET efficiency by measuring the decrease in donor lifetime, as the acceptor provides an additional de-excitation route. The presence of a donor, however, would not normally affect the lifetime of the acceptor, but in the case of these quenching constructs, the fluorescence lifetime of Atto 647N decreases with increased AV overlap, likely due to contact interactions with Cy3B. Thus, the fluorescence decay of four constructs with varying AV overlap was measured under acceptor excitation. By measuring the fluorescence decay of the acceptor instead of the donor, the effects of FRET on the decay are avoided, and only photophysical changes arising from contact interactions are measured. The fluorescence lifetime of the donor was not studied here as these decay curves are dominated by the fast energy transfer from FRET.

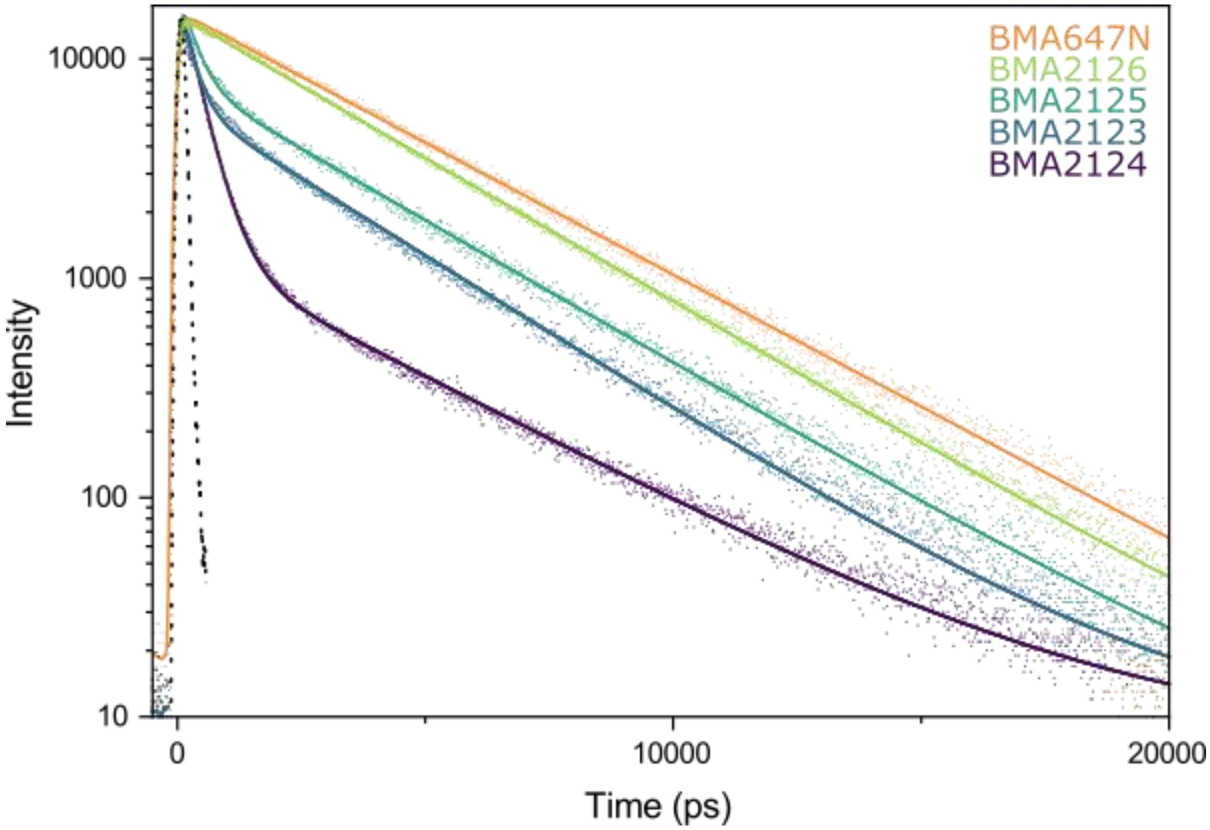


Fig. 3.3.16 TCSPC data showing the fluorescence decay of Atto647N for 4 quenching constructs and an Atto647N only sample, intensity is on a log scale. Quenching samples are shown in increasing compound overlap from purple to green. Atto647N only sample is in orange, instrument response function is shown as a dotted line. The data are fit with a bi-exponential decay model, the parameters of which are in Table 3.3.2.

As can be seen in Fig. 3.3.16, the apparent fluorescence lifetime of Atto 647N decreases for quenching constructs with increased AV overlap with Cy3B, when compared to Atto 647N alone. The fluorescence decay of Atto 647N conjugated to DNA alone appears to be a single exponential with a lifetime of 4.00 ns, however when close enough to Cy3B as to make contact interactions, the decay has a bi-exponential characteristic. The parameters of a bi-exponential fit can be seen in Table. 3.3.2, showing that the longer lived lifetime (τ_{slow}) has a similar decay to that of Atto 647N alone, whereas the "quenched" state exhibits a shorter lifetime, approximately 10 times faster than that of the unquenched Atto 647N.

By plotting the relative amplitudes of these two exponentials against the compound overlaps of the constructs (Fig. 3.3.17), we can again see a similar pattern of increased quenching with increased AV overlap, with a plateau somewhere around 20% compound overlap.

Table 3.3.2 Compound overlaps (O_c) and fitting parameters for the data shown in Fig. 3.3.16

Construct Name	O_c	τ_{fast}	τ_{slow} (ns)	A_{fast}	A_{slow}
BMA647N	0	-	4.00	-	1
BMA2126	0.016	0.76	3.83	0.06	0.94
BMA2125	2.353	0.31	3.98	0.6	0.4
BMA2123	26.95	0.26	3.56	0.65	0.35
BMA2124	35.50	0.38	3.70	0.94	0.06

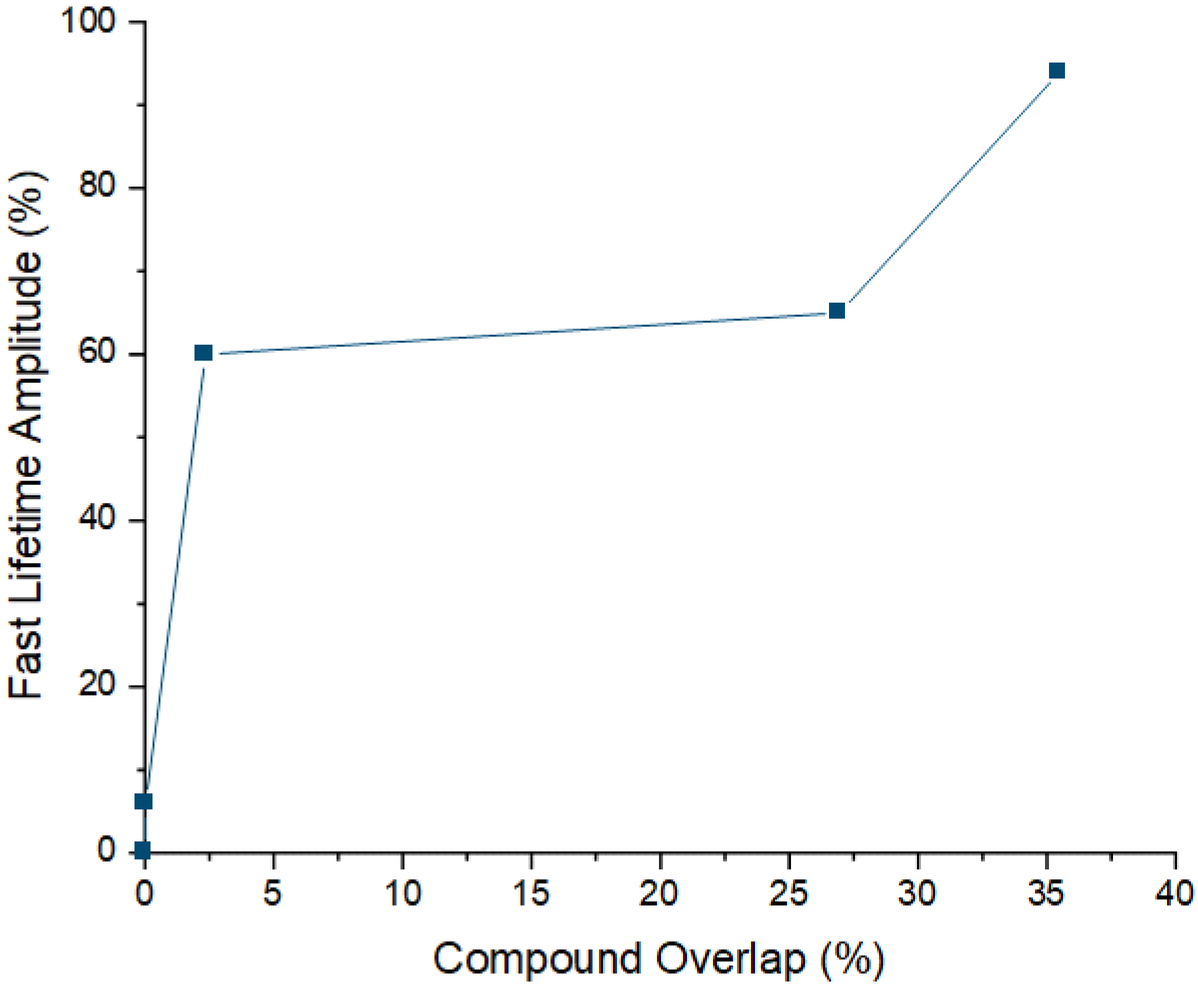


Fig. 3.3.17 Amplitudes of the fast lifetime decay of Atto647N plotted against compound overlap.

3.4. Discussion

Using confocal smFRET ALEX, I have shown that the contact quenching effect previously used to investigate binary conformational states can be sufficiently sensitive to distinguish varying degrees of quenching, rather than simply presence or absence of quenching. The compound overlap metric, based on the degree of intersection between accessible volumes, represents a good predictor of quenching which can potentially be used to build a kinetic model of the process. The addition of a viscogen (50% glycerol) to the system reduces the degree of quenching as expected, and the treatment of this data with a simple kinetic model of the quenching process would suggest that the glycerol decreases the rate of diffusion into the bound state relative to the dissociation rate of the dyes, although it is not clear if this is statistically significant and more data may be required to confirm it. However, the same treatment also suggests that the degree of quenching when in the bound state also decreases, which would not be expected from an increase in viscocity alone, such that some of the decrease in quenching upon addition of glycerol arises from some other interaction.

The spectral shifts of the dyes, both in excitation and emission, suggests that the quenching effect seen at the single-molecule level has a photophysical origin comparable to H-aggregates and J-aggregates seen in bulk precipitates of cyanine dyes[125]. An H-aggregate like origin has been previously suggested for a Cy3 homodimer formation on DNA constructs[126], however in the work presented here the two interacting dye molecules are different from one another, and exhibit spectral shifts in opposite directions. Additionally, the TCSPC data shows a marked decrease in the fluorescence lifetime of the acceptor in the presence of a sufficiently proximal donor, something not typically seen in FRET systems. The shift in spectral peaks coupled with the effect on lifetime of the acceptor corroborates the hypothesis that the contact quenching effect arises from more than simply energy transfer between the dyes.

The smFRET ALEX data presents distinct populations of fluorescence stoichiometry, indicating that the quenched and unquenched states are stable on the millisecond timescale of the experiment. This, paired with the H/J-aggregate like spectral shift, strongly suggests that the quenching phenomenon arises from the dyes “sticking” in a bound state, on the order of milliseconds. However, as the fluorescence spectra have so far only been measured in the ensemble, it is unclear whether the sticking of the dyes can produce two (or more) different photophysical states which would present different single-molecule stoichiometries whilst each contributing to the spectral shifts of either dye, or; a single state which exhibits both spectral shifts simultaneously. In either case, the quenched state(s) must have a low brightness of emission in red when excited by green light, such that the mid-S population is reduced on quenching.

The shifts in ensemble fluorescence spectra and TCSPC curves both represent unique ways of monitoring the contact quenching effect, both at the ensemble level, but also as potential means of observing the phenomenon in greater detail at the single-molecule level. Using pulsed interleaved excitation (PIE), one could obtain the lifetime of the acceptor at the single-molecule level. This information is typically unused in PIE experiments, as the acceptor lifetime does not usually change with FRET, however this may give further insight into the origins of the quenching mechanism, or even a better way of measuring it. Furthermore, by interleaving pulses with very short wavelength separation (perhaps by passing a white light laser through a bandpass filter, then splitting the band down the middle with a longpass/shortpass filter) one could preferentially excite either peak of the quenched and unquenched states. For example, the Cy3B excitation peak appears to shift from 563 to 548 nm, with the 563/548 ratio changing from 0.7 to 1.3 in the quenched state (Fig. 3.3.12). If this could be replicated at the single molecule level by interleaving two pulsed at or around those wavelengths, then the quenched donor might be distinguishable from the unquenched donor.

The qqFRET techniques shown here represent a new and useful addition to the smFRET toolkit, using the same sample preparation and equipment as a regular smFRET experiment, even shorter distance scales are now accessible. Where before, only indistinguishably high FRET efficiencies arise from extreme dye proximity, some degree of quantitative proximity measurement is available to the experimenter. However, to advance the precision of qqFRET, and to achieve measurements that are absolutely quantitative (rather than just relative degrees of quenching) as with accurate FRET correction, a number of questions still need to be answered.

All atom molecular dynamics (MD) may shed light on the association and dissociation of the dyes. Atto 647N and Cy3B, as well as several other FRET dyes, have been parameterised for the CHARMM force field[127]. Furthermore, predictions of spectral properties may be made by performing time dependent density functional theory (TD-DFT) calculations on MD simulation snapshots of the dyes in bound and unbound states (Fig. 3.4.1). Comparison of real fluorescence spectra to simulated spectra from the dyes in different conformations could help determine the photophysical origin of the observed spectral shifts, both in ensemble and seen at the single-molecule level.

In the form presented here, qqFRET is measured by the loss of doubly labelled molecules relative to acceptor only molecules, however, the amount of acceptor only species in a sample will always be non-zero, and sometimes not easily known. In order to access absolute quantities of quenching, either the amount of intrinsically acceptor only species in the sample must be reliably quantified, or an alternative means of measuring qqFRET should be explored. In a sample which can be switched between conformations, this may be possible, as forcing the molecule of interest into a state with zero quenching will give a baseline A_{only} read which can safely assumed to be the same for the conformation exhibiting quenching behaviour, as they are the same sample. Measuring the loss of doubly labelled molecules against the donor only population would also be possible, but introduce similar problems.

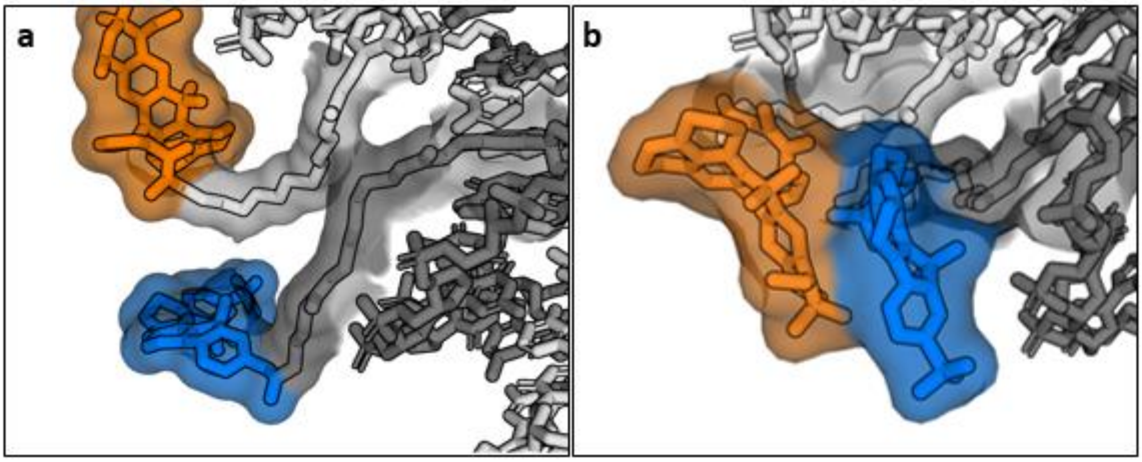


Fig. 3.4.1 Snapshots of a short (50 ns) preliminary all atom MD simulation of Cy3B (blue) and Atto 647N (orange) tethered to a double stranded DNA (grey). **a.** The simulation begins with the dyes apart, however after a few nanoseconds of diffusion they collide and **b.** remain in contact with each other for the remaining duration of the simulation.

Planned work to test this idea involves the maltose binding protein MalE. MalE switches between an open and closed conformation in the presence of maltose[106]. By labelling MalE such that in the closed conformation the dyes are sufficiently close as to quench, the baseline proportion of acceptor only species can be determined in the open state (absence of maltose) and then subtracted from the measurement in the closed state (presence of maltose), to be compared to the expected QR from AV modelling. Not only would this validate a correction based approach, but would test the qqFRET method on a protein based system, as the experiments presented here are only on DNA.

With further work exploring the origins of the contact quenching effect, and development of the means to both predict and measure it, qqFRET techniques may provide an invaluable asset to the family of smFRET methods. Investigations into much smaller protein and nucleic acid systems could proceed without the need for constrictive labelling strategies which avoid dye-dye interaction, and efforts to resolve biomolecular structures by restraining models with accurate FRET measurements could be enhanced with shorter distance measurements wherever necessary. The extreme sensitivity of qqFRET to local structure, as seen with the easily distinguished 1bp distances at 3' and 5' separation, could be exploited to investigate DNA allostery, placing both dyes downstream of a protein binding site or damage region, such that their AV overlap will be altered by small scale DNA conformational changes.

Chapter 4. Applications of the smfBox in DNA-Protein systems

Chapter Overview

In this chapter I use the smfBox to perform smFRET experiments in four different DNA-Protein systems in collaboration with other laboratories. In conjunction with nuclear magnetic resonance (NMR) work, conformations of flap endonuclease-1 (FEN1) are monitored by labelling the arch and saddle regions of the enzyme and measuring distances between them in response to the flapped DNA substrate it is specific to. Alongside molecular dynamics (MD) simulations, the conformational landscape of the helicase RepA is measured, to help elucidate the complex sequence of motions the protein undergoes in order to perform its function of displacing other proteins from single-stranded DNA. DNA substrates for the damage recognition proteins OGG1 and ATL1 are fluorescently labelled and measured to understand how nucleotide level damage affects DNA conformational structure, and recognition by DNA repair systems.

Contributions to work in this chapter

The FEN1 protein used in this chapter was made, labelled and provided by the Grasby lab (University of Sheffield). The RepA protein was provided by the Leake lab (University of York). The ATL1 protein used here was provided by the Williams lab (University of Sheffield) and OGG1 was provided by the Helleday lab (University of Sheffield)

The MD simulations used for AV analysis here were carried out by Antoinette Alevropoulos-Borrill from the Noy lab (University of York).

Publications from work in this chapter

- Bennet, I. A. et al. Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1. *Nucleic Acids Res.* **46**, 5618–5633 (2018).

4.1 Introduction

DNA binding Proteins

As discussed at length in Chapters 1 and 2, single-molecule FRET is a powerful tool for investigating structure and dynamics of biomolecules such as DNA and proteins. By accessing the full conformational landscape of a biomolecule, rather than just the average structure, smFRET can reveal multiple states coexisting at equilibrium, and provide insight into how this equilibrium shifts under various conditions. Additionally, the determination of accurate FRET efficiencies for conversion to absolute (rather than relative) distances can complement data from other structural techniques such as x-ray crystallography, NMR, or molecular dynamics simulations[128,129].

Here I present smFRET data on four DNA-protein systems. In the first two systems (FEN1 and RepA), the behaviour of proteins involved in DNA replication is monitored by dyes conjugated to the proteins themselves. In the next two systems (ATL1 and OGG1), the conformational states of DNA are monitored in response to aberrant nucleotide adducts, and the proteins which recognise them.

FEN1

Fig. 4.1.1 Crystal structure 5UM9[130] of FEN1 in complex with a flapped DNA substrate. DNA is shown in black.

Flap-endonuclease-1 (FEN1) is a human nuclease with 5’ flap endonuclease activity and 5’-3’ exonuclease activity, and is integral to both DNA replication and DNA repair pathways[131,132]. As nucleases like FEN1 cut DNA, it is vital that they have a high degree of specificity to their intended substrate in order to prevent off target activity which could have deleterious consequences for the organism. As such, the question of how nucleases exhibit such a high degree of specificity is an important area of research, as mutations in such proteins can affect genomic stability, having consequences for cancer and neurodegenerative diseases[133].

In the following work, smFRET experiments on FEN1 were used in complement with NMR data (gathered by other researchers in the study) to shed light on the conformational ensemble populated by FEN1 and elucidate the role of arch movements in reaching a catalytically active state specific to its substrate.

RepA

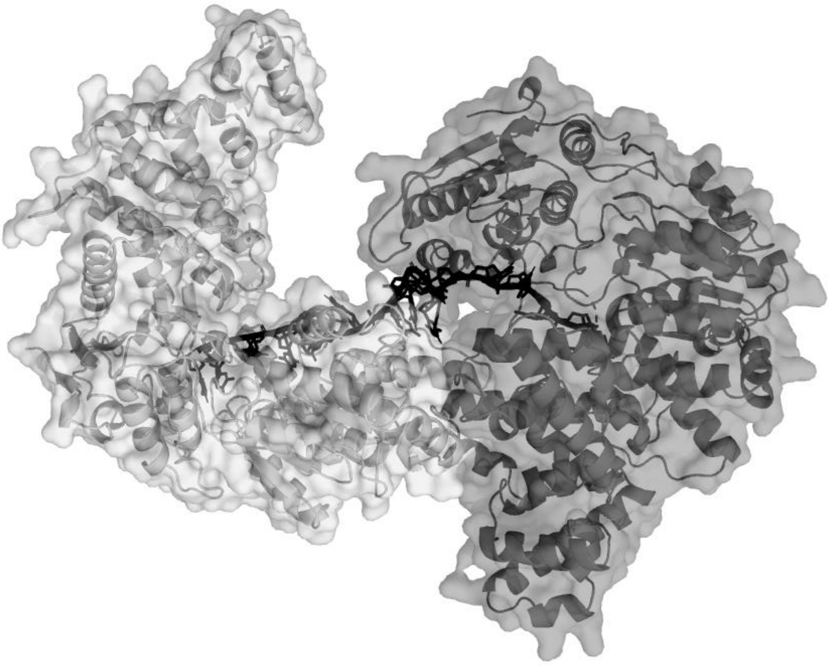


Fig. 4.1.2 Crystal structure 1UAA[134] of RepA helicase in complex with ssDNA (black). The crystal structure contains two copies of the RepA protein; one in an open conformation (light grey), the other in a closed conformation (dark grey) on the same ssDNA. Note the large separation between subdomains in the open conformation on the left.

RepA is an accessory replicative helicase from *Escherichia coli.* During DNA replication, helicases like RepA remove DNA bound proteins ahead of the replication fork which can interfere with the progression of replication machinery[135–137]. RepA moves from 3'-5' along single-stranded DNA on the template leading strand of the replication fork, removing proteins which block replication[136].

RepA has four subdomains, one of which rotates around a hinge to move a significant distance away from the rest of the protein (Fig. 4.1.2). Helicases like RepA hydrolyse nucleotide triphosphates (in the case of RepA; ATP) to translocate along DNA[138], in some cases producing enough force to disrupt even a streptavidin-biotin interaction[139]. However, it is not well understood how conformational changes in the helicases themselves generate this force.

Here I present data on RepA as part of a wider effort to use smFRET techniques and molecular dynamics simulations to understand the conformational landscape of RepA, and how it generates force along DNA.

ATL1 and OGG1

Alkyltransferase-like protein-1 (ATL1) is a *Schizosaccharomyces pombe* protein which recognises O^6-alkylated guanine lesions in DNA[140]. ATL1 does not itself remove these lesions, but rather recruits other proteins to initiate nucleotide excision repair[141]. Removal of alkyl adducts from guanine is crucial to genome stability, as alkylated guanine can mismatch with thymine (rather than cytosine), resulting in sustained mutation after replication. In order to flag up the presence of aberrant alkylation, ATL1 "flips out" bases[142,143] (Fig. 4.1.3a), into a binding pocket for recognition via contacts with chemical groups on the nucleotide that would otherwise be obscured by base pairing interactions on the opposite DNA strand.

8-oxo-guanine DNA glycosylase-1 (OGG1) is a human protein which recognises *and* excises 8-oxo guanine lesions in DNA, arising from oxidation of guanine by naturally occurring reactive oxygen species[144,145]. Unrepaired 8-oxo-guanine can cause mutations, by mismatching to adenine during replication, so repair by 8-oxo-guanine is crucial to genome stability. Like with ATL1, OGG1 flips out bases to recognise its target lesion (Fig. 4.1.3b).

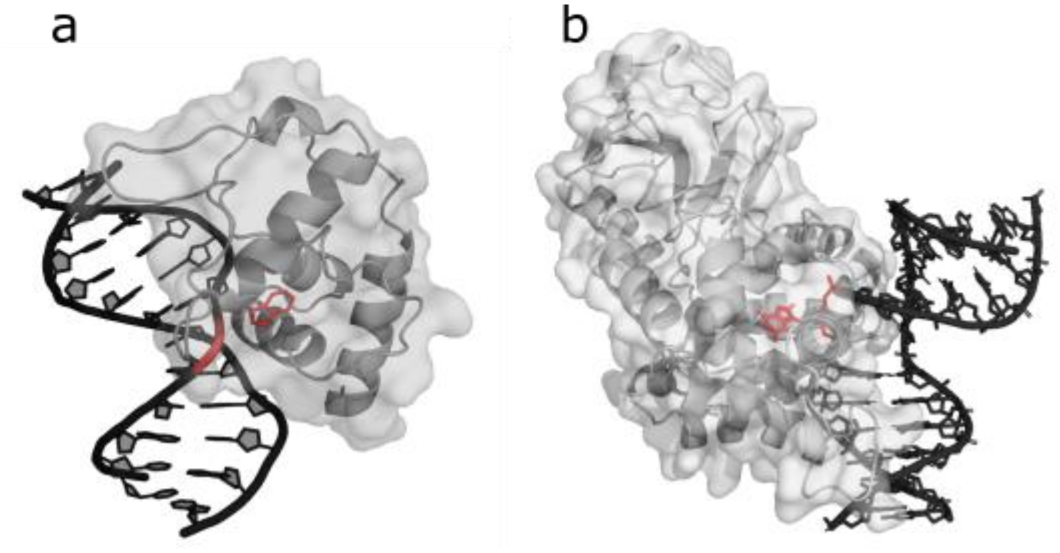


Fig. 4.1.3.a Crystal structure 3GX4[146] of ATL1 (grey) in complex with a DNA construct (black) containing an O^6-methylguanine damage lesion (red) **b.** Crystal structure 2XHI[147] of OGG1 (grey) in complex with a DNA construct (black) containing an 8-oxo-guanine damage lesion

Both ATL1 and OGG1 substrates; DNA with base level damage, appear to exhibit sharp bends, as seen in x-ray crystal structures of these DNAs in complex with their proteins (Fig.4.1.3). An important question in DNA damage recognition is that of whether DNA damage is recognised solely by the local structure of the damage adduct itself, or whether the damage induces a larger conformational change in the double helix structure of the surrounding DNA which drives recognition and repair machinery to the site of the lesion[148–150]. If base level damage induces a change in the conformational landscape of DNA even in the absence of proteins, then this could go some way to explain the rapid speed at which DNA damage recognition proteins can find their targets, seemingly without having to flip out every base along the way[151].

smFRET is an excellent tool to investigate such questions due to its ability to monitor the full conformational landscape of a molecule, without ensemble averaging, and report absolute distances which can be compared to 3D structures from other techniques like x-ray crystallography. In this work I present smFRET data on DNAs containing O^6-methylguanine and 8-oxo-guanine damage, both in the absence and presence of the proteins which recognise such damage.

4.2 Methods and Materials

DNA Constructs

Fluorescently labelled DNA constructs for the ATL1 and OGG1 studies were made by reacting oligonucleotides (Table 4.2.1) obtained from Eurogentec, with internal C6-amino-thymine residues with NHS dyes, purifying, and annealing as described in Chapter 2. In the case of the ATL1 substrate, the oligonucleotide was ordered with an O^6-methyl-guanine base, and for the OGG1 substrate the oligonucleotide contains an 8-oxo-guanine. The amino-modified bases to be labelled were chosen using the AV modelling software Olga[122] so as to produce the greatest possible change in FRET efficiency upon bending. The flap DNA substrate for FEN1 (DF5,1) was obtained from the Grasby lab, made of 3 constituent oligonucleotides Table (4.2.1) annealed in a 1:1:1 ratio.

Table 4.2.1: DNA oligonucleotides used in this chapter to make the DNA constructs in Table 4.2.2

Oligo Name	Sequence	Label
EA1	5'-ACTGACTGACTGACTGCCATG-**O6-Me-dG**-CTAGTAAC-**T(C6-Amino)**-GACTGACTGACTG-3'	Atto 647N on C6 linker
EA2	5'-ACTGACTGACTGACTGCCATGGCTAGTAAC-**T(C6-Amino)**-GACTGACTGACTG-3'	Atto 647N on C6 linker

EA3	5’-CAGTCAGTCAGTCAGTTACTAGCCATGGCAG-**T(C6-Amino)**-CAGTCAGTCAGT-3’	Cy3B on C6 linker
OGGG1-FRET-OG	5’-GCTAGC-**T(C6-Amino)**-ATATACGTCCA-**(8-oxo-G)**-GTCTACTATATACACTCC-3’	Atto 550 on C6 linker
OGGG1-FRET-CTRL	5’-GCTAGC-**T(C6-Amino)**-ATATACGTCCA-**G**GTCTACTATATACACTCC-3’	Atto 550 on C6 linker
OGGG1-FRET-TEMP	5’-GGAGTG-**T(C6-Amino)**-ATATAGTAGACCTGGACGTATATAGCTAGC-3’	Atto 647N on C2 linker
DHPS1	5′-TGAAAGGCAGAGCGCTAGCTCTGCCTTTCGAGCGAAGCTCC3-′	-
F1	5′-dTTTTTACAAGGACTGCTCGACAC-3′	-
T1	5′-GTGTCGAGCAGTCCTTGTGACGACGAAGTCGTCC-3′	

Table 4.2.2: DNA constructs used in this chapter.

Construct Name	Donor strand	Acceptor strand	
ATL1 Undamaged Substrate	EA3	EA2	
ATL1 Damaged Substrate	EA3	EA1	
OGG1 Undamaged Substrate	OGGG1-FRET-CTRL	OGGG1-FRET-TEMP	
OGG1 Damaged Substrate	OGGG1-FRET-OG	OGG1-FRET-TEMP	
DF5,1	DHPS1	F1	T1

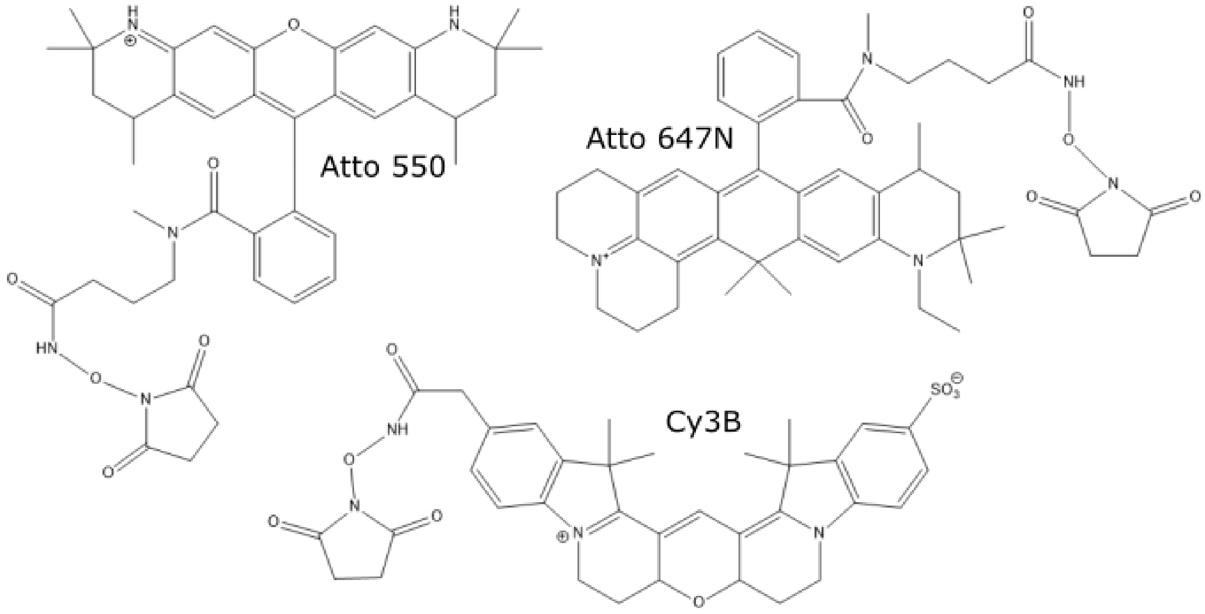


Fig. 4.2.1: Dye NHS esters used in this chapter for labelling DNA.

Proteins

Human FEN1 protein labelled with Cy3B and Atto 647N was provided by the Grasby Lab. Wild type FEN1 contains two solvent accessible cysteines (C235 and C311), thus a quadruple mutant was created with two target residues (E120 and S293) mutated to cysteines and the two solvent accessible cysteines mutated to alanine. Quadruple mutant FEN1 (E120C/C235A/S293C/C311A) was then labelled stochastically with Cy3B and Atto 647N maleimide. To check that the mutations did not seriously affect the function of the protein, activity assays were performed on the quadruple mutant. The quadruple mutant had an initial rate of substrate degradation of more than half that of the wild type. Whilst the activity was lowered by the mutation, it was considered to be acceptably similar to wild type rate so as to still draw meaningful conclusions from experiments.

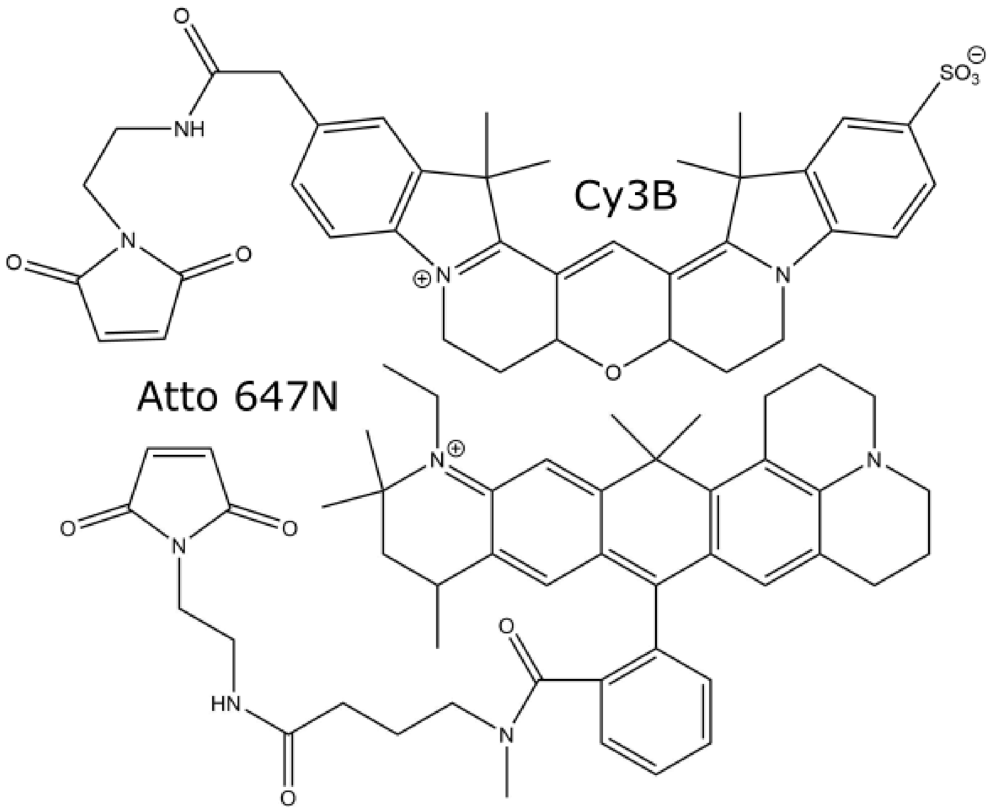


Fig. 4.2.2: Maleimide dyes used for labelling FEN1

Escherichia coli RepA helicase protein labelled with Alexa Fluor 546 and Alexa Fluor 647 was provided by the Leake Lab. Target sites A97 and A473 were mutated to cysteine (A96C/A473C) and labelled stochastically with Alexa Fluor 546 and Alexa Fluor 647 maleimide.

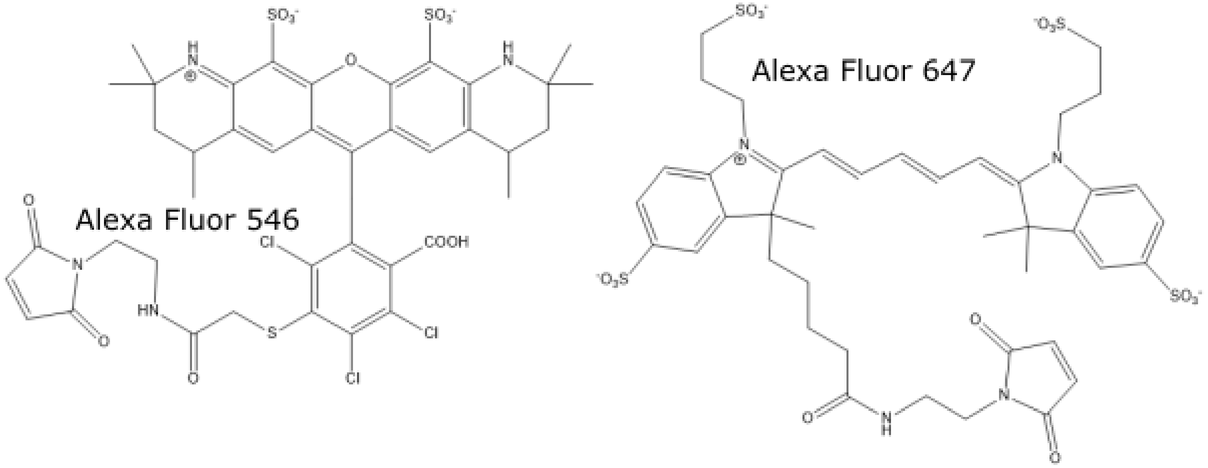


Fig. 4.2.3: Maleimide dyes used for labelling RepA

Schizosacharomyces pombe ATL1 protein was provided by the Williams Lab. Catalytically inactive Human OGG1 protein (triple mutant K249C/C253K/D268N) was provided by the Helleday Lab.

AV modelling

For initial experimental design and prediction of FRET parameters, accessible volume (AV) models were generated by using Olga[122], modelling dye positions onto structures determined by x-ray crystallography. For FEN1, Cy3B and Atto 647N maleimides were modelled onto the structure 5UM9[130] with FRET efficiencies checked for both dye labelling isomers (S293C-A647N/E120C-Cy3B and S293C-Cy3B/E120C-Atto647N), and an R_0 of 62.4 Å was used to calculate predicted FRET efficiencies.

For RepA, Atto 550 (as an analogue of Alexa Fluor 546) and Cy5 (analogue of Alexa Fluor 647) were modelled onto the crystal structure 1UAA[134] which exhibits both an open and closed conformation as a dimer. An R_0 of 68 Å was used to calculate predicted FRET efficiencies. This same procedure was used on MD simulation clusters of RepA generated and provided by the Noy Lab.

For OGG1 and ATL1, straight DNA duplexes were first modelled onto crystal structures (PDB accession codes 2XHI[147] and 3GX4[146] respectively) of the protein with a bound nucleotide substrate, to extend the outwards for labelling. Cy3B NHS and Atto 647N NHS were modelled onto the ATL1 substrate (both a straight DNA, and the bent DNA from the crystal structure). Atto 550 NHS and Atto 647N NHS were modelled onto the OGG1 substrate (both a straight DNA, and the bent DNA from the crystal structure). For ATL1 an R_0 of 62.4 Å for the Cy3B/Atto 647N dye pair was used to calculate predicted FRET efficiencies and for OGG1 and R_0 of 63.2 Å was used for the Atto 550/Atto 647N dye pair.

smFRET Measurements

Single-molecule FRET experiments were conducted using the smfBox as described in Chapter 2. Samples were prepared to approximately 100 pM so as to give 1-5 bursts per second.

For FEN1, the sample was diluted into FEN1 binding buffer (55 mM HEPES-KOH pH 7.5, 110 mM KCl, 8 mM $CaCl_2$, 0.1 mg ml^{-1} BSA, 1 mM DTT) with or without 20 nM DF5,1 substrate. Glass coverslips were pre-passivated with 1 mg ml^{-1} BSA for 5 minutes prior to each measurement, and 90 minutes of data were acquired for both conditions.

For RepA, the sample was diluted into Rep buffer (10 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 0.1 mg ml^{-1} BSA) with either 10 mM or 500 mM NaCl. Coverslips were pre-passivated with 2 mg ml^{-1} BSA prior to each measurement, and 7 hours of data were acquired for each condition.

For ATL1, the DNA substrates were diluted into ATL1 binding buffer (100 mM Tris HCl pH 8.0, 10 mM NaCl, 0.2 mg ml^{-1} BSA, 2 mM DTT, 0.1 mg ml^{-1}) either with or without 50 nM ATL1 protein. 1 hour of data was acquired for both conditions.

For OGG1, the DNA substrates were diluted into OGG1 buffer (50 mM Tris HCl pH 7.5, 50 mM NaCl, 0.1 mg ml^{-1} BSA, 1 mM DTT) either with or without 10 µM OGG1 protein. 1 hour of data was acquired for both conditions.

smFRET Analysis

smFRET data were saved as photon-HDF5[99] files and analysed using FRETBursts[20] Jupyter notebooks described in Chapter 2.

For FEN1, spectral crosstalk parameters were determined by selecting donor only and acceptor only bursts from within the doubly labelled data set ($\alpha = 0.255$, $\delta = 0.084$), and gamma was calculated using DNA standards also labelled with Cy3B and Atto 647N ($\gamma = 1.15$) due to an absence of low FRET species within the FEN1 data and an acceptor heavy aggregate which obfuscated stoichiometry correction. After applying background subtraction and accurate FRET correction parameters, a dual channel burst search (m=10, F=15) was used initially, and then bursts were selected with >40 photons under green excitation, >10 photons under red excitation, and <300 photons in total to remove large aggregates from analysis. Bursts with intermediate stoichiometry (between 0.5 and 0.85) were then plotted as a 1D FRET histogram and fitted with an unrestrained double Gaussian function.

For RepA, spectral crosstalk parameters were determined by selecting donor only and acceptor only bursts from within the doubly labelled data set ($\alpha = 0.051$, $\delta = 0.0769$). Due to the absence of clearly distinct populations to fit 2D Gaussians to, gamma and beta were determined by combining data sets for both salt concentrations and fitting through all molecules ($\gamma = 0.488$, $\beta = 0.790$). After applying background subtraction and accurate FRET correction parameters, a dual channel burst search was used initially (m=10, F = 20), and bursts with >50 photons under green excitation and >50 photons under red excitation were selected. Bursts were then plotted as 1D FRET histograms and fitted with a quadruple Gaussian distribution, with parameters for Gaussian centres and widths linked between both conditions, and amplitudes fitted freely.

For ATL1, spectral crosstalk parameters were determined by selecting donor only and acceptor only bursts from within the doubly labelled data set ($\alpha = 0.158$, $\delta = 0.120$) , and gamma was calculated using DNA standards also labelled with the Cy3B and Atto 647N due to an absence of low and high FRET species ($\gamma = 0.547$, $\beta = 0.496$). After applying background subtraction and accurate FRET correction parameters, a dual channel burst search was used initially (m=10, F = 20), and bursts with >50 photons under green excitation and >50 photons under red excitation were selected. Bursts were then plotted as 1D FRET histograms and fitted with single Gaussian distribution with all parameters free.

For OGG1 background was determined and subtracted, a dual channel burst search was used initially (m=10, F = 20), and bursts with >50 photons under green excitation and >50 photons under red excitation were selected. Bursts were then plotted as 1D FRET histograms and fitted with single Gaussian distribution with all parameters free. Data was not accurate FRET corrected.

4.3 Results

FEN1

Accessible volume (AV) modelling on a crystal structure of FEN1 in complex with a flapped DNA substrate was used to find an optimum location on the mobile “arch” region which could be labelled and give a change in FRET efficiency should it move toward or away from the rest of the protein. The second labelling site was chosen on the “saddle” region, believed to be immobile, ensuring that changes in inter-dye distance (and FRET efficiency) could be attributed only to the movement of the arch.

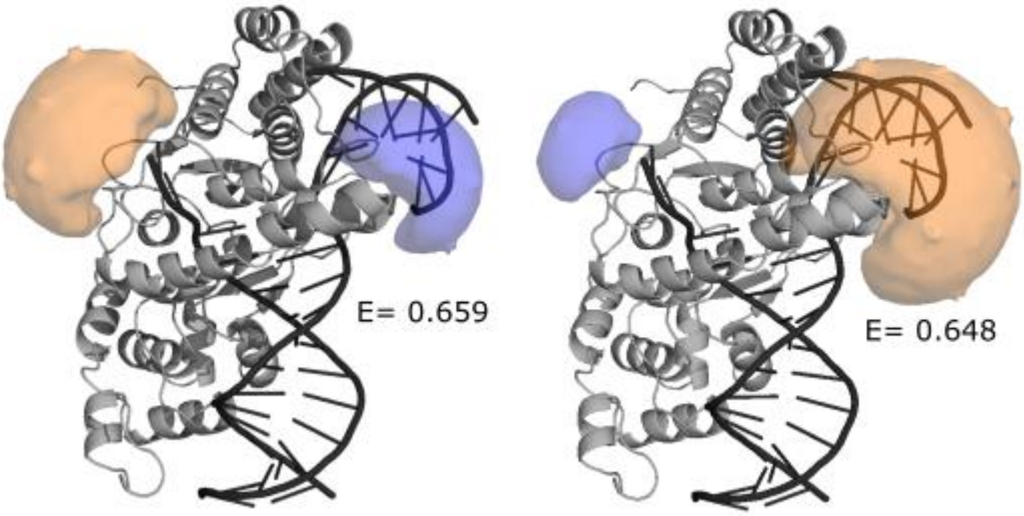


Fig. 4.3.1 FEN1 with AV clouds modelled on at both possible labelling orientations. E120C-Cy3B/S293C-Atto647N is shown on the left and E120C-Atto647N/S293C-Cy3B is on the right. The Atto 647N AV’s are shown in orange, and the Cy3B AV’s are shown in blue, which is much smaller due to the shorter linker length of Cy3B maleimide. Both orientations of the dyes have very similar predicted FRET efficiencies; 0.659 and 0.648.

As the protein is labelled stochastically, such that either dye could react with either labelling site, two such possible labelling orientations were modelled. A double donor, and double acceptor may also occur, however these are rejected from the final data by ALEX, so are irrelevant to FRET calculations. The two AV clouds are of very different sizes, as the linker of Cy3B maleimide is much shorter than that of Atto 647N maleimide, so the two possible doubly labelled species may have different FRET efficiencies if the larger AV cloud is constricted at one labelling site but not the other. The E120C-Cy3B/S293C-Atto647N orientation gives a predicted FRET efficiency of 0.659, whereas the E120C-Atto647N/S293C-Cy3B labelling orientation gives a predicted FRET efficiency of 0.648 (Fig. 4.3.1). The negligible difference between these two FRET efficiencies helpfully ensures that any heterogeneity present in the data will have arisen from structural dynamics rather than the different labelling sites available to the dye conjugation reaction.

Next, labelled FEN1 was measured by single-molecule FRET, both in the absence and presence of the flapped DNA substrate. FRET efficiencies were corrected (as described in Chapter 2), such that FRET peaks in the experimental data can be reliably compared to predictions from AV modelling.

Rather than a single homogenous population under either condition, under both conditions FEN1 displays a broad distribution of FRET efficiencies (Fig. 4.3.2) with a peak at 0.876, much higher than predicted from the AV models on the crystal structure. However, fitting a double Gaussian distribution to both data sets reveals that not only do both conditions exhibit different amplitudes of the same population (Table. 4.3.1), but that the lower FRET population is at 0.645, which is in excellent agreement with the prediction from crystal structure of 0.648. This suggests that the low FRET population is the conformation exhibited by the crystal structures, and both with and without the DNA substrate the arch region can move closer to the centre of the protein.

In the presence of the DNA substrate, the distribution shifts to give slightly more of this high FRET state. The presence of heterogeneity in the distribution at all suggests that there are conformational interconversions on the order of milliseconds (the timescale of a measurement of one molecule) or slower, as the states would average to give a single peak if the rates of interconversion were faster. This is consistent with NMR data which show motions in the arch on a similar time scale[152].

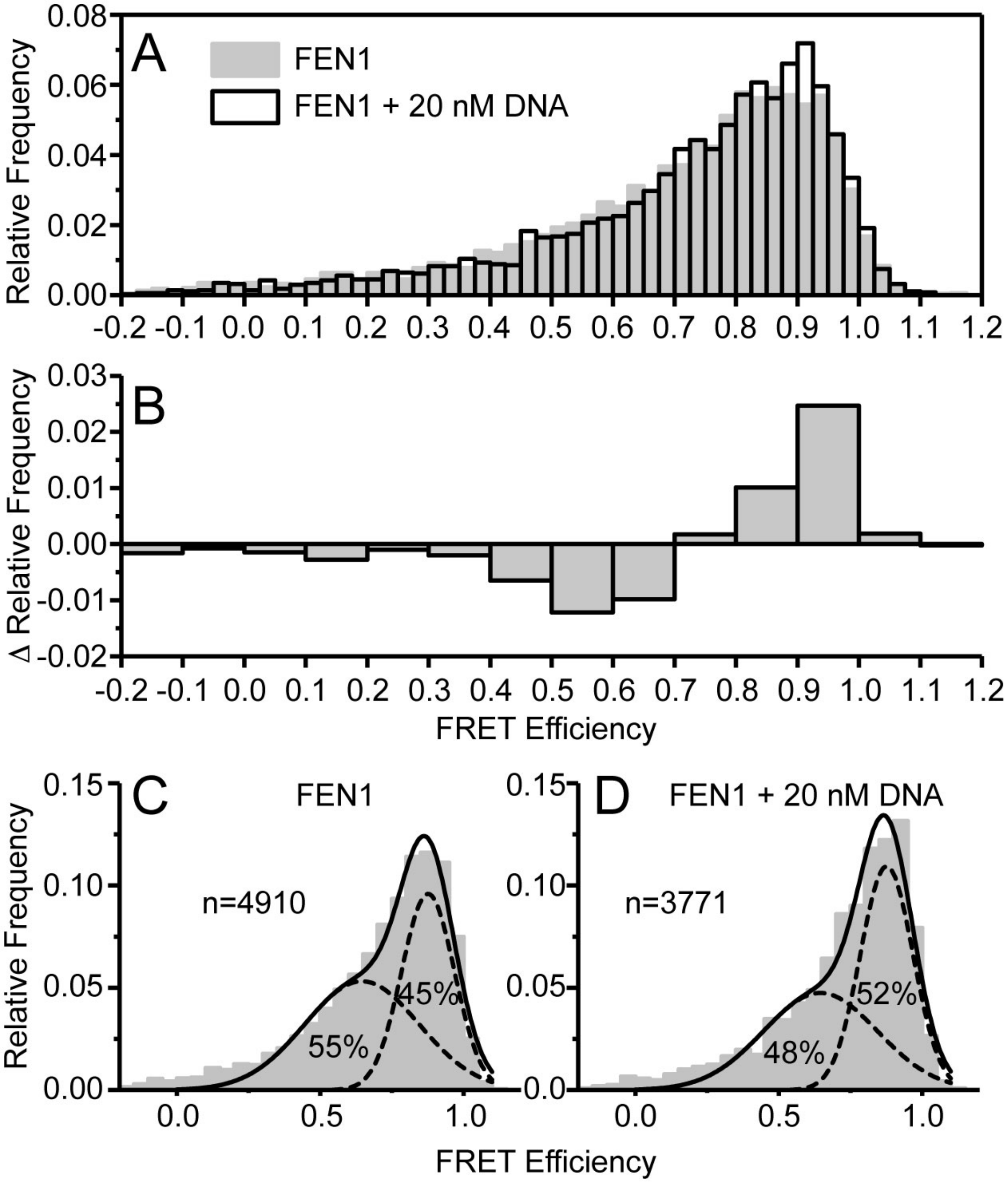


Fig. 4.3.2 FRET Efficiency histograms of FEN1 with and without the DNA substrate. **a.** Relative frequencies of FRET efficiencies for FEN1 alone (grey shaded region) and in complex with 20 nM DNA (black lined region). **b.** The difference in relative frequencies of FRET efficiencies upon addition of 20nM DNA. **c.** FEN1 alone and **d.** FEN1 with 20 nM DNA showing the unrestrained fit to the sum of two Gaussian functions (see Table. 4.3.1 for fitting parameters) to show that upon addition of DNA the lower FRET population decreases and the higher FRET population increases.

Table. 4.3.1: Gaussian fit parameters for FEN1 data shown in Fig. 4.3.2

Data	Mean E	Sigma	Amplitude
FEN1 alone low FRET	0.6455	0.1989	0.5304
FEN1 alone high FRET	0.8762	0.0924	0.4428
+20 nM DNA low FRET	0.6450	0.1990	0.4732
+20 nM DNA high FRET	0.8760	0.0920	0.5044

RepA Helicase

AV modelling initially on a crystal structure of RepA helicase revealed sites which exhibit an extreme change in FRET efficiency upon opening and closing of the protein (Fig. 4.3.3). Whilst the protein is labelled stochastically as with FEN1, Alexa Fluor 546 and Alexa Fluor 647 have very similar sized AV's, so differences in possible labelling isomers should not be a significant source of heterogeneity in FRET. For the crystal structures this was checked, giving 0.293 and 0.295 for the open state, and 0.960 for both combinations in the closed state.

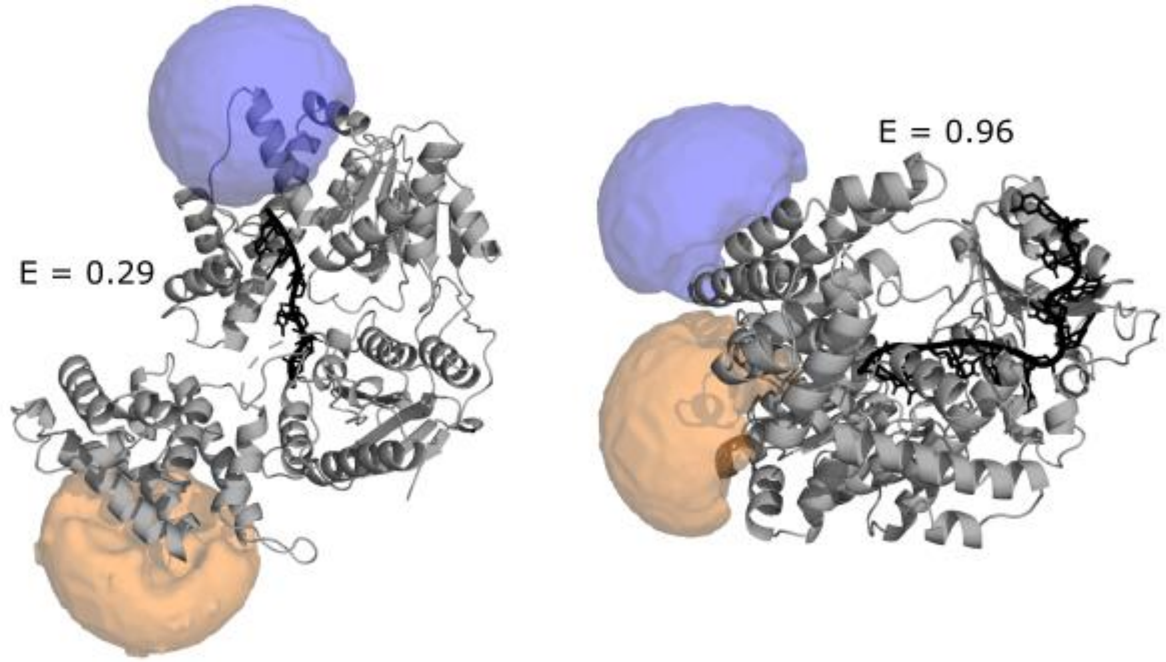


Fig. 4.3.3: AV clouds modelled onto RepA helicase crystal structure (1UAA[134]) in both the open and closed conformation. In the open conformation the donor Alexa Fluor 546 (blue) is distant from the acceptor Alexa Fluor 647 (orange), giving a predicted FRET efficiency of 0.29. In the closed state the dyes are much closer, giving a predicted FRET efficiency of 0.96.

In addition to crystal structures, all-atom molecular dynamics simulations of RepA with implicit solvent were run by the Noy Lab, and clusters were obtained for additional AV modelling to contextualize the data. Whilst the MD simulations also returned a cluster with a very high FRET efficiency in the closed state, and a large number of clusters in open states with low FRET efficiencies, it also returned clusters with intermediate FRET efficiencies at 0.501, 0.550, and 0.858.

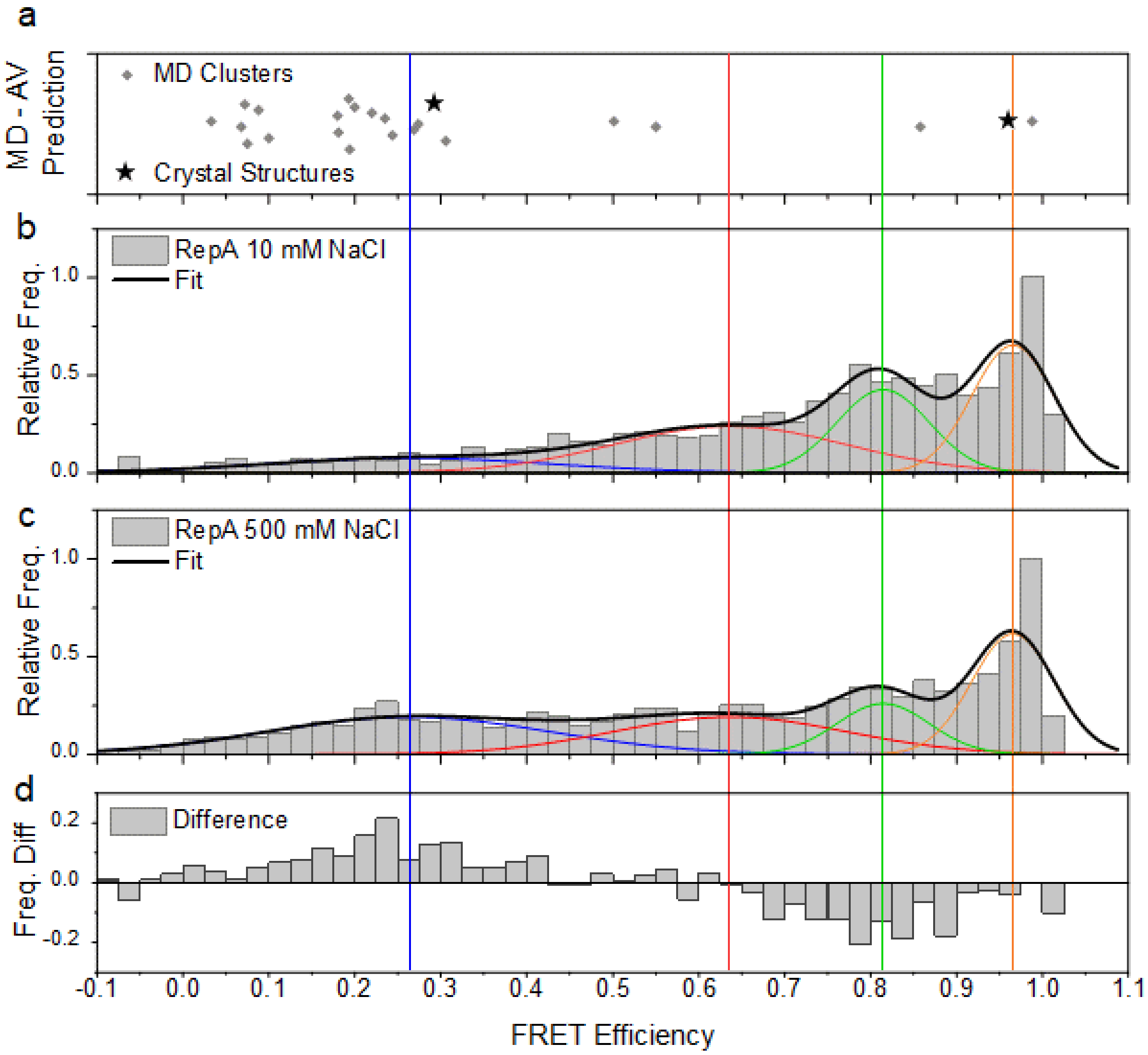


Fig. 4.3.4: FRET efficiency plots of RepA helicase. **a.** Predicted FRET efficiencies from AV's modelled onto MD simulation clusters (grey diamonds) and crystal structures (black stars). **b.** FRET efficiency histogram of RepA in 10 mM NaCl and **c.** 500 mM NaCl, the quadruple Gaussian fits are shown as black lines with component gaussians shown in blue, red, green, and orange. Vertical coloured lines show the center point of each Gaussian. Fit parameters are reported in Table. 4.3.1 **d.** Difference histogram of FRET efficiencies between 10 mM and 500 mM NaCl.

smFRET data were then acquired of RepA helicase both at low and high NaCl concentrations (Fig. 4.3.4)so as to drive the protein into its open and closed states, not only to get sufficient statistics on them but also to aid in accurate FRET correction (as gamma and beta determination works optimally with a wide spread of FRET efficiencies). A quadruple Gaussian distribution was fit to both data sets, with centre and width parameters shared between both.

Not only are the low and high FRET population centres (0.26 and 0.97) in good agreement with prediction from crystal structures, but the two mid FRET populations (0.64 and 0.81) lend credence to the mid FRET clusters shown in the MD simulations. The population at E=0.64 appears to decrease upon increasing salt concentration, and the E=0.26 population increases, whereas the other two remain approximately the same. This would suggest an interconversion between the 0.26 and 0.64 states, the rates of which are altered by increased salt concentration.

Table. 4.3.2: Parameters for quadruple Gaussian fit shown in Fig. 4.3.3

Parameter	Peak1	Peak2	Peak3	Peak4
Center	0.263	0.636	0.814	0.966
Width	0.165	0.136	0.054	0.048
Amplitude (10 mM NaCl)	0.076	0.239	0.426	0.654
Amplitude (500 mM NaCl)	0.191	0.194	0.260	0.617

ATL1

A FRET labelled DNA substrate was designed for ATL1 experiments, starting with a crystal structure of ATL1 in complex with a short DNA with an O^6-methylguanine adduct. The short DNA in the crystal structure was extended by aligning straight duplex DNA structures to either end, revealing a distinct bend (Fig. 4.3.5). AV’s were then modelled onto the two aligned straight DNAs to find labelling positions which give an optimum change in FRET efficiency between an undamaged duplex DNA and the bent state. Note that as each strand of the DNA is labelled separately (see Chapter 2 methods for protocol), the reverse labelling orientation is not possible as with the protein systems described above, and so do not need to be checked.

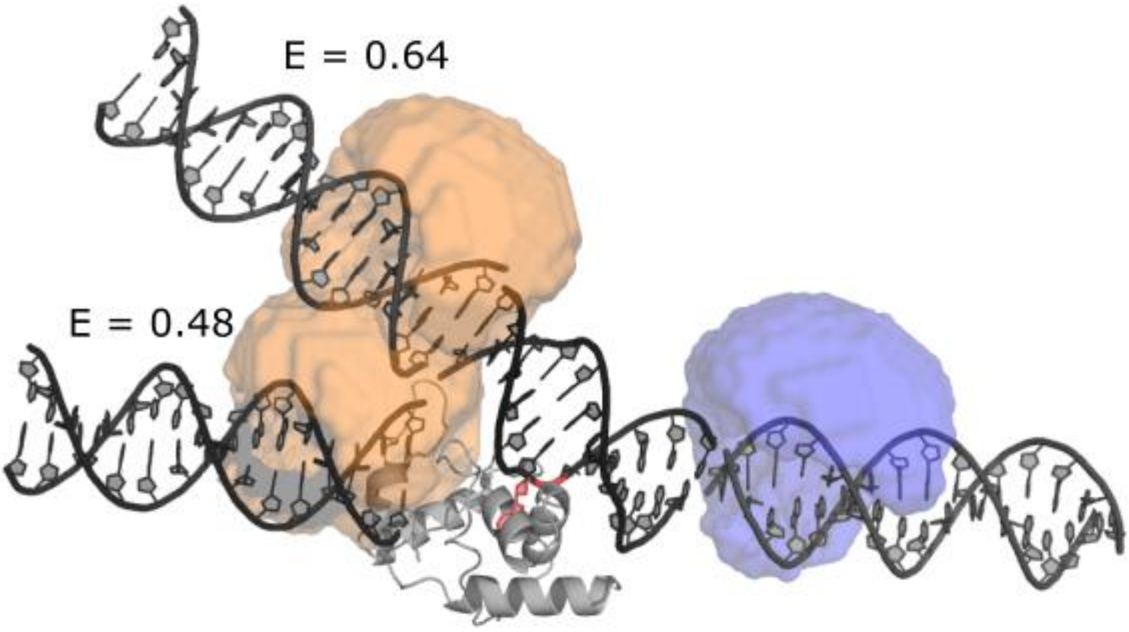


Fig. 4.3.5 AV clouds modelled onto the ATL1 DNA construct in both the bent and straight conformation. The straight DNA is shown here with the middle section hidden. In the bent conformation shown in the ATL1 crystal structure, the acceptor Atto 647N (orange) moves closer to the donor Cy3B (blue). The predicted FRET efficiencies for the straight and bent states are 0.48 and 0.64 respectively.

Next, smFRET was used to measure the labelled DNA construct containing an O^6-methylguanine, with and without ATL1, with an undamaged duplex DNA used as a control. The three conditions each displayed single FRET efficiency peaks (Fig. 4.3.6), however these may be distinct populations so close as to appear as a single broad peak, or populations which interconvert faster than the timescale of the experiment and so average out, or rather some combination of both.

The undamaged duplex DNA displays a FRET efficiency of 0.512, rather than the 0.48 predicted by AV modelling, however this is within a comparable level of error for a duplex DNA to the standards used in Chapter 2. Curiously, the O^6mG damage causes a decrease in FRET efficiency, whereas the protein increases the FRET efficiency, as expected from the AV modelling, but not as high (0.558 vs 0.64).

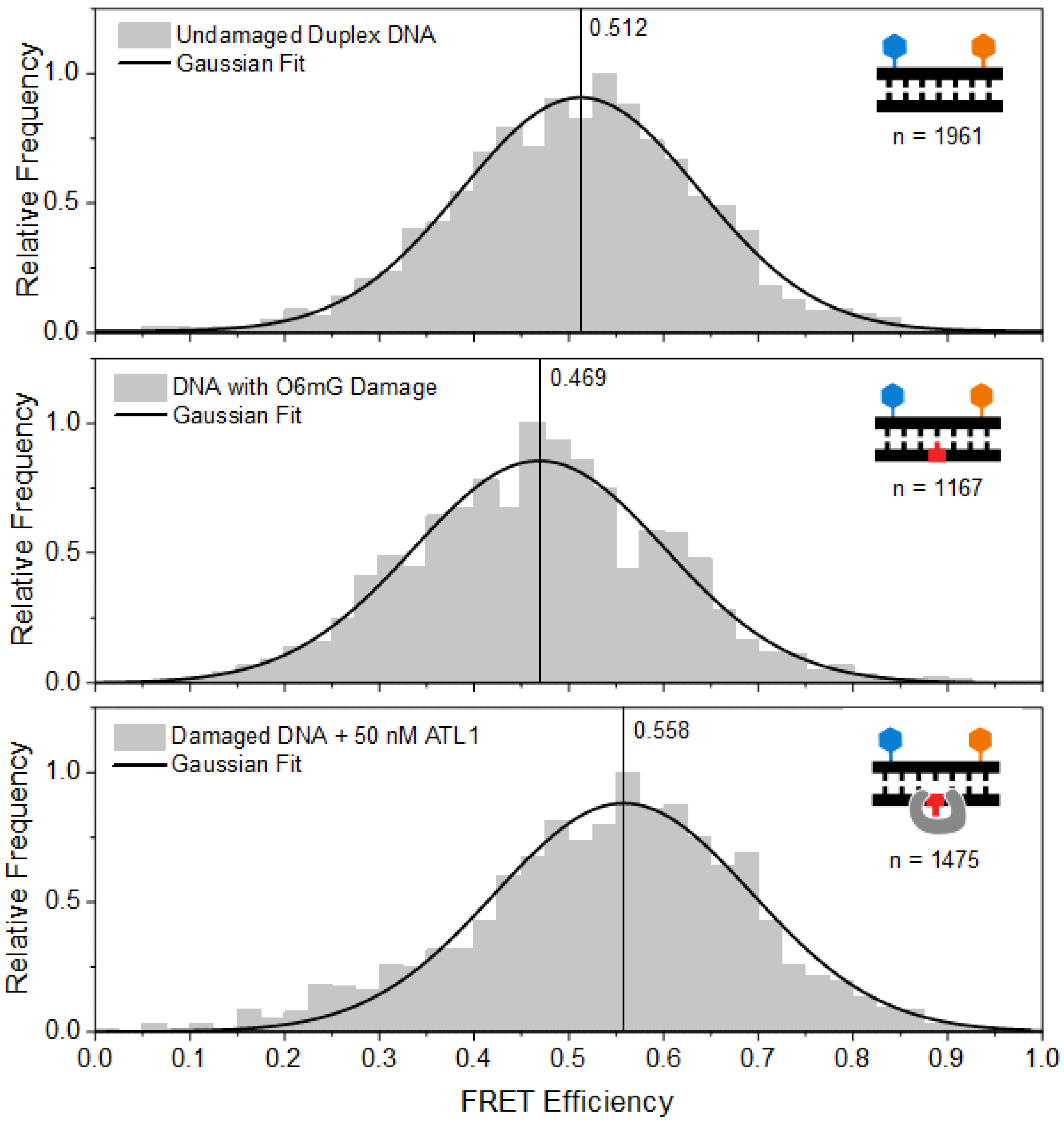


Fig. 4.3.6 Accurate FRET efficiency histograms for an undamaged duplex DNA (top), DNA with O^6-methylguanine damage (middle), and the same damaged DNA in the presence of 50 nM ATL1. Single Gaussian fits are shown in thick black lines, and vertical lines show the centre point of the fitted Gaussian.

The much smaller than expected change in FRET efficiency upon ATL1 binding could be for any number of reasons. The substrate may interconvert between a bent and straight structure within the ATL1 complex, thus giving an intermediate peak as discussed above. Additionally, the AV models from the crystal structures may underestimate the amount of additional coiling (or uncoiling) induced by the protein. Whilst a bend in the DNA toward the donor will bring the acceptor closer, a twist may move it further away. In any case, the decreased FRET efficiency for the damaged DNA, even in the absence of protein, strongly demonstrates that an O^6-methylguanine induces a larger scale structural rearrangement in the double helix.

OGG1

A FRET labelled DNA substrate was designed for OGG1 in much the same way as for ATL1. A crystal structure of a bent DNA substrate in complex with OGG1 was extended by aligning duplex DNAs. Optimal labelling positions for Atto 550 and Atto 647N were found, giving a FRET efficiency of 0.23 for a straight DNA and 0.62 in the bent state displayed in the crystal structure (Fig 4.3.7).

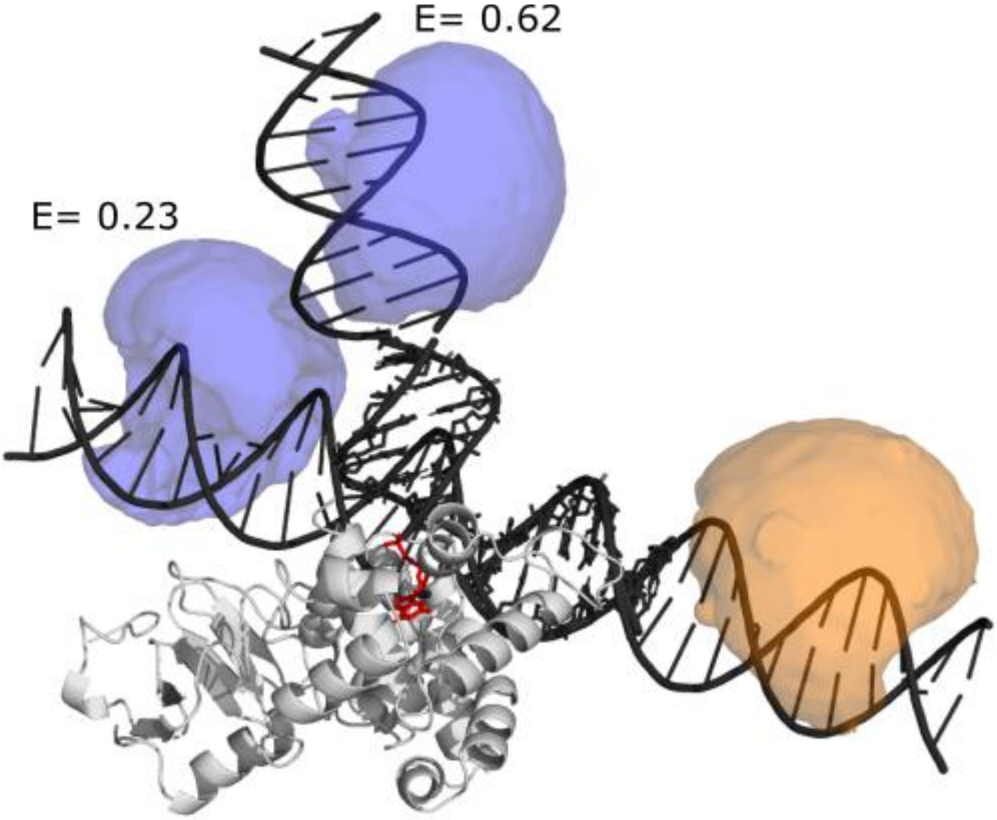


Fig. 4.3.7 AV clouds modelled onto the OGG1 DNA construct in both the bent and straight conformation. The conformation of the bent DNA is created by aligning two duplex DNA's to the ends of the DNA substrate in the crystal structure. In the bent conformation shown in the OGG1 crystal structure, the acceptor donor Atto 550 (blue) moves closer to the acceptor Atto 647N. The predicted FRET efficiencies for the straight and bent states are 0.23 and 0.62 respectively.

smFRET data for these labelled DNA substrates were then acquired, for the undamaged duplex DNA and dNA with 8-oxo-G damage lesion, both with and without OGG1 (Fig. 4.3.8). Single Gaussian curves were fit to the data to quantify both the peak FRET efficiency and distribution width. None of the four conditions showed any significant increase in FRET efficiency as expected from the AV modelling (0.62), and the damaged DNA showed no apparent change at all compared to the undamaged duplex, even with the OGG1 protein. Interestingly however, the undamaged DNA appeared to show an increase both in FRET efficiency and distribution width in the presence of the OGG1 protein.

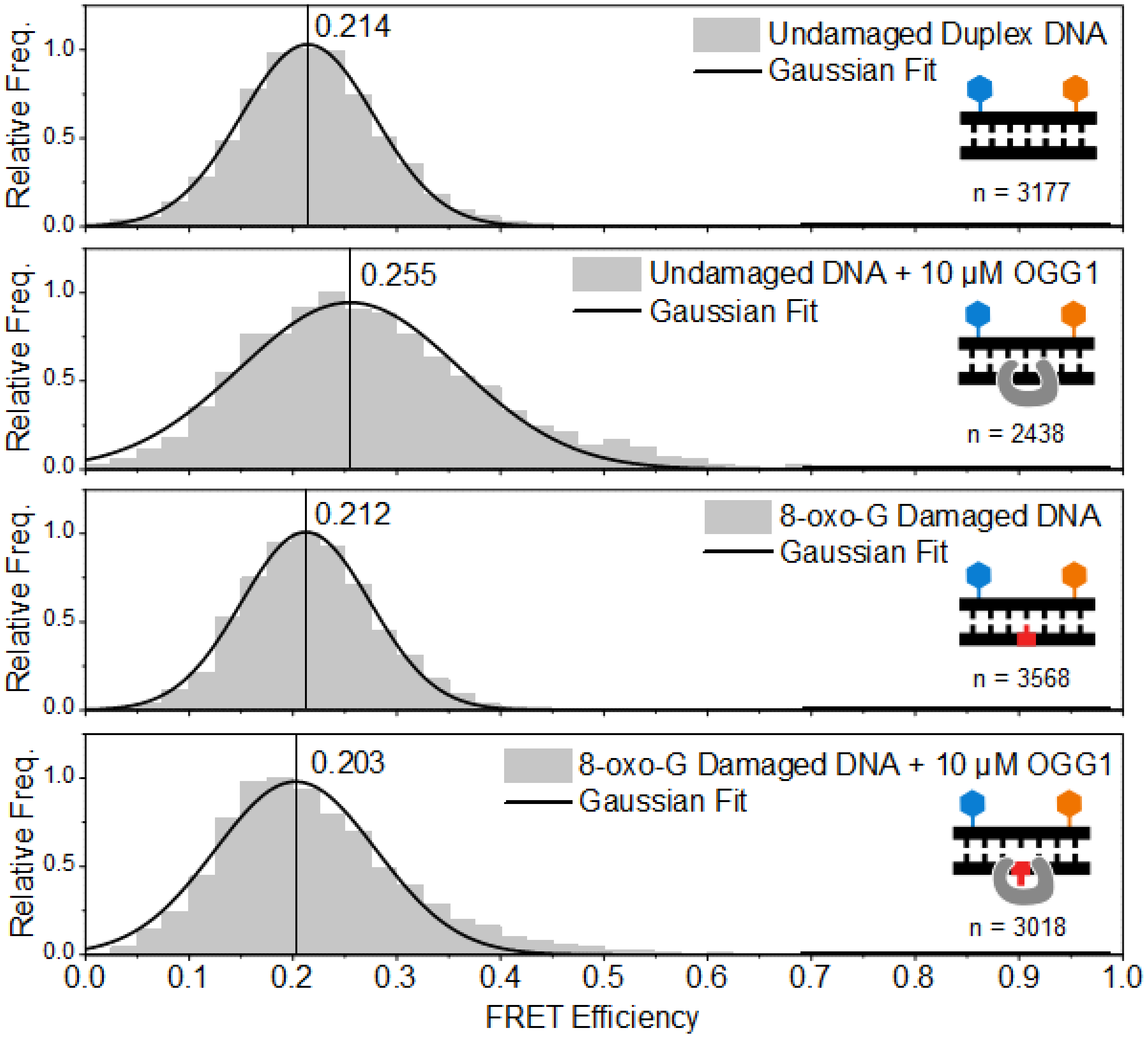


Fig. 4.3.8 FRET efficiency histograms for OGG1 substrates. From top to bottom; Undamaged DNA alone, Undamaged DNA with 10 µM OGG1, Damaged DNA alone, and Damaged DNA with 10 µM OGG1. Single Gaussian fits are shown in thick black lines, and vertical lines show the centre point of the fitted Gaussian.

The lack of any change at all with the 8-oxo-G damage may mean that the presence of the lesion simply does not affect the larger structure of the DNA double helix, or as described above, it may induce both a twist and a bend which cancel out to give no detectable change. The unexpected conformational heterogeneity displayed by the undamaged DNA in complex with OGG1 potentially suggests that the protein will also induce a bend in DNA that does not contain damage, but this bend is transient and reaches a third conformation not detectable by this assay if the lesion is recognised by the protein.

4.4 Discussion

In this chapter I have presented a wide array of smFRET data on different biomolecular systems. Not only does this display the applicability of the smfBox as a powerful system for investigating DNA-Protein systems, but adds to the growing list of applications of smFRET techniques in general.

In particular, the accurate FRET determined distances recovered in this work, in FEN1 and RepA, match not only predictions from existing x-ray structures, but also reveal additional states in their conformational landscapes which would otherwise be hidden by ensemble measurements. As smFRET works on molecules in liquid solutions at room temperature, it can capture transient states which may not even crystallise at all, making accurate FRET an excellent tool to cover the blind spots in crystallography work.

In FEN1 I have revealed the FRET efficiencies corresponding to open and closed states of the arch, which coexist in equilibrium, both alone and with DNA bound. This corroborates NMR data indicating arch flexibility, however it is not clear whether the shift in equilibrium upon DNA binding occurs due to an increase in the forward rate, or a decrease in the backward rate out of the closed conformation. Furthermore, if there are subtle differences in the closed state with absence and presence of protein which are further responsible for the specificity of its enzymatic activity, then the labelling positions used here have not revealed them. To this end, surface immobilisation experiments may be useful in capturing the dwell times in either state, and additional labelling positions may further restrain models of arch movement over the active site.

In RepA, there appears to be four populations of FRET efficiencies. The existence of four separate states is corroborated by TIRF-smFRET data on UvrD (another similar helicase) which also shows four states[153]. Accurate FRET distances from two of these states correspond well to x-ray structures of RepA in the open and closed states, lending credibility to the distances inferred by the other two states, which in turn correspond well to clusters seen in MD simulations. Going forward, the assignment of each FRET efficiency population to clusters seen in simulations will aid in interpretation of further smFRET measurements of RepA in the presence of nucleotides (ATP, and non-hydrolysable analogues) and various DNA substrates.

The work shown here with ATL1 and OGG1 show that confocal smFRET may be a useful tool to capture conformational changes induced by base level DNA adducts. In the case of ATL1 and O^6-methylguanine, the damaged DNA even alone exhibited a significant change in FRET efficiency when compared to an undamaged DNA with the same sequence. This would strongly suggest that the aberrant base does indeed induce a change on the wider conformational landscape of the surrounding DNA structure even before the binding of ATL1. However, upon binding of ATL1, the change in FRET efficiency is not quite as large as would be expected from crystal structures, suggesting that the crystal structure alone does not capture the full conformational ensemble of the complex. The OGG1 data is less straightforward to interpret, with no significant changes in FRET efficiency at all, except for the undamaged DNA in the presence of the protein. This could be for any number of reasons, as discussed above, there may be both twists and bends involved in the conformational changes which prevent the labelling positions used here from being sensitive to the change. With regard to the undamaged DNA in the presence of protein, it may be the case that in the absence of an aberrant base to bind to, the protein continues to interrogate all other bases on the sequence, resulting in the heterogeneity seen here.

In the case of both ATL1 and OGG1, measurements of substrates with alternative labelling positions may provide more sensitivity to structural changes. With enough labelling positions, accurate FRET distances could be used to restrain the three dimensional structure of the DNA duplex under each condition, or validate MD structures in a more integrative manner as with the RepA work presented here.

Chapter 5. Concluding Remarks

5.1 Impact of this work

In the previous three chapters I have presented a new open-source instrumentation platform for confocal smFRET, used it to advance a new single-molecule fluorescence technique, and applied it to a number of DNA protein systems.

With other FRET labs around the world, we have shown that smFRET can be used to reliably determine accurate FRET efficiencies for absolute distances rather than just for relative distances changes. The fluorescently labelled DNA standards used in this work have also been sent out to other laboratories not involved in the original study, as a benchmark against which they can test the capabilities of their own instruments. This work has not only validated the accurate FRET technique, but also proven that the smfBox, a cost-effective and custom built microscope can reproduce such values as well as other commercially available microscopes. The wider biophysics community now has access to confocal smFRET techniques at just a fraction of the cost of purchasing a purpose built confocal system. It is hoped that the smfBox will present research opportunities not just to our own lab, but to other laboratories around the world, some of whom have already been in contact with us to build smfBoxes of their own.

The work on quantitative quenching FRET (qqFRET) shown in Chapter 3 represents a new tool in the arsenal of techniques available to single-molecule fluorescence spectroscopists for investigating biomolecular systems. The contact quenching phenomenon can be reliably measured with ALEX enabled confocal microscopes like the smfBox, and uses exactly the same sample preparation as a standard single-molecule FRET experiment, only altering the distance between dyes. In fact, the ability to discriminate between distances at extremely close range with qqFRET essentially means that placing dyes sufficiently far apart so as to make them distinguishable by FRET becomes less of an issue, as qqFRET becomes more sensitive in the same region that FRET becomes less sensitive. Whilst qqFRET to some extent abandons the sensitivity to heterogeneity given by regular smFRET (as QR is calculated as an average of the whole sample), for single conformations it represents an extension of the effective range of an smFRET experiment. The kinetic model and AV overlap metric I have presented to explain the contact quenching effect, whilst likely over-simplistic, will form the basis of further investigation to fully understand the relationship between dye labelling positions and observed quenching.

By employing the smfBox system to investigate various DNA-protein systems in collaboration with other biophysics labs, I have not just furthered research in those areas, but also demonstrated the applicability of an open-source microscope for serious research in single-molecule fluorescence microscopy.

The presented work on FEN1 has corroborated data obtained by NMR, revealing how complex regional flexibility in the arch of the protein contributes towards it's specificity towards flapped DNA substrates, and prevents off target effects hydrolysing phosphodiester bonds in normal duplex DNA. In RepA, I have experimentally confirmed the existence of two new states seen in molecular dynamics simulations in addition to the two seen in structures from x-ray crystallography. This will pave the way for further experiments determining how conformational rearrangements in RepA generate the force needed to process along single stranded DNA and perform its function of displacing blockers to replication machinery.

Absolute FRET efficiencies measured on O^6-methylguanine containing DNA's have provided new evidence that base-level DNA damage adducts can affect the conformational landscape of DNA structure, which may be involved in recognition by DNA repair proteins. Similarly, changes in the conformational structure of an undamaged duplex DNA in response to OGG1 have revealed possible bending of the DNA by OGG1 even in the absence of damage. However, no significant changes are seen with damage, either alone or with the OGG1, so this system may require alternative labelling strategies to reveal the bent state seen in x-ray structures. These data, while only a starting point for more comprehensive investigations, have strong implications for DNA damage recognition mechanisms, potentially suggesting a much more active role in the structure of DNA itself in flagging up damage to host repair machinery.

5.2 Future work

Work is already underway to replicate the results of the accurate FRET benchmarking study in a protein system. The protein MalE, exhibits very different conformations in the presence and absence of maltose, and so labelling this protein gives access to not only two distances to be tested for benchmarking per labelling combination, but also allows for validating measurements of kinetic rates at select maltose concentrations. This will bolster the confidence of the wider biophysics community that accurate FRET works just as well in protein as it does on nucleic acids, whilst also providing some indication of the replicability of kinetics measurements with smFRET. Whilst the hairpin data shown with the smfBox in Chapter 2 shows the measurement-to-measurement precision of kinetic rates on the same instrument, a multi-lab study will reveal to what degree the use of *different* instruments will affect these rates.

The MalE system planned to be used for further inter-lab benchmarking will also represent a good test system for the qqFRET technique presented in Chapter 3. Preliminary AV calculations have revealed available labelling positions on MalE which will yield transitions to and from zero quenching to high quenching, and between varying degrees of quenching, upon addition of maltose. This will allow for qqFRET measurements with different degrees of AV overlap, that are free of variation in singly-labelled populations; as it will be the same labelled protein under different buffer conditions.

In addition to testing in protein systems, molecular dynamics simulations of DNAs labelled with Cy3B and Atto 647N will further aid in understanding of the quenching system. The motions of the dyes when tethered to DNA will aid in refining AV overlap calculations (as well as being of interest to any application of AV's in general), and snapshots of these simulations will act as starting points for TD-DFT calculations, providing insight into the photophysical origins of the contact quenching effect. Predictions from these calculations can be compared against the steady state spectra already obtained and shown here, or against transient absorption spectroscopy measurements planned to be taken in future experiments.

The qqFRET technique, with further development, may prove instrumental in investigating DNA protein systems like those described above. The extreme sensitivity of qqFRET to subtle differences in local structure could make it an excellent tool in the field of nucleic acid allostery, reporting on structural rearrangements of DNA and RNA downstream of protein binding or genomic damage.

Further development of the smfBox may also prove to be equally important both to qqFRET and DNA damage recognition studies. The expansion of the smfBox to include pulsed lasers and additional detectors will allow for PIE and MFD, as well as a host of additional techniques in the wider field of single-molecule fluorescence spectroscopy. Further expansions could provide additional, better ways to measure contact quenching, and allow for measuring fast conformational changes in DNA structure, via techniques such as nano-second FCS.

All in all, single-molecule biophysics has exciting times ahead, with new techniques for measuring shorter distances, more accessible instrumentation for implementing them, and a growing list of DNA-protein systems in which to use them.

References

1. Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. (Kluwer Academic/Plenum Publishers, 1999).
2. Lerner, E. *et al.* Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **359**, (2018).
3. Lerner, E. *et al.* FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *eLife* **10**, e60416 (2021).
4. Cario, G. & Frank, J. Über Entstehung wahrer Lichtabsorption un scheinbare Koppelung von Quantensprügen. *Z Phys.* **10**, 185–199 (1922).
5. Perrin, F. Théorie quantique des transferts d'activation entre molécules de même espèce. Cas des solutions fluorescents. *Ann Phys Paris* **17**, 283–314 (1936).
6. Förster, T. Energie Transport und Fluoreszenz. *Naturewissenschaften* **6**, 166–175.
7. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Phys* **2**, 55–75 (1948).
8. Muñoz-Losa, A., Curutchet, C., Krueger, B. P., Hartsell, L. R. & Mennucci, B. Fretting about FRET: Failure of the Ideal Dipole Approximation. *Biophys. J.* **96**, 4779–4788 (2009).
9. VanDerMeer, B. W. Kappaphobia is the elephant in the fret room. *Methods Appl. Fluoresc.* **8**, 030401 (2020).
10. VanBeek, D. B., Zwier, M. C., Shorb, J. M. & Krueger, B. P. Fretting about FRET: Correlation between κ and R. *Biophys. J.* **92**, 4168–4178 (2007).
11. van der Meer, B. W. Kappa-squared: from nuisance to new sense. *Rev. Mol. Biotechnol.* **82**, 181–196 (2002).
12. Eilert, T., Kallis, E., Nagy, J., Röcker, C. & Michaelis, J. Complete Kinetic Theory of FRET. *J. Phys. Chem. B* **122**, 11677–11694 (2018).
13. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 719–726 (1967).

14. Hellenkamp, B. *et al.* Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat. Methods* **15**, 669–676 (2018).

15. Plochowietz, A., El-Sagheer, A. H., Brown, T. & Kapanidis, A. N. Stable end-sealed DNA as robust nano-rulers for in vivo single-molecule fluorescence. *Chem. Sci.* **7**, 4418–4422 (2016).

16. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci.* **105**, 11176–11181 (2008).

17. Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M. & Eaton, W. A. Polyproline and the “spectroscopic ruler” revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci.* **102**, 2754–2759 (2005).

18. Lee, N. K. *et al.* Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation. *Biophys. J.* **88**, 2939–2953 (2005).

19. Ha, T. Single-Molecule FRET. *Single Mol.* **2**, 283–284 (2001).

20. Ingargiola, A., Lerner, E., Chung, S., Weiss, S. & Michalet, X. FRETBursts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET. *PLOS ONE* **11**, e0160716 (2016).

21. Hohlbein, J., Craggs, T. D. & Cordes, T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* **43**, 1156–1171 (2014).

22. Lee, S., Lee, J. & Hohng, S. Single-Molecule Three-Color FRET with Both Negligible Spectral Overlap and Long Observation Time. *PLOS ONE* **5**, e12270 (2010).

23. Hohng, S., Joo, C. & Ha, T. Single-Molecule Three-Color FRET. *Biophys. J.* **87**, 1328–1337 (2004).

24. Shrestha, D., Jenei, A., Nagy, P., Vereb, G. & Szöllősi, J. Understanding FRET as a Research Tool for Cellular Studies. *Int. J. Mol. Sci.* **16**, 6718–6756 (2015).

25. Piston, D. W. & Kremers, G.-J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32**, 407–414 (2007).

26. Broussard, J. A. & Green, K. J. Research Techniques Made Simple: Methodology and Applications of Förster Resonance Energy Transfer (FRET) Microscopy. *J. Invest. Dermatol.* **137**, e185–e191 (2017).

27. Pollok, B. A. & Heim, R. Using GFP in FRET-based applications. *Trends Cell Biol.* **9**, 57–60 (1999).

28. Marvin, J. S. *et al.* An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat. Methods* **10**, 162–170 (2013).

29. Jin, S., Ellis, E., Veetil, J. V., Yao, H. & Ye, K. Visualization of human immunodeficiency virus protease inhibition using a novel Förster resonance energy transfer molecular probe. *Biotechnol. Prog.* **27**, 1107–1114 (2011).

30. Gomes, G.-N. & Gradinaru, C. C. Insights into the conformations and dynamics of intrinsically disordered proteins using single-molecule fluorescence. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1865**, 1696–1706 (2017).

31. Hohlbein, J. *et al.* Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nat. Commun.* **4**, 2131 (2013).

32. Nagy, J. *et al.* Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Commun.* **6**, 6161 (2015).

33. Jahn, M. *et al.* The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function. *Proc. Natl. Acad. Sci.* **111**, 17881–17886 (2014).

34. Aznauryan, M., Søndergaard, S., Noer, S. L., Schiøtt, B. & Birkedal, V. A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Res.* **44**, 11024–11032 (2016).

35. Kilic, S. *et al.* Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. *Nat. Commun.* **9**, 235 (2018).

36. Dolino, D. M., Adariani, S. R., Shaikh, S. A., Jayaraman, V. & Sanabria, H. Conformational Selection and Submillisecond Dynamics of the Ligand-binding Domain of the N-Methyl-d-aspartate Receptor *. *J. Biol. Chem.* **291**, 16175–16185 (2016).

37. Lamb, D. C. Single-Pair FRET: An Overview with Recent Applications and Future Perspectives. in *Single Particle Tracking and Single Molecule Energy Transfer* 97–129 (John Wiley & Sons, Ltd, 2009). doi:10.1002/9783527628360.ch5.

38. Betzig, E. & Chichester, R. J. Single Molecules Observed by Near-Field Scanning Optical Microscopy. *Science* **262**, 1422–1425 (1993).

39. Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).

40. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci.* **93**, 6264–6268 (1996).

41. Kapanidis, A. N. *et al.* Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci.* **101**, 8936–8941 (2004).

42. Deniz, A. A. *et al.* Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *Proc. Natl. Acad. Sci.* **96**, 3670–3675 (1999).

43. Müller, B. K., Zaychikov, E., Bräuchle, C. & Lamb, D. C. Pulsed Interleaved Excitation. *Biophys. J.* **89**, 3508–3522 (2005).

44. Kühnemuth, R. & Seidel, C. A. M. Principles of Single Molecule Multiparameter Fluorescence Spectroscopy. *Single Mol.* **2**, 251–254 (2001).

45. Nettels, D., Gopich, I. V., Hoffmann, A. & Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci.* **104**, 2655–2660 (2007).

46. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008).

47. Choi, U. B., Weninger, K. R. & Bowen, M. E. Immobilization of Proteins for Single-Molecule Fluorescence Resonance Energy Transfer Measurements of Conformation and Dynamics. *Methods Mol. Biol. Clifton NJ* **896**, 3–20 (2012).

48. Zelger-Paulus, S., Hadzic, M. C. A. S., Sigel, R. K. O. & Börner, R. Encapsulation of Fluorescently Labeled RNAs into Surface-Tethered Vesicles for Single-Molecule FRET Studies in TIRF Microscopy. *Methods Mol. Biol. Clifton NJ* **2113**, 1–16 (2020).

49. Lamichhane, R., Solem, A., Black, W. & Rueda, D. Single Molecule FRET of Protein-Nucleic Acid and Protein-Protein complexes: Surface Passivation and Immobilization. *Methods San Diego Calif* **52**, 192–200 (2010).

50. Bruschini, C., Homulle, H., Antolovic, I. M., Burri, S. & Charbon, E. Single-photon avalanche diode imagers in biophotonics: review and outlook. *Light Sci. Appl.* **8**, 87 (2019).

51. Yasuda, R. *et al.* The ATP-waiting conformation of rotating F1-ATPase revealed by single-pair fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci.* **100**, 9314–9318 (2003).

52. Pirchi, M. *et al.* Photon-by-Photon Hidden Markov Model Analysis for Microsecond Single-Molecule FRET Kinetics. *J. Phys. Chem. B* **120**, 13065–13075 (2016).

53. Di Fiori, N. & Meller, A. The Effect of dye-dye interactions on the spatial resolution of single-molecule FRET measurements in nucleic acids. *Biophys. J.* **98**, 2265–2272 (2010).

54. Wilson, H. & Wang, Q. ABEL-FRET: tether-free single-molecule FRET with hydrodynamic profiling. *bioRxiv* 786897 (2019) doi:10.1101/786897.

55. Doose, S., Heilemann, M., Michalet, X., Weiss, S. & Kapanidis, A. N. Periodic acceptor excitation spectroscopy of single molecules. *Eur. Biophys. J.* **36**, 669–674 (2007).

56. Fischer, C. J. & Lohman, T. M. ATP-dependent translocation of proteins along single-stranded DNA: models and methods of analysis of pre-steady state kinetics. *J. Mol. Biol.* **344**, 1265–1286 (2004).

57. Hwang, H., Kim, H. & Myong, S. Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7414–7418 (2011).

58. Hwang, H. & Myong, S. Protein induced fluorescence enhancement (PIFE) for probing protein–nucleic acid interactions. *Chem. Soc. Rev.* **43**, 1221–1229 (2014).

59. Levitus, M. & Ranjit, S. Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Q. Rev. Biophys.* **44**, 123–151 (2011).

60. Ploetz, E. *et al.* Förster resonance energy transfer and protein-induced fluorescence enhancement as synergetic multi-scale molecular rulers. *Sci. Rep.* **6**, 33257 (2016).

61. Morten, M. J., Lopez, S. G., Steinmark, I. E., Rafferty, A. & Magennis, S. W. Stacking-induced fluorescence increase reveals allosteric interactions through DNA. *Nucleic Acids Res.* **46**, 11618–11626 (2018).

62. Morten, M. J., Steinmark, I. E. & Magennis, S. W. Probing DNA Dynamics: Stacking-Induced Fluorescence Increase (SIFI) versus FRET. *ChemPhotoChem* **4**, 664–667 (2020).

63. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163–180 (2001).

64. Ingargiola, A. *et al.* Parallel multispot smFRET analysis using an 8-pixel SPAD array. in (eds. Enderlein, J., Gryczynski, Z. K., Erdmann, R., Koberling, F. & Gregor, I.) 82280B (2012). doi:10.1117/12.909470.

65. Ingargiola, A. *et al.* Optical crosstalk in SPAD arrays for high-throughput single-molecule fluorescence spectroscopy. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **9**, 255–258 (2018).

66. Nir, E. *et al.* Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006).

67. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

68. Norman, D. G., Grainger, R. J., Uhrín, D. & Lilley, D. M. J. Location of Cyanine-3 on Double-Stranded DNA: Importance for Fluorescence Resonance Energy Transfer Studies. *Biochemistry* **39**, 6317–6324 (2000).

69. Xie, Q., Soutto, M., Xu, X., Zhang, Y. & Johnson, C. H. Bioluminescence resonance energy transfer (BRET) imaging in plant seedlings and mammalian cells. *Methods Mol. Biol. Clifton NJ* **680**, 3–28 (2011).

70. Cole, N. B. Site-Specific Protein Labeling with SNAP-Tags. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al* **73**, 30.1.1-30.1.16 (2013).

71. Chakraborty, A., Wang, D., Ebright, Y. W. & Ebright, R. H. Azide-specific labeling of biomolecules by Staudinger-Bertozzi ligation phosphine derivatives of fluorescent probes suitable for single-molecule fluorescence spectroscopy. *Methods Enzymol.* **472**, 19–30 (2010).

72. Barth, A., Voithenberg, L. V. von & Lamb, D. C. Quantitative Single-Molecule Three-Color Förster Resonance Energy Transfer by Photon Distribution Analysis. *J. Phys. Chem. B* (2019) doi:10.1021/acs.jpcb.9b02967.

73. Lee, J. *et al.* Single-Molecule Four-Color FRET. *Angew. Chem. Int. Ed.* **49**, 9922–9925 (2010).

74. Holden, S. J. *et al.* Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophys. J.* **99**, 3102–3111 (2010).

75. Greenfeld, M., Pavlichin, D. S., Mabuchi, H. & Herschlag, D. Single Molecule Analysis Research Tool (SMART): An Integrated Approach for Analyzing Single Molecule Data. *PLOS ONE* **7**, e30024 (2012).

76. Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. iSMS: single-molecule FRET microscopy software. *Nat. Methods* **12**, 593–594 (2015).

77. Schrimpf, W., Barth, A., Hendrix, J. & Lamb, D. C. PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data. *Biophys. J.* **114**, 1518–1528 (2018).

78. Eggeling, C., Fries, J. R., Brand, L., Günther, R. & Seidel, C. a. M. Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proc. Natl. Acad. Sci.* **95**, 1556–1561 (1998).

79. Kalinin, S. *et al.* A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9**, 1218–1225 (2012).

80. Schmid, S., Götz, M. & Hugel, T. Single-Molecule Analysis beyond Dwell Times: Demonstration and Assessment in and out of Equilibrium. *Biophys. J.* **111**, 1375–1384 (2016).

81. Kalinin, S., Sisamakis, E., Magennis, S. W., Felekyan, S. & Seidel, C. A. M. On the Origin of Broadening of Single-Molecule FRET Efficiency Distributions beyond Shot Noise Limits. *J. Phys. Chem. B* **114**, 6197–6206 (2010).

82. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating Structural Heterogeneities from Stochastic Variations in Fluorescence Resonance Energy Transfer Distributions via Photon Distribution Analysis. *J. Phys. Chem. B* **110**, 6970–6978 (2006).

83. Kalinin, S., Valeri, A., Antonik, M., Felekyan, S. & Seidel, C. A. M. Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States. *J. Phys. Chem. B* **114**, 7983–7995 (2010).

84. Santoso, Y., Torella, J. P. & Kapanidis, A. N. Characterizing single-molecule FRET dynamics with probability distribution analysis. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* **11**, 2209–2219 (2010).

85. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J. & Kapanidis, A. N. Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* **100**, 1568–1577 (2011).

86. Tomov, T. E. *et al.* Disentangling Subpopulations in Single-Molecule FRET and ALEX Experiments with Photon Distribution Analysis. *Biophys. J.* **102**, 1163–1173 (2012).

87. Craggs, T. D. *et al.* Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase. *Nucleic Acids Res.* **47**, 10788–10800 (2019).

88. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Methods* **5**, 965–971 (2008).

89. Hellenkamp, B., Wortmann, P., Kandzia, F., Zacharias, M. & Hugel, T. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat. Methods* **14**, 174–180 (2017).

90. Martens, K. J. A. *et al.* Visualisation of dCas9 target search in vivo using an open-microscopy framework. *Nat. Commun.* **10**, 3552 (2019).

91. Pitrone, P. G. *et al.* OpenSPIM: an open-access light-sheet microscopy platform. *Nat. Methods* **10**, 598–599 (2013).

92. Rosenegger, D. G., Tran, C. H. T., LeDue, J., Zhou, N. & Gordon, G. R. A High Performance, Cost-Effective, Open-Source Microscope for Scanning Two-Photon Microscopy that Is Modular and Readily Adaptable. *PLOS ONE* **9**, e110475 (2014).

93. Almada, P. *et al.* Automating multimodal microscopy with NanoJ-Fluidics. *Nat. Commun.* **10**, 1223 (2019).

94. Schröder, D. *et al.* Cost-efficient open source laser engine for microscopy. *Biomed. Opt. Express* **11**, 609–623 (2020).

95. Sharkey, J. P., Foo, D. C. W., Kabla, A., Baumberg, J. J. & Bowman, R. W. A one-piece 3D printed flexure translation stage for open-source microscopy. *Rev. Sci. Instrum.* **87**, 025104 (2016).

96. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).

97. Culley, S. *et al.* Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat. Methods* **15**, 263–266 (2018).

98. Müller, M., Mönkemöller, V., Hennig, S., Hübner, W. & Huser, T. Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ. *Nat. Commun.* **7**, 10980 (2016).

99. Ingargiola, A., Laurence, T., Boutelle, R., Weiss, S. & Michalet, X. Photon-HDF5: An Open File Format for Timestamp-Based Single-Molecule Fluorescence Experiments. *Biophys. J.* **110**, 26–33 (2016).

100. Ambrose, B., Cully, J., Baxter, J. & Craggs, T. D. *craggslab/smfBox: Submission Release*.

101. Müller, C. B. *et al.* Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL Europhys. Lett.* **83**, 46001 (2008).

102. Nagy, A., Wu, J. & Berland, K. M. Characterizing observation volumes and the role of excitation saturation in one-photon fluorescence fluctuation spectroscopy. *J. Biomed. Opt.* **10**, 44015 (2005).

103. Farooq, S. & Hohlbein, J. Camera-based single-molecule FRET detection with improved time resolution. *Phys. Chem. Chem. Phys.* **17**, 27862–27872 (2015).

104. Schwille, P. & Haustein, E. Fluorescence Correlation Spectroscopy. 33.

105. Lapham, J., Rife, J. P., Moore, P. B. & Crothers, D. M. Measurement of diffusion constants for nucleic acids by NMR. *J. Biomol. NMR* **10**, 255–262 (1997).

106. Mächtel, R., Narducci, A., Griffith, D. A., Cordes, T. & Orelle, C. An integrated transport mechanism of the maltose ABC importer. *Res. Microbiol.* **170**, 321–337 (2019).

107. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P. J. & Seidel, C. A. M. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* **475**, 455–514 (2010).

108. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie* **73**, 1391–1396 (1991).

109. Hosur, M. V. *et al.* Influence of stereochemistry on activity and binding modes for C2 symmetry-based diol inhibitors of HIV-1 protease. *J Am Chem Soc* **116**, 847–855 (1994).

110. Craggs, T. D. *et al.* Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase. *Nucleic Acids Res.* **47**, 10788–10800 (2019).

111. Tsytlonok, M. *et al.* Dynamic anticipation by Cdk2/Cyclin A-bound p27 mediates signal integration in cell cycle regulation. *Nat. Commun.* **10**, 1676 (2019).
112. LeBlanc, S. J. *et al.* Coordinated protein and DNA conformational changes govern mismatch repair initiation by MutS. *Nucleic Acids Res.* **46**, 10782–10795 (2018).
113. Kalinin, S. *et al.* A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9**, 1218–1225 (2012).
114. Zilles, A. Crystals of different fluorescent ATTO dyes gradually dissolving in ethanol. 47.
115. Özkumur, E. *et al.* Quantification of DNA and Protein Adsorption by Optical Phase Shift. *Biosens. Bioelectron.* **25**, 167–172 (2009).
116. Brune, R., Doose, S. & Sauer, M. Analyzing the influence of contact-induced quenching processes on Foerster resonance energy transfer. in *Biophotonics 2007: Optics in Life Science* vol. 6633 66331M (International Society for Optics and Photonics, 2007).
117. Yang, H. *et al.* Protein Conformational Dynamics Probed by Single-Molecule Electron Transfer. *Science* **302**, 262–266 (2003).
118. Zhu, P., Clamme, J.-P. & Deniz, A. A. Fluorescence Quenching by TEMPO: A Sub-30 Å Single-Molecule Ruler. *Biophys. J.* **89**, L37–L39 (2005).
119. Cordes, T. *et al.* Sensing DNA opening in transcription using quenchable Förster resonance energy transfer. *Biochemistry* **49**, 9171–9180 (2010).
120. Robb, N. C. *et al.* The Transcription Bubble of the RNA Polymerase–Promoter Open Complex Exhibits Conformational Heterogeneity and Millisecond-Scale Dynamics: Implications for Transcription Start-Site Selection. *J. Mol. Biol.* **425**, 875–885 (2013).
121. Robb, N. C. *et al.* Single-molecule FRET reveals the pre-initiation and initiation conformations of influenza virus promoter RNA. *Nucleic Acids Res.* **44**, 10304–10315 (2016).
122. Dimura, M. *et al.* Automated and optimally FRET-assisted structural modeling. *Nat. Commun.* **11**, 5394 (2020).
123. *The PyMOL MOlecular Graphics System, Version 2.3.2, Schrödinger, LLC.*

124. Höfig, H., Gabba, M., Poblete, S., Kempe, D. & Fitter, J. Inter-Dye Distance Distributions Studied by a Combination of Single-Molecule FRET-Filtered Lifetime Measurements and a Weighted Accessible Volume (wAV) Algorithm. *Molecules* **19**, 19269–19291 (2014).

125. Bricks, J. L., Slominskii, Y. L., Panas, I. D. & Demchenko, A. P. Fluorescent J-aggregates of cyanine dyes: basic research and applications review. *Methods Appl. Fluoresc.* **6**, 012001 (2017).

126. Nicoli, F. *et al.* Proximity-induced H-aggregation of cyanine dyes on DNA-duplexes. *J. Phys. Chem. A* **120**, 9941–9947 (2016).

127. CHARMM-DYES: Parameterization of Fluorescent Dyes for Use with the CHARMM Force Field | Journal of Chemical Theory and Computation. https://pubs.acs.org/doi/abs/10.1021/acs.jctc.0c00721.

128. Dimura, M. *et al.* Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* **40**, 163–185 (2016).

129. Sanders, J. C. & Holmstrom, E. D. Integrating single-molecule FRET and biomolecular simulations to study diverse interactions between nucleic acids and proteins. *Essays Biochem.* (2021) doi:10.1042/EBC20200022.

130. Tsutakawa, S. E. *et al.* Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nat. Commun.* **8**, 15855 (2017).

131. Balakrishnan, L. & Bambara, R. A. Flap endonuclease 1. *Annu. Rev. Biochem.* **82**, 119–138 (2013).

132. Grasby, J. A., Finger, L. D., Tsutakawa, S. E., Atack, J. M. & Tainer, J. A. Unpairing and gating: sequence-independent substrate recognition by FEN superfamily nucleases. *Trends Biochem. Sci.* **37**, 74–84 (2012).

133. Liu, Y., Kao, H.-I. & Bambara, R. Flap Endonuclease 1: A Central Component of DNA Metabolism. *Annu. Rev. Biochem.* **73**, 589–615 (2004).

134. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. Major Domain Swiveling Revealed by the Crystal Structures of Complexes of E. coli Rep Helicase Bound to Single-Stranded DNA and ADP. *Cell* **90**, 635–647 (1997).
135. Paul, T. *et al.* E. coli Rep helicase and RecA recombinase unwind G4 DNA and are important for resistance to G4-stabilizing ligands. *Nucleic Acids Res.* **48**, 6640–6653 (2020).
136. Boubakri, H., de Septenville, A. L., Viguera, E. & Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J.* **29**, 145–157 (2010).
137. Ivessa, A. S. *et al.* The Saccharomyces cerevisiae Helicase Rrm3p Facilitates Replication Past Nonhistone Protein-DNA Complexes. *Mol. Cell* 12.
138. von Hippel, P. H. & Delagoutte, E. A general model for nucleic acid helicases and their 'coupling' within macromolecular machines. *Cell* **104**, 177–190 (2001).
139. Morris, P. D. & Raney, K. D. DNA Helicases Displace Streptavidin from Biotin-Labeled Oligonucleotides. *Biochemistry* **38**, 5164–5171 (1999).
140. Wilkinson, O. J. *et al.* Alkyltransferase-like protein (Atl1) distinguishes alkylated guanines for DNA repair using cation–π interactions. *Proc. Natl. Acad. Sci.* **109**, 18755–18760 (2012).
141. Onodera, T. *et al.* Role of alkyltransferase-like (ATL) protein in repair of methylated DNA lesions in Thermus thermophilus. *Mutagenesis* **26**, 303–308 (2011).
142. Yang, C.-G., Garcia, K. & He, C. Damage Detection and Base Flipping in Direct DNA Alkylation Repair. *ChemBioChem* **10**, 417–423 (2009).
143. Hitomi, K., Iwai, S. & Tainer, J. A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair* **6**, 410–428 (2007).
144. Fleming, A. M., Ding, Y. & Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci.* **114**, 2604–2609 (2017).

145. Seifermann, M. & Epe, B. Oxidatively generated base modifications in DNA: Not only carcinogenic risk factor but also regulatory mark? *Free Radic. Biol. Med.* **107**, 258–265 (2017).

146. Tubbs, J. L. *et al.* Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459**, 808–813 (2009).

147. Dalhus, B. *et al.* Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase. *Struct. Lond. Engl. 1993* **19**, 117–127 (2011).

148. Chakraborty, S. *et al.* Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. *Nucleic Acids Res.* **46**, 1240–1255 (2018).

149. Chen, X. *et al.* Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat. Commun.* **6**, 5849 (2015).

150. Kong, M. *et al.* Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process. *Mol. Cell* **64**, 376–387 (2016).

151. Tafvizi, A., Mirny, L. A. & Oijen, A. M. van. Dancing on DNA: Kinetic Aspects of Search Processes on DNA. *ChemPhysChem* **12**, 1481–1489 (2011).

152. Bennet, I. A. *et al.* Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1. *Nucleic Acids Res.* **46**, 5618–5633 (2018).

153. Nguyen, B., Ordabayev, Y., Sokoloski, J. E., Weiland, E. & Lohman, T. M. Large domain movements upon UvrD dimerization and helicase activation. *Proc. Natl. Acad. Sci.* **114**, 12178–12183 (2017).