

Development of Single-Molecule DNA Hybridisation Tools for High-Content Analysis of Protein-DNA Interactions and Gene Expression

Rebecca Andrews

Wolfson College

University of Oxford



A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at the University of Oxford

Hilary Term 2021

Disclaimer

I hereby declare that the work in this thesis is that of the candidate alone, except where indicated in the text, and as described below.

- Software used for analysis of transient binding on the surface was developed by Horst Steuer.
- A modified catabolite activator protein was expressed from a plasmid made by Hafez El Sayyed.
- Catabolite activator protein was labelled by Abhishek Mazumder.
- Chromosomal insertion of the repeated binding array was performed by Hafez El Sayyed.

Rebecca Andrews
January 2021

Abstract

This thesis presents results in the development of two single-molecule methods which use DNA hybridisation, *in vitro* and *in vivo*, to investigate protein-DNA interactions and processes.

Single-molecule DNA sequencing methods have made important contributions to genomics, diagnostics and functional analysis of genetic processes. Despite the fact that there are a large variety of single-molecule sequencing methods, there is currently no systematic way to connect the functional properties, with regards to protein interactions, of a single DNA molecule (single-molecule phenotype) with DNA sequence. In this thesis, I describe proof-of-principle experiments for ‘Gap-Seq’, a short-read single-molecule sequencing method that interrogates DNA sequence through the transient binding of short fluorescent DNA strands to surface-immobilised gapped-DNA substrates. A way to identify a base can be the dwell times of the short DNAs on the immobilised substrates: the correct base shows comparatively long dwells of the short DNAs on the substrates, whereas, non-complementary bases show no binding or comparatively shorter dwells on the substrates. I also show how single base DNA modifications affect protein-DNA interactions and a method for how Gap-Seq can be implemented to evaluate the DNA sequence at the protein binding site. Once developed further, Gap-Seq will fill a significant void in single-molecule DNA/RNA sequencing and will be important for both mechanistic work and screening libraries of nucleic acids against targets of biomedical importance.

I have also used DNA hybridisation to demonstrate the proof-of-principle for a new type of fluorescence *in situ* hybridisation (FISH) assay in live bacterial cells. Specifically, my work introduces ‘*in vivo* single-molecule FISH’ (*in vivo* smFISH), where short doubly labelled single-stranded DNAs are electroporated into cells and subsequently bind transiently or stably to target RNA. I establish that the electroporated probes exhibit site-specific binding and produce bright fluorescence signals that should lead to direct measurements of the kinetics of gene transcription on the chromosome, and advance the understanding of many fundamental processes involving RNA.

Acknowledgements

I am incredibly grateful to my supervisor Achilles Kapanidis for his ongoing support, mentoring, guidance and encouragement over the course of my DPhil studies. Thank you for making me feel like a valued member of the group.

I would like to thank all the members of the Kapanidis group for all their help over the years, especially Abhishek Mazumder and Horst Steuer for their help with my in vitro studies and Hafez El Sayyed for his patience in teaching a physicist microbiology, for collaborating with me with and for all the laughs! Thank you to my collaborators on the Gap-Seq project Tom Brown and Afaf El-Sagheer. I would also like to thank Nicole Robb for all her help and guidance, I am very grateful. Thank you to my office mates Maabur Sow, Andrew McMahon, Nicolas Shiaelis, Heesoo Uhm, Oliver Pambos, Christof Hepp, Mirjam Kummerlin, Aleksander Zagajewski and Anna Wang for the great chat and lab snacks in times gone by! I am grateful to the EPSRC for funding my time in the group.

It does not feel right to ignore SARS-CoV-2. No thanks for the pandemic, social restrictions, anxiety and turmoil you have caused in my life but thank you for reminding me that research in the life sciences is so incredibly valuable.

To my parents: Thank you for your love and support throughout.

To Nik: Thank you for always being there for me.

Rebecca Andrews
January 2021

Publications

- **Andrews, R.**, Steuer, H., Mazumder, A., Shivalingam A., El-Sagheer, A.H., Brown, T., Kapanidis, A.N. Short-read single-molecule DNA sequencing for highly parallel analysis of protein-DNA interactions. (Manuscript in preparation)
- **Andrews, R.**, DNA hybridisation kinetics using single-molecule fluorescence imaging. *Essays Biochemistry* 2021 **65**(1) 27–36 (2021).
- Patent Application: **Andrews, R.**, Kapanidis, A.N. PCT/GB2018/052598. 2018.

Contents

1	Introduction	1
1.1	Nucleic Acids	2
1.1.1	Structures of Hybridised Nucleic Acids	4
1.1.2	Degenerate and Universal Bases	6
1.2	Fluorescence	8
1.2.1	Fluorophores	11
1.2.2	Förster Resonance Energy Transfer (FRET)	13
1.2.3	Photophysical Effects	14
1.3	Microscopy	15
1.3.1	TIRF Microscopy	16
1.3.2	Epi-Fluorescence Microscopy	18
1.3.3	HILO Microscopy	18
1.3.4	Bright Field Microscopy	19
1.4	Single-Molecule Nucleic Acid Hybridisation	20
1.4.1	Experimental Influences on Observing DNA Hybridisation	22
1.4.2	Intrinsic Factors Affecting DNA Hybridisation Kinetics	28
1.4.3	External Factors Affecting DNA Hybridisation Kinetics	31
1.4.4	Coaxial Stacking	34
1.4.5	DNA-PAINT	34
1.5	DNA Sequencing	36
1.5.1	Single-Molecule DNA Sequencing	37
2	Gapped-DNA Hybridisation for Gap-Seq	39
2.1	Introduction	39
2.2	Results	43
2.2.1	Observing DNA Hybridisation in Gapped-DNA	43
2.2.2	Detecting a Single Base Pair Mismatch	47
2.2.3	Shorter Seals for Increased Discrimination	53
2.2.4	Fluorescence Quenching of Seals	55
2.2.5	A-T Exclusive Regions	60
2.2.6	Interrogation of an Unknown Gap	62

2.3	Discussion	67
3	Protein-DNA Interactions	70
3.1	Introduction	70
3.2	Results	74
3.2.1	CAP Binding	74
3.2.2	Forming a Gapped-DNA In Situ From dsDNA	79
3.2.3	CAP Binding Followed by Gap-Seq	84
3.2.4	Mixed dsDNA on the Surface Followed by Gap-Seq	89
3.3	Discussion	93
4	In Vivo Single-Molecule FISH	95
4.1	Introduction	95
4.2	Results	99
4.2.1	Electroporation of DNAs	99
4.2.2	Site-Specific Binding With EUB338	104
4.2.3	Plasmid Expression	107
4.2.4	Chromosomal Expression	113
4.3	Discussion	122
5	Materials and Methods	125
5.1	Chapter 2	125
5.1.1	DNAs	125
5.1.2	Buffers	126
5.2	Chapter 3	126
5.2.1	DNAs	126
5.2.2	Protein Expression, Purification and Labelling	127
5.2.3	Buffers	128
5.2.4	Assay for Denaturing and Reannealing In Situ	129
5.2.5	Assay of CAP Binding Followed by Gap-Seq	129
5.2.6	Mixed DNA on the Surface Followed by Gap-Seq	130
5.3	Chapter 2 & 3 Data Collection	131
5.3.1	Microscopy and Data Analysis	131
5.4	Chapter 4	133
5.4.1	DNAs	133
5.4.2	Electroporation of Cells	133
5.4.3	Imaging	133
5.4.4	Plasmid Creation	133
5.4.5	Chromosome Expression	134
5.5	DNA Sequences	134

Chapter 1

Introduction

Despite being critical to life, there remain many unanswered questions about critical processes and interactions with DNA inside cells. To truly understand these cellular processes, scientists have for years studied them at the ensemble level which averages behaviour across many molecules. However, with the invention of single-molecule microscopy, the true heterogeneity of such processes is becoming known. To answer specific questions about key interactions and processes on DNA, novel methods need to be invented to allow them to be observed with single-molecule microscopy.

In this thesis I discuss the development of single-molecule DNA hybridisation tools for high-content analysis of protein-DNA interactions and gene expression in prokaryotes. First, in Chapter 2, I describe experiments conducted to show the proof-of-principle for Gap-Seq: a high-throughput short-read single-molecule sequencing method based on transient DNA hybridisation to gapped-DNAs. In Chapter 3, I show that Gap-Seq can be implemented to study protein phenotype in sequence dependent protein-DNA interactions. In the final results chapter,

Chapter 4, I demonstrate how the transient hybridisation of short labelled DNAs in vivo can be used to investigate gene expression inside live cells.

The following introduction provides an overview of the biology and methodologies relevant for the whole thesis, with each research chapter containing an introduction to the relevant topics.

1.1 Nucleic Acids

In 1869, Friedrich Miescher became the first scientist to isolate DNA during his study of human white blood cells, labelling it ‘nuclein’ [1]. The name nuclein later morphed into ‘nucleic acid’ and is still used today. The term nucleic acid refers not only to deoxyribonucleic acid (DNA) but also ribonucleic acid (RNA). DNA and RNA consist of long chains of nucleotides where a nucleotide is made of a base, a phosphate and a sugar, which is either deoxyribose in DNA or ribose in RNA, see Figure 1.1. The base binds to the 1’ position on the sugar carbon ring with the phosphate binding to the 5’ position.

The model with nucleic acids made from nucleotides, aligned to make a phosphate backbone with repeating bases, was proposed by Phoebus Levene in 1909 [2]. The nucleotides bind in varying sequences of four bases: thymine (T), adenine (A), guanine (G) and cytosine (C) for DNA with uracil (U) replacing thymine (T) in RNA. Bases can be divided into two classes: pyrimidines and purines, where purines have two carbon rings and pyrimidines a single carbon ring, see Figure 1.2.

The bases in nucleotides serve as registers for binding of DNA/RNA strands. When two single stranded DNAs (ssDNAs) bind to make a double stranded DNA (dsDNA) the base content of the dsDNA follows Erwin Chargaff’s rules with

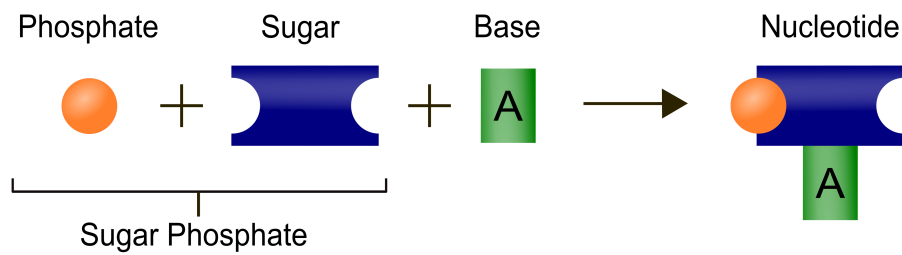


Figure 1.1: A nucleotide is made from three components: a phosphate molecule, a sugar (either deoxyribose in DNA or ribose in RNA) and a base.

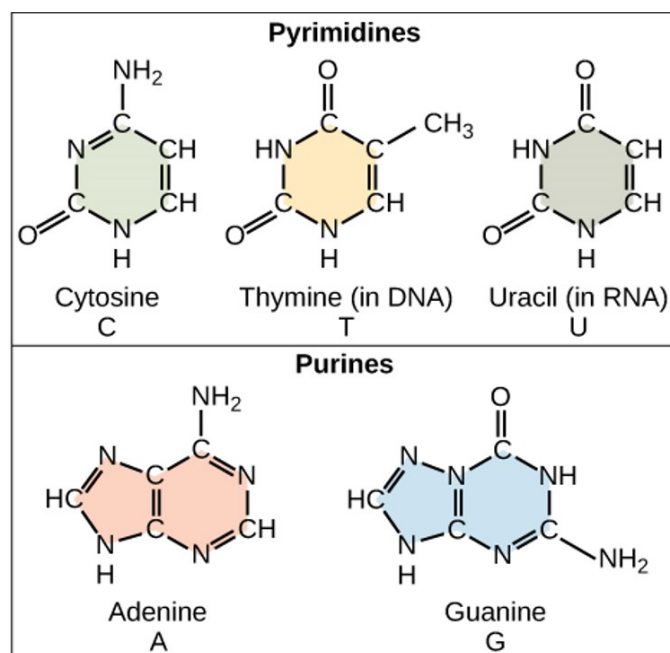


Figure 1.2: Chemical structures of bases which can be split into two classes, pyrimidines and purines. With a single carbon ring, cytosine (C), thymine (T) and uracil (U) are classed as pyrimidines and adenine (A) and guanine (G), with two carbon rings, are classed as purines. [3]

$[C] = [G]$ and $[A] = [T]$ [4].

1.1.1 Structures of Hybridised Nucleic Acids

The exact way in which bases are bound and the structure of dsDNA was only understood in 1953 when three crucial papers were published by James Watson & Francis Crick [5], Maurice Wilkins [6] and Rosalind Franklin [7]. From these discoveries, the double helical nature of dsDNA is now known along with the exclusivity of C-G or A-T/U binding.

Duplex DNA exists as two ssDNAs bound together through complementary base binding via hydrogen bonds, see Figure 1.3. A-T binding consists of two hydrogen bonds, whereas, three hydrogen bonds are present in G-C binding.

The hydrogen bonds between complementary bases do not solely stabilise the dsDNA, forces due to stacking of the hydrophobic bases are also responsible [9], [10]. Stacking occurs when the aromatic rings of neighbouring base pairs interact through Van der Waals forces [11] with some inter-strand base stacking also. The way the bases are arranged allow for multiple electrostatic and hydrophobic interactions between rings creating the majority of the stabilising force of dsDNA. The melting temperature of DNA (T_m) is the temperature at which the probability of a dsDNA being denatured equals 0.5. The nearest neighbour model allows the theoretical calculation of the melting temperature of DNA from the enthalpy and the entropy associated with tandem pairs in different sequences, accounting for hydrogen bonding between bases but is largely due to the Van der Waals forces from stacking of the bases [12], [13], [14], [15].

DNA exists most commonly as B-type DNA - a right-handed helical structure

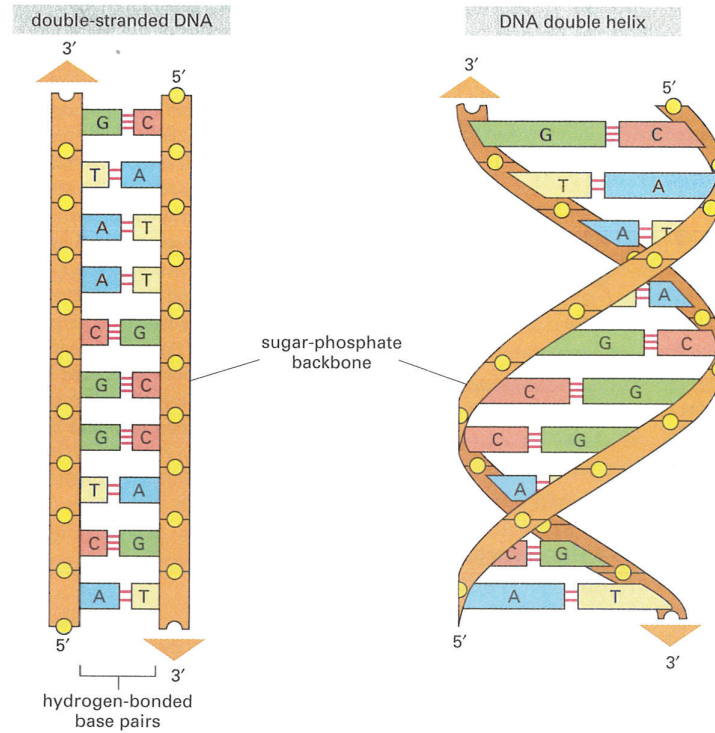


Figure 1.3: Two complementary ssDNA strands form a dsDNA when hydrogen bonds form between the corresponding bases on each strand. Naturally a dsDNA takes the shape of a double helix, shown in the right image. [8]

with anti-parallel strands, which rotate every 10 base pairs with a distance or pitch of $34 \text{ \AA}$ [5]. B-DNA has a distance of $3.4 \text{ \AA}$ [5] and a rotation of 36° [16] between adjacent bases. Due to the arrangement of the bases there exists a major and minor groove in double stranded nucleic acids.

A-DNA exists as a dehydrated DNA but also in some RNA-RNA/DNA duplexes [17]. In A-type DNA the bases are closer together, only $2.3 \text{ \AA}$ separation and tilted with respect to the central axis of the DNA [17]. Although wider and more compressed than B-DNA, A-DNA is still a right-handed anti-parallel duplex [17], see Figure 1.4.

Although DNA and RNA have a very similar make-up, RNA usually exists

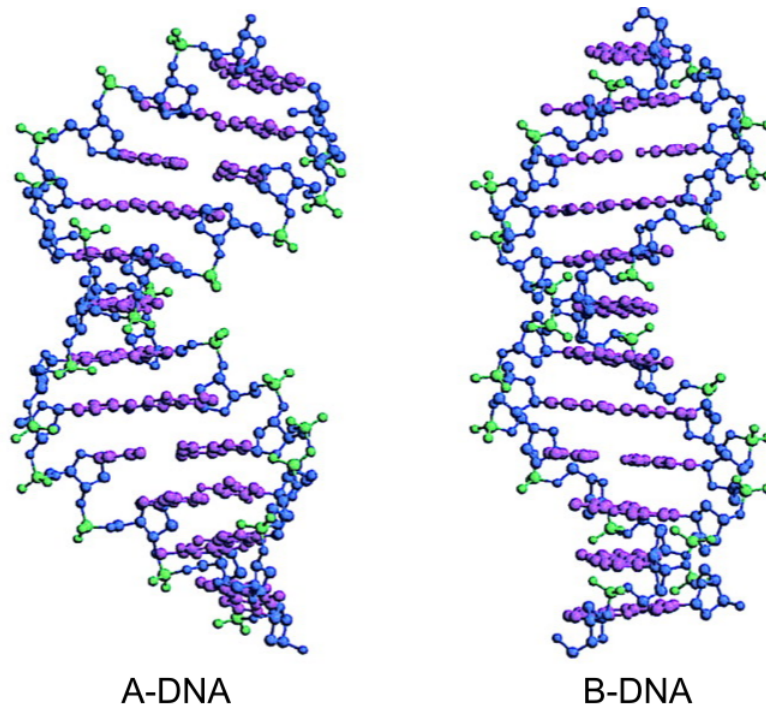


Figure 1.4: Side structure of A-DNA and B-DNA, both right-handed anti-parallel duplexes with A-DNA dehydrated and compressed compared to more common B-DNA. Image adapted from [18].

in a single-stranded form. However, as mentioned previously, RNA can hybridise to form duplexes with other RNAs or DNAs. Single-stranded RNA is also liable to forming secondary structures by hybridisation between different sections of the RNA.

1.1.2 Degenerate and Universal Bases

Nucleic acids occur naturally, however, it is possible to synthesise nucleic acids with specific sequences and characteristics. A bases or bases that can bind indiscriminately to any other base may be desired when designing DNAs to be hybridised with unknown DNA sequences. Two ways to create such a base are through the

use of degenerate bases or through the use of universal bases.

A degenerate base is created when it is possible for any one of the four bases to be incorporated at a single position on a ssDNA. This method creates 4^n nucleotides with different sequences, with n as the number of degenerate base positions created [19]. The relative concentration of the DNA of the full desired sequence is reduced by 4^n . Therefore, degenerate bases are not always desirable in situations where the overall oligonucleotide concentration is capped, such as with fluorescently labelled oligonucleotides (short DNAs) in single-molecule fluorescence experiments.

Another option is the use of universal bases. Universal bases are defined as bases that will bind indiscriminately to any of the four DNA bases. One such base analogue is 5-nitroindole [20]. 5-Nitroindole has been shown to bind with a similar strength to all of the natural bases, with a range of 3 °C in the melting temperature, when two complementary 17 mer oligonucleotides hybridise - with the universal base as the central base with 8 bases either side on one of strands [20]. The lack of large discriminations in binding between different bases, by definition, means 5-nitroindole can be used as an universal base analogue. Of the natural bases, 5-nitroindole can be most likened to purines, with a main two ring aromatic structure (Figure 1.5).

5-Nitroindole does not form hydrogen bonds within its ‘Watson-Crick’ base pair but binds through intercalating between the base opposite and one of its adjacent bases [22]. Therefore, 5-nitroindole solely stacks within the duplex during hybridisation without distorting the double helical structure of the duplex [22]. The effect of stacking energies on duplexes containing 5-nitroindole has been shown, with the base adjacent to the 5-nitroindole having a measurable impact on the melting tem-

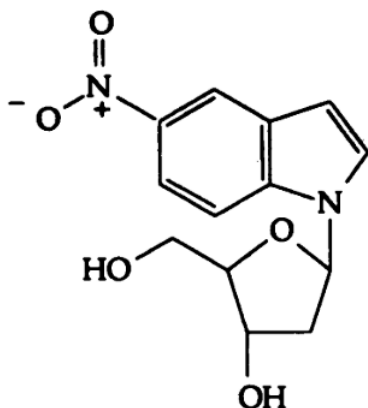


Figure 1.5: Chemical structure of 5-nitroindole attached to deoxyribose forming a deoxynucleoside [21].

perature. An adjacent T or C to a 5-nitroindole can result in a 3 °C increase in the melting temperature compared to an adjacent G or A [23].

1.2 Fluorescence

Fluorescence from an atom occurs when an incident photon (λ_{exc}) is absorbed by an electron in the atom, exciting it from the ground singlet state (S_0) to a higher energy singlet state (S_1) or (S_2). The electron then experiences non-radiative loss of some energy through vibrations to move to the lowest energy level in S_1 on a time scale of around 10^{-12} seconds [24]. As the spin of the electron is still paired with the electron in S_0 , the electron can quickly relax back to S_0 by radiating energy as a photon, see Figure 1.6. The fluorescence lifetime (time from excitation to emission) is approximately 10^{-8} seconds [24], much faster than phosphorescence which occurs when the electron enters a triplet state (T_1/T_2) where the electrons are no longer spin-coupled.

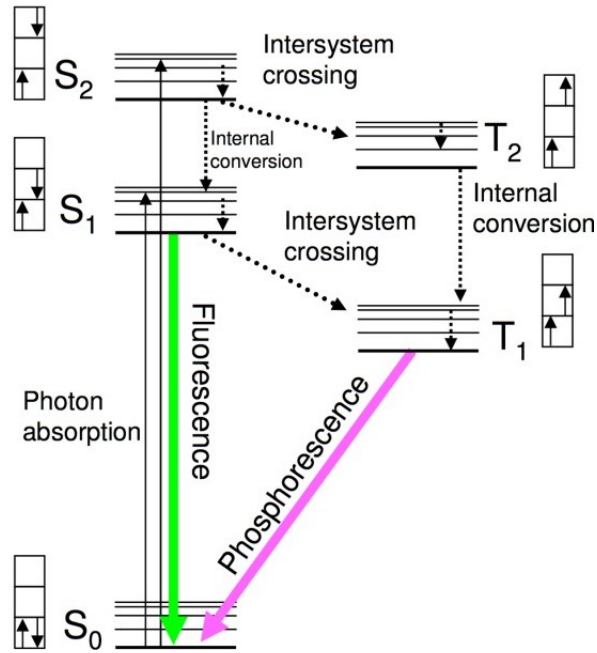


Figure 1.6: Jablonski diagram of the possible excited states of an electron. The electron starts in the ground singlet state S_0 and is excited by photon absorption to either the first S_1 or second S_2 singlet states. The electron experiences non-radiative energy losses and moves to the lowest energy level of S_1 . The electron then relaxes back to S_0 radiatively releasing fluorescence. Alternatively, when in S_1/S_2 , the electron can undergo intersystem crossing into a triplet state T_1/T_2 where it is no longer spin coupled with the electron in S_0 . Over a longer time period than fluorescence, the electron will radiatively decay back to S_0 known as phosphorescence. Adapted from [25].

In comparison to the incident photon, the emitted photon has a lower energy due to vibrational losses, hence, a longer wavelength (λ_{em}). The difference between the excitation wavelength λ_{exc} and the emission wavelength λ_{em} is known as the Stokes shift, first observed in quinine by Sir GG. Stokes, 1852 [26].

The Stokes shift in λ is exploited in single-molecule microscopy through the use of wavelength dependent filters to separate λ_{em} from λ_{exc} to enable imaging of the fluorescence from the sample. The results discussed in this thesis centre on DNA hybridisation, however, DNA is not intrinsically highly fluorescent. Therefore, DNA can be labelled with fluorescent molecules to allow DNA to be visualised in fluorescence microscopy.

In fluorescence microscopy, there is the desire to reduce the amount of background fluorescence from fluorophores in solution to improve the signal to background ratio. One method to reduce fluorescence intensity is fluorescence quenching. Two forms of quenching are collisional quenching and static quenching. Collisional quenching occurs when a fluorophore interacts with a molecule (a quencher) causing an excited electron to cross to a triple state and then return to the ground state S_0 via a non-radiative process [24]. Static quenching occurs when multiple fluorophores or a fluorophore and a quencher form a complex which cannot be excited [24], therefore, the electron never leaves the ground state S_0 . DABCYL (4-(Dimethylamino)-azobenzene-4'-carboxylic acid) is an example of a non-fluorescent molecule that can be used as a static quencher [27].

1.2.1 Fluorophores

Quinine, the first known fluorophore, was discovered by Sir JFW Herschel in 1845 [28]. There are two key measures by which a fluorophore can be characterised, the quantum yield of the fluorophore and the fluorescence lifetime of the fluorophore. The quantum yield is a ratio of emitted photons from the fluorophore to the number of photons absorbed, whilst the fluorescence lifetime measures the length of time that a fluorophore spends in the excited state before fluorescing (of the order 10^{-9} seconds) [24].

Small-Molecule Organic Fluorophores

There exist small-molecule organic fluorophores that are often < 1 kDa. Small-molecule organic fluorophores are attractive for single-molecule microscopy as they can be attached to biological molecules of interest without too much hindrance of the molecule due to their small nature.

There are multiple small-molecule organic fluorophores used in this thesis; for reference they are detailed in Table 1.1. Two fluorophores of special interest are Cy3B and ATTO674N as they are commonly used as a FRET pair (see FRET below).

Cy3 is a green cyanine dye [36] that can be linked to DNA, however, it has a low quantum yield due to the availability of a variety of molecular conformations. Cy3B has a modified structure to reduce vibrational losses within the molecule, hence, an increased quantum yield (see Table 1.1). For Cy3B, the maximum excitation and emission peaks are very slightly red-shifted from those of Cy3 as shown in Table 1.1. The excitation and emission spectra for Cy3B can be seen

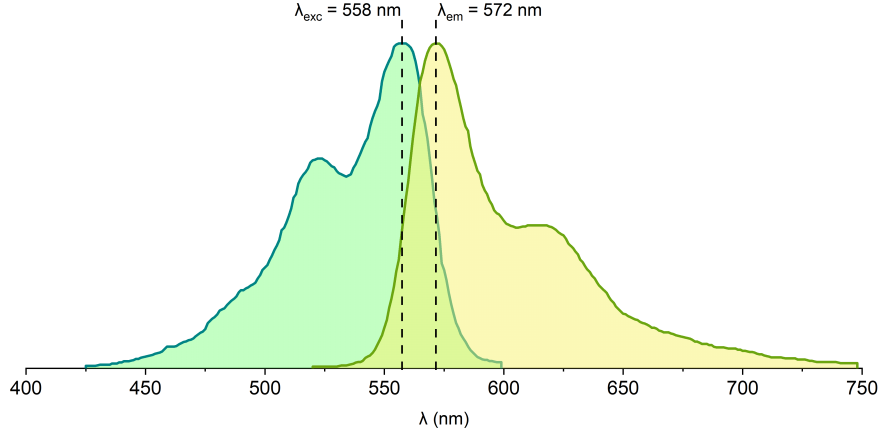


Figure 1.7: Excitation (blue) and emission (green) spectra for Cy3B with $\lambda_{exc} = 558$ nm and $\lambda_{em} = 572$ nm. Spectra data from [34].

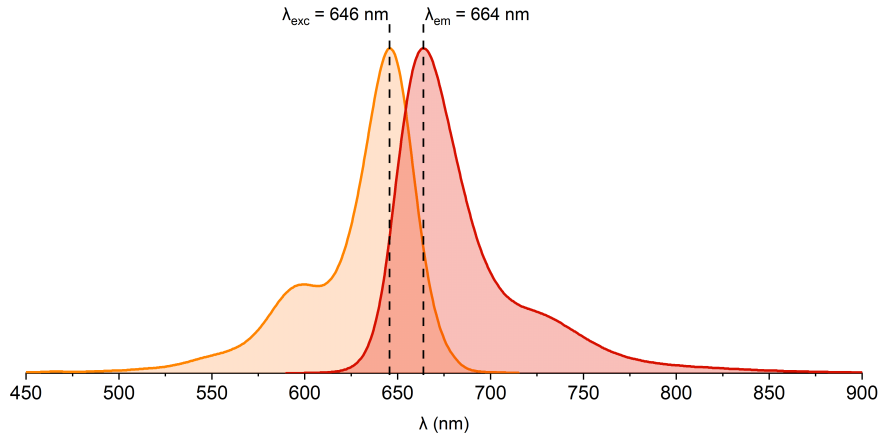


Figure 1.8: Excitation (orange) and emission (red) spectra for ATTO647N with $\lambda_{exc} = 646$ nm and $\lambda_{em} = 664$ nm. Spectra data from [35].

Small-molecule Organic Fluorophore Properties				
Fluorophore	Maximum λ_{exc} (nm)	Maximum λ_{em} (nm)	Quantum Yield	Fluorescence Lifetime (ns)
Cy3 [29]	550	570	0.04	0.3
Cy3B [29]	558	572	0.67	2.8
ATTO647N [30]	646	664	0.65	3.5
Cy5 [31] [32]	649	670	0.28	1.0
ALEXA647 [33]	650	668	0.33	1.0

Table 1.1: Properties of the fluorophores used in this thesis. From left to right: maximum excitation wavelength (nm); maximum emission wavelength (nm); quantum yield; fluorescence lifetime (ns).

in Figure 1.7 with $\lambda_{exc} = 558$ nm and $\lambda_{em} = 572$ nm. ATTO647N, developed by ATTO-TEC, lies within the red region of the visible part of the electromagnetic spectrum with $\lambda_{exc} = 646$ nm and $\lambda_{em} = 664$ nm, see Figure 1.8.

1.2.2 Förster Resonance Energy Transfer (FRET)

Cy3B and ATTO647N are commonly used together as a donor and an acceptor pair for Förster resonance energy transfer (FRET). FRET occurs when energy is transferred through dipole-dipole interaction between an excited donor molecule and an unexcited acceptor molecule. Once excited through FRET, the acceptor molecule can fluoresce. The energy transferred between the two molecules is non-radiative and is heavily distance dependent. Therefore, measuring the strength of the FRET efficiency allows the distances between the donor molecule and the acceptor molecule to be calculated. The FRET efficiency can be calculated using the following equation:

$$E = \frac{1}{1 + (R/R_0)^6} \quad (1.1)$$

where R is the distance between the two dipoles and R_0 is the Förster radius (distance where $E = 0.5$ for the chosen fluorophores).

E^* which denotes the apparent FRET is commonly used in experiments to quantify changes in the observed fluorescence intensity. E^* is defined as:

$$E^* = \frac{F_{D_{exc}}^{A_{em}}}{F_{D_{exc}}^{A_{em}} + F_{D_{exc}}^{D_{em}}} \quad (1.2)$$

where $F_{D_{exc}}^{A_{em}}$ is the fluorescence intensity of the acceptor when the donor is excited and $F_{D_{exc}}^{D_{em}}$ is the fluorescence intensity of the donor when it is directly excited [37].

1.2.3 Photophysical Effects

One behaviour that fluorophores can exhibit is blinking - a temporary lack of fluorescence. Blinking occurs when the electron is excited from the ground state to an excited state and then, crosses into a triplet state (see Figure 1.6). When in the triplet state, the spins of the ground and excited electron are no longer paired. This leads to a slower relaxation of the electron as Pauli exclusion principle prevents unpaired spins from being in the same energy state. The delay in relaxation can possibly be the time taken for a chemical reaction or interaction to flip the spin of the electron. Hence, there is a dark gap before the electron can be excited again once it is back in the ground state. These temporary dark periods of time followed by fluorescence are characteristic of blinking [38].

Fluorophores can also undergo a permanent loss of fluorescence known as photobleaching, when excited fluorophores undergo irreversible structural changes that cause permanent non-fluorescent molecules. Some fluorophores appear to fluoresce

longer than others before photobleaching. This characteristic can be described as the fluorophore's photon budget and can range from 10^2 - 10^6 photons [24].

If an excited fluorophore interacts with oxygen this can lead to photo-oxidation, therefore, photobleaching of the fluorophore [24]. An enzymatic oxygen scavenging system can be implemented in single-molecule experiments to stabilise the fluorophores by reducing blinking and prolonging the time before photobleaching. Glucose oxidase and catalase (gloxy) allow oxygen to react with glucose in the imaging buffer, reducing the amount of oxygen in proximity to the fluorophores. Although this oxygen scavenger system works at reducing oxygen, the products are acidic which needs to be accounted for when working with pH sensitive biomolecules. Another addition that can be made to reduce blinking and photobleaching is Trolox [39]. Trolox reacts with oxygen in the imaging buffer to reduce blinking through quenching of the long-lived triple states [40].

1.3 Microscopy

When molecules are labelled with fluorophores and excited by laser light, at wavelength within the excitation spectrum, fluorescence is produced. However, when there are many molecules together under a fluorescence microscope, individual fluorophores cannot be imaged, only a diffuse fluorescence. In order to be able to image single-molecules, it is essential to only excite a few fluorophores at a time. This can be done by reducing the concentration of the fluorescently labelled fluorophores, quenching fluorophores and, also, by exciting only a small volume of the sample.

1.3.1 TIRF Microscopy

Total internal reflection fluorescence (TIRF) microscopy is a type of illumination which relies on a small excitation volume within a sample to reduce diffuse background fluorescence. In an objective based TIRF microscope, the incident collimated laser beam, after passing through the objective, is totally internally reflected at the glass-sample boundary. Total internal reflection occurs when the excitation beam is incident at the boundary at an angle above the critical angle, at which there is no longer transmission through the boundary. Due to boundary conditions, the total internal reflection produces an evanescent light field past the boundary [41], see Figure 1.9. The evanescent field is a non-propagating field, which exponentially decays with distance from the boundary. Therefore, only a very small depth of the sample above the boundary is illuminated, approximately 150-200 nm from the surface [24]. Only fluorophores within the excitation volume are able to fluoresce, reducing the diffuse fluorescence seen when illuminating all the sample. Single-molecules can therefore be observed when close to or at the surface.

Some of experiments described here were conducted on a widefield TIRF microscope with two excitation wavelengths and two emission channels, see Figure 1.10. Excitation wavelengths of 532 nm and 640 nm were used to excite wavelength matched fluorophores. Inside the microscope the two lasers are directed and collimated along the same beam path so that both wavelengths can illuminate the same area of the sample. The lasers can be used for continuous wave excitation either alone or together, but also for alternating laser excitation (ALEX) where the sample is exposed to both excitation wavelengths temporally separated by

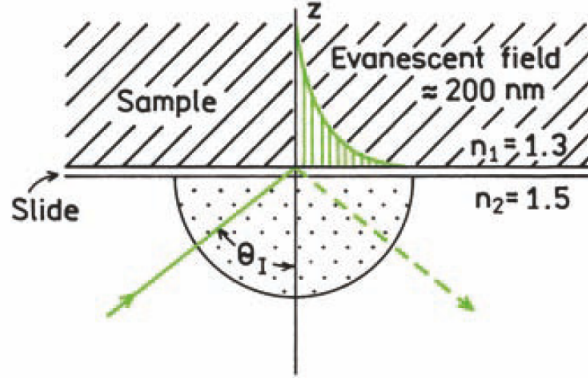


Figure 1.9: Total internal reflection occurs at the boundary between glass ($n_2 = 1.5$) and water ($n_1 = 1.3$) when the incident angle of the light is greater than the critical angle. Due to boundary conditions there is a resulting evanescent field which exponentially decays into the sample at a distance of approximately 150-200 nm [24].

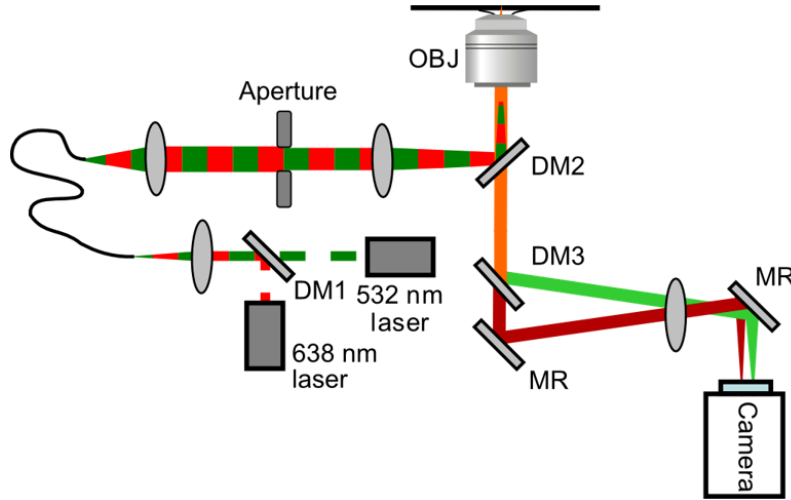


Figure 1.10: Objective based total internal reflection fluorescence (TIRF) microscope: 532 nm and 640 nm excitation light is combined at the first dichroic mirror (DM1), collimated and directed for total internal reflection at the sample boundary. The resulting evanescent wave illuminates the sample. The emission light is collected, filtered to remove excitation light at DM2 and imaged separated based on wavelength at DM3 and imaged separately (green and red channels) on the camera. Image adapted from [42].

10-100 ms [43]. ALEX is particularly useful in the detection of FRET, as both fluorophores in a FRET pair can be directly excited (Section 1.2.2). The fluorescence emitted by the fluorophores is then collected through the objective and separated from the excitation light by a dichroic mirror with reflective bands at 532 nm and 640 nm. This allows only the emission light to be collected by the camera. The light is further separated by a long pass dichroic mirror into green fluorescence and red fluorescence and focused separately on a sCMOS camera. The two different fluorescence channels show the same field of view with the emission light split by wavelength.

1.3.2 Epi-Fluorescence Microscopy

Epi-fluorescence microscopy uses a microscope set-up similar to total internal reflection fluorescence (TIRF) microscopy (Section 1.3.1), however, the incident light is not reflected at the surface of the sample but instead directly excites the sample, see Figure 1.11. The emission fluorescence is collected back through the objective and filtered and directed on to the camera. Unlike TIRF, most, if not all, of the depth of the sample is illuminated, with the sample experiencing a higher intensity of illumination light. Conventional epi-fluorescence does not lend itself to single-molecule studies due to the higher intensity background fluorescence from a large excitation volume.

1.3.3 HILO Microscopy

One way to reduce the high intensity background fluorescence of epi-fluorescence microscopy, whilst maintaining a higher imaging intensity than TIRF microscopy,

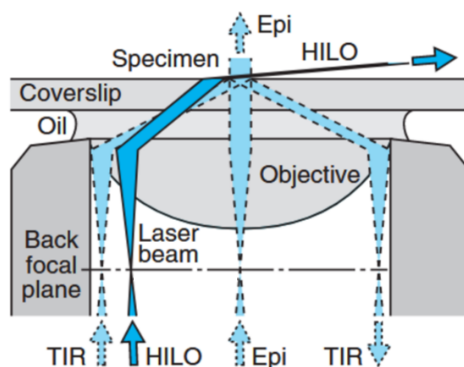


Figure 1.11: Optics for total internal reflection (TIR) fluorescence microscopy, epi-fluorescence microscopy and highly inclined and laminated optical sheet (HILO) microscopy. TIR: the excitation beam is totally internally reflected at the coverslip-specimen interface. Epi: the excitation beam is angled straight into the sample perpendicular to the coverslip. HILO: the excitation beam enters the specimen at an angle as a thin optical sheet. Image adapted from [44].

is by using highly inclined and laminated optical sheet (HILO) microscopy [44]. The excitation beam in HILO microscopy enters the sample at an angle rather than directly as in epi-fluorescence microscopy, see Figure 1.11. The angled illumination allows the excitation intensity to be high whilst minimising the excitation volume, leading to a better signal-to-background ratio than epi-fluorescence microscopy by approximately eight-fold [44]. In comparison to TIRF microscopy, the higher excitation intensity makes HILO microscopy more favourable for exciting single molecules inside cells but with a better signal-to-background ratio than epi-fluorescence microscopy.

1.3.4 Bright Field Microscopy

Bright field microscopy is a simple form of microscopy where the sample is directly illuminated, for example from above, with the light passing through the

sample and collected by the objective. Bright field images are low resolution and typically consist of dark areas, where the incident light was absorbed, on a white background. Bright field imaging is exclusively used in this thesis to image the location of cells.

1.4 Single-Molecule Nucleic Acid Hybridisation

Nucleic acids, created by repeating nucleotides, are able to bind to other nucleic acids through complementary base pairing. This is seen most commonly in DNA duplexes, where two single stranded DNAs (ssDNAs) hybridise to form a double stranded DNA (dsDNA) molecule. Nucleic acid hybridisation can be modelled as a nucleation step with one to a few bases binding between two single-stranded nucleic acids, then further binding of the rest of the bases from the nucleation site outwards until the construct is completely hybridised [45], [46]. When modelling hybridisation, a three-state model, rather than a two-state model, appears to fit most accurately. The three stages being: two single-stranded nucleic acids that do not interact, unstable intermediates with some level of binding between the two ssDNAs and stable fully bound double stranded nucleic acid constructs [47], [48].

Hybridisation does not always result in stable constructs, as nucleic acids can dissociate after even fully hybridising. This can happen repeatedly with the same or different nucleic acids and is known as transient hybridisation. Transient hybridisation of DNA can be quantified by an equilibrium constant, K_{eq} :

$$K_{eq} = \frac{k'_{on}}{k'_{off}} \quad (1.3)$$

where k'_{on} is the association rate and k'_{off} is the dissociation rate. From Equation 1.3, transient binding can be viewed as two separate processes, annealing of the nucleic acids described by an association rate k'_{on} and denaturing of the double stranded nucleic acid described by a dissociation rate k'_{off} . The association rate, k'_{on} , is dependent on the concentration of the fluorescently labelled ssDNAs available for binding. Another way the binding can be quantified, taking into consideration the concentration of the ssDNAs, is by the association rate constant k_{on} ($M^{-1}s^{-1}$). For low labelled-ssDNA concentrations below 300 nM, one can assume that there are no reactions between the labelled ssDNAs, making the dissociation rate, k'_{off} , equivalent to the dissociation rate constant, k_{off} (s^{-1}) [49]. Further, under the assumption of no photobleaching of the fluorescent dyes, the association rate constant, k_{on} , can be written as $k_{on} = \tau_{unbound}^{-1} [ssDNA]^{-1}$ and the dissociation rate constant, k_{off} , can be written as $k_{off} = \tau_{bound}^{-1}$, where $\tau_{unbound}$ is the average time between hybridisation events and τ_{bound} is the average time of a hybridisation event [50].

Single-molecule fluorescence microscopy allows individual DNA hybridisation events to be observed directly, therefore enabling the investigation of the kinetics of such reactions. Commonly, in surface-based studies that use TIRF microscopy, ssDNAs are labelled with fluorescent dyes and recorded for seconds to minutes as they bind to a complementary ssDNAs appropriately spaced on a solid support. If the duration of binding events is plotted as a cumulative distribution, two distinct average times for the duration of hybridisation can be extracted. The ‘short’ lifetime can be assumed to be partial hybridisation events (stage two of hybridisation) with the ‘long’ lifetime associated with full hybridisation (stage three of hybridisation) [51].

1.4.1 Experimental Influences on Observing DNA Hybridisation

Surface-Immobilised DNA Hybridisation

Commonly, surface-immobilised DNA hybridisation experiments are conducted using TIRF microscopy, where labelled or unlabelled ssDNAs (probes) hybridise with ssDNAs immobilised on a passivated glass surface. Immobilisation is usually achieved through a biotin-neutravidin linker; however, there are alternative ways to immobilise DNA on the surface, such as by using ssDNAs immobilised on DNA origami [52], on surface-tethered nanoparticles [53] or on long tethers [54]. Figure 1.12a shows some different methods used to surface-immobilised DNA hybridisation using TIRF microscopy. A large advantage of studying DNA hybridisation with immobilised DNA, is the ability to observe the same DNA molecule during multiple hybridisation events; however, the local environment for surface-immobilised DNA is drastically different than DNA free in solution.

From single-molecule fluorescence movies of surface-immobilised DNA the kinetics of hybridisation can be calculated through the changes in the intensity of fluorescence measured over time - the exact changes depend on the experimental fluorophore design. For example, a fluorophore can be used for localisation of a binding site either on the immobilised DNA, or in close proximity, and a differently coloured fluorophore can be directly excited on the hybridising DNA with a recorded increase in fluorescence upon hybridisation. A similar design uses a pair of fluorophores, one on each ssDNA, to observe binding through the use of Förster resonance energy transfer (FRET) between the two fluorophores in close proximity. Another method of fluorophore labelling, is creating one set of the

ssDNAs with a fluorescence quencher, a non-fluorescent molecule that absorbs energy, and when the quencher strand hybridises with a fluorescently labelled strand the reduction/absence of fluorescence indicates binding. Figure 1.12c shows a field of view from a movie of labelled ssDNAs being directly excited when binding to surface-immobilised ssDNAs that were previously localised using a different coloured fluorophore, first example in Figure 1.12a. For each single-molecule a fluorescence intensity vs. time trace can be plotted for the length of the movie. By measuring the duration of binding events (t_{bound}) and the time between binding events (t_{unbound}) frequency histograms can be created and fitted by a decaying exponential function when assuming a two-state model. From the decaying exponential fit, the average time of hybridisation (τ_{bound}) can be calculated from the t_{bound} histogram and the average time between hybridisation events (τ_{unbound}) can be calculated from the t_{unbound} histogram, as shown in Figure 1.12d.

Compared to solution-based experiments, ssDNAs that hybridise with surface-immobilised DNAs experience a greater repulsive electrostatic force [55], [56], [57]. When modelled, the electrostatic repulsion from surface-immobilised ssDNAs is the main reason for reduced k_{on} when compared to solution-based models [58]. Crowding of the DNA at high densities [59] and non-specific interactions of the probes with the surface [60] also contribute to a reduction of k_{on} . Intriguingly, single-molecule measurements have shown that after non-specific binding, probes perform a search process on the surface but the result is a low yield of hybridisation with the immobilised DNA, therefore, there is no increase of k_{on} [59], [61]. On the other side of the reaction, single-molecule surface-based studies are able to show k_{off} is a combination of multiple distinct dissociation rates, corresponding to different dissociation behaviours, rather than a single average rate as seen in ensemble

studies [61], [62]. Interestingly, the type of surface used can also affect k_{off} with hydrophilic surfaces decreasing the average k_{off} when compared to hydrophobic surfaces [51].

There are alternative single-molecule fluorescence methods for surface-based experiments which do not directly tether the DNA, such as trapping or confinement of the hybridising DNAs. Surface-tethered lipid vesicles (around 100 nm in diameter), which contain fluorescently labelled biomolecules, are an example of confinement [63]; within each vesicle, the DNAs can freely diffuse and interact without being modified or hindered by tethering, allowing DNA hybridisation to be observed [64], as shown in Figure 1.12a. Since the membrane of lipid vesicles acts as a barrier to exchanging molecules with the exterior of the vesicles, considerable effort has been taken to make them porous [65], [66], [67]. Even with such alternatives, there is no universal method of surface immobilisation which allows DNA hybridisation kinetics to be observed as they would be free in solution.

DNA Hybridisation in Solution

Traditionally, ensemble fluorescence experiments of DNA hybridisation were conducted in solution; currently, this is also a method that makes this possible at the single-molecule level. Through the use of a confocal microscope (Figure 1.12b), a small volume of a sample (≈ 0.2 fl) [68] can be fluorescently illuminated, enabling imaging of single molecules as they diffuse in solution. However, standard confocal microscopy only provides a brief snapshot (≈ 1 ms) as each molecule randomly diffuses through the excitation volume. Therefore, standard confocal microscopy is not well suited to observing multiple hybridisation events on a single molecule over long time periods, which is needed to gather data on hybridisation kinetics.

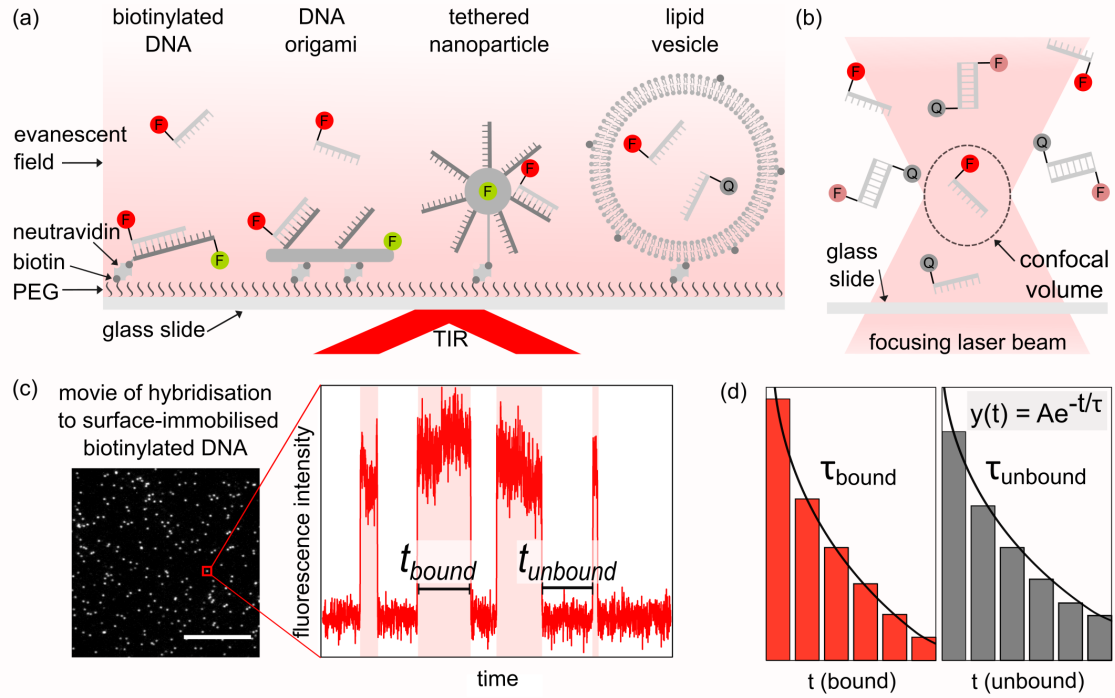


Figure 1.12: a) Methods for surface-immobilised single-molecule DNA hybridisation measurements. From left to right: single stranded DNA (ssDNA) immobilised via biotin/neutravidin, ssDNAs immobilised via DNA origami, ssDNAs immobilised via a tethered nanoparticle, ssDNAs confined via immobilised lipid vesicle. Polyethylene glycol (PEG) on surface of glass slide to allow immobilisation via biotin and neutravidin. In sample, ssDNA labelled with fluorophores (F) or fluorescence quenchers (Q) excited by evanescent field from total internal reflection (TIR) of incident laser beam. b) Confocal microscopy for single-molecule imaging of hybridisation between fluorescently labelled ssDNAs and ssDNAs labelled with a quencher within the confocal volume, shown by dashed ellipse. c) Example of field of view from movie of hybridisation of fluorescently labelled ssDNAs to surface-immobilised DNA. Scale bar 10 μm . Red box: Example fluorescence intensity vs. time trace for a single unlabelled surface-immobilised DNA undergoing transient hybridisation with fluorescently labelled ssDNAs. A hybridisation event is characterised as a rise in fluorescence. The time for hybridisation can be measured as t_{bound} and the time between binding events measured as t_{unbound} . d) Histograms of t_{bound} and t_{unbound} , fitted with a decaying exponential function ($y(t) = Ae^{-t/\tau}$, where A is a constant, τ is the calculated average time) to calculate the average bound time, τ_{bound} and the average unbound time τ_{unbound} .

A new confocal microscopy method, 3D single-molecule tracking (3D-SMT), has been developed to observe transient DNA hybridisation in solution over longer time periods (≈ 1 s) [69]. As the total maximum acquisition time is short, only fast transient hybridisation events between very short DNAs (8 nucleotides or less) can be recorded. Another novel approach for longer observations of DNA hybridisation in solution uses fluorescence to monitor the position of a labelled DNA whilst in the presence of an electric field. The changes in the DNA's diffusion and drift reveals the kinetics of hybridisation [70]. Clearly, measuring DNA hybridisation in solution would avoid the changes in the kinetic rates seen for surface immobilised assays, but currently there is no solution-based method able to provide as high a throughput on hybridisation kinetics over longer periods of time as surface-immobilised assays.

Effect of Fluorophores on Hybridisation Kinetics

DNA hybridisation can simply be modelled as the interaction of two ssDNAs, however, in fluorescence microscopy one or both of the DNAs is modified for the inclusion of a small-molecule fluorescent dye. Fluorescent dyes are usually chosen specifically for their optical properties with less consideration for how the fluorophore interacts locally. When using an assay with fluorescently labelled DNAs, dye-DNA and dye-dye interactions during DNA hybridisation can potentially alter hybridisation kinetics.

At the ensemble level, dye-dye or dye-quencher interactions, where a quencher is a non-fluorescent molecule that absorbs energy, between labelled strands have been shown to reduce k_{off} , therefore, creating a stabilising effect for hybridisation [71], [72]. As well as dye-dye interactions, dyes attached at the end of a DNA, also known as terminal dyes, reduce k_{off} due to stacking interactions with the adjacent end DNA bases [73].

Single-molecule fluorescence measurements have been able to look further at the changes in hybridisation kinetics due to the presence of fluorophores. DNA hybridisation between two fluorescently-labelled ssDNAs, or within a fluorescently-labelled DNA hairpin, found dye-dye interactions to increase the stability of the hybridisation and reduce k_{off} [64], [74]. Similarly, single-molecule experiments show terminal labelling with a cyanine fluorescent dye (Cy3) stabilises hybridisation by stacking with terminal bases, reducing k_{off} , but also leads to an increase in k_{on} of oligonucleotides [60], [64], [75]. Terminal dyes not only stabilise hybridisation through stacking but electrostatically interact with the DNA in dye-DNA interactions, with positively charged dyes showing a larger stabilising effect than

negatively charged dyes [74]. Fluorescently labelling DNA clearly shows stabilising effects during hybridisation leading to a decreased k_{off} but can also increase k_{on} in terminally labelled DNA. Therefore, the transient binding of fluorescently labelled ssDNAs needs to be considered a modified interaction compared to unlabelled DNAs. However, fluorescence based measurements of DNA hybridisation are still incredibly powerful and useful tools in biotechnology.

1.4.2 Intrinsic Factors Affecting DNA Hybridisation Kinetics

DNA Sequence

A DNA duplex exists as two ssDNAs bound together through complementary base binding via hydrogen bonds and stacking interactions between bases. For base-pairing, adenine (A) - thymine (T) binding consists of two hydrogen bonds and is, therefore, weaker than guanine (G) - cytosine (C) binding which forms three hydrogen bonds. It is well known that a DNA sequence with a higher G/C content will hybridise more stably than a sequence containing mostly A/T bases, consequently, playing a determining factor in the dissociation rate of transient hybridisation. Intriguingly, modelling of sequence dependent effects on DNA hybridisation shows that DNA sequence can, in fact, affect both the association and dissociation rates [76]. At the single-molecule level, where DNA hybridisation kinetics can be directly imaged, the DNA sequence used for multiple binding sites has been seen to affect the k_{on} of individual probes. If multiple side-by-side binding sequences are used, there is a linear increase in the total number of binding events to that entire binding site compared to a single binding sequence, as would be logically

expected. If the binding sequence is designed as periodic sequence motifs, such as a repeating TCC sequence, rather than the whole binding sequence repeating side-by-side, the same linear increase in k_{on} with an increase in the number of binding sites is observed for the entire binding area [77]. Compared to side-by-side repeats, periodic sequence motifs provide the same number of individual binding sites, however, they are not all available at the same time due to the overlapping nature of the binding sites. Therefore, periodic sequence motifs increase k_{on} for individual probes as predicted by theory, as there are more ways to correct misalignment before hybridisation of the entire probe [76].

Number of Bases

The actual sequence of bases is important in the hybridisation kinetics but also is the number of bases involved in hybridisation. From ensemble experiments, it is well established that long complementary ssDNAs bind more stably than short complementary ssDNAs during DNA hybridisation. Single-molecule experiments are able to give greater insight into the exact changes in the kinetics of DNA hybridisation when the length of the DNA is changed. For short DNAs (7-12 nucleotides), an extra base in length can dramatically change the duration of time for which the ssDNA hybridises before it thermally dissociates [49], [78], [79], with k_{off} drastically decreasing with increasing length [49], [80], [81], [82]. In fact, there is a negative single exponential relationship between k_{off} and the length of DNA [49]. However, during such hybridisation events, k_{on} does not show any noticeable change when the ssDNA is free of secondary structures. Understanding the changing kinetics of DNA hybridisation at the level of an additional base is essential for the development of biotechnologies which depend on the repeated

binding of short ssDNAs.

Single Base-Pair Mismatch

Although the length of a DNA can be an indication of the stability of its binding during hybridisation it is necessary to know the DNA sequence and its complementarity with the other DNA involved in hybridisation. Base-pair mismatches have a destabilising effect on DNA hybridisation and the shorter the DNA the larger the effect of even a single base mismatch on the stability. Single-molecule experiments are able to show clearly that a single base-pair mismatch can dramatically decrease the time a DNA is hybridised due to increases in k_{off} [78], [83], [84], [85], [86]. Commonly, single-molecule fluorescence experiments use a central mismatch for discrimination from the complementary [84], [85], [86], [87], with a minority of experiments placing mismatches away from the centre [78], [83]. Further, the type of mismatch can also affect the change in the dissociation rate; e.g., C-C mismatches are more destabilising than G-G mismatches [64]. The effect of mismatches on k_{on} is much less clear. Multiple studies have found that k_{on} decreases in the case of a single-base mismatch [64], [84], [85], [86]; however, other studies claim there to be no change [83], or in fact, an increase [64]. When looking specifically at k_{on} of short DNAs, seven contiguous bases are necessary for the smallest decrease in k_{on} , indicating terminal mismatches have the least impact [64].

Secondary Structure of ssDNA

Formation of a secondary structure due to transient hybridisation within a single DNA strand can occur when it diffuses in solution or when it is tethered on a solid support; such secondary structures can affect hybridisation kinetics. Specif-

ically, both ensemble and simulated DNA hybridisation experiments showed that secondary structures in DNAs can drastically reduce k_{on} during hybridisation [48], [88], [89], [90], [91]. Single-molecule experiments, which are capable of measuring small distances, such as between two points on a DNA, are well suited to study the negative effect of secondary structures on k_{on} . A reduction in k_{on} is observed for secondary structures in the hybridising strand, such as non-competing duplex regions within the strand [92], and also due to strands internally forming hairpins. When comparing DNAs of the same hybridising sequence those which formed hairpin secondary structures had a reduced k_{on} compared to unstructured DNAs of a similar length [51]. This decrease in k_{on} , due to secondary structures, can even be seen for short DNAs [54], [93]. For technological applications, where fast repeated hybridisation is desired, such as DNA-PAINT [80], a reduced k_{on} is unfavourable. To avoid this, a DNA sequence with no internal complementary bases can be used to eliminate most secondary structures that reduce k_{on} [94].

1.4.3 External Factors Affecting DNA Hybridisation Kinetics

Salts

The overall negative charge that nucleic acids carry, due to the negatively charged phosphate groups in the DNA backbone, results in repulsive electrostatic forces during DNA hybridisation which must be overcome for successful hybridisation. This repulsion applies to nucleic acids in solution but even more so to a situation where a local field is produced, such as nucleic acids immobilised on a surface – a key method used for single-molecule fluorescence DNA hybridisation studies.

A way to shield (or “screen”) the electrostatic forces, and consequently increase the rate of hybridisation, is to add cations in the form of salts to the local environment. Monovalent and divalent cations, such as Na^+ and Mg^{2+} , can be used to shield the electrostatic repulsion between two ssDNAs, hence facilitating and stabilising hybridisation. At the ensemble level, which cation or combination of cations is the most effective is debated [56], [95], [96] but it is universally recognised that cations play a large role in shielding the electrostatic repulsion between ssDNAs, hence encouraging duplex formation.

As one of the key controllable variables in single-molecule DNA hybridisation, the effects of salt type and concentration have been examined extensively. It is generally seen that k_{on} positively correlates with the salt concentration in solution independent of which salt cation [49], [51], [54], [64], [94]. As an alternative, there are also other molecules that can provide electrostatic shielding such as cationic conjugated polymers (CCPs) [85]. In terms of hybridisation, the increased k_{on} due to the presence of cations is due to increasing the proportion of successful binding events [51] as the cations alter the structure of the ssDNAs making hybridisation more favourable [49]. This trend in increasing k_{on} , however, is not seen for structured ssDNAs such as hairpins [51], as the opening rate of hairpins slows with increasing salt concentrations due to stabilisation of the secondary structure [97]. Although it is clear that an increase in salt concentration has a positive effect on k_{on} for unstructured ssDNAs, any correlation with k_{off} is not. Some studies report a slight decrease in k_{off} with an increase in the salt concentration [49]; however, surprisingly there is commonly very little or no change in k_{off} with increasing salt concentrations [51], [54], [64].

Concentration

An ever-present external factor in DNA hybridisation is the concentration of the ssDNAs in solution. Single-molecule studies show that the association rate, k'_{on} , linearly increases with increasing concentration of hybridising ssDNAs in solution [49], [75], [80], [93], [98]. On the other side of the reaction, there is no known dependence between the dissociation rate k'_{off} and the concentration of the same ssDNAs in solution [80], [93]. The maximum concentration of labelled ssDNAs in single-molecule TIRF microscopy measurements is limited to a maximum of 50-100 nM, due to the fluorescence background produced by unbound DNA, which if too high, can prevent the detection of single molecules.

Temperature

Another external factor during DNA hybridisation is the temperature, and as transient hybridisation is thermally driven, the temperature of the measurement will affect the kinetics of DNA hybridisation. A high temperature provides short DNAs with more thermal energy to escape a hybridised state, as shown in the observed increase of k_{off} in single-molecule imaging experiments [80]. When investigated further with single molecules the behaviour can be described by Eyring transition state theory [99]. Interestingly, k_{on} has been seen to slightly decrease with increasing temperature [80], possibly due to increased events where the ssDNA dissociates whilst in the process of trying to hybridise – known as abortive hybridisation.

1.4.4 Coaxial Stacking

Most of the discussion on nucleic acid hybridisation here has focused on two ssDNAs that anneal to form a dsDNA construct. However, it is possible for ssDNA to hybridise with many types of constructs such as a single stranded overhang of a dsDNA or within the gap of a gapped dsDNA. The region upon which the ssDNA binds is itself single-stranded but may be next to or surrounded by double-stranded regions. If the ssDNA were to bind adjacent to a double-stranded region, the stacking interactions and the change in the local structure of the DNA [100] would stabilise the binding of the ssDNA and reduce the dissociation rate. This is known as coaxial stacking. In the case of binding to gapped-DNA, for a ssDNA that is the same length as the gap, there would be interactions at both ends of the ssDNA having an additive effect. This coaxial stacking stabilises the ssDNA leading to the increased stabilisation; therefore, a decreased dissociation rate when compared to ssDNA annealing with another ssDNA [101], [102].

1.4.5 DNA-PAINT

DNA-PAINT (Points Accumulation for Imaging in Nanoscale Topography) is a super-resolution imaging technique which utilises the binding kinetics of DNA hybridisation to avoid common issues with photophysics of the fluorophores in imaging [80]. A ssDNA (docking strand) is immobilised on DNA origami and complementary labelled ssDNAs (seals) are in solution surrounding the docking strand. When a seal transiently binds with the docking strand the fluorophore enters the illumination region of a TIRF (total internal reflection fluorescence) microscope. The binding is indicated with a sharp increase in the fluorescence

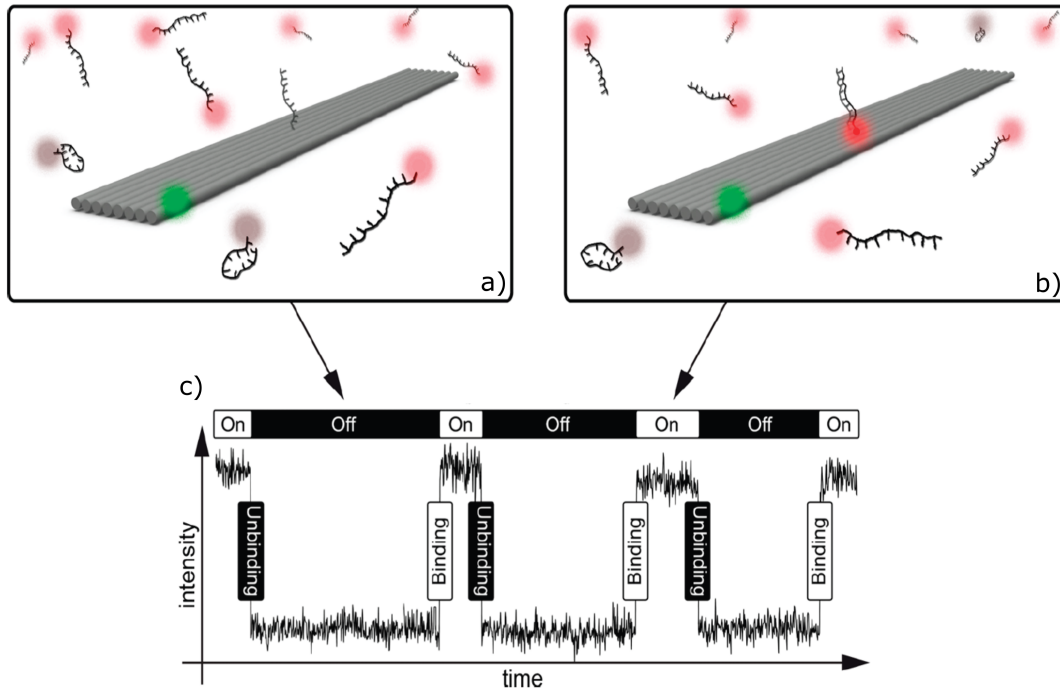


Figure 1.13: Schematic of DNA-PAINT. a) Surface immobilised DNA origami construct with a protruding single strand DNA docking strand - labelled complementary seals are in solution. b) A labelled complementary seal has bound to the docking strand. The fluorophore on the seal fluoresces as it is within the excitation region of the laser. c) Intensity-time trace of the detection of transient binding of the seal. A binding event is identified as a sudden large increase in fluorescence intensity with the fluorescence decreasing suddenly upon dissociation of the seal. Image adapted from [80].

signal which reduces back to the background fluorescence when the seal dissociates, see Figure 1.13.

The labelled DNA seals are short (≤ 10 nucleotides), therefore, due to thermal energy the seals will dissociate from the docking strand after a short period of time, being replaced by another seal from solution. The continuous interrogation of the docking strand in DNA-PAINT allows the errors due to photobleaching to be eliminated from the experiment allowing long acquisitions. The ability avoids

photophysical errors and, therefore, imaging for prolonged periods of time allows accurate position information to be gathered. Chen et al. [103] showed that DNA-PAINT can be used with a spatial accuracy of 30 nm.

Since its invention, DNA-PAINT has been implemented alongside many different super-resolution microscopy techniques in order to image targets inside cells, such as structured illumination microscopy (SIM) [104], stimulated emission depletion (STED) microscopy [105], stochastic optical reconstruction microscopy (STORM) [106], [107] and spinning disk confocal (SDC) microscopy [108]. DNA-PAINT has so far had a wide range of biological applications such as imaging synaptic proteins [109], imaging forces inside live cells [110], creating 3D images of internal cell structures [107], [111] and immunostaining of neuronal cells and tissues [104], [105].

1.5 DNA Sequencing

Only within the last forty years have successful methods to sequence DNA been developed [112], [113]. The first methods of DNA sequencing were developed in 1977, the chain termination method by Sanger et al. [112] and the chemical degradation method by Maxam and Gilbert [113].

Sequencing by hybridisation is an established sequencing technique which was first proposed in 1988 by various groups [114], [115], [116]. Sequencing by hybridisation is conducted through the binding of oligonucleotides (oligos) of a specific length to an unknown DNA to read its sequence. When a labelled oligo from solution binds to the immobilised DNA, a region of the unknown DNA sequence which is complementary to the oligo, is identified. Although well established, sequencing

by hybridisation methods do not provide position specific sequences.

There are many different methods for DNA sequencing, some involve hybridisation whilst others involve synthesis but the majority involve the amplification of the DNA. Amplification can introduce errors during the replication process, making methods such as single-molecule sequencing, that do not employ amplification, desirable.

1.5.1 Single-Molecule DNA Sequencing

Within the last decade, single-molecule DNA sequencing has grown in importance with there now being a large commercial market. Although single-molecule sequencing can eliminate amplification errors, current single-molecule methods have only been developed for long-read sequencing. Pacific Biosciences and Oxford Nanopore Technologies are two examples of assays that dominate the commercial single-molecule long-read sequencing market.

Pacific Biosciences use an immobilised single polymerase within a zero-mode waveguide to incorporate fluorescently labelled nucleotides as DNA synthesis proceeds [117]. Once a labelled nucleotide is incorporated the fluorophore is excited and imaged allowing a single base to be read in real time (Figure 1.14a). Oxford Nanopore Technologies' DNA sequencing method instead utilises the potential difference across a protein pore [118]. Each nucleotide produces a signature shift in the pore voltage allowing individual bases to be read as a ssDNA is passed through the pore (Figure 1.14b).

Both of the single-molecule sequencing techniques implemented by Pacific Biosciences and Oxford Nanopore Technologies utilise a surface immobilised protein

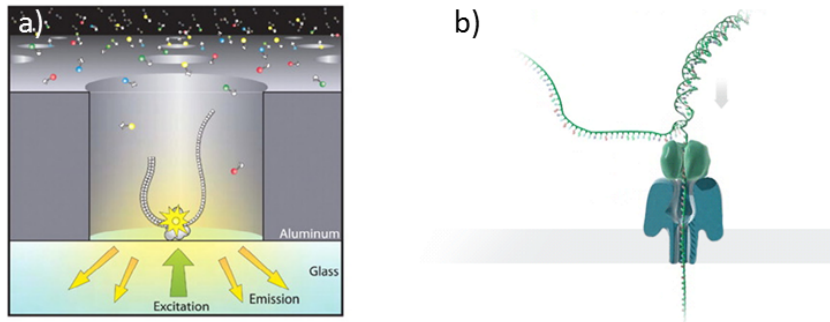


Figure 1.14: a) PacBio: RNAP immobilised in a zero-mode waveguide incorporates a fluorescently labelled nucleotide which is excited through total internal reflection fluorescence allowing the base to be read. [119]. b) Oxford Nanopore: as it passes through the nanopore each nucleotide produces a signature potential difference allowing the base to be called. [120]

to sequence DNA that is in solution. The immobilised protein prevents single-molecule investigations of different protein-DNA interactions which are usually conducted with surface immobilised DNA. Therefore, current single-molecule sequencing methods are incompatible with investigating protein-DNA interactions.

Chapter 2

Gapped-DNA Hybridisation for Gap-Seq

2.1 Introduction

Driven by the human genome project [121], DNA sequencing has evolved from the early chain termination [112] and chemical degradation methods [113] of the 1970s to what is now labelled ‘next generation sequencing’. Although next generation sequencing has revolutionised DNA sequencing, the majority of next generation sequencing methods rely on the amplification of DNA, introducing base errors. The desire for high accuracy has led to single-molecule sequencing, free from the DNA amplification that leads to sequencing errors.

Single-molecule DNA sequencing methods have made important contributions to genomics, diagnostics and functional analysis of genetic processes. Despite the fact that there are a large variety of single-molecule sequencing methods, there is currently no systematic way to connect the functional properties of a single

DNA molecule with DNA sequence. Here, experiments are described towards the development of ‘Gap-Seq’, a short-read single-molecule sequencing method that interrogates DNA sequence through the transient binding of short fluorescent DNA strands to surface immobilised gapped-DNA substrates, see Figure 2.1.

The proposed first stage in the Gap-Seq method is to have many gapped-DNA of different sequences immobilised on the surface via PEG, biotin and neutravidin. Fluorescently labelled short DNAs (seals) that are designed to probe the sequence in the gapped-DNA are then placed in solution. The first base of the seal will be known allowing sequencing of the first base in the gap and the rest of the seal will contain degenerate (D) or universal bases enabling the correct seal to be fully complementary to the rest of the gap sequence. A degenerate base is formed when a mixture of the four possible bases is used during the synthesis of DNA and a universal base is a synthetic base designed to bind to all bases. Seals that are fully complementary to the gap sequence will bind more stably than seals that form any mismatches. Fluorescence intensity vs. time traces for each gapped-DNA will show transient hybridisation events indicated by the increase in fluorescence upon seal binding and decrease when the seal dissociates. These fluorescence intensity vs. time traces can be analysed and the length of time for hybridisation (dwell time) and the time between hybridisation events (off time) measured. The measured times can then be plotted in a frequency histogram and mathematically fitted to find the average time for hybridisation and the average time between hybridisation events. The fully complementary seal will have a longer average dwell time than any seals with a mismatch, therefore, if the seals are coded via a fluorophore the first base of the seal can be identified and the first base in the gap sequenced. Gap-seq could then be used to sequence further bases in the gapped-DNAs by

moving the position of the known base in the seal. As Gap-Seq aims to sequence a single base at a time it will become a single-molecule short-read sequencing method. Gap-Seq merges both the single-molecule sequencing and the sequencing by hybridisation fields, with advancements in the ease of implementation to existing set-ups and the ability to gain position specific information. Inspiration for Gap-Seq was taken from DNA-PAINT [80], a super-resolution imaging technique that uses the repeated binding and dissociation of labelled oligonucleotides (oligos) to ssDNA binding points to create super-resolution images of samples.

One of the main applications of Gap-Seq will be to directly read the sequence of protein-DNA/RNA interactions and reactions as there are many different varieties. Some proteins interact with DNA/RNA by simply binding, such as the bacterial transcription factor catabolite activator protein (CAP) [122]. Other proteins when they bind to specific DNA sequences instigate reactions, for example RNA polymerase (RNAP) initiating transcription. The sequence specificity of protein binding is also utilised in aptamers (three-dimensional DNA/RNA) that are used to capture specific proteins with biomedical applications. Clearly, there exists a vast array of protein-DNA/RNA sequence dependent interactions - the motivating factor behind developing an assay to link DNA phenotype to sequence.

This chapter features proof-of-principle experiments in the development of ‘Gap-Seq’: a new short-read single-molecule DNA sequencing method based on transient DNA hybridisation. To investigate if a single base could be read at a time through a noticeable change in hybridisation kinetics, the dwell times for fully complementary seals and those with a single base mismatch were examined. The robustness of the method was tested for A/T rich sequences and fluorescence quenching was implemented to allow a higher working concentration of seals. Fi-

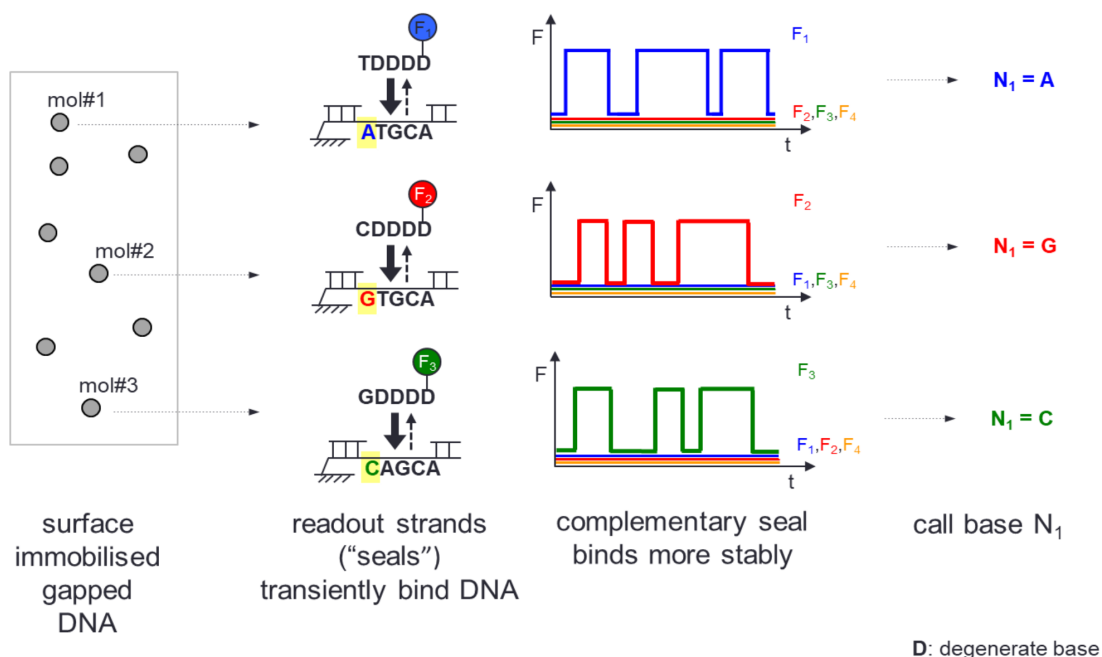


Figure 2.1: Schematic of the proposed high-throughput Gap-Seq assay. Gapped-DNA of varying sequences are immobilised on the surface. Fluorescently labelled ($F_1/F_2\ldots$) oligonucleotides (seals) that are designed to probe the first base (N_1) in the gapped-DNA transiently hybridise in the gaps. To interrogate the first base in the gap, the first base of the seal is known followed by degenerate (D) or universal bases. Seals that are fully complementary to the sequence in the gap will bind more stably than seals that form any mismatches. Fluorescence intensity (F) vs. time traces of the transient binding events (where increase in fluorescence intensity signifies a binding event) can be extracted for each gapped-DNA. Analysis of the kinetics of hybridisation allow the complementary seal to be identified and, therefore, informing on the first base in the sequence of the gap.

nally, experiments were conducted with degenerate and universal bases to show how Gap-Seq can be used to sequence unknown DNA sequences.

2.2 Results

2.2.1 Observing DNA Hybridisation in Gapped-DNA

First, it was important to establish the ability to view DNA hybridisation to gapped-DNA. In order to do this, a gapped-DNA hairpin construct, labelled with Cy3B was surface-immobilised via PEG-biotin/neutravidin linkage to biotin on the hairpin, see Materials and Methods. The gapped-DNA hairpin construct was formed through the hybridisation of a 78 nt biotinylated hairpin and complementary 27 nt Cy3B labelled DNA. The gap was 15 nucleotides (nt) long with individual experiments conducted for complementary ATTO647N labelled ssDNAs (seals) of different lengths in solution with approximately 1000 hairpins immobilised on the surface per field of view, see Figure 2.2. Individual experiments were conducted for a 12 nt seal (GS12-compl-T^{ATTO647N,P28}), a 9 nt seal (GS9-compl-T^{ATTO647N,P28}) and a 6 nt seal (GS6-compl-T^{ATTO647N,P28}), explanation of the nomenclature in Materials and Methods Chapter. Cy3B is a green fluorophore, therefore, excited by 532 nm laser excitation. Under 532 nm illumination a binding event was characterised by co-localised fluorescent spots in both the green emission channel (Cy3B fluorescence) and the red emission channel (ATTO647N fluorescence) due to FRET from Cy3B to ATTO647N (Fig. 2.2). The presence of FRET between the two fluorophores is possible because of the close proximity of the dyes when a seal is bound in the gap and no FRET is observed when seals

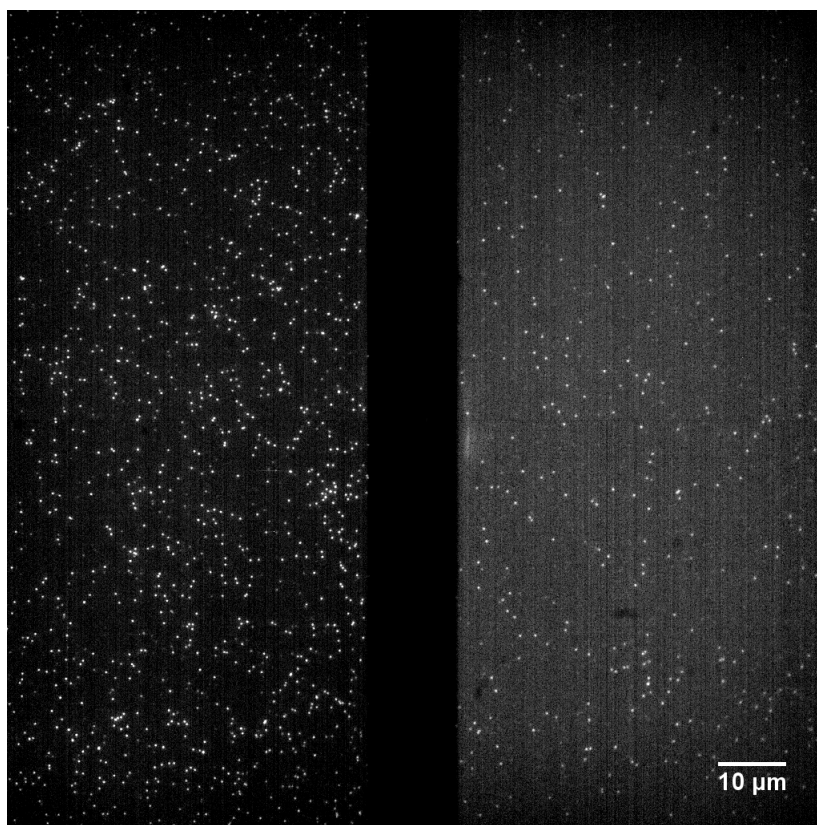


Figure 2.2: Example raw data of approximately 1000 biotinylated gapped-DNA constructs immobilised on the surface and 9 nt labelled seals in solution. Left-hand panel (green emission channel) shows fluorescence from individual Cy3B fluorophores attached to the gapped-DNA constructs when directly excited at 532 nm. Right-hand panel (red emission channel) shows fluorescence, due to FRET, from individual ATTO647N labelled seals bound in the gapped-DNAs.

are diffusing in solution. Seals were designed to bind immediately at the 5' end of the gap with a 9 bp separation between the fluorophores resulting in a theoretical FRET efficiency of ≈ 0.9 when the donor fluorophore was excited.

Transient binding events were observed for 75 nM GS12-compl- $T^{\text{ATTO647N,P28}}$ and 30 nM GS9-compl- $T^{\text{ATTO647N,P28}}$, see Figure 2.3 and Figure 2.4. A binding event can be identified as a sharp increase in red fluorescence intensity due to acceptor fluorophore on the seal undergoing FRET from the donor fluorophore

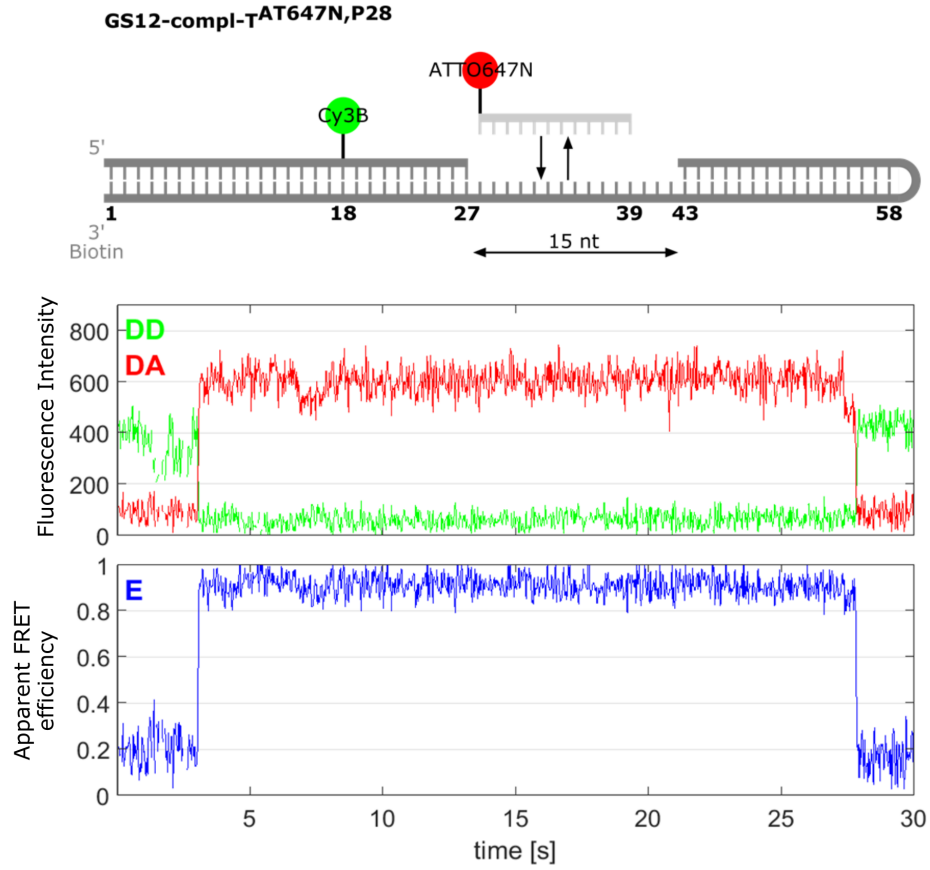


Figure 2.3: Representative time traces showing transient binding of GS12-compl-T_{ATTO647N,P28} to biotinylated hairpin gapped-DNA construct with 15 nt gap. Top traces: Fluorescence-time traces of a transient binding event when excited at 532 nm. Green: fluorescence in the green emission channel due to Cy3B donor fluorophore (DD - donor excitation, donor emission) red: fluorescence in the red emission channel, due to ATTO647N acceptor fluorophore (DA - donor excitation, acceptor emission). Bottom trace: FRET-time trace with apparent FRET efficiency (E) = 0.9 for binding events.

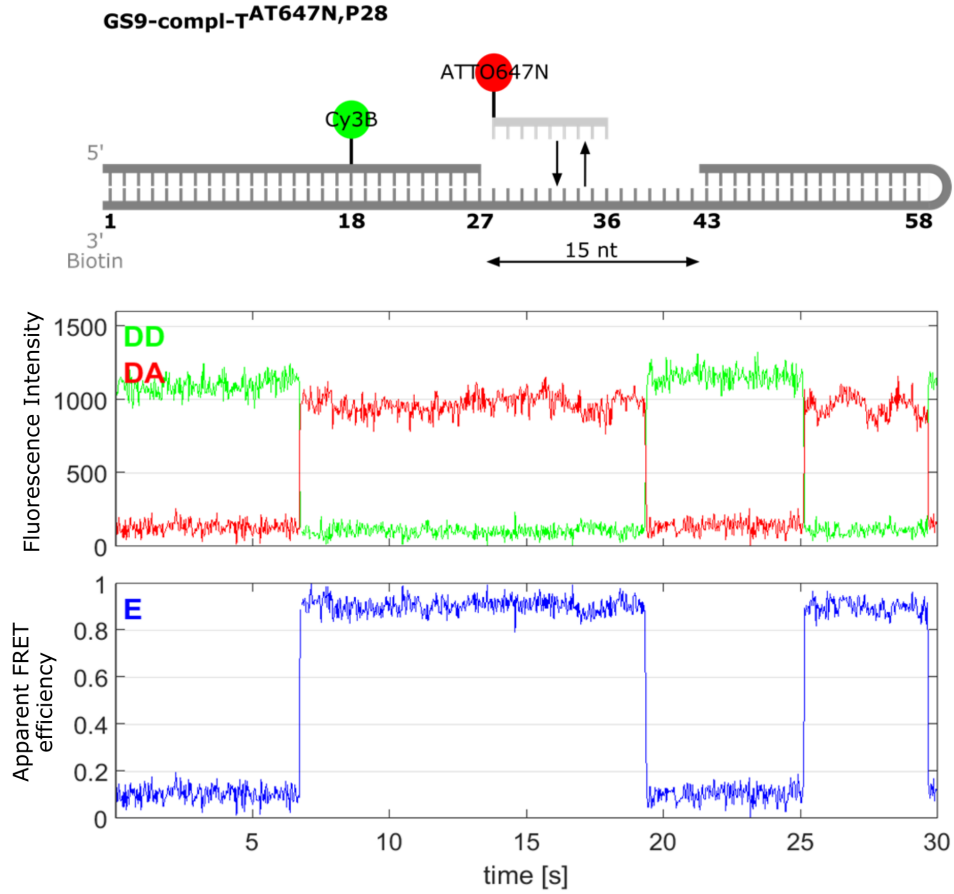


Figure 2.4: Transient binding of GS9-compl-T^{ATTO647N,P28} to biotinylated hairpin gapped-DNA construct with 15 nt gap. Top traces: Fluorescence-time traces of transient binding events when excited at 532 nm. Green: fluorescence in the green emission channel; red: fluorescence in the red emission channel. Bottom trace: FRET-time trace with apparent FRET efficiency (E) = 0.9 for binding events.

on the gapped-DNA. The fluorescence intensity vs. time traces show an anti-correlation between the fluorescence observed from Cy3B and ATTO647N, with an apparent FRET efficiency of ≈ 0.9 . However, no transient binding events were recorded for the 6 nt seal (GS6-compl-T^{ATTO647N,P28}). As reducing the length of the DNA reduced the stability of binding it is assumed that transient binding of the 6 nt seal was faster than the minimum exposure time of the camera of 10 ms.

The initial experiments of transient DNA hybridisation to gapped-DNA show that transient binding events to the gapped-DNA can be observed through the evidence of FRET-time traces. The experiments showed successful observation of repeated transient binding of 12 nt and 9 nt seals, within the 15 nt gapped-DNA, with seals of 6 nt thought to dissociate too quickly to be detected.

2.2.2 Detecting a Single Base Pair Mismatch

Creating Adjustable Gaps

To create an adjustable sized DNA gap, the hairpin was redesigned as two separate strands, a biotinylated bottom strand and an unlabelled fully complementary top strand (right flanking) with the left flanking labelled top strand remaining the same. The selection of the length of the right flanking top strand allowed the size of the gap to be altered with the ability to perform experiments with co-axial stacking, where the seal fits completely into the size of the gap. The whole construct was lengthened to 65 bps by the addition of 5 bases at the 3' end. Seals were designed to be shorter in length than previously ones to allow for faster repeated binding as previously GS12-compl-T^{ATTO647N,P28} bound quite stably in comparison to the observation time.

Using FRET to detect binding to the gapped-DNA confirmed DNA hybridisation to the gapped-DNA, however, the donor fluorophore is liable to bleaching, limiting the length of observation. Instead of using FRET, Cy3B can be excited at 532 nm and imaged to create a localisation and then ATTO647N on the seals directly excited at 638 nm. Fluorescence intensity vs. time traces for such show an increase in the red fluorescence when a seal binds to the position where a gapped-DNA was detected. When the seal thermally dissociates, the fluorescence intensity decreases creating a similar shaped ‘binding event’ as seen in the donor excitation acceptor emission fluorescence-time traces. As there are many fluorescently labelled seals in solution, this circumvents the problem of photobleaching similar to DNA-PAINT [80].

Fluorescence intensity vs. time traces of individual gapped-DNAs are then plotted. Hidden Markov Modelling (HMM) analysis is then used to measure the length of binding events and the time between events. Dwell time histograms can be plotted for the length of binding events and an off time histogram plotted for the time between binding events. Dwell time histograms are fitted with a double exponential decay fit rather than a single exponential decay fit as this is a more accurate fit, as seen in other experiments [75]. The binding of a seal is not a two state model (bound/unbound), therefore, the multiple exponential fit accounts for partial hybridisation or dissociation events, seen as high fluorescence, before the seal has fully bound. Partial hybridisation or dissociation events are given an average dwell time by T_1 with the fully bound average dwell time described by T_2 , with T_2 being used for comparisons between different seals. The off time histograms were fitted with a single exponential as absence of fluorescence represents the one state where there is no hybridisation.

Complementary Seal

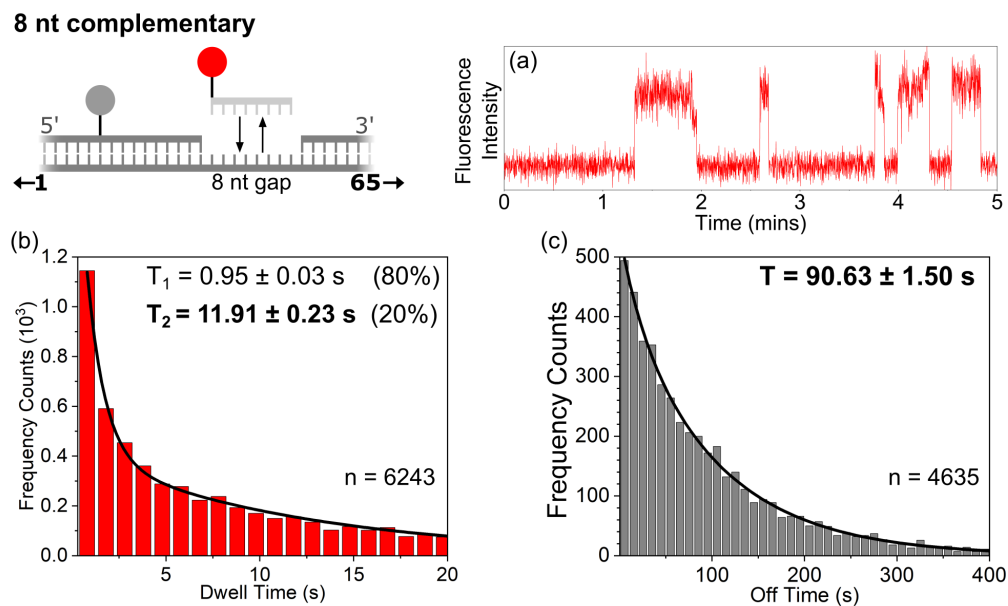


Figure 2.5: Fully complementary 8 nucleotide (nt) seal binding in a 8 nt gapped-DNA. a) Example fluorescence intensity vs. time trace. b) Dwell time histogram with an average dwell times of $T_1 = 0.95 \pm 0.03$ s (80%) and $T_2 = 11.91 \pm 0.23$ s (20%); number of events (n) = 6243. c) Off time histogram with average off time of $T = 90.63 \pm 1.50$ s; number of events (n) = 4635. Recorded with 100 ms exposure time and 10 nM seal.

The first experiments were performed with surface immobilised gapped-DNA with a gap of 8 nts and 8 nt fully complementary seals in solution. To monitor the multiple binding events of the seals without excess background fluorescence, the seals were suspended at 10 nM in the solution surrounding the gapped-DNAs. As can be seen from Figure 2.5a, at 100 ms exposure time binding events of the 8 nt fully complementary seals were detected. Dwell time analysis was then performed giving average dwell times for the fully complementary seal of $T_1 = 0.95 \pm 0.03$ seconds (80%) and $T_2 = 11.91 \pm 0.23$ seconds (20%), see Figure 2.5b. A histogram was also created for the time between binding events with the average

off time of $T = 90.63 \pm 1.50$ seconds, see Figure 2.5c.

Mismatch Seals

It is known that the length of an oligo [49] and the number of consecutive complementary base pairs (bp) [64] can determine the strength of binding of an oligo. Therefore, a seal with a single bp mismatch would be expected to bind for a shorter period of time than a fully complementary seal, especially, if the mismatch breaks a large number of consecutive bases. If the difference in hybridisation kinetics can be observed when a fully complementary seal is compared to a single bp mismatch seal it is a potential way to sequence a DNA base.

Experiments with 8 nt gapped-DNAs but now with single base mismatched DNA seals were performed next. Seals with single internal mismatches were created by substituting cytosine for guanine creating a G-G mismatch at two different positions, allowing the position of the mismatch and the effect on the dwell time to be probed. A strong bp was chosen to be mismatched to heighten the increase in the dissociation rate. GS8-mis5G-T^{ATTO647N,P28} with a position 5 mismatch was designed and another seal, GS8-mis3G-T^{ATTO647N,P28} with a position 3 mismatch. As can be seen in Figures 2.6 & 2.7, transient binding was observed for both of the mismatch seals.

Dwell time analysis was conducted showing the two mismatch seals exhibiting a shorter average dwell time for full hybridisation compared with the fully complementary seal. GS8-mis5G-T^{ATTO647N,P28} with an average dwell time of $T_2 = 9.47 \pm 0.18$ seconds (15%) (Figure 2.6b) and GS8-mis3G-T^{ATTO647N,P28} with an average dwell time of $T_2 = 2.22 \pm 0.01$ seconds (26%) (Figure 2.7b). Interestingly, as both the single base mismatch seals use a G-G mismatch, the large difference

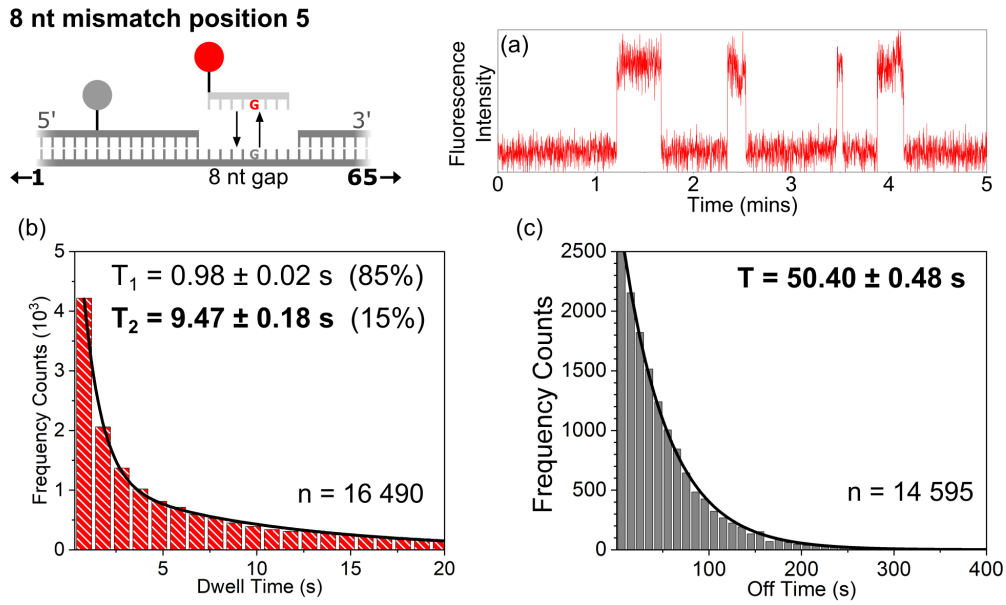


Figure 2.6: Single base-pair mismatch seal binding to 8 nt gapped-DNA with with G-G mismatch at position 5 in the seal. a) Example fluorescence intensity vs. time trace. b) Dwell time histogram with average dwell times of $T_1 = 0.98 \pm 0.02$ s (85%) and $T_2 = 9.47 \pm 0.18$ s (15%); number of events (n) = 16 490. c) Off time histogram with average off time of $T = 50.40 \pm 0.48$ s; number of events (n) = 14 595. Recorded with 100 ms exposure time and 10 nM seal.

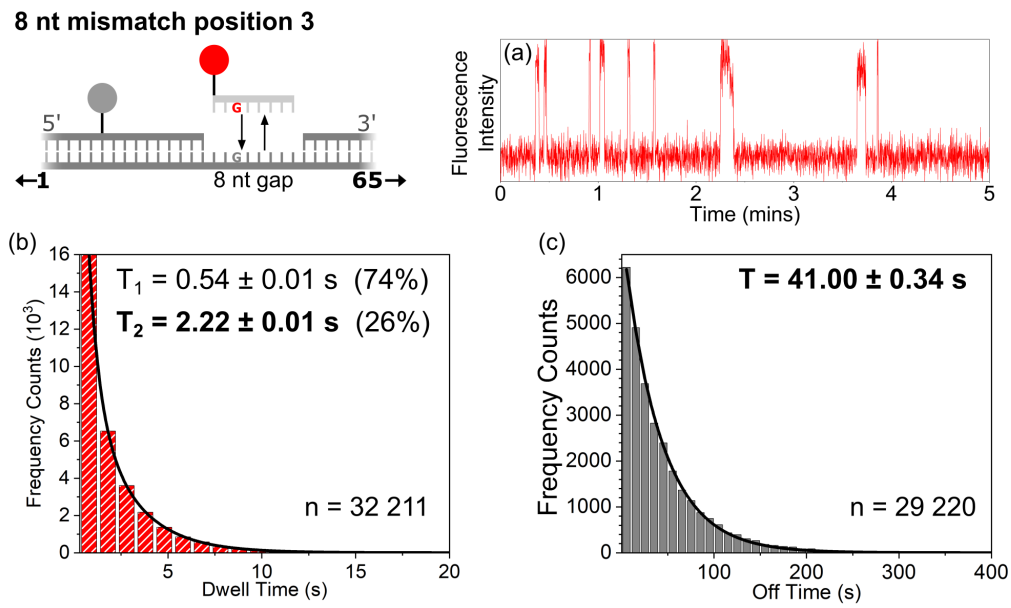


Figure 2.7: Single base-pair mismatch seal binding to 8 nt gapped-DNA with G-G mismatch at position 3 in the seal. a) Example fluorescence intensity vs. time trace. b) Dwell time histogram with average dwell times of $T_1 = 0.54 \pm 0.01$ s (74%) and $T_2 = 2.22 \pm 0.01$ s (26%) number of events (n) = 32 211. c) Off time histogram with average off time of $T = 41.00 \pm 0.34$ s; number of events (n) = 29 220. Recorded with 100 ms exposure time and 10 nM seal.

in average bound times between the two can be assumed due to the position of the mismatch within the seal. Surprisingly, GS8-mis3G-T^{ATTO647N,P28} shows that for these seals a mismatch closer to the terminus creates more instability than an internal mismatch. As expected, figures 2.6c & 2.7c show a reduction in the average off time with the introduction of a single mismatch base in the seal with $T = 50.40 \pm 0.48$ seconds for GS8-mis5G-T^{ATTO647N,P28} and $T = 41.00 \pm 0.34$ seconds for GS8-mis3G-T^{ATTO647N,P28}.

There is discrimination between the three seals for the average dwell time T_2 and for the average off time. For further experiments only the average dwell time is used as errors can be easily introduced into the calculated off time by varying seal concentrations between experiments. Therefore, results for further experiments feature the average dwell times as this can be kept more standard across multiple experiments as the key factors that affect this are concentration of salt and ambient temperature. From the average dwell times observed, complementary seals bind more stably than seals with a single-base mismatch, showing that the kinetics of transient binding show a promising way to identify a single base at a known position.

2.2.3 Shorter Seals for Increased Discrimination

As demonstrated with the 8 nt seals, removing a single base pairing through mismatching bases reduces the dwell time. A reduced dwell time is favourable as this allows more binding events to occur in a set time period, hence, increasing the interrogation rate and speed of the assay to identify a base. In order to test heightened discrimination at a shorter seal length a fully complementary seal of 6

6 nt complementary

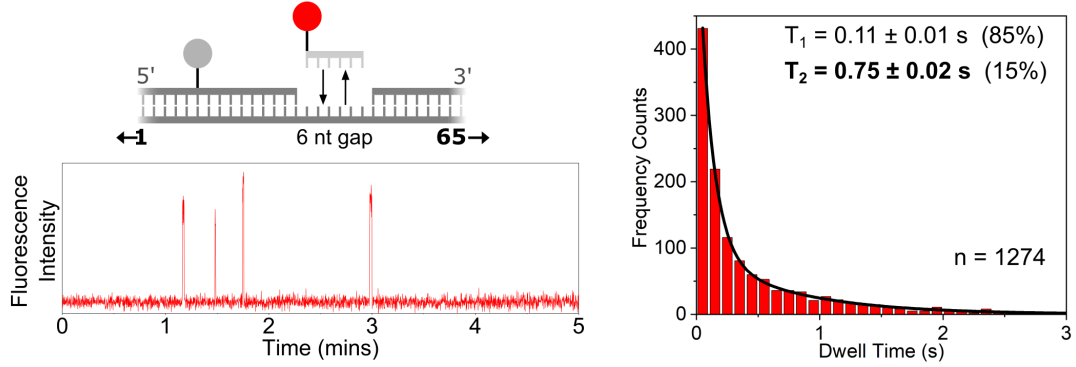


Figure 2.8: Fully complementary 6 nt seal, GS6-compl- $T^{\text{ATTO647N,P28}}$, at 30 nM hybridising in a 6 nt gapped-DNA with average dwell times of $T_1 = 0.11 \pm 0.01$ s (85%) and $T_2 = 0.75 \pm 0.02$ s (15%) recorded at 10 ms exposure time; number of events (n) = 1274.

nt (GS6-compl- $T^{\text{ATTO647N,P28}}$) and the corresponding mismatch with cytosine for guanine at position 3 (GS6-mis3G- $T^{\text{ATTO647N,P28}}$) were designed to be used with a 6 nt gapped-DNA.

Gapped-DNAs with a gap of 6 nt were immobilised on the surface and experiments conducted with both seals. For GS6-compl- $T^{\text{ATTO647N,P28}}$ there were 1274 binding events with average dwell times of $T_1 = 0.11 \pm 0.01$ seconds (85%) and $T_2 = 0.75 \pm 0.02$ seconds (15%) recorded at 10 ms exposure time, see Figure 2.8. However, only a hand full of transient hybridisation events were recorded for the mismatch seal GS6-mis3G- $T^{\text{ATTO647N,P28}}$. The decrease in binding stability of the seal, due to the mismatch, leads to the hypothesis that transient hybridisation was scarcely observed as seals bound on average for a shorter time than the exposure time of the camera. The results here prove the initial hypothesis that shorter seals provided better single base discrimination, with clear discrimination between the fully complementary seal GS6-compl- $T^{\text{ATTO647N,P28}}$ and the single base mismatch

seal GS6-mis3G-T^{ATTO647N,P28}. Shorter seals could prove instrumental in Gap-Seq by providing a simpler way to use the kinetics of transient hybridisation to identify a single base at a known position.

2.2.4 Fluorescence Quenching of Seals

Single molecules that are fluorescently labelled can easily be lost in diffuse background fluorescence at high concentrations. In single-molecule experiments, low concentrations of labelled biomolecules are used to reduce the background fluorescence. Another way to also reduce the background fluorescence is to implement the quenching of fluorophores when not involved in the activity of interest. There are many ways fluorophores can be quenched and one approach used here is static quenching. Previous experiments in this chapter were conducted with seals labelled at the 5' terminus with a single fluorophore. Here, the use of additional fluorophores and dark quenchers is explored with a view to reducing the diffuse background fluorescence. A reduction in the background fluorescence will allow a higher working concentration of seals, hence, increasing the association rate of seals hybridising to the gapped-DNAs. For the previous experiments in this chapter, where the seal was singly labelled and there was no fluorescence quenching implemented, the working concentration was 10-30 nM.

Unbound seals in solution are known to form random coils, a property which could allow fluorescence quenchers at the ends of the seals to be in close proximity leading to contact fluorescence quenching [24]. Upon binding to a gapped-DNA, the seals no longer exist as random coils allowing the two ends of the seal to be apart, therefore, any contact induced fluorescence quenching would

cease. Two modified 8 nt fully complementary seals were designed to test fluorescence quenching in solution, both labelled with ATTO647N at 5' terminus: GSeq-8compl-T^{ATTO647N P28, DABCYL P35} with the fluorescence quencher DABCYL at the 3' terminus and GSeq-8compl-T^{ATTO647N P28,P35} with an ATTO647N at 3' terminus. 8 nt seals have an approximate radius of gyration of 1 nm when unbound and a length of approximately 3 nm when hybridised, assuming the polymer physics model of a ssDNA being a random-coil polymer holds at short lengths. The distance between the fluorophores, which are placed at either end of the DNA, therefore changes when unbound or bound allowing the fluorophores to quench each other or fluoresce respectively.

When illuminated at 640 nm (Fig. 2.9) both seals showed a reduced background fluorescence when compared to singly labelled seals allowing the working concentration to be increased 3 fold to 100 nM. Comparing the two quenching regimes at 30 nM, the average (signal-to-noise ratio) SNR for GSeq-8compl-T^{ATTO647N P28,P35} was 1.24, whereas, the average SNR for GSeq-8compl-T^{ATTO647N P28, DABCYL P35} was 1.14. At 500 nM, bound seals on the surface are still visible for GSeq-8compl-T^{ATTO647N P28,P35} providing an upper limit for the concentration of the seal. However, at 500 nM it is not possible to distinguish single-molecules from the diffuse background fluorescence for GSeq-8compl-T^{ATTO647N P28, DABCYL P35}. Both seals show similar levels of fluorescence quenching, but due to the increased signal as a consequence of the second fluorophore, the seal labelled with ATTO647N at both termini is the more favourable quenching design.

As the Gap-Seq assay is reliant on transient hybridisation to gapped-DNA it is important that the presence of a quencher or secondary fluorophore does not inhibit seal binding to the gapped-DNA and the contact quenching is only present when

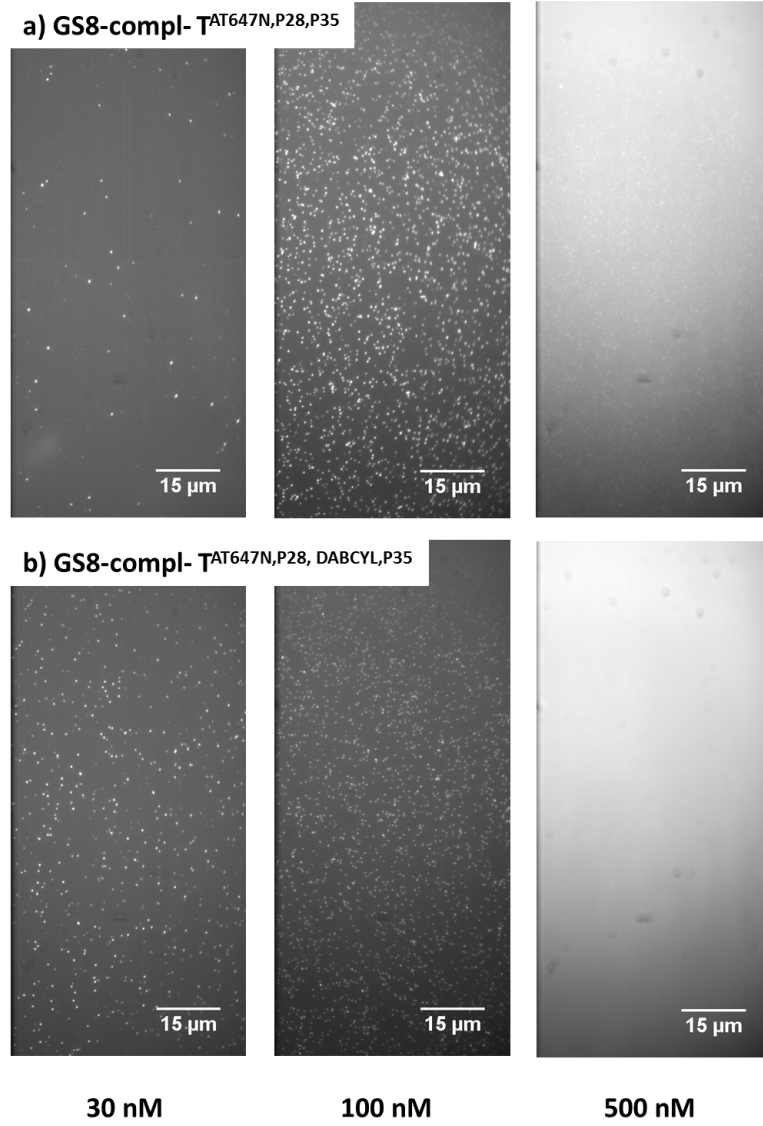


Figure 2.9: Raw microscope data (red emission channel) of seals and surface immobilised 8 nt gapped-DNA construct and illuminated at 638 nm. a) GSeq-8compl-T^{ATTO647N P28,P35} at 30 nM, 100 nM and 500 nM. b) GSeq-8compl-T^{ATTO647N P28,DABCYL,P35} at 30 nM, 100 nM and 500 nM.

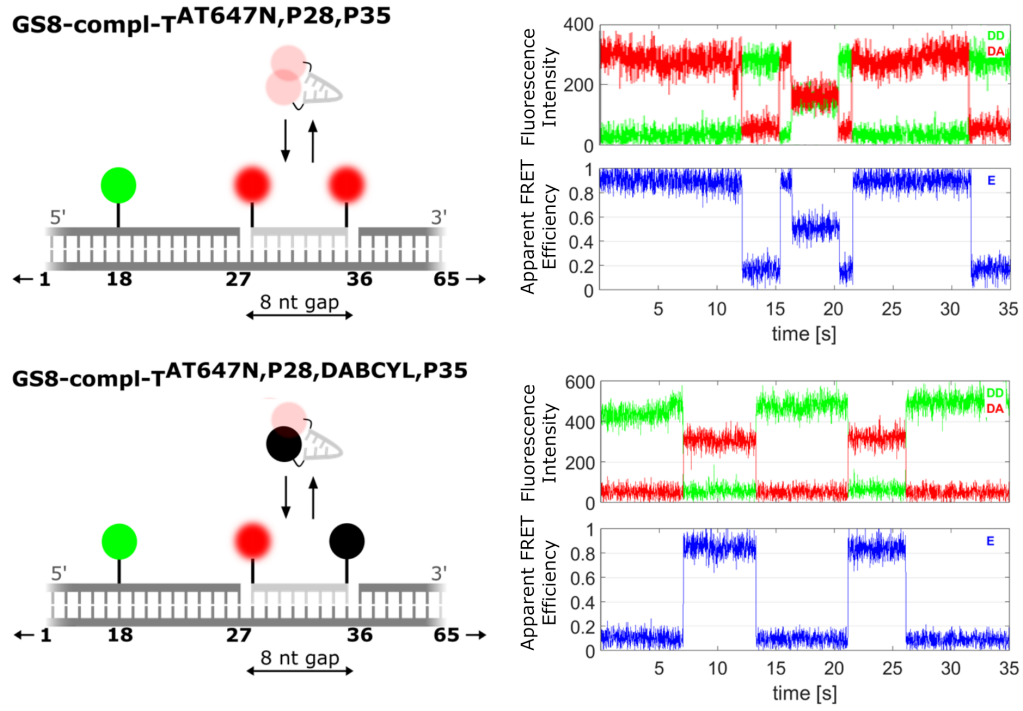


Figure 2.10: Schematics of the fluorescence quenching of GSeq-8compl-T^{ATTO647N P28,P35} and GSeq-8compl-T^{ATTO647N P28, DABCYL P35} in solution and unquenching upon binding. Fluorescence intensity and FRET time traces of 100 nM seals transiently binding to a 9 nt gapped-DNA construct. DD - donor excitation, donor emission; DA - donor excitation, acceptor emission. GSeq-8compl-T^{ATTO647N P28,P35} FRET trace shows apparent FRET efficiency $E = 0.5$ corresponding to fluorescence from secondary fluorophore only. Exposure time 100 ms.

the seals are unbound. Both GSeq-8compl-T^{ATTO647N P28,P35} and GSeq-8compl-T^{ATTO647N P28, DABCYL P35} samples were illuminated at 532 nm to test for transient binding to the 8nt surface immobilised gapped-DNAs using FRET. Transient binding events were observed for both seals (Fig. 2.10) proving that contact quenching was terminated when the seals bound to the gapped-DNA constructs.

Interestingly, GSeq-8compl-T^{ATTO647N P28,P35} showed an intermediate FRET efficiency of 0.5 after $E = 0.9$ for some binding events (Fig. 2.10). A value of $E = 0.5$

is expected for fluorophores of 16-20 bp separation. The ATTO647N, at the 3' terminus of the seal was 16 bps from the Cy3B fluorophore on the gapped-DNA. Hence, this intermediate FRET efficiency can be attributed to FRET between the Cy3B at position 18 and ATTO647N at position 35, when the ATTO647N at position 28 could no longer participate in FRET.

Both of the doubly labelled seals, GSeq-8compl-T^{ATTO647N P28,P35} and GSeq-8compl-T^{ATTO647N P28, DABCYL P35}, showed successful implementation of fluorescence quenching with no hindrance in binding for the seals. Further experiments were designed to incorporate seals terminally labelled with two fluorophores, allowing 100 nM to be used as the new increased working concentration increasing the on rate of hybridisation to the gap.

After the success of quenching and increased signal with doubly fluorescently labelled seals a 7nt seal was designed to include three ATTO647Ns, one at each terminus (as previously) and the third fluorophore in the centre of the seal at position 4. The seal showed fluorescence quenching properties, however, unlike the doubly labelled seals no binding events were observed. The absence of binding events indicates that the third fluorophore either prevents binding to the gap or more likely that when the seal is bound the fluorophores contact induced quenching continues as the fluorophores are in too close proximity. Therefore, unfortunately, triply labelled short seals are not viable to observe transient binding of seals to gapped-DNA.

2.2.5 A-T Exclusive Regions

The previous gapped-DNA binding sequences had a majority of strong bps, however, Gap-Seq needs to be applicable to all possible DNA sequences, particularly areas known for weaker hybridisation, such as A/T rich sequences. To test the ability to observe transient hybridisation for regions of weak binding a 7 nt gapped-DNA construct was designed to have a binding sequence of only A or T bases, inspired by a common promoter region sequence 5'-TATAATA-3'.

To investigate this, 100 nM GS7-ATrich-T^{ATTO647N,P28,P34} seals, which are A-T exclusive, were suspended in the standard imaging buffer that were fully complementary to surface immobilised 7 nt gapped-DNAs. Short transient binding events were observed through the use of FRET between the donor fluorophore on the gapped-DNA and the acceptor fluorophores on the seal (Figure 2.11a). However, only 133 binding events were recorded over 3 fields of view to be able to create a histogram that can be fitted to find an average dwell time, suggesting that some binding events could not be observed due to a limit of exposure time of 10 ms (Figure 2.11b).

To increase the number of detected binding events by increasing the stability of binding, an experiment was conducted to determine the effect of tetramethylammonium chloride (TMAC). TMAC is known to increase duplex stability for A-T rich regions by equalising the melting temperature between majority A-T regions and majority G-C sequences [123]. Again for the same 7 nt seals and gapped-DNA constructs, 3 M TMAC was added to the standard imaging buffer. Transient hybridisation was again observed (Figure 2.11c), with an increase in the number of events observed ($n = 554$) and a dwell time histogram was created from analysis

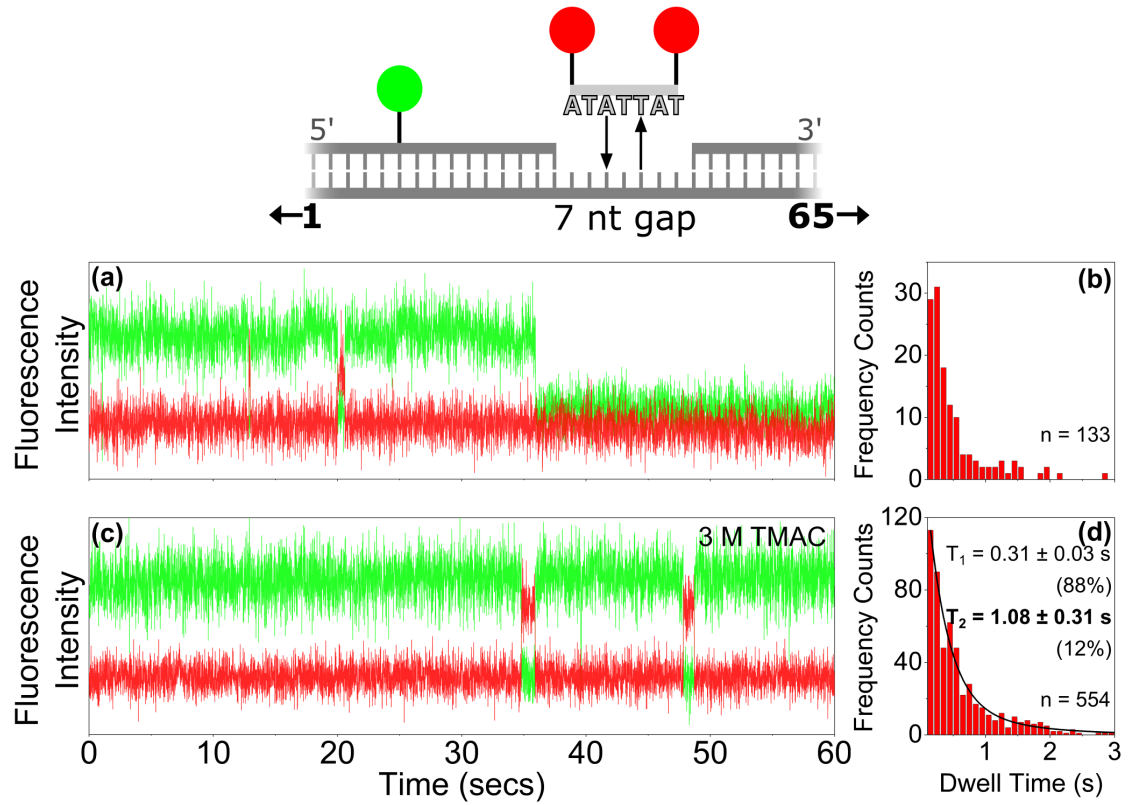


Figure 2.11: Binding to A/T rich sequences to test the applicability of the method to weak binding sequences. A 7 nt seal, exclusively A/T bases, binding in a 7 nt gapped DNA excited by 532 nm laser to produce FRET. a) Fluorescence intensity vs. time trace for the seal binding to the gapped DNA under standard imaging conditions. b) Dwell time histograms of the seal bound in the gap, fitting not possible due to low number of events; number of events (n) = 133. c) Fluorescence intensity vs. time trace for the seal binding to the gapped DNA in the presence of 3 M tetramethylammonium chloride (TMAC). d) Dwell time histogram of binding in the presence of 3 M TMAC, fitted to give $T_1 = 0.31 \pm 0.03$ s (88%) and $T_2 = 1.08 \pm 0.31$ s (12%); number of events (n) = 554.

of the fluorescence intensity vs. time traces (Figure 2.11d). Fitting the dwell time histogram with a decaying biexponential gives the average times of $T_1 = 0.31 \pm 0.03$ seconds (88%) and $T_2 = 1.08 \pm 0.31$ seconds (12%).

Here, transient hybridisation events were observed for ‘weak’ binding sequences for 7 nt seals with the ability to enhance the stability of the binding of the seals through the use of 3 M TMAC.

2.2.6 Interrogation of an Unknown Gap

The experiments earlier in this chapter focus on known sequences for both the gap and the seals, however, if Gap-Seq is to be useful for sequencing, the ability to interrogate unknown sequences is critical.

Degenerate Bases

A degenerate base is formed when a mixture of the four possible bases is used during the synthesis of DNA. Therefore, for every degenerate base (n) there are 4^n variations of the sequence of the DNAs. The resulting concentration of the seal with the fully complementary sequence is $1/4^n$ of the overall concentration of the seals.

Here, seals were designed to simulate investigating an unknown region of three bases with the base being ‘sequenced’ in the centre of the two degenerate bases, see Figure 2.12. Gapped-DNAs with a 7 nt gapped were immobilised on the surface and 10 nM of degenerate seals were placed in solution. Two experiments were conducted, one with the complementary base in the middle position GS7-degen^{P30,P32}-T^{ATTO647N,P28,P34} and the other with a T-T mismatch in the central

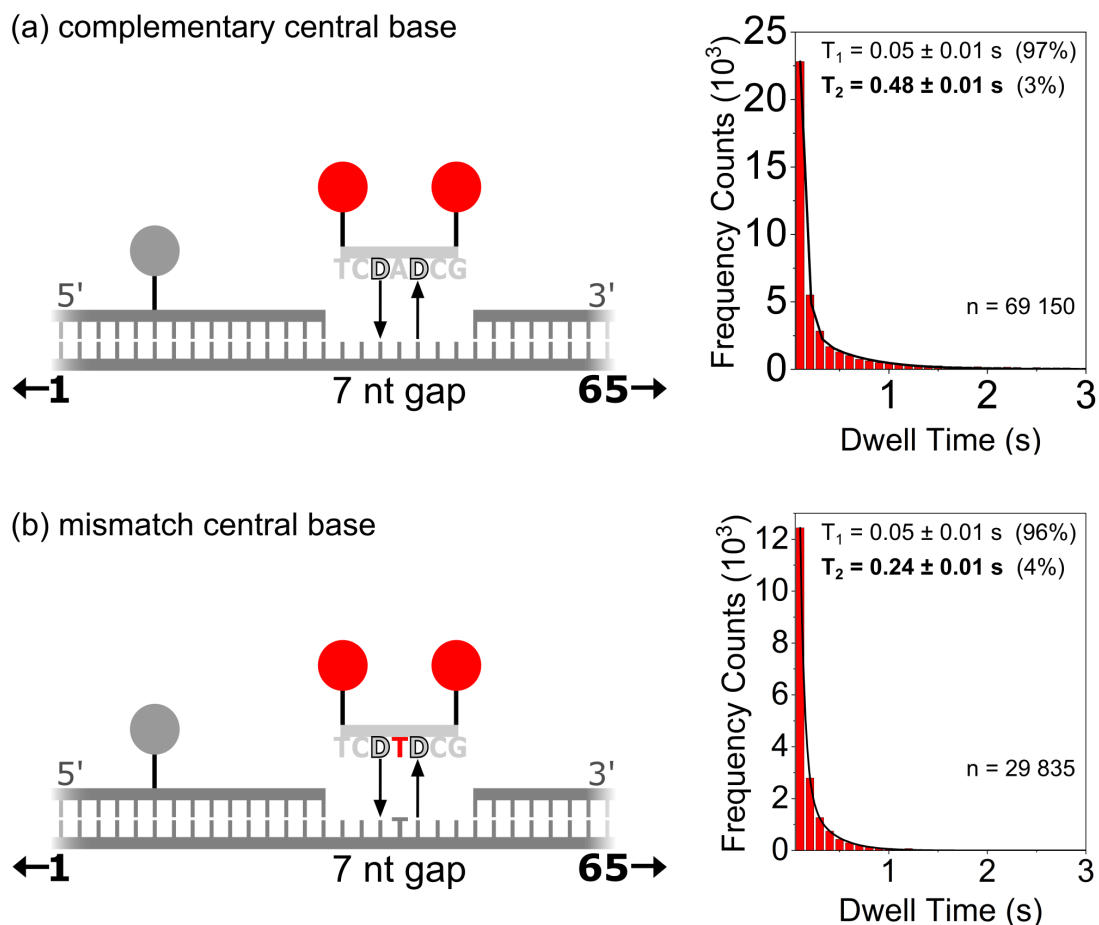


Figure 2.12: Hybridisation of a 7 nt seal doubly labelled with ATTO647N featuring two degenerate bases at position 3 and 5 in the seal to a 7nt gapped-DNA. a) Central base (position 5) is known to be complementary to the corresponding base in the gap. Dwell time histogram of events provides average dwell times of $T_1 = 0.05 \pm 0.01$ s (97%) and $T_2 = 0.48 \pm 0.01$ s (3%); number of events (n) = 69 150. b) Central base (position 5) is known to create a T-T mismatch in the gap. Average dwell times of mismatch seal $T_1 = 0.05 \pm 0.01$ s (96%) and $T_2 = 0.24 \pm 0.01$ s (4%); number of events (n) = 29 835. Both experiments were excited at 640 nm with 10 ms camera exposure time.

position GS7-mis4T-degen^{P30,P32}-T^{ATTO647N,P28,P34}.

Short transient binding events were observed for both the complementary and

the mismatch seal with an average fully hybridised dwell time of $T_2 = 0.48 \pm 0.01$ (3%) seconds for the complementary central base and $T_2 = 0.24 \pm 0.01$ (4%) seconds for the central mismatch base. From the average dwell times (T_2), there is evidence that the mismatch seal exhibited weaker binding compared to the complementary seal. Being able to differentiate a correct base in an ‘unknown’ sequence shows a possible way to use Gap-Seq to investigate unknown sequences. However, there were some issues with using the degenerate seals. When producing degenerate bases, a reduced concentration of the complementary seals is formed as fully complementary seals are only a proportion of the overall seal concentration. The reduced concentration of the fully complementary seal has a knock-on effect of reducing the association rate of hybridisation and increasing the time needed to observe multiple binding events. As interrogating unknown sequences is a key element of a sequencing assay, another method of sequencing an unknown gap was investigated.

Universal Bases - 5-Nitroindole

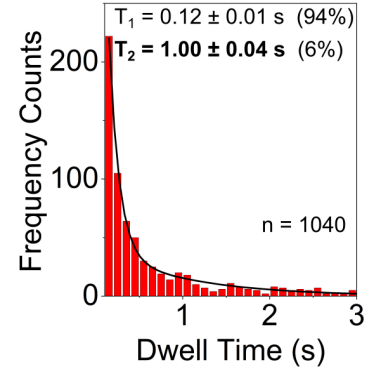
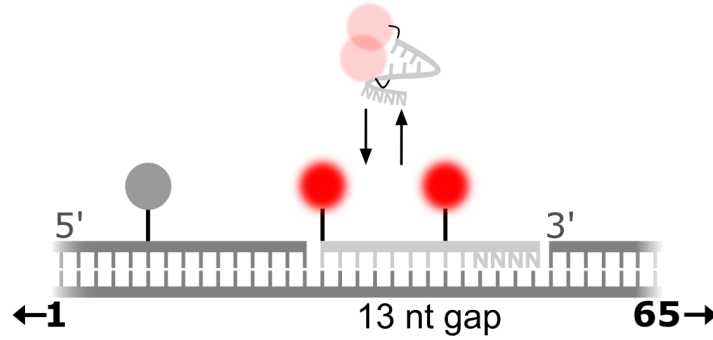
Universal bases are synthetic DNA bases which are designed to bind to all bases, therefore, exhibit complementary binding to any sequence. Here, the universal base 5-nitroindole is investigated as an alternative to the use of degenerate bases when binding to unknown sequences.

A 13 nt seal was designed with four 5-nitroindoles at the 3’ end of the seal as 5-nitroindole when placed towards the terminals of oligonucleotides produces more stable binding than using 5-nitroindole as an internal base close to the centre of the oligonucleotide [20]. The choice of four contiguous 5-nitroindoles was made as multiple 5-nitroindole substitutions that are contiguous have been shown to be

less disruptive than spaced nitroindoles when positioned towards the ends of the duplex [20], [21]. Contiguous bases are less disruptive compared to two or more groups of multiple 5-nitroindole as separate groups favour binding to each other creating secondary structures within an oligo - a hinderance for hybridisation. The optimal number of contiguous 5-nitroindoles is four, with $n \geq 5$ leading to the formation of hairpins [124]. The length of the oligo was chosen to be 13 nt as when used a substitute for natural bases 5-nitroindole introduces instability in the binding. Instead of substituting for natural bases, the 5-nitroindoles can be used to extend the end of an oligonucleotide to prevent a large disruption in stability of hybridisation. Therefore, the optimal design of an oligonucleotide for transient binding to an unknown DNA sequence would contain a single group of contiguous four 5-nitroindoles as an addition to one of the terminus of oligonucleotide that is already known to bind stably enough for single-molecule detection. Therefore, a 13 nt seal was designed with four 5-nitroindoles at the 3' end of the seal doubly labelled with ATTO647N called GS13-N^{10,11,12,13}-T^{AT647N,P28,P35}. This design allows for a 5 base region in the gapped-DNA to be interrogated (the known base preceding the four universal bases, position 9 in the seal). A second corresponding seal was designed to create a G-G mismatch at the known base position (position 9), GS13-mis9G-N^{10,11,12,13}-T^{AT647N,P28,P35}. The mismatch seal allows investigation of any difference in the average dwell time for the seal that is complementary at the known position and a seal with a mismatch at the known position.

First, with surface immobilised gapped-DNA with a 13 nt gap, the complementary seals at 100 nM were placed in imaging buffer containing 0.25 M NaCl and imaged at 640 nm with an exposure time of 10 ms. Transient binding events were observed with dwell time of $T_1 = 0.12 \pm 0.01$ seconds (94%) and $T_2 = 1.00 \pm 0.04$ sec-

(a) fully complementary



(b) mismatch position 9

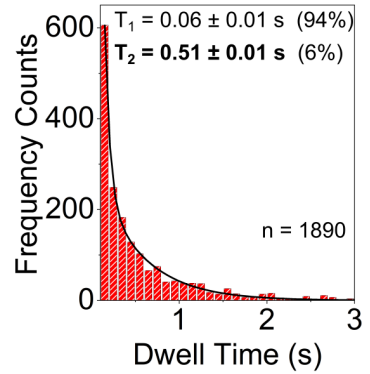
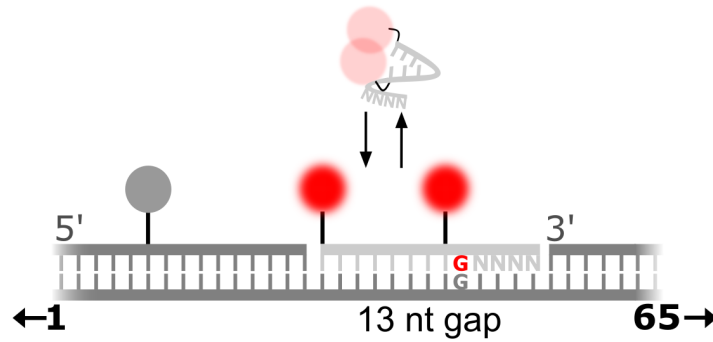


Figure 2.13: Hybridisation of a 13 nt seal doubly labelled with ATTO647N featuring four universal bases (5-nitroindole) at positions 10-13 in the seal to a 13 nt gapped-DNA. a) The seal, GS13-N^{10,11,12,13}-T^{AT647N,P28,P35}, is fully complementary to the sequence in the gap. Dwell time histogram with average dwell times of $T_1 = 0.12 \pm 0.01$ s (94%) and $T_2 = 1.00 \pm 0.04$ s (6%); number of events (n) = 1040. b) Mismatch seal, GS13-mis9G-N^{10,11,12,13}-T^{AT647N,P28,P35}, has a G-G mismatch (position 9) to the sequence in the gap. Average dwell time times of $T_1 = 0.06 \pm 0.01$ s (94%) and $T_2 = 0.51 \pm 0.01$ s (6%); number of events (n) = 1890. Both experiments excited at 640 nm with 10 ms camera exposure time.

onds (6%) for 1040 binding events. For comparison, the mismatch seals at 100 nM in 0.25 M NaCl were imaged binding to gapped-DNA with a 13 nt gap. Average dwell times of $T_1 = 0.06 \pm 0.01$ seconds (94%) and $T_2 = 0.51 \pm 0.01$ seconds (6%) were recorded for the mismatch seal with 1890 binding events. There is a clear difference in the average dwell time T_2 , $\approx 49\%$, between the fully complementary seal and the mismatch seal with the average mismatch seal. These experiments show how universal bases, such as 5-nitroindole, can be used in the Gap-Seq method to interrogate unknown sequences. Specifically, these experiments are a proof-of-principle for identifying the first base in a series of five unknown bases using Gap-Seq.

2.3 Discussion

This chapter shows the proof-of-principle results for Gap-Seq - a high throughput single-molecule DNA sequencing method that uses the binding kinetics of labelled oligonucleotides binding to surface immobilised gapped-DNAs. Although there are multiple single-molecule sequencing methods, no single method allows the observation of protein binding followed by sequencing.

The first experiments of binding to hairpin gapped-DNA showed it was possible to measure the kinetics of seals transiently binding in gapped DNAs. Next, experiments showed it was possible to identify the difference between a fully complementary seal and a seal with a single base-pair mismatch for 8/6-nt gapped-DNAs. Decreasing the length of the seal and the size of the gap in the DNA allowed for clearer discrimination between the two seals highlighting a method to read the sequence of gap a single base at a time. The implementation of fluo-

rescence quenching through a secondary fluorophore on the seal showed that the working concentration can be increased by tenfold allowing a faster rate for hybridisation, therefore, a faster interrogation of the gapped-DNAs. The robustness of observing DNA transient hybridisation for weaker binding sequences that are A/T rich was shown and the use of tetramethylammonium chloride (TMAC) created more stability for such constructs. Finally, proof-of-principle experiments showed that Gap-Seq would be able to function with unknown sequences both through the use of degenerate bases or universal bases.

Gap-Seq is advantageous compared to commercial fluorescence single-molecule DNA sequencing methods as it can be conducted on a widefield TIRF microscope without the addition of advanced equipment such as zero-mode waveguides. However, currently Gap-Seq does not use fluidics, which would reduce the labour intensive process of washing and loading wells. Future installation of fluidics would make the experiments easier to perform; however, loading of the samples is not the time limiting step. Currently, the data collection takes 30 minutes and the data analysis around >3 hours using custom built software. Further work to automating time-trace selection in the analysis step would dramatically reduce the time to perform the whole assay, for example, with the use of artificial intelligence which could be trained on the data already analysed by hand.

In the future, further experiments need to be conducted to establish Gap-Seq as a single-molecule sequencing technique. The first experiment will be to directly read a base. Currently, experiments show discrimination between a fully complementary seal and a seal where the base of interest is mismatched. However, experiments have yet to be conducted with the complementary and the corresponding three mismatch seals. The experiments will either be carried out sequentially, not-

ing the order the seals are added in, or conducted simultaneously through colour coding the fluorophores with respect to the known base used for interrogation as shown in the Gap-Seq assay in Figure 2.1. Such a method is only affective for sequencing a few bases due to its time intensive nature. Once fully developed, Gap-Seq will be an incredibly useful high-throughput single-molecule short-read DNA sequencing method that can, but not limited to, be used to get a deeper understanding of sequence dependent protein-DNA interactions.

Chapter 3

Protein-DNA Interactions

3.1 Introduction

Transcription is a key DNA process for gene expression and one of the main protein-DNA sequence dependent interactions. During transcription, the sequence of the DNA template strand is copied by the enzyme RNA polymerase (RNAP) producing messenger ribonucleic acid (mRNA). Ribosomes then translate mRNA to produce proteins, the products of gene expression. Transcription is regulated by transcription factors which interact with DNA. There are two main classes of transcription factors; activators and repressors. In *E. coli*, activators can bind (in most cases, specifically) to DNA sequences upstream of the promoter region to help stabilise the binding of RNAP, whereas repressors commonly bind to the promoter region blocking the binding of RNAP.

An activator protein of interest is catabolite activator protein (CAP) which is a transcription factor associated with the lac operon. Transcription of the lac operon produces polycistronic mRNAs containing the genes for three different enzymes

that are necessary to transport lactose into the cell, break down lactose and help to remove toxic by-products. For transcription initiation on the lac operon, there have to be two jointly occurring factors, a lack of glucose (the favoured source of carbon) and the presence of lactose. Under such conditions, the transcription sigma factor 70 (σ^{70}) of RNAP recognises the DNA sequences at the two promoter regions at the -35 and -10 positions on the DNA and binds the multi-subunit RNAP to the DNA. CAP increases the affinity of RNAP for the lac promoter, and accelerates the opening of the transcription bubble by RNAP. CAP is a dimer formed from two identical subunits and recognises a twofold symmetric 22-base pair consensus binding sequence 5'- AAATGTGATCTAGATCACATTT-3' [122] at position -61.5 upstream from the transcription start site, only binding when cyclic adenosine monophosphate (cAMP) is present. When CAP binds at the upstream enhancer element it bends the DNA by approximately 80° , with a primary and secondary kink allowing further points of contact between the side of the protein and the DNA [125], see Figure 3.1.

There are a vast array of sequence dependent reactions within bacterial transcription alone and are a clear justification for developing a single-molecule sequencing method to investigate such events. Gap-Seq was developed to be used as the last step in a larger assay to link phenotype to sequence for protein-DNA interactions. The assay will consist of three individual steps (see Figure 3.2): Step 1 - single-molecule protein interactions with surface immobilised double stranded DNAs (dsDNAs) are observed, Step 2a - the dsDNAs are denatured leaving immobilised single stranded DNAs (ssDNAs), Step 2b - gapped-DNA constructs are formed on the immobilised ssDNAs by annealing in situ of left flanking top strands and right flanking top strands to known sequences either side of the (unknown)

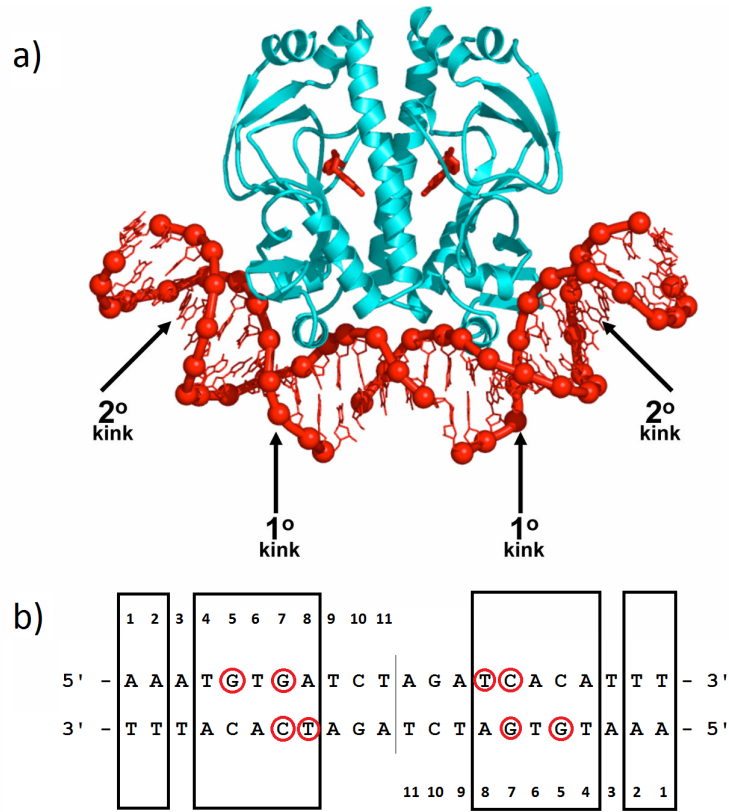


Figure 3.1: a) Structural model of catabolite activator protein (CAP) bound to DNA. CAP bends DNA $\approx 80^\circ$ through primary (1°) and secondary (2°) kinks. b) CAP recognises a twofold symmetric 22 bp consensus binding sequence. Black boxes indicate the regions of the sequence which highly affect CAP binding, red circles indicate CAP-base contacts. Adapted from [125].

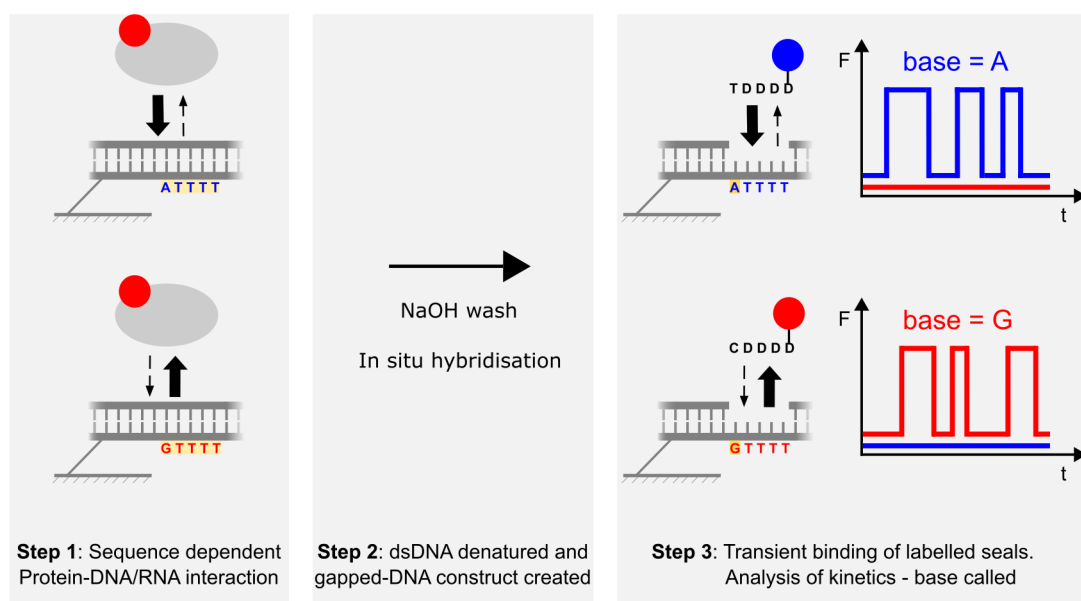


Figure 3.2: The principle of sequencing protein-DNA/RNA interactions using Gap-Seq. Step 1: Protein-DNA interactions observed at the single-molecule level. Step 2: dsDNA denatured using NaOH and gapped-DNA formed from hybridisation in situ. Step 3: Gap-Seq assay is performed with the transient binding of fluorescently labelled DNAs to gapped-DNAs. Step 3 can be repeated with different seals to interrogate subsequent bases (D is a degenerate or universal base).

sequence of interest, Step 3 - sequencing the gapped-DNA (Gap-Seq) through analysing of the kinetics of transiently binding fluorescently labelled oligos (seals). By using this assay, it is hoped that protein-DNA sequence dependent reactions can be studied at high throughput giving important insight into DNA-related processes.

This chapter first examines how sequence, especially single base changes can affect the behaviour of sequence specific binding proteins, particularly CAP. In order to apply Gap-Seq to protein-DNA interactions the assay to produce gapped-DNA in situ is optimised. The final experiments show how protein-DNA interactions can be linked to sequence at the single-molecule level using Gap-Seq and how the

method can be used to examine multiple surface immobilised sequences.

3.2 Results

3.2.1 CAP Binding

The first experiments conducted aimed to examine the sequence dependent nature of CAP binding at the single-molecule level. A 49-nucleotide dsDNA was immobilised on the surface via PEG, biotin and neutravidin (biotin at 5' end of DNA), with the 22 base-pair consensus CAP binding sequence within the dsDNA. The DNA was labelled with Cy3B at the 3' end allowing for the localisation of the dsDNA when excited at 532 nm and viewed in the green emission channel of the microscope, left panel in Figure 3.3. Each of the two CAP subunits was singly labelled with Alexa647, allowing for excitation at 640 nm and emission in the red emission channel of the microscope. As the imaging was conducted in TIRF, labelled single CAP proteins can be seen in the red emission channel when bound at the surface (right panel in Figure 3.3). From this, fluorescence intensity vs. time traces can be produced allowing analysis of CAP binding kinetics.

The consensus sequence was surface immobilised and 1.5 nM of CAP was placed in solution with 0.2 mM cAMP to activate binding. As can be seen in Figure 3.4, single CAP molecules were observed binding to the consensus binding sequence dsDNA. HMM analysis was conducted on the fluorescence intensity vs. time traces and a dwell-time histogram was created (number of binding events = 1239). The histogram was fitted with a decaying biexponential fit to calculate the average dwell times. As can be seen in Figure 3.4, a single decaying exponential fit is an

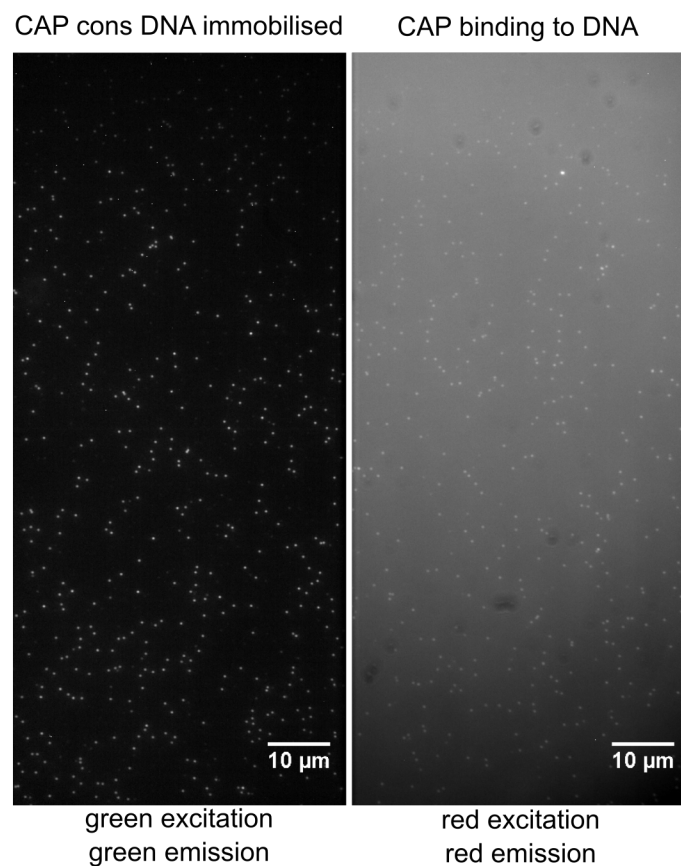


Figure 3.3: Left: Surface immobilised Cy3B 49 nt dsDNA containing 22 base-pair CAP consensus binding sequence excited at 532 nm and emission in the green channel of the microscope. Right: Alexa647 labelled CAP binding to the surface immobilised DNA excited at 640 nm and emission detected in the red emission channel of the microscope. Both images taken in TIRF regime, 100 ms exposure.

inaccurate fit of the data - consistent with the theory that this is not a simple two-state process. From the biexponential fit, the average long dwell time was observed as $T_2 = 8.45 \pm 0.34$ seconds (32%) and the average short dwell time as $T_1 = 1.25 \pm 0.09$ seconds (68%). As with the labelled seals binding the gapped-DNA in the previous chapter, the binding of the protein should not be thought of as a simple two-state model of fully bound and fully unbound CAP. Here, the long dwell time relates to fully binding to the consensus binding sequence with the shorter dwell time due to partial binding followed by dissociation of the protein.

To assess the effect of changing one base in a sequence, a single base at position 5 in the CAP binding sequence was flipped (C changed to G). The alteration was made on both strands of the DNA creating a fully hybridised construct. Position 5 was chosen as it is a key contact for CAP when binding. This sequence was labelled as '1 x mutant sequence', given in Figure 3.5. Again, transient binding events were observed when the CAP bound to the immobilised 1 x mutant sequence and fluorescence intensity vs. time traces for three example molecules can be seen in Figure 3.5. By plotting a dwell time histogram ($n = 1031$) average dwell times of $T_1 = 1.46 \pm 0.09$ seconds (78%) and $T_2 = 5.18 \pm 0.47$ (22%) seconds were observed. When comparing T_2 values, a single base change reduces the bound time of the CAP by $\approx 39\%$ and shows clearly that even small changes to sequence can have an effect on the phenotype of a protein.

The CAP binding sequence is symmetric, as CAP is a dimer of identical subunits, to observe the full effect of a single base change the change in sequence should be symmetric. Therefore, a second construct was designed and named '2 x mutant sequence' where the bases at both position 5 and position 18 were flipped. Again, full complementary base pairing was conserved for the construct. Single-molecule

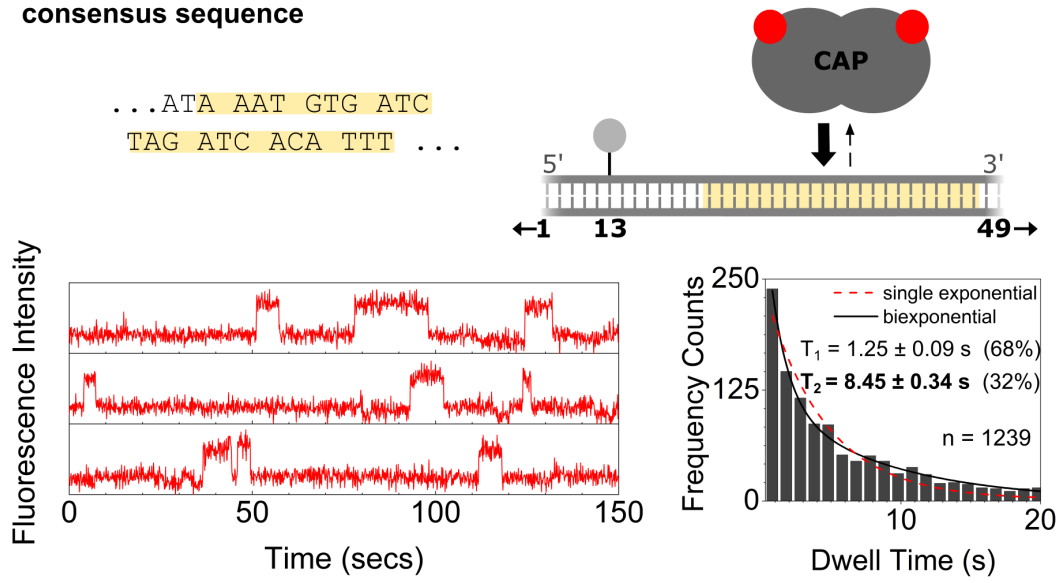


Figure 3.4: CAP binding to the consensus binding sequence. Top left: Consensus binding sequence (single strand) highlighted in yellow. Top right: Alexa647 labelled CAP binding to dsDNA (consensus binding region in yellow). dsDNA is labelled at position 13 with Cy3B used for localisation of the DNA. Bottom left: Fluorescence intensity vs. time traces for three separate dsDNA localisations with binding of CAP molecules characterised by large increases of fluorescence intensity and large decreases with dissociation. 100 ms exposure time, 1.5 nM CAP. Bottom right: Dwell time histogram of observed binding events ($n = 1239$) fitted with a single decaying exponential fit (red dashed) giving a poor fit to the data and a decaying biexponential fit (black) giving $T_1 = 1.25 \pm 0.09$ s (68%) and $T_2 = 8.45 \pm 0.34$ s (32%).

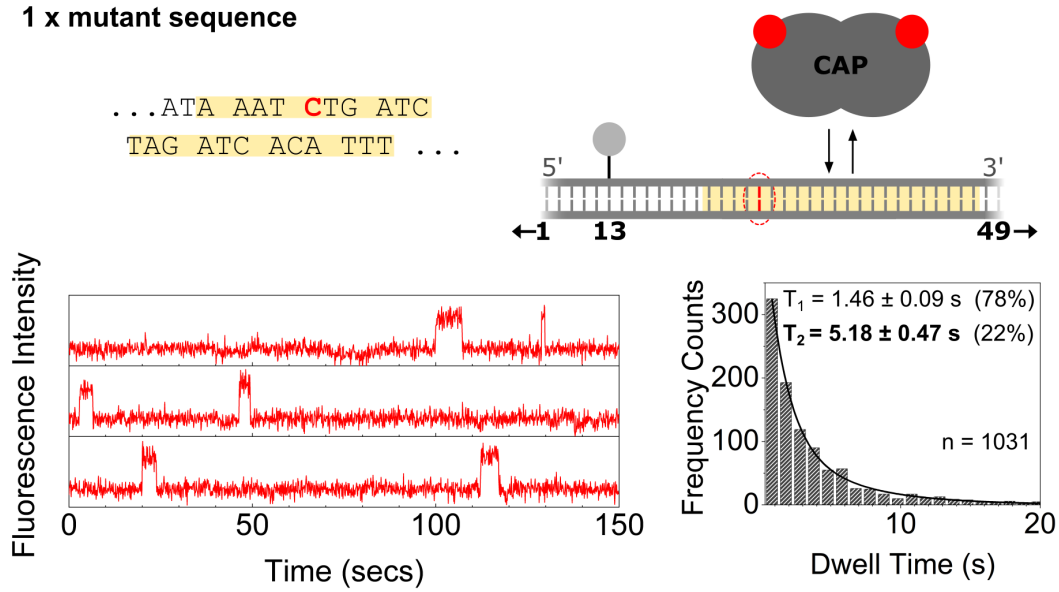


Figure 3.5: CAP binding to the 1 x mutant binding sequence. Top left: 1 x mutant binding sequence (single strand) highlighted in yellow. Top right: Alexa647 labelled CAP binding to dsDNA (binding region in yellow, mutation in red ellipse). dsDNA is labelled at position 13 with Cy3B used for localisation of the DNA. Bottom left: Fluorescence intensity vs. time traces for three separate dsDNA localisations with binding of CAP molecules characterised by large increases of fluorescence intensity and large decreases with dissociation. 100 ms exposure time, 1.5 nM CAP. Bottom right: Dwell time histogram of observed binding events ($n = 1031$) fitted with a decaying exponential fit giving $T_1 = 1.46 \pm 0.09$ s (78%) and $T_2 = 5.18 \pm 0.47$ s (22%).

binding events were observed for the 2 x mutant binding sequence, see Figure 3.6 for example fluorescence intensity vs. time traces. A dwell time histogram was created from 2177 events and average dwell times of $T_1 = 0.94 \pm 0.08$ seconds (60%) and $T_2 = 3.31 \pm 0.14$ seconds (40%) were observed. Here, T_1 was reduced compared to the consensus and 1 x mutant binding sequences which is to be expected due to a likely increase in abortive binding events. When comparing T_2 values for all three constructs there is a clear drop in the stability of binding with only slight mutations to sequence, with $\approx 39\%$ decrease in average bound time from the consensus sequence to the 1x mutant sequence and $\approx 60\%$ decrease in the average bound time for CAP between the consensus sequence and the 2 x mutant sequence. Also, it was observed that there was a slight increase in the time for association of CAP with the addition of a single mutation in the binding sequence, however, a slight decrease in the time for association for the doubly mutated sequence. The changes in the time for association of CAP were not large and due to the association time having a CAP concentration dependence, the time for association was not used as a measure for direct comparison between the different binding sequences. As shown here, single base changes can have drastic effects on the observed behaviour of DNA binding proteins and is the motivation to be able to link phenotype to sequence at the single-molecule level.

3.2.2 Forming a Gapped-DNA In Situ From dsDNA

In order for Gap-Seq to be performed after a protein-DNA assay the surface-immobilised DNA must change from dsDNA to gapped-DNA. As the desired outcome is to probe directly the sequence in the gap a method needed to be developed

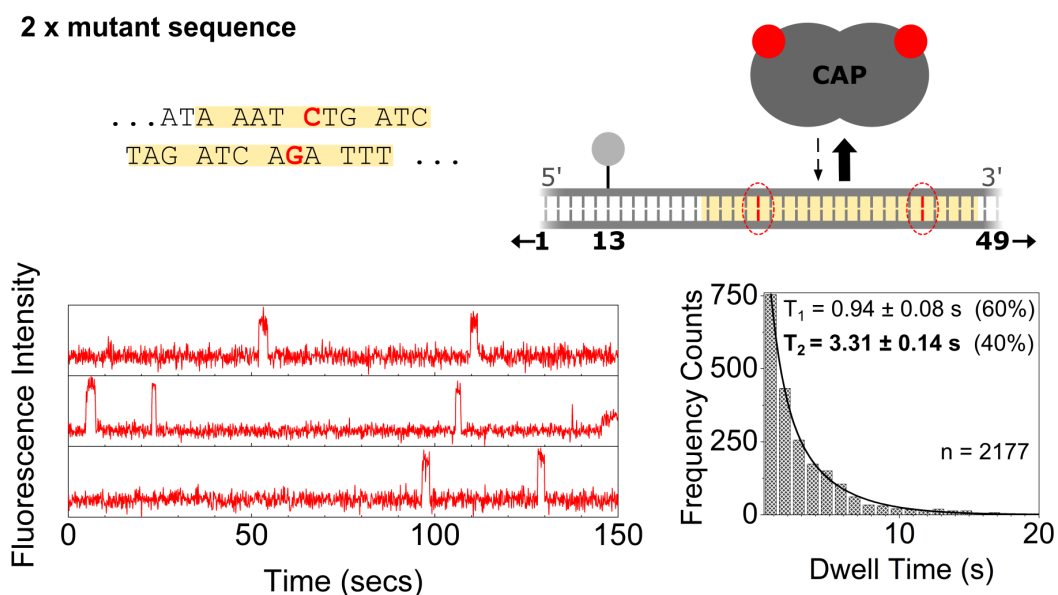


Figure 3.6: CAP binding to the 2 x mutant binding sequence. Top left: 2 x mutant binding sequence (single strand) highlighted in yellow. Top right: Alexa647 labelled CAP binding to dsDNA (binding region in yellow, mutations in red ellipses). dsDNA is labelled at position 13 with Cy3B used for localisation of the DNA. Bottom left: Fluorescence intensity-time traces for three separate dsDNA localisations with binding of CAP molecules characterised by large increases of fluorescence intensity and large decreases with dissociation. 100 ms exposure time, 1.5 nM CAP. Bottom right: Dwell time histogram of observed binding events ($n = 2177$) fitted with a decaying exponential fit giving $T_1 = 0.94 \pm 0.08$ s (60%) and $T_2 = 3.31 \pm 0.14$ s (40%).

to keep the biotinylated bottom strand of the DNA and replace the top strand with two separate strands: the right flanking strand and the left flanking strand. This gap formation needs to be performed in situ leading to two steps: a way to denature the dsDNA and a way to then anneal oligos to create a gap in situ to the immobilised bottom strand.

Denaturing and Reannealing dsDNA FRET Standards In Situ

The first DNA construct used to investigate denaturing and re-annealing of surface immobilised DNA was a 45 nt dsDNA FRET standard, see Figure 3.7. The FRET standard was biotinylated on the bottom strand at the 3' end and labelled with Cy3B at the 5' end. The top strand was labelled with ATTO647N at position 28, relative to the biotin, creating a 17 base-pair gap between the two dyes - apparent FRET efficiency of $E \approx 0.4-0.5$.

An experiment was conducted to denature and re-anneal dsDNA in situ by observing FRET efficiency changes under 532 nm illumination. First, the DNA FRET standards were imaged allowing doubly labelled DNA to be identified ($n = 370$) and a majority $E = 0.45$ recorded (Figure 3.7a). After a 1 minute incubation with 10 mM sodium hydroxide (NaOH) the same field of view was imaged again, with only 2 molecules of the doubly labelled DNA FRET standards remaining (Figure 3.7b). No distribution of E at 0.45 was recorded proving the success in denaturing the DNA FRET standard. The sample was then incubated with 50 nM of the top strand, STD45-T^{AT647N,P28}, in the presence of 0.5 M NaCl for 1 minute, washed and imaged. Under illumination at 532 nm, 136 annealed DNA FRET standards were observed (Figure 3.7c) along with a distribution of E at 0.45. The reappearance of FRET at $E = 0.45$ shows the success of re-annealing the DNA FRET

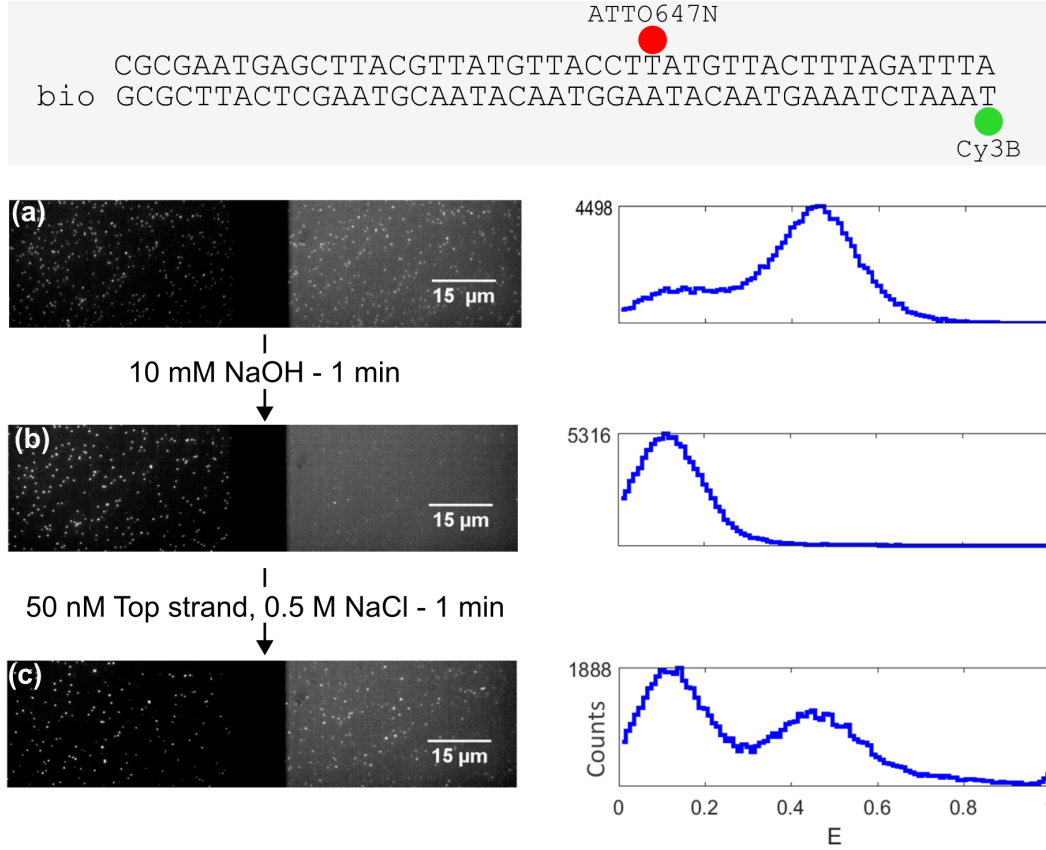


Figure 3.7: FRET standard construct created from STD45-B^{Bio,P1,Cy3B,P45} and STD45-T^{AT647N,P28} with a 17 bp separation between Cy3B and ATTO647N. a) Surface immobilised DNA FRET standards before NaOH wash. Cropped raw image of green (left) and red (right) emission channels of the fluorescence of doubly labelled DNA FRET standards ($n=370$). FRET distribution showed a majority distribution at $E=0.45$. b) After 1 minute 10 mM NaOH only 2 DNA FRET standards remained. FRET distribution showed a single peak at 0.15 due to noise. c) After re-annealing with 50 nM STD45-T^{AT647N,P28} in 0.5 M NaCl doubly labelled FRET standards were observed ($n=136$) and a FRET efficiency distribution at 0.45 was recorded. All samples illuminated at 532 nm.

standard. E distributions at 0.15 reflect the donor-only peak, and indicate the fraction of incomplete hybridisation. Not all DNA FRET standards re-annealed, with a re-annealing efficiency of $\approx 35\%$ observed. Quantification is hindered by the photobleaching of the donor fluorophore attached to the immobilised bottom strand as photobleaching prevents detection of re-annealed dsDNA.

Denaturing CAP dsDNA and Gap Hybridisation In Situ

After establishing a basic proof-of-principle assay for the denaturing and reannealing of the dsDNA FRET standard, an assay was developed to denature surface immobilised dsDNA used for CAP binding and anneal flanking strands to the remaining immobilised bottom ssDNA to create gapped-DNA.

The same dsDNA CAP binding construct was used as described in the previous section: 49 nt dsDNA, including the consensus CAP binding sequence, biotinylated at the 5' end and labelled with Cy3B on the top strand. Here, immobilised dsDNAs were recorded in the green emission channel and the absence of the signal after the denaturing step indicated successful denaturing of the dsDNA. After denaturing the DNA, a gapped-DNA construct was created by incubating the remaining surface-immobilised ssDNA with complementary right and left flanking strands in annealing buffer (0.5 M NaCl). Optimisation of the denaturing protocol was carried out including varying the concentration of NaOH used (1 mM to 25 mM) and the length of time it was present in the sample for (2 seconds to 2 minutes). Similarly, the protocol for hybridised a gap in situ was optimised through varying the concentration of the flanking strands (1 nM to 10 nM) and the incubation time (2 minutes to 5 minutes).

As can be seen in Figure 3.8, a single field of view originally containing ds-

DNA containing the consensus CAP binding sequence ($n = 1296$) was treated with 25 mM NaOH (2 minute incubation) leaving only 3% ($n = 40$) of the single-molecules on the surface. Following the denaturing of the DNA, a hybridisation mix was introduced for 5 minutes which contained 10 nM of each of the top flanking strand necessary to form a gapped-DNA. The left-hand flanking strand was labelled with Cy3B and under 532 nm excitation 860 single-molecules were observed on the surface. It was not possible to quantify from this observation alone that all 860 molecules represent complete gapped-DNA as there may have been some partial gapped-DNA that were missing the right-hand flanking strand and also some fluorescent molecules could have been due to non specific interactions with the surface. This experiment was analysed further in the next section where transient binding was observed in the newly formed gapped-DNAs and 450 fluorescence intensity vs. time traces were picked for analysis, therefore, 35% of the localisations of dsDNA showed transient DNA hybridisation. This is, however, not the maximum number of gapped-DNAs that are formed, only those that produce fluorescence intensity vs. time traces where there was measurable transient DNA hybridisation were selected for analysis. As experiments were conducted some further optimisation of the protocol was made. The final protocol found here to be the most effective was: 25 mM NaOH - 2 seconds, hybridisation mix 10 nM - 5 minutes. A description of the full assay can be found in Materials and Methods.

3.2.3 CAP Binding Followed by Gap-Seq

The motivation for the development of Gap-Seq was to observe sequence dependent protein-DNA interactions and then sequence the DNA in the same experiment,

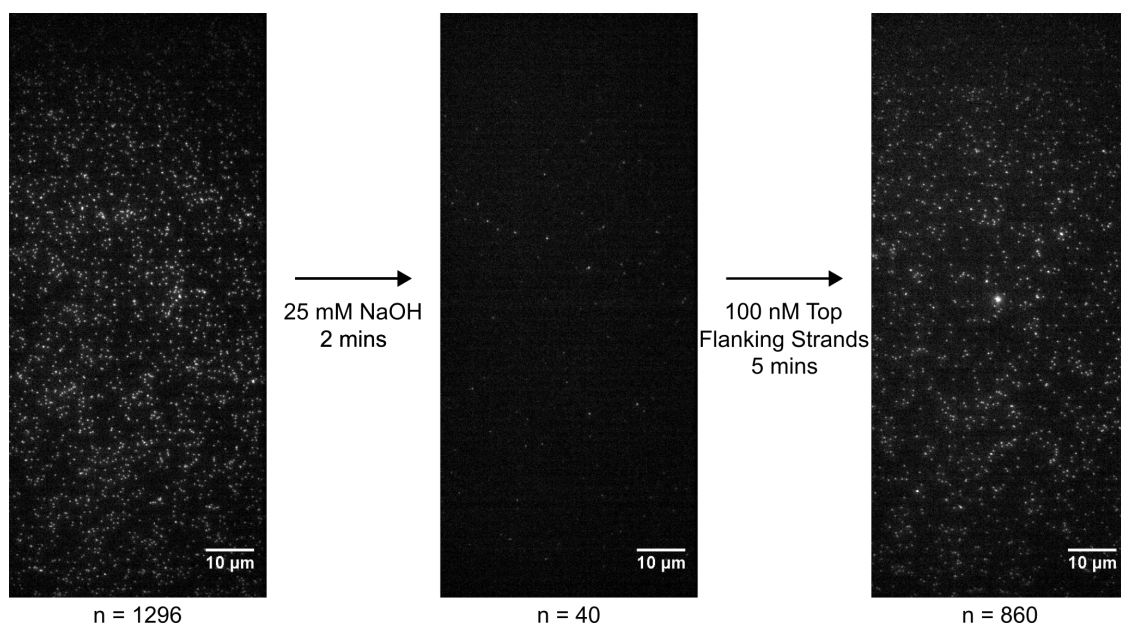


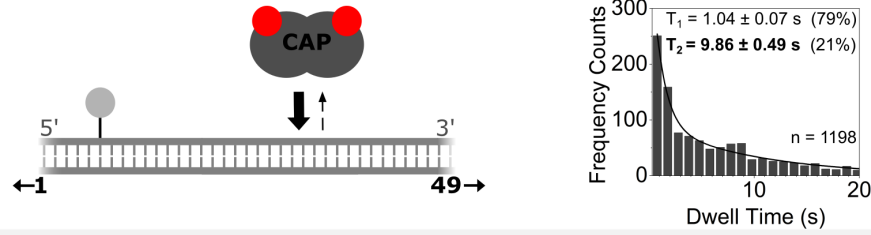
Figure 3.8: Denaturing dsDNA DNA and forming gapped-DNA in situ for the same field of view. Left panel: Green emission channel with 1296 dsDNA labelled with Cy3B immobilised on the surface. Middle panel: After 25 mM NaOH incubation for 2 minutes, there are only 40 single-molecules visible in the green emission channel. Right panel: After incubation with hybridising mix of 10 nM of right hand flanking strand and 10 nM of left hand flanking stranded (labelled with Cy3b) for 5 minutes to form gapped-DNA, 860 single-molecules can be observed on the surface. All illumination at 532 nm, 100 ms exposure.

with high throughput allowing many different sequences to be analysed in parallel. Here, as proof-of-principle that an assay can be conducted to observe protein-DNA interactions followed by probing the sequence with Gap-Seq, the following experiments were conducted, see Figure 3.9.

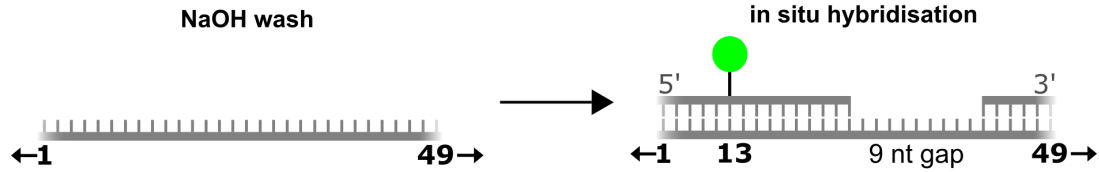
- Step 1: CAP binding assay (consensus sequence) with 1.5 nM CAP was performed and localisations of dsDNA identified.
- Step 2: Two minute denaturing of dsDNA with 25 mM NaOH, hybridisation of the 9 nt gap in situ with the gap hybridisation mix of the top flanking strands at 10 nM for five minutes.
- Step 3: Transient binding, at the localisations identified in Step 1, of either the complementary 9 nt probe GS9-CAP-T^{AT647N,P21,P29} or the single base mismatch probe GS9-CAP-mis5C-T^{AT647N,P21,P29} at 100 nM.

Step 1 allowed the behaviour of CAP binding to the consensus binding sequence to be characterised and average bound times of $T_1 = 1.04 \pm 0.07$ (79%) seconds and $T_2 = 9.86 \pm 0.49$ (21%) seconds were recorded. Step 2 denatured the dsDNA and hybridised gapped-DNA in situ (example of this stage is shown in Figure 3.8). Step 3 was conducted with one seal, either complementary 9 nt seal GS9-CAP-T^{AT647N,P21,P29} or the single base mismatch seal GS9-CAP-mis5C-T^{AT647N,P21,P29}. When using the complementary seal average bound times of $T_1 = 0.63 \pm 0.01$ seconds (93%) and $T_2 = 3.29 \pm 0.32$ seconds (7%) were observed. The single base mismatch probe gave average bound times of $T_1 = 0.41 \pm 0.01$ seconds (88%) and $T_2 = 1.70 \pm 0.06$ seconds (12%). Comparing T_2 times (full hybridisation) for both the complementary and the single mismatch seal there is clear difference

Step 1: CAP binding to consensus sequence



Step 2: Denature and hybridised GAP insitu



Step 3: Add either complementary or mismatch seal

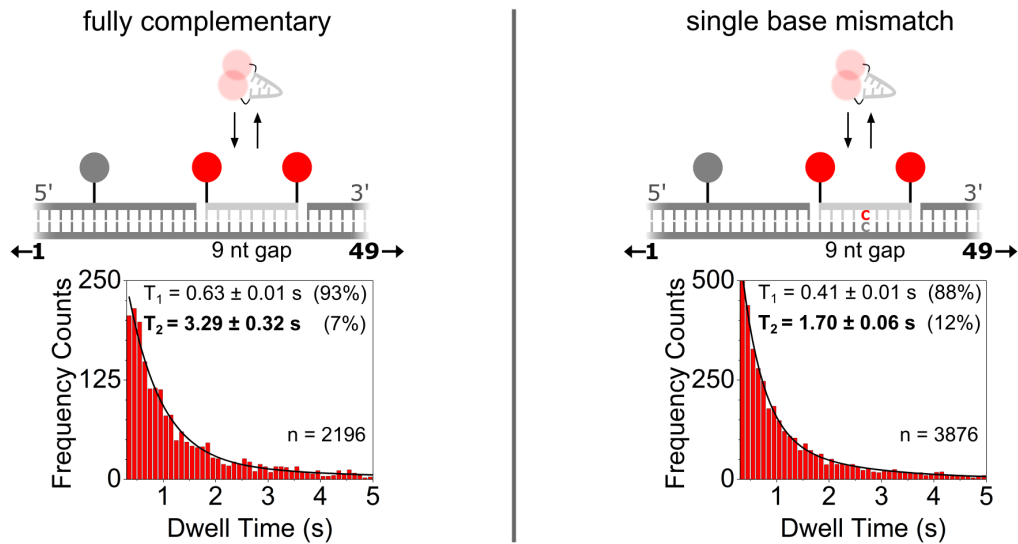


Figure 3.9: Full proof-of-principle assay, with experimental data, to link phenotype to sequence consisting of three steps. Step 1: 1.5 nM CAP doubly labelled with Alexa647 binding to CAP consensus sequence dsDNA construct. The average dwell times observed as $T_1 = 1.04 \pm 0.07$ s (79%) and $T_2 = 9.86 \pm 0.49$ s (21%). Step 2: DNA is denatured using 25 mM NaOH (2 mins). 9 nt gapped-DNA formed in situ with 10 nM of left and right flanking strands in annealing buffer (gap hybridisation mix). Step 3: In two separate experiments 100 nM GS9-CAP- $T^{\text{AT647N,P21,P29}}$ (left) and 100 nM GS9-CAP-mis5C- $T^{\text{AT647N,P21,P29}}$ (right) seals were observed binding to a 9 nt gapped-DNA and dwell time histograms plotted. The average dwell times were $T_1 = 0.63 \pm 0.01$ s (93%) and $T_2 = 3.29 \pm 0.32$ s (7%) for the complementary probe and $T_1 = 0.41 \pm 0.01$ s (88%) and $T_2 = 1.70 \pm 0.06$ s (12%) for the single base mismatch probe.

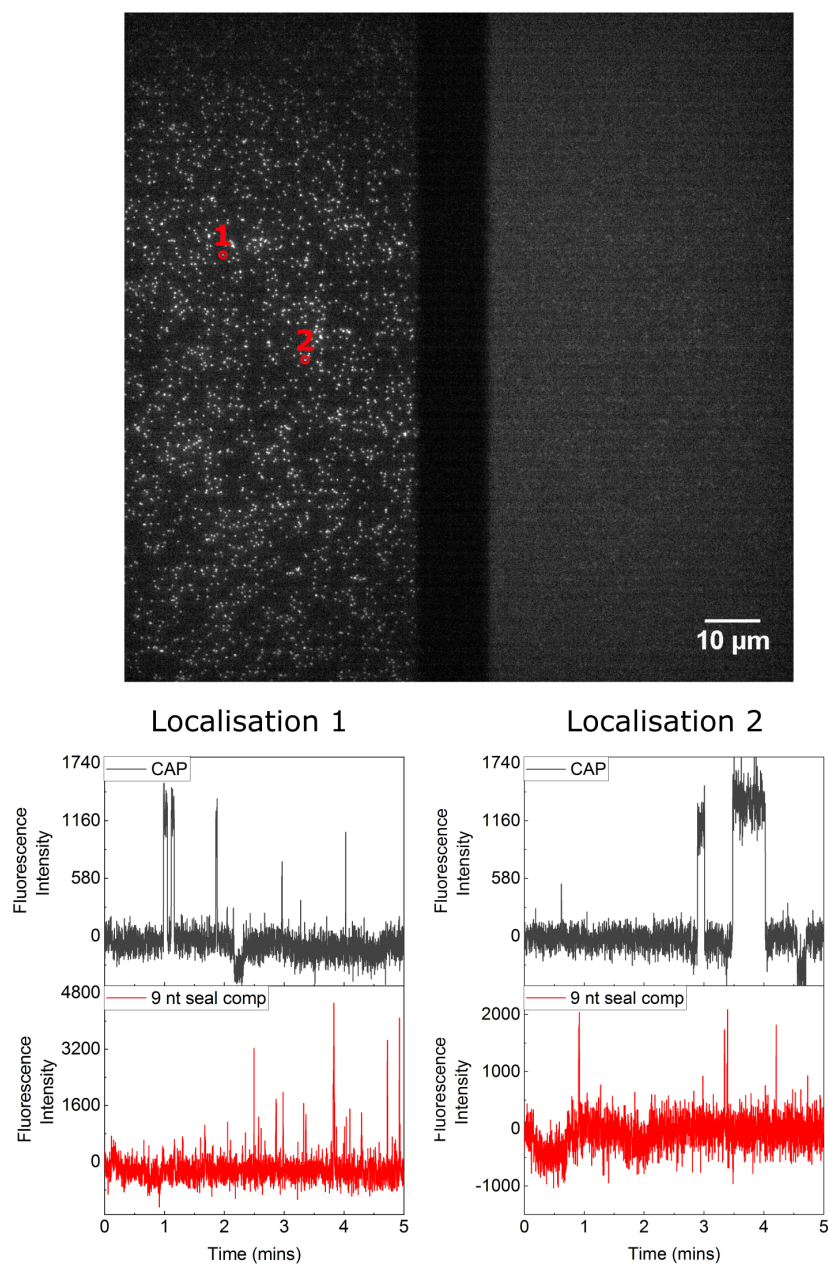


Figure 3.10: Top: Two example localisations of consensus CAP binding sequence dsDNA immobilised on the field of view for an experiment excited at 532 nm, 100 ms exposure. Bottom: Fluorescence intensity vs. time traces for CAP binding (grey) to the localisations and GS9-CAP-T^{AT647N,P21,P29} binding (red) to gapped-DNA formed at the localisations. All traces under 640 nm excitation, 100 ms exposure.

with a reduction of $\approx 48\%$ in the dwell time for the single base mismatch. Figure 3.10 shows how for single dsDNA localisations the fluorescence intensity vs. time traces for both CAP binding in Step 1 to the same localisation (grey traces) and seal binding in Step 3 in the gapped-DNA formed at the localisation (red traces) can be observed allowing phenotype to be linked to sequence for single molecules. Negative drops from the baseline fluorescence in the fluorescence intensity vs. time traces were due to the background fluorescence correction and were ignored by the HMM analysis. These experiments show that it is feasible to conduct a single-molecule fluorescence protein-DNA sequence dependent interaction followed by the Gap-Seq assay.

3.2.4 Mixed dsDNA on the Surface Followed by Gap-Seq

Gap-Seq is being developed to be a high throughput method to be used with different DNA sequences randomly distributed on the surface. All previous experiments were conducted with only a single DNA sequence immobilised on the surface. Here, two DNA sequences were mixed on the surface, the consensus CAP binding dsDNA labelled with Cy3B (green fluorophore) and the 1 x mutant CAP binding DNA labelled with Cy5 (red fluorophore). The consensus dsDNAs were localised in the green emission channel and the 1 x mutant dsDNAs in the red emission channel. Also present on the surface were gold nanoparticles labelled with green and red fluorophores to act as fiducial markers used in data analysis to correct for lateral drift of the field of view [126], see Materials and Methods Chapter. An example of the field of view of such a sample when excited at 532 nm and 640 nm can be seen in Figure 3.11.

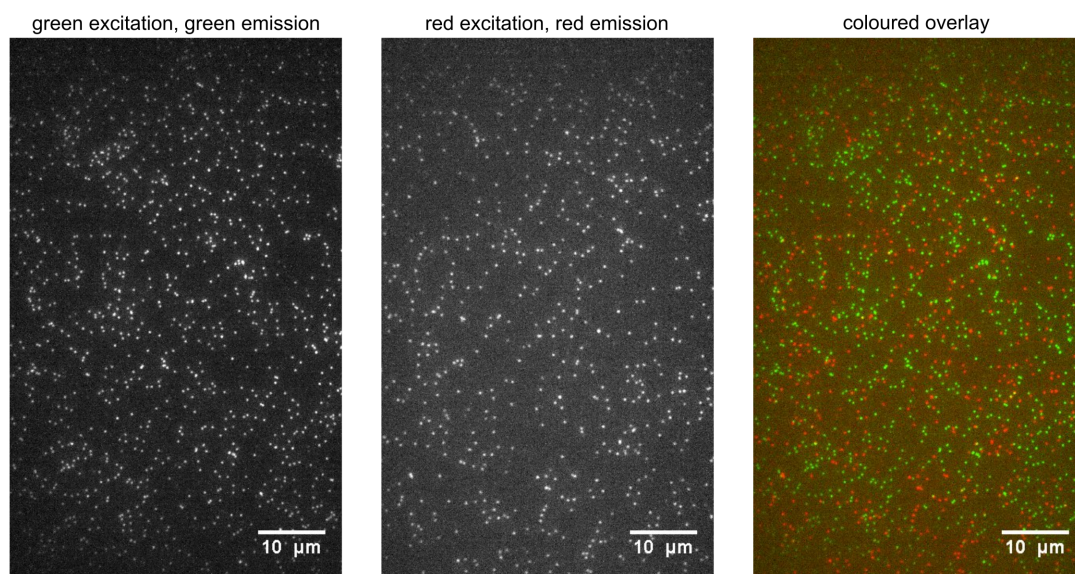


Figure 3.11: Two different DNA sequences immobilised on the same surface (Step 1). Left: The consensus CAP binding dsDNA, labelled with Cy3B, imaged in the green emission channel when excited at 532 nm. Middle: The 1 x mutant CAP binding dsDNA, labelled with Cy5, imaged in the red emission channel when excited at 640 nm. Right: Coloured overlay of the two emission channels to show the true field of view.

Experiments were conducted following the same assay as shown in Figure 3.9 with some slight adjustments. Step 1 localised the dsDNAs on the surface, however, CAP was not introduced for assessment of its binding kinetics. CAP binding to the same dsDNA sequences is documented earlier in the chapter in Figures 3.4 & 3.5. In Step 2, the NaOH wash was performed for 2 seconds instead of 2 minutes and the gap hybridisation mix contained 100 nM of the top flanking strands. Gapped-DNAs with 9 nt gaps were formed and a seal was introduced in Step 3. Hybridisation kinetics of the seal to the two different sequences was analysed to assess if a difference in the dwell time of the seal to the two gap sequences could be observed. With two DNA sequences differing by one base on the surface, the introduction of one seal sequence produces both binding to a complementary sequence and binding to a single base mismatch sequence.

GS9-CAP-T^{AT647N,P21,P29} is fully complementary to a gapped-DNA formed from the consensus CAP binding DNA and is a single mismatch to the gapped-DNA formed from the 1 x mutant CAP binding DNA. By using such an experimental design, both complementary and mismatching can be assessed in the same field of view under the same experimental conditions, allowing any changes in kinetics to be directly linked to the sequence without any interfering external factors.

The first experiments used the seal GS9-CAP-T^{AT647N,P21,P29}, labelled as seal 1 in Figure 3.12. When binding with full complementarity to the consensus CAP sequence an average fully hybridised dwell time of $T_2 = 6.58 \pm 0.25$ seconds (5%) was observed. In comparison, when seal 1 bound to the 1 x mutant CAP sequence gapped-DNA forming a single base mismatch the average fully hybridised dwell time was $T_2 = 1.45 \pm 0.05$ seconds (10%). There was a clear $\approx 78\%$ reduction in the average fully hybridised time for the single base mismatch hybrid compared to

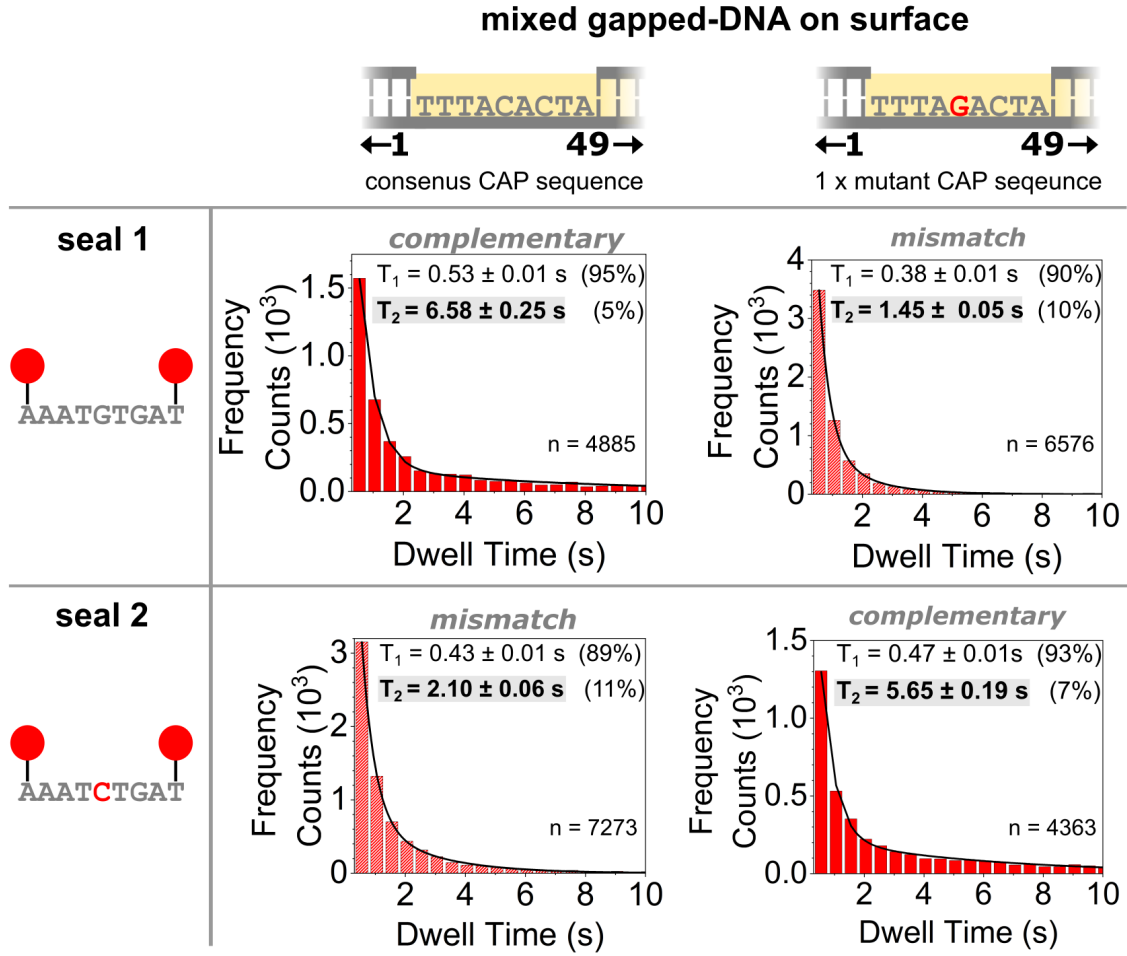


Figure 3.12: Dwell time histograms of the Gap-Seq (Step 3) part of the assay with mixed DNA on the surface. For the complementary binding of seal one $T_1 = 0.53 \pm 0.01 \text{ s}$ (95%) and $T_2 = 6.58 \pm 0.25 \text{ s}$ (5%). For the mismatch binding of seal 1: $T_1 = 0.38 \pm 0.01 \text{ s}$ (90%) and $T_2 = 1.45 \pm 0.05 \text{ s}$ (10%). Likewise, for the complementary binding of seal two $T_1 = 0.47 \pm 0.01 \text{ s}$ (93%) and $T_2 = 5.65 \pm 0.19 \text{ s}$ (7%). For the mismatch binding of seal 1: $T_1 = 0.43 \pm 0.01 \text{ s}$ (89%) and $T_2 = 2.10 \pm 0.06 \text{ s}$ (11%). n is number of binding events.

a fully complementary bound seal showing that Gap-Seq can be used for multiple DNA sequences on a surface.

Also, experiments were conducted with a second seal GS9-CAP-mis5C-T^{AT647N,P21,P29} that is fully complementary to the gapped-DNA formed from the 1 x mutant CAP

binding dsDNA and, hence, has a single base mismatch to the gapped-DNA formed from the consensus CAP binding DNA. Once again, there was clear discrimination between the two sequences. The complementary hybrid formed between seal 2 and the 1 x mutant sequence gapped-DNA had an average fully hybridised dwell time of $T_2 = 5.65 \pm 0.19 \text{ s}$ (7%). When compared to the mismatch hybrid formed between seal 2 and the consensus CAP sequence gapped DNA, which had an average dwell time of $T_2 = 2.10 \pm 0.06 \text{ s}$ (11%), there is once again a clear difference $\approx 63\%$ reduction in the average dwell time.

Both experiments conducted with mixed DNA sequences have shown that Gap-Seq is capable of detecting single base changes in the sequence of surface immobilised DNA, shown in earlier experiments to affect CAP behaviour. In the future, the whole assay could be developed to screen large varieties of DNA sequences on a surface and observe the changes in phenotype of DNA binding proteins, providing key insights into sequence specific protein-DNA interactions.

3.3 Discussion

This chapter has shown how even a single base change can alter the phenotype of catabolite activator protein (CAP) and how Gap-Seq can be applied to study such changes in behaviour. The first experiments in this chapter showed that changing the base at position 5 in the CAP consensus binding sequence reduced the average bound time of CAP molecules by $\approx 39\%$. As CAP is a dimer of identical subunits, consequently, altering position 18 in the symmetric binding sequence shows the true extent of single base changes by reducing the average bound time by $\approx 60\%$ compared to the consensus binding sequence. The next experiments investigated

denaturing surface-immobilised dsDNA whilst retaining the bottom strand of the DNA attached to the surface. Upon this single strand DNA hybridisation was performed in situ allowing gapped-DNA constructs to be formed. Once gapped-DNA formation from dsDNA was successful Gap-Seq could be implemented creating a three-step assay to link protein-DNA interactions to sequence. Experiments were performed to show how CAP binding to consensus binding sequence dsDNA could be observed at the single-molecule level followed by Gap-Seq. The difference in binding kinetics for complementary seals and those with a single base mismatch was observed for the gapped-DNA formed in situ showing how Gap-Seq can be applied in the future. Finally, ideally Gap-Seq will be applied to screen multiple DNA sequences on the surface, therefore, experiments were conducted with two dsDNA sequences randomly immobilised on the same surface. After the formation of gapped-DNA in situ, analysis showed that sequences differing by a single base caused a different average dwell time for the same seal, proving that the differences in average dwell time are truly due to sequence. In the future this assay can be developed further, to allow many DNA sequences to be investigated in parallel, as well as their corresponding effect on sequence dependent protein-DNA interactions.

Chapter 4

In Vivo Single-Molecule FISH

4.1 Introduction

Fluorescence in situ hybridisation (FISH) is a method by which nucleic acids inside fixed cells can be located by the hybridisation of complementary fluorescently labelled oligonucleotides (probes). Since the first example of FISH in 1977 [127], it has been used for many applications from fundamental biology to medicine and can be used to localise genes within the cells, especially in the case of eukaryotic cells and can be used to detect and quantify differential gene expression amongst cells in the same tissues.

In bacteria, FISH has been widely used for bacterial identification by using DNA probes that are unique to certain bacterial subtypes. FISH is also used to detect actively transcribing genes at the both ensemble and at the single-molecule level [128], [129]. FISH is applied to fixed cells to avoid cell degradation and to maintain the integrity of FISH targets, which are susceptible to degradation by deoxyribonucleases (DNases) and ribonucleases (RNases), which are present in

cells to degrade unwanted RNA. The cells are chemically treated to permeabilise the cell walls allowing fluorescently-labelled probes (<30 nucleotides) to enter the cells and specifically bind to RNA/DNA targets inside the cells. Excess unbound probes are then removed leaving only those attached to the target RNA/DNA allowing for localisation via fluorescence microscopy, see Figure 4.1. FISH images can be created from multiple different fluorophores allowing the identification of numerous targets [130]. A major disadvantage of using FISH to image cells is that fixed cells are no longer alive, therefore, FISH is used as a passive way to localise and/or study gene expression of the cell at the time of fixing and real time cell processes can not be observed.

An example target for RNA FISH in bacteria is the 16S ribosomal RNA (rRNA). 16S rRNA is expressed to form the 30S small subunit of ribosomes and is responsible for messenger RNA (mRNA) binding during translation. The EUB338 FISH probe is used to identify 16S rRNA by complementary binding to the 338-355 region of the rRNA [131]. EUB338 is often used as a bacterial marker as the 16S rRNA is present in the majority of bacteria.

There have been some attempts to perform a RNA detection technique similar to FISH but on live eukaryotic cells in order to observe internal processes. Vargas et al. [133] in 2005 injected 50 nucleotide FISH probes that bound to a messenger RNA (mRNA) to investigate transport in the nucleus of eukaryotic cells. It was not possible to wash out unbound probes as with FISH, therefore, to reduce the background fluorescence in the cell molecular beacons were implemented. In the molecular beacon, the secondary structure of the DNA allows a quencher to be in close proximity to the fluorophore to suppress fluorescence when the probe was unbound. Another technique used to increase the target fluorescent signal was

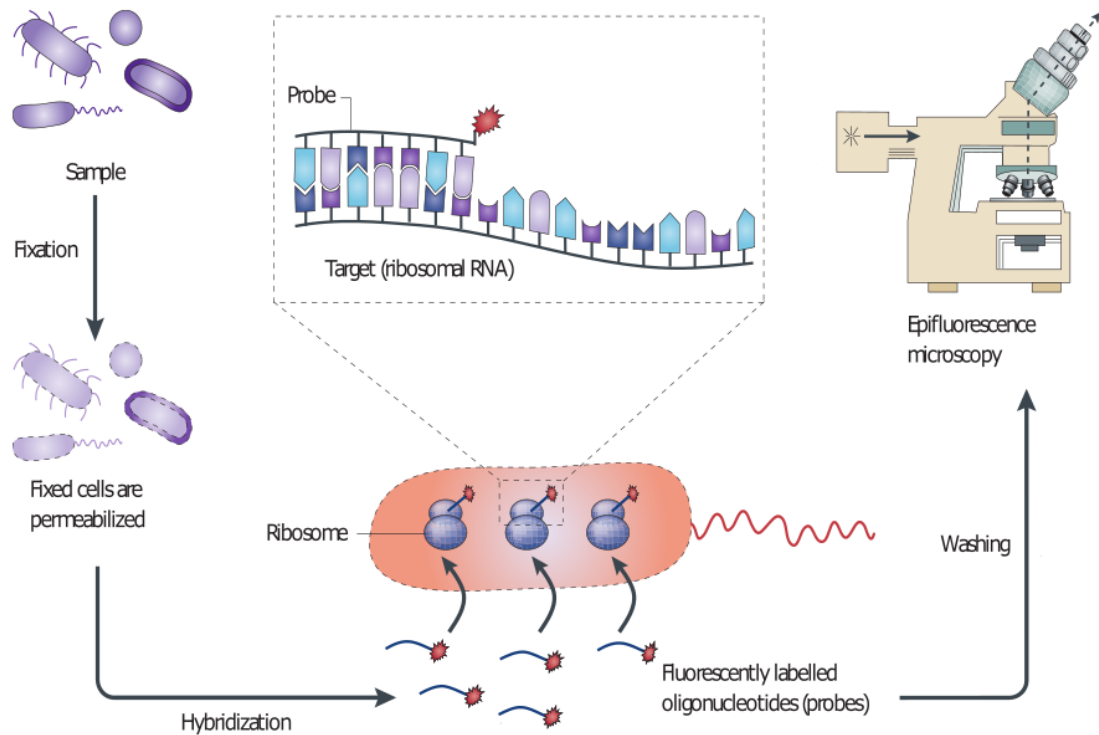


Figure 4.1: RNA FISH protocol. Cells are fixed and permeabilised before fluorescently-labelled oligonucleotides (probes) are introduced that specifically bind to target RNA which can be imaged using fluorescence microscopy. Adapted from [132].

repeated binding sites on the target mRNA with a maximum of 96 probes able to bind to the mRNA.

Zhang et al. [134] also used 96 repeats of binding sites in order to create a bright localisation of the target RNA in live eukaryotic cells. Here, the experiment consisted of electroporating the cells in a buffer containing ratiometric bimolecular beacons instead of injection in the case of Vargas et al. [133]. Compared to standard molecular beacons, ratiometric bimolecular beacons are designed to allow nuclear export reducing the number of false signals observed and feature a reference

dye to allow tracking of delivery into cells.

Although there have been many experiments where FISH has been shown in live eukaryotic cells, there are very few examples of live cell FISH in prokaryotes. The first attempt at FISH in non-fixed bacteria was in 2005 [135]. The method used quenched probes which, when bound to the target RNA, became fluorescent due to ligation at the site of binding. The oligonucleotide probes were able to enter the cells through semi-permeable cell walls that had been treated chemically. However, the semi-permeablising treatment lead to an almost non-existent yield of live cells for microscopy drawing criticism of the assay [132].

In 2019, it was reported by Batani et al. [136] that FISH had been performed on live bacteria cells, with an application to sort cell types, with the FISH probe delivered through heat shock of the cells. Heat shock involves exposing cells to sudden high temperatures, creating a temporary porous plasma membrane allowing nucleic acids to enter the cell and crucially, post-treatment, the live cells can recover. Batani et al. used traditional unquenched FISH probes (singly labelled oligonucleotides) for target binding inside the cell leading to diffuse fluorescence within cells. Therefore, the probes used were sufficient to detect different strains of bacteria, however, the assay does not give any spatial information as to the location of the binding of the probes within the cells. Live FISH in bacteria for localisation and observation of live processes such as transcription and internal movement of mRNA is clearly desirable, however, there have been very few advancements making the technique viable.

The results presented in this chapter are based around developing a novel assay to conduct single-molecule FISH (smFISH) in vivo (*E. coli*) through the transient binding of labelled oligonucleotides. Taking inspiration from Gap-Seq,

doubly labelled oligos (probes) which transiently bind are introduced into cells. The transient nature of binding allows for continuous probing of target RNA, avoiding photobleaching and also preventing blocking of processes occurring on the RNA. As with Gap-Seq, the probes are able to self-quench when unbound, reducing background fluorescence and allowing single binding events to be detected. The proof-of-principle technique was developed with an aim to observe the location of transcription for single genes, track the mobility of nascent RNA and as a method to link transcription to downstream processes such as RNA degradation inside live bacteria.

4.2 Results

4.2.1 Electroporation of DNAs

To conduct experiments involving fluorescently-labelled DNA *in vivo*, a feasible delivery method to introduce the DNA inside the cells needed to be chosen. It has been previously shown that electroporation can be used to introduce fluorescently labelled probes into live cells [137]. The first experiment was conducted to check the viability of electroporating 8 nt probes that were either singly or doubly labelled with ATTO647N into MG1655WT *E. coli* cells. As a negative control, cells were also electroporated in the absence of probes to establish the level of background fluorescence that may arise from auto-fluorescent components that are present in the recovery media post electroporation.

As can be seen in Figure 4.2a, in the absence of probes, there was no fluorescence recorded in the electroporated cells when excited at 640 nm and a small

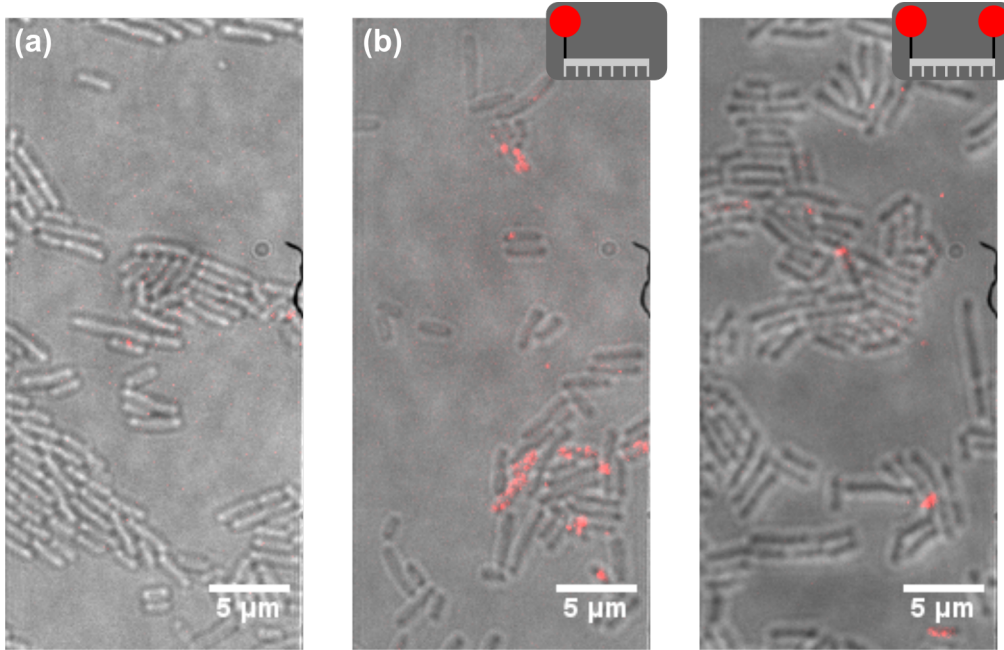


Figure 4.2: Overlaid bright field and maximum intensity fluorescence images for the electroporation of MG1655WT cells excited at 640 nm, exposure time 50 ms. a) No DNA present during electroporation. b) Sample electroporated with 8 mer, singly labelled with ATTO647N, GS8-compl- $T^{\text{ATTO647N,P28}}$ 20 nM. c) Sample electroporated with 8 mer, doubly labelled with ATTO647N, GS8-compl- $T^{\text{AT647N,P28,P35}}$ 20 nM.

amount of fluorescence due to contaminants outside of the cells. For cells electroporated in the presence of GS8-compl-T^{ATTO647N,P28}, a singly labelled 8 nt probe, fluorescence could be clearly observed inside approximate 15% of the cells, see Figure 4.2b. When the doubly labelled probe GS8-compl-T^{AT647N,P28,P35} was introduced, some fluorescence could be observed inside cells more than the negative control but less than the GS8-compl-T^{ATTO647N,P28} sample. The doubly labelled probe used here is known to undergo static quenching when in solution, see Chapter 2. Therefore, the presence of fluorescence in the two samples containing the probes suggests that the probes entered the cells with quenching seen for the doubly labelled probe. These initial experiments showed that it is possible to introduce short fluorescently labelled DNA into live *E. coli* and the diffusing fluorescence due to free unbound DNA could be observed.

Characterisation of Probe Degradation In Vivo Using FRET Labelled Probes

Enzymes, such as nucleases, are able to degrade linear DNA, therefore, the next step was to check for degradation of the labelled probes that are introduced by electroporation. A doubly labelled probe carrying a FRET donor on one end and a FRET acceptor on the other end, Cy3B and ATTO647N respectively, named GS7-mis4T-T^{Cy3B,P28,AT647N,P34} was chosen. The two fluorophores are a known FRET pair as their emission and excitation spectra overlap allowing them to undergo FRET if in close proximity.

Figure 4.3a shows the MG1655WT *E. coli* cells after electroporation in the presence of 20 nM of GS7-mis4T-T^{Cy3B,P28,AT647N,P34}. When the cells were illuminated at 532 nm (Figure 4.3b) some autofluorescence can be seen inside the cells

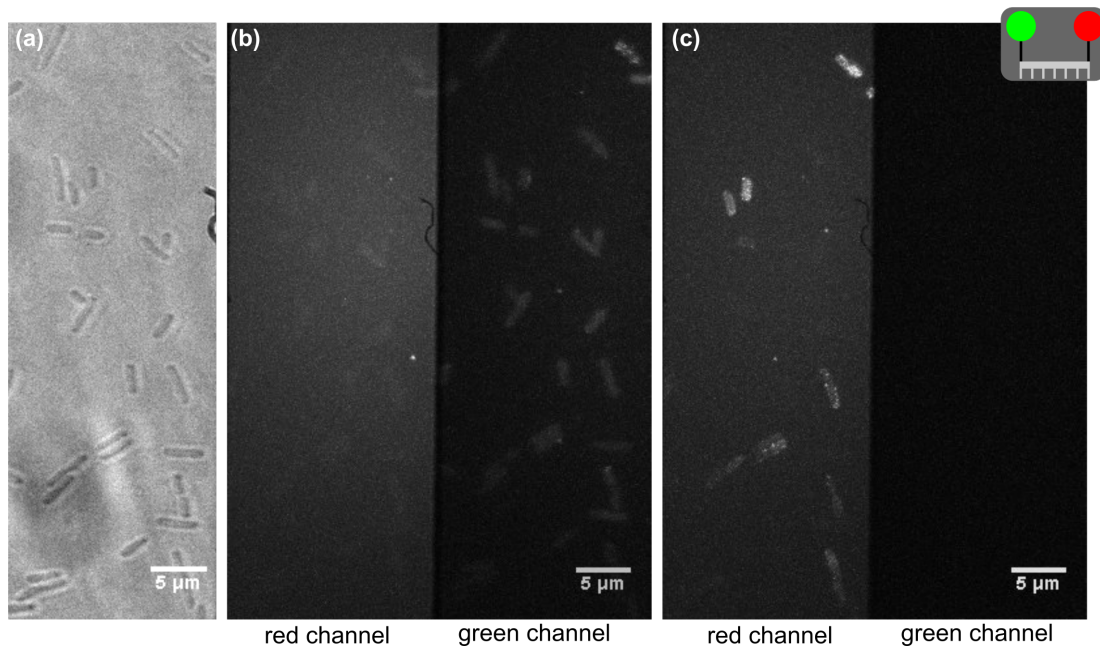


Figure 4.3: MG1655WT *E. coli* cells electroporated with GS7-mis4T-T^{Cy3B,P28,AT647N,P34} 20 nM. a) Bright field image of cells. b) Summation of maximum fluorescence intensity of both the red (left) and green (right) emission channels under 532 nm excitation. c) Summation of maximum fluorescence intensity of both the red (left) and green (right) emission channels under 640 nm excitation.

in the green emission channel. However, there is very little fluorescence from the Cy3B fluorophores and no obvious FRET is seen in the red emission channel. The absence of green fluorescence indicates that the Cy3B fluorophores were quenched by the other fluorophore on the DNA, suggesting that there was no degradation of the probes. When excited at 640 nm there is red fluorescence detected in some of the cells. The presence of fluorescence suggests that the ATTO647N fluorophores are not being well quenched by the Cy3B fluorophores. The fluorescence does show that the labelled probes entered the cells and the ATTO647N is stable and able to absorb radiation and fluoresce.

For a single cell from the field of view, shown in Figure 4.4a, the average fluorescence intensity in the green and red emission channels can be plotted over time under green and red excitation respectively (Figure 4.4b). The measurements were recorded using alternating laser excitation (ALEX) allowing for both 532 nm and 640 nm excitation. As can be seen in Figure 4.4b, the red fluorescence due to 640 nm excitation reduced over time due to photobleaching and the green emission started at a low level and decayed very slightly at the beginning of the acquisition due to the photobleaching of the autofluorescence of the cells. If the probes had been degraded, there would have been an equal concentration of red and green fluorophores diffusing inside the cell; therefore one would expect the intensity levels and the photobleaching pattern to be similar. The absence of this suggests that the fluorophores were still in close enough proximity to static quench the Cy3B fluorophore, with the ATTO647N fluorophore not fully quenched. Also, as there was not a rise in green fluorescence with the decay in red fluorescence, the quenching can, therefore, be attributed to contact quenching rather than quenching of Cy3B through FRET with ATTO647N. The lack of free fluorophores and corresponding

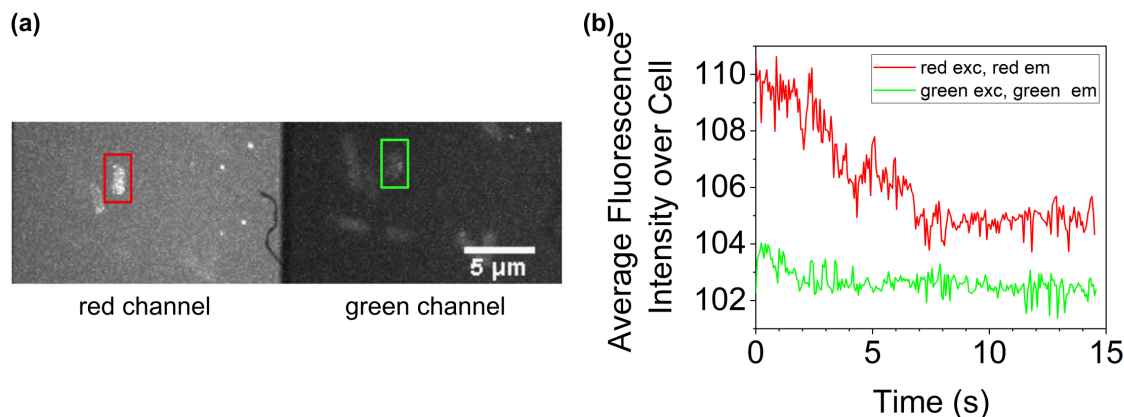


Figure 4.4: Fluorescence intensity for a single cell. a) Summation of maximum fluorescence intensity under 640 nm (red emission channel only) and 532 nm excitation (green emission channel only). Boxes indicate cell corresponding to the fluorescence intensity vs. time plot. b) Plot of average fluorescence intensity over time the single cell. Green excitation (532 nm) and green emission in green line (bottom) and red excitation (640 nm) and red emission in red line (top) recorded using ALEX, 50 ms exposure time.

fluorescence intensity profiles is evidence that the probes are not degraded inside the cell, possibly protected from degradation due to their short nature and the terminal fluorophores protecting the ends of the probe from nucleases.

4.2.2 Site-Specific Binding With EUB338

After establishing successful delivery of the labelled probes and the preservation of the probes in vivo, the next step was to observe specific binding of probes in vivo. A 18 mer oligo singly labelled with Cy3 (EUB338) [131] was used as a probe as it is known to bind to 16S ribosomal RNA (rRNA) and EUB338 probes are commonly used for FISH in fixed cells. The 16S rRNA is a component of the 30S subunit of the bacterial ribosome and fluorescence microscopy has localised ribosomes at the cell poles, with detection of ribosomes in mid-cells regions for cells in the later

stages of chromosome replication in live *E. coli* [138], [139], [140]. Electroporating EUB338 into MG1655WT *E. coli* should allow for any binding to 16S rRNA to be visualised in live cells.

EUB338 at 200 nM was electroporated into MG1655WT cells and imaged at 532 nm with 50 ms exposure time. Figure 4.5a shows the summation of intensity over 200 frames for four fields of view of electroporated cells under 532 nm illumination. Cells outlined in red showed bright fluorescence at the cell poles. Most other cells showed diffuse fluorescence with very little bright fluorescence in specific locations, with only a handful of cells showing bright fluorescence in the centre of the cell.

The bright fluorescence and, therefore, longer binding of EUB338 to 16S rRNA at the poles can be seen in Figure 4.5b. Here, a single cell from the field of view (yellow box) is shown over time via 5 second summations of the fluorescence intensity. Over the time of the acquisition, bright fluorescent signals were recorded at both poles that appeared and disappeared. As fluorescent signals appeared during the acquisition time, see Figure 4.5c, sticking of non-electroporated probes outside the cells at those locations can be ruled out. Fluorescence from probes stuck on the outside of the cells would be expected to be observed from the beginning of the acquisition and decay over time due to photobleaching. Here, the bright fluorescence signals at the cell poles are due to the binding of the fluorescently labelled EUB338 to 16S rRNA, which is an integral part of the ribosome - shown to gather at cell poles [138], [139]. It is, however, uncertain whether the probes are targeting 16S rRNA before or after the assembly into ribosomes. The electroporation of EUB338 inside live cells has shown site-specific binding to 16S rRNA indicating that in vivo FISH can be used for localisation of target RNA inside live cells.

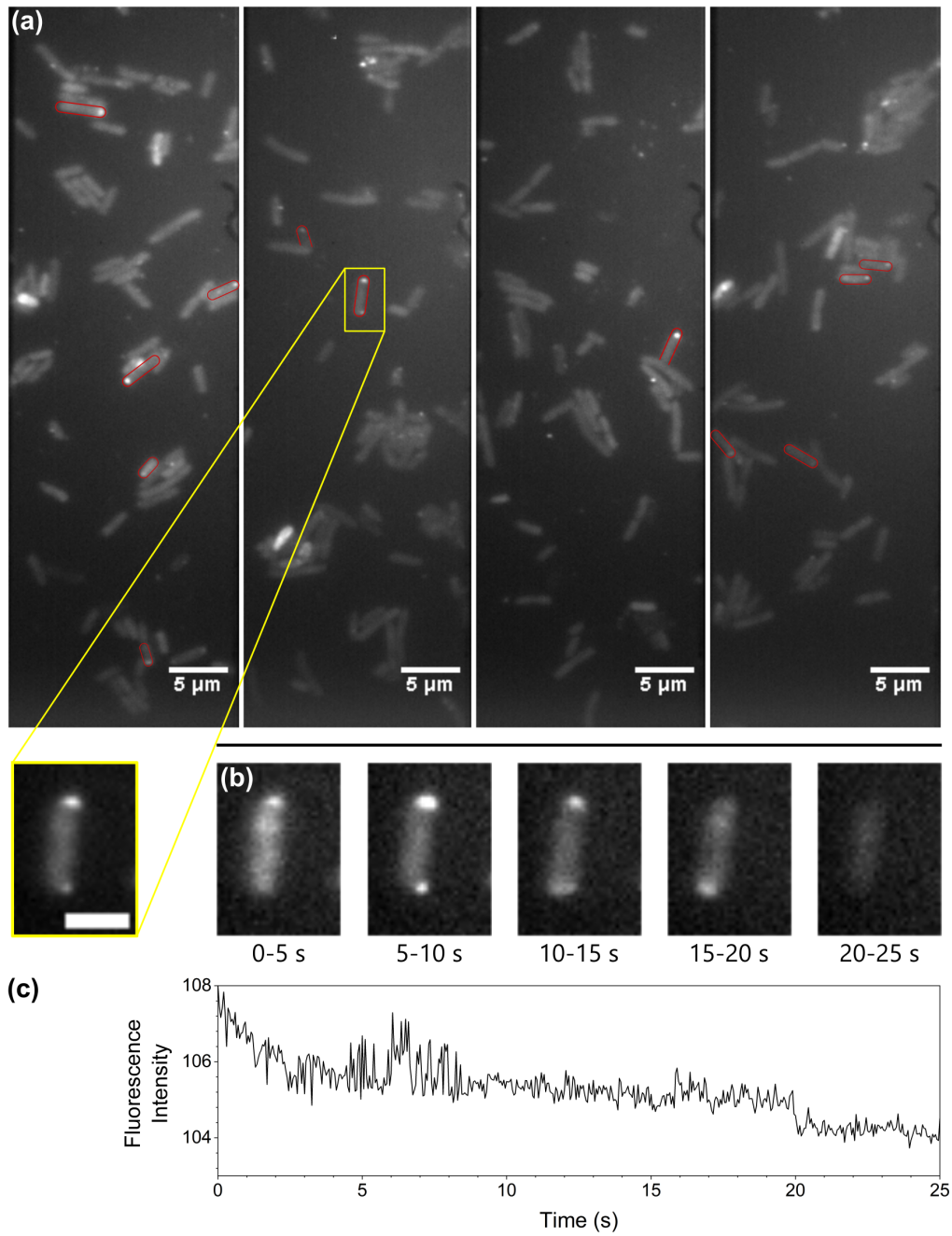


Figure 4.5: a) Four different fields of view of MG1655WT *E. coli* cells electroporated with EUB338 200 nM (known to bind to 16S rRNA), 200 frames summations of maximum fluorescence intensity (50 ms). Red outlines: cells bright fluorescence at the poles. b) 100 frame summations (5 seconds) of fluorescence intensity for a single cell over time (yellow box). Bright fluorescence at the cell poles indicating binding to 16S rRNA. Scale bar 2 μm . c) Fluorescence intensity vs. time trace over entire cell for 0-25 seconds.

4.2.3 Plasmid Expression

After the success of site-specific binding using the EUB338 probe the next experiments were conducted to test and characterise the ability to use probes in vivo to any selected target RNA. In order to do this a custom-standard target needed to be designed that could be added to any gene's RNA sequence, making it a target for in vivo FISH. As a proof-of-principle, a mini-gene was designed to contain six iterations of the binding site for the doubly ATTO647N labelled 8 nt probe GS8-compl-T^{AT647N,P28,P35}.

In order to express the mini-gene, the plasmid pBEC1 was generated by Gibson assembly where the plasmid backbone was amplified by PCR from the plasmid pACBSR [141] and the mini-gene insert was assembled by hybridising two long PCR primers. The pBEC1 plasmid was engineered to express target RNA that included six binding sites for the 8 nt seals with 6 nt spaces between each binding site. The mini-gene insert was put under the control of an arabinose-inducible promoter, that produces a high number of the probe targets upon arabinose addition [142], and was followed by a terminator creating an RNA of ≈ 100 -102 bases. The full insert on the plasmid can be seen in Figure 4.6, the six binding sites for the 8 nt probes shown in orange and the 6 nt spacers shown in green. The expressed RNA is short and hybridisation of the probes can only occur after the RNA exits from the RNAP exit channel, therefore, detectable transcription of the mini-gene is ≈ 2 seconds long if the speed of transcription elongation is assumed to be 42 nts^{-1} [143]. The expression of the RNA is controlled by a pBAD promoter, therefore, in the presence of arabinose and the absence of glucose over expression of the RNA is achieved, providing many target RNA with complementary binding sites for the

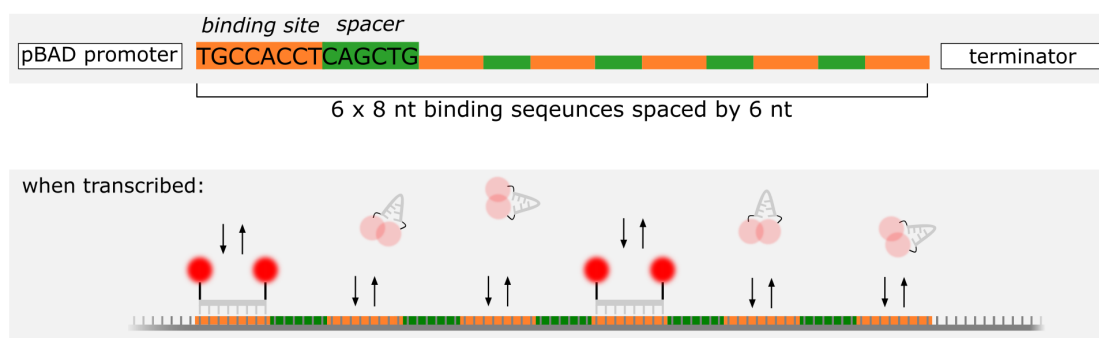


Figure 4.6: Top: Diagram of the insert used to create plasmid pBEC1 from the cropped and linearised pACBSR plasmid. The sequence includes six binding sites for GS8-compl-T^{AT647N,P28,P35} (orange) with 6 nt spaces (green) expression under the control of a pBAD promoter. Bottom: Schematic of GS8-compl-T^{AT647N,P28,P35} binding to target RNA containing the repeated binding site transcribed from the insert.

probes.

The plasmid pBEC1 was transformed into MG1655WT cells and cultured. The resulting cells made electro-competent by successive washes to remove salts and ions that might interrupt the electric field during electroporation. The cells were then electroporated in the presence of 20 nM GS8-compl-T^{AT647N,P28,P35}. Figure 4.7 shows the composite images of bright field images of the electroporated cells and fluorescence images from the red emission channel of the microscope for WT cells and those containing pBEC1, expressing the target RNA. When compared side-by-side very little fluorescence was observed inside the WT cells that did not express the target RNA, whereas, the cells containing pBEC1, over expressing the target RNA (grown in the presence of 0.1% arabinose), showed bright fluorescence in the majority of the cells in the field of view. The double labelled probe is known to mostly self-quench when unbound, therefore, fluorescence observed in the cells indicated site-specific binding of the seals to the target RNA.

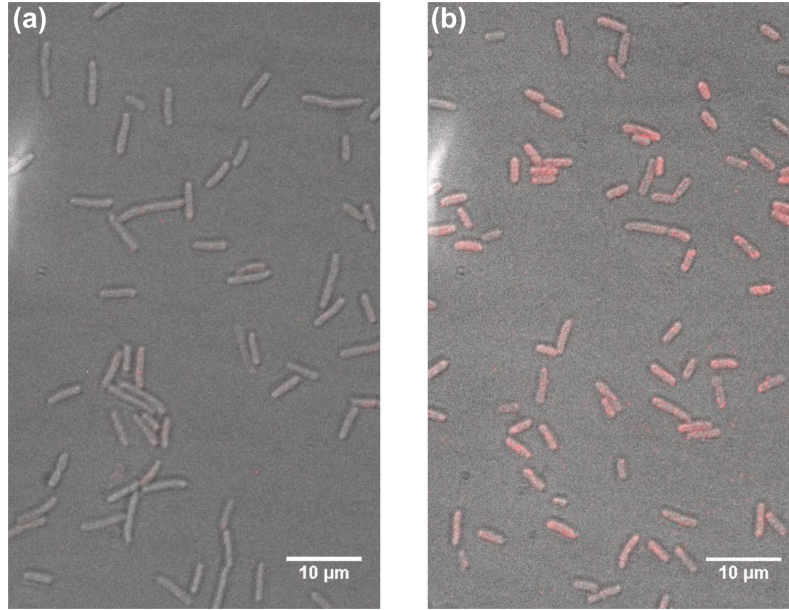


Figure 4.7: Composite images under bright field illumination and fluorescence emission for 640 nm excitation, 50 ms exposure time. a) MG1655WT cells, that do not express the target RNA, electroporated with 20 nM of GS8-compl- $T^{AT647N,P28,P35}$ showed very little fluorescence inside cells. b) Cells containing the plasmid pBEC1, over expressing the target RNA, electroporated with 20 nM of GS8-compl- $T^{AT647N,P28,P35}$ showed high levels of fluorescence inside most cells indicating site-specific binding to target RNA.

As can be seen from Figure 4.8, the location of fluorescence observed inside the pBEC1 cells changed over time as bright spots (foci) were seen slowly diffusing inside the cells. Slow photobleaching of the probes was seen over the acquisition with some initial bleaching of the cell autofluorescence and fluorescent dyes observed in the first 10 seconds. Figure 4.8 shows slowly diffusing bright foci of fluorescence are observed that indicate binding of the fluorescently labelled probes to the target RNA. The foci were present at the start of recording and appeared and disappeared throughout the movie indicating ongoing transient hybridisation of the probes to the target RNA.

Figure 4.9 shows the fluorescence intensity over time for individual cells. The fluorescence intensity vs. time traces show a general trend of slow photobleaching but there are step function-like increases and decreases in fluorescence. These step-like fluctuations are similar to the fluorescence signature of the seals binding to gapped-DNA as seen in earlier chapters. If examined closely, the fluorescence intensity vs. time trace for the cell in the blue box shows sequential increasing in the fluorescence intensity from approximately 55 seconds onwards, indicating multiple probes binding to targets possibly due to a burst of transcription. Figure 4.9 clearly shows binding and dissociation of the probes over time to the expressed RNA targets. The experiments here have shown that the 8 nt doubly labelled probe, GS8-compl-T^{AT647N,P28,P35}, can be used to indicate the presence of target RNA, through site-specific binding, when target RNA is expressed from a plasmid.

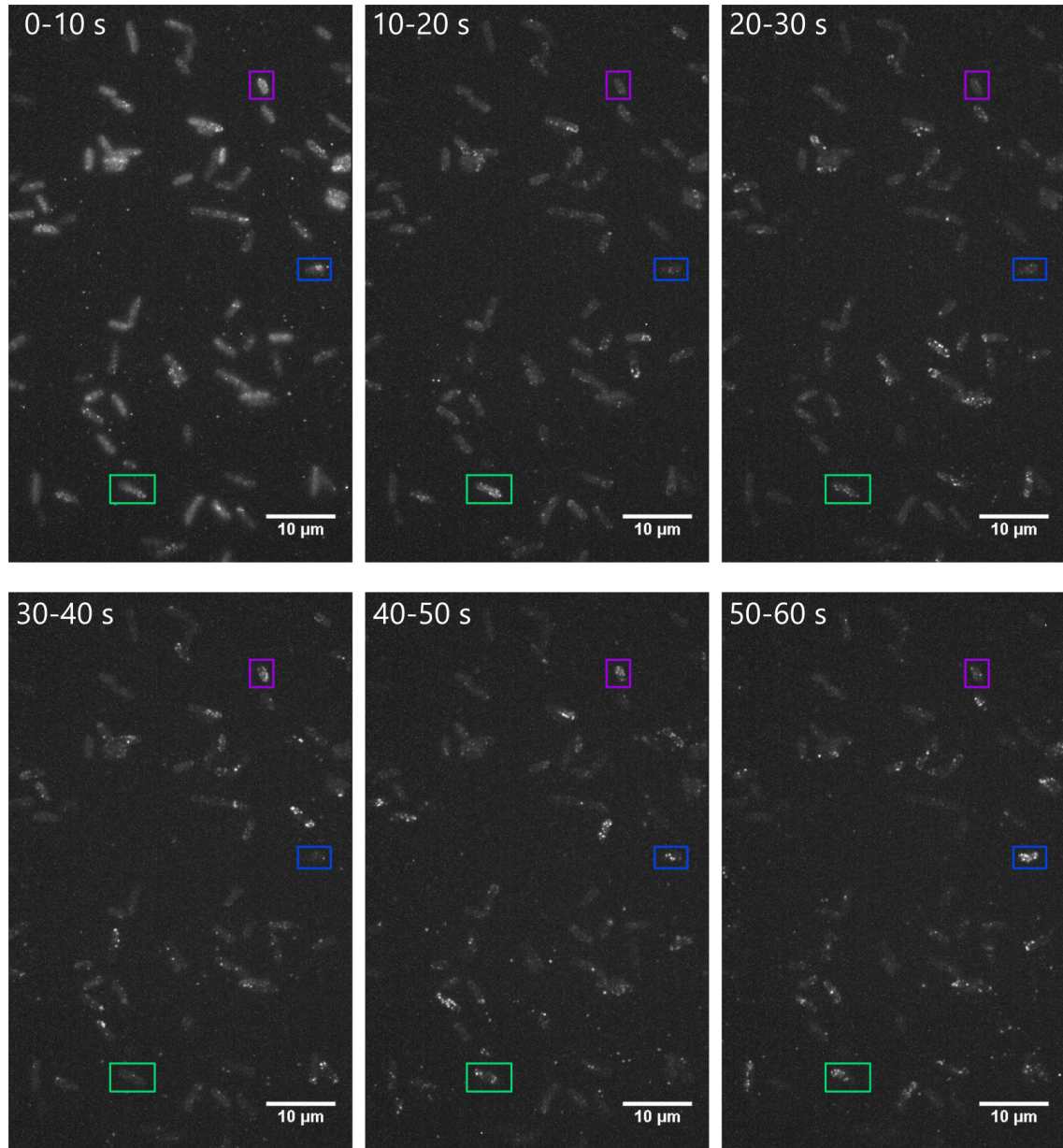


Figure 4.8: Red fluorescence images of MG1655 cells containing pBEC1 over expressing the target RNA, electroporated in the presence of 20 nM GS8-compl- $T^{AT647N,P28,P35}$. A series of summations of the maximum intensity of fluorescence in intervals of ten seconds over a 60 second recording with 50 ms exposure. Fluorescence due to binding of probes was observed in most cells as bright foci that slowly diffused.

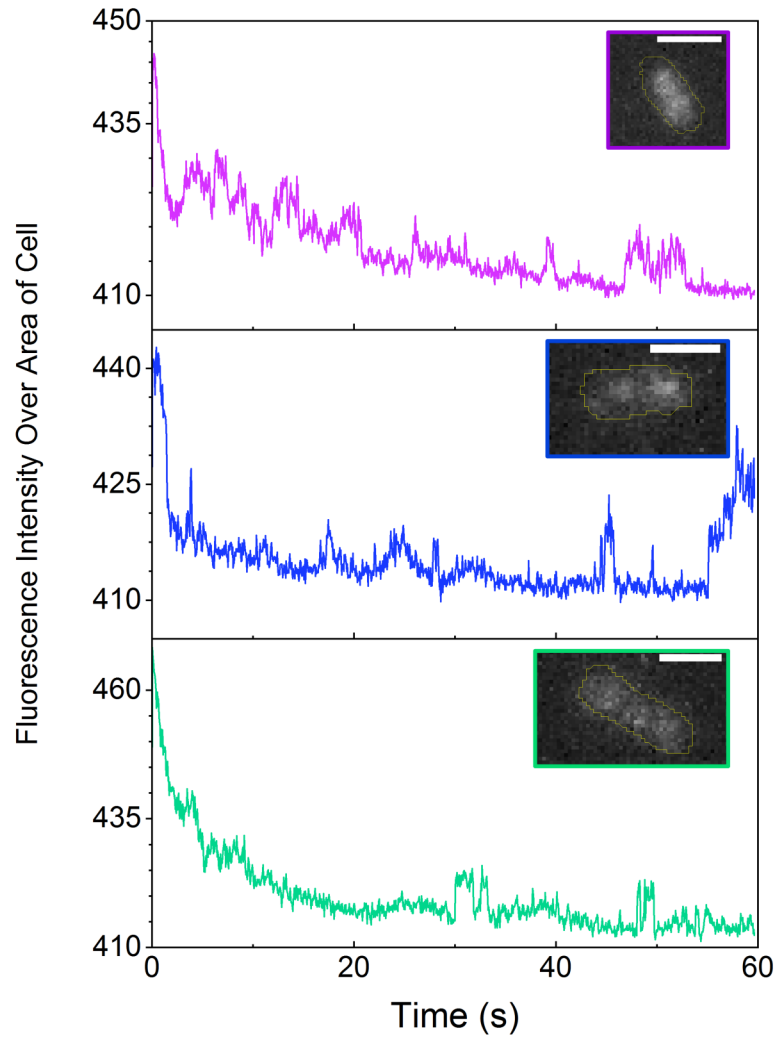


Figure 4.9: Fluorescence intensity vs. time traces for the area of three cells (insets - summations of fluorescence over 60 seconds) selected from the field of view of the previous figure. Scale bars $2\ \mu\text{m}$. Overall decrease in fluorescence intensity due to photobleaching of probes, however, the step-like increases and decreases in fluorescence showed binding of fluorescently labelled probes to target RNA.

4.2.4 Chromosomal Expression

Having demonstrated site-specific binding to a RNA target expressed from the pBEC1 plasmid, the next step was to insert the previously characterised probe binding repeats (see Figure 4.6) upstream of a long transcript on the chromosome and measure transcription events. Observing target RNA on the chromosome potentially allows localisation of single genes compared to many genes that are synthetically amplified by the presence of tens of plasmids copies per cells. The expression site can be localised and measurement of the time taken to transcribe the RNA arising from one gene.

To generate the target RNA carrying the repeated 8 nt probe binding site array on the chromosome, first the array was inserted by Gibson assembly on a plasmid pTARA [144] carrying a T7 RNA polymerase gene under the control of the arabinose inducible promoter pBAD. The ribosome binding site was replaced by the 8 nt probe array which, when transcribed, would produce a long RNA transcript with the 8 nt probe array binding site at the start of the transcript. Then the plasmid was used as a template to generate a lambda red substrate that allowed the targeted genetic change of inserting the whole construct on the chromosome between *glnA* and *typA* genes in MG1655 E.coli. Insertion of the binding array on the chromosome was conducted by Dr. El Sayyed, Kapanidis Group. The target RNA was ≈ 3 kilobase (kb) with the oligo binding sites at the start of the transcript available for binding once exited from the RNAP exit channel. For a transcript of 3kb, a transcription duration of ≈ 71 seconds is expected, if the speed of transcription elongation is assumed to be 42 nts^{-1} [143]. A longer RNA target was used as it was hoped that transcription could be observed

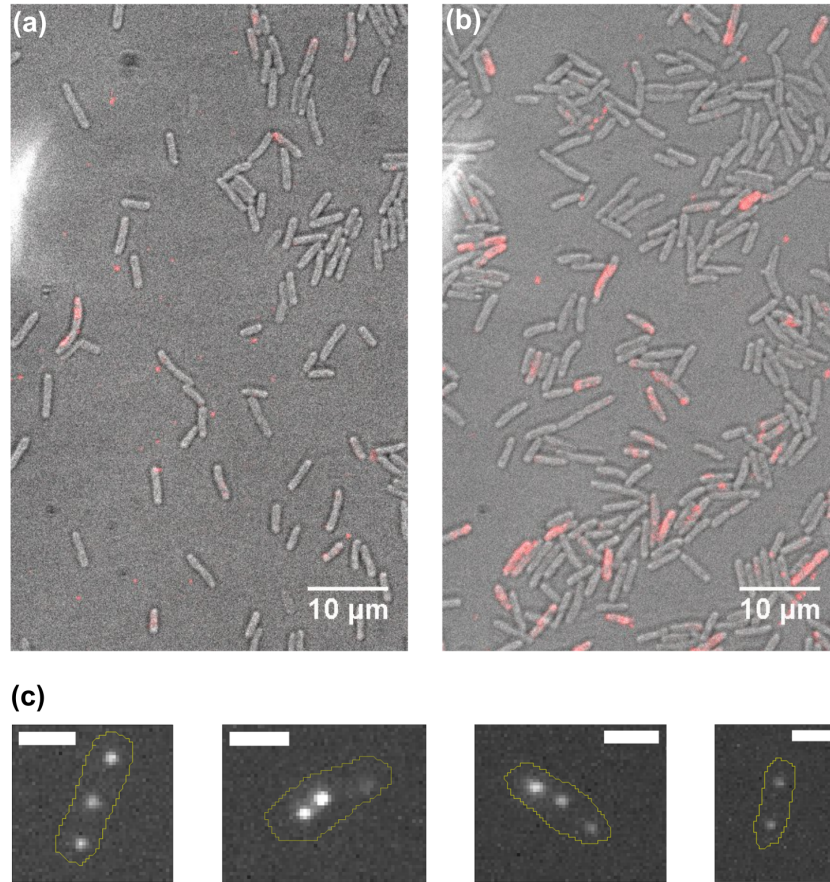


Figure 4.10: a) and b) Composite image of bright field and fluorescence emission under 640 nm excitation *E. coli* cells electroporated in the presence of 20 nM GS8-compl-T^{AT647N,P28,P35}. a) MG1655WT cells with little internal fluorescence indicating a lacking of site-specific binding of the probes. b) Cells modified to express 3 kb target RNA, with 6 binding sites for the probes, on the chromosome. Bright fluorescence inside cells indicates binding of probes to target RNA. c) Cells with target RNA chromosomal expression electroporated in the presence of 20 nM GS8-compl-T^{AT647N,P28,P35} showing bright fluorescent foci due to specific binding of the probes to target RNAs. Each image is a single frame, 50 ms exposure time. Scale bars 2 μm.

at a static localisation before the nascent RNA was released. For probes of 8 nt, the duration of transient hybridisation (dwell time) would be expected to be in the range of milliseconds to seconds, based on in vitro data previously recorded for the probes. As seen in Chapter 2, the 8 nt seals bound to gapped-DNA, in the presence of 0.5 M NaCl, for an average long dwell time of ≈ 11 seconds. In vivo there is no coaxial stacking, which is known to stabilise binding of the seals, as well, physiological salt concentrations are lower than those used in vitro for the gapped-DNA experiments. Both the lack of coaxial stacking and low salt concentrations suggests that the 8 nt probes should bind in vivo for a shorter time than observed in the gapped-DNA experiments, even if in vivo macromolecular crowding is considered.

Figure 4.10 shows MG1655WT cells and cells that express the chromosomal target RNA both electroporated in the presence of 20 nM GS8-compl-T^{AT647N,P28,P35} imaged in bright field and 640 nm illumination. As can be seen in Figure 4.10a, there is very little fluorescence seen inside the wild type cells, where the target RNA is not expressed, indicating no specific binding of the probes. In contrast, when the target RNA is expressed on the chromosome there is fluorescence observed in approximately 10% of the cells due to the site-specific binding of the probes. The RNA expressed in these experiments can be described as leaky expression of the promoter as the experiments were conducted in the absence of arabinose and the target RNA is regulated by a pBAD promoter which restricts expression in the absence of arabinose. The leaky expression regime used here provides a lower number of target RNAs in the cell making it easier to see binding to individual RNAs. Figure 4.10c shows a selection of cells expressing the target RNA that have been electroporated in the presence of 20 nM GS8-compl-T^{AT647N,P28,P35}. There are

clear fluorescent foci within the cells, approximately 2 or 3 on average per cell. The cells seen in Figure 4.10c are represented by single frames as the foci can be either stationary or mobile, therefore, the foci are easily blurred with frame summations.

In Figure 4.11 an example cell is selected that initially had three fluorescence foci. By frame 54 (2.7 seconds later) there were four foci. The focus at the top of the cell remained stationary during this period with the other two foci diffusing inside the cell. During the rest of the acquisition, the number of foci fluctuated between two and zero. If the foci corresponded to only unbound fluorophores being photobleached over time one would expect the number of foci to decrease. An increase in the number of foci suggests that there were oligos transiently binding.

During the first 50 frames of the acquisition (2.5 seconds), when one would expect the overall fluorescence to decrease due to photobleaching, the fluorescence intensity in fact increased, see Figure 4.11, indicating binding of probes. Also, between 100 to 350 frames there were repeated step-like increases and decreases in fluorescence which is signature of transient binding of fluorescently labelled probes. The fluorescence intensity vs. time trace in Figure 4.11 shows similar changes in fluorescence intensity when compared to the fluorescence intensity vs. time traces for plasmid expression of the target RNA in Figure 4.9.

Also, the foci for the cell shown in Figure 4.11 show two different types of behaviour, static foci and slowly diffusing foci, with foci able to move between the two classifications. The bright focus at the top of the cell in Figure 4.11 remained static for the first 54 frames, whereas, the other foci slowly diffused inside the cell. The bottom focus (seen in frame 68 through to frame 157) can also be classed as a static focus. The binding sites for the probes are first transcribed followed by the rest of the T7 RNA polymerase sequence as the probe binding sequence lies

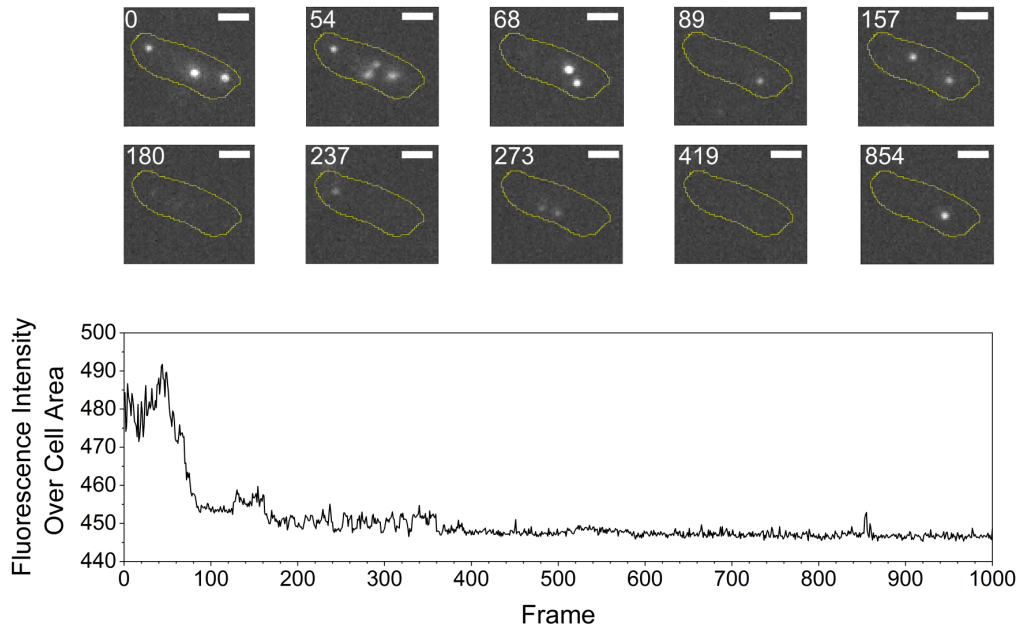


Figure 4.11: A single cell expressing target RNA on the chromosome electroporated in the presence of 20 nM GS8-compl- $T^{AT647N,P28,P35}$. Single frames (50 ms) show the evolution of fluorescent foci over time with foci appearing and disappearing. Scale bars 2 μm . The fluorescence intensity of the cell shows increases in the fluorescence intensity during acquisition indicative of site-specific binding of the fluorescently labelled probes.

at the start of the transcript, creating tethered binding sites before the transcript is released after ≈ 71 seconds. It can be hypothesised that the stationary foci are where transcription of the gene is happening with the slowly diffusing foci produced by probes bound to released RNA.

To understand the static foci, one can look more closely at the top focus in the cell shown in Figure 4.11. Figure 4.12 looks at the fluorescence intensity of the single focus over the first 65 frames (3.25 seconds). From frame 0 to frame 57 (2.9 seconds) the focus remained static and from the fluorescence intensity vs. time trace during this time there were intriguing changes in the fluorescence intensity.

Between frame 40 and frame 54 (0.7 seconds) there was an increase in fluorescence intensity at the single focus. The target RNA is known to have multiple binding sites for probes, therefore, it is possible that the increase was due to a second bound oligo. Examining the fluorescence intensity vs. time trace in Figure 4.12 there are jumps in fluorescence intensity in the trace of similar sizes. From Figure 4.12 the increase in fluorescence intensity between frames 40 and 54 was due to a doubling of fluorescence at the focus. A doubling of fluorescence at a focus can easily be interpreted as a second probe binding to the target RNA and as the target RNA has multiple probe binding sites this is to be expected. Also, if the time for binding of the probes is examined this is roughly in the range predicted from in vitro studies of milliseconds to seconds. An alternative argument is that the first level of fluorescence is from only one fluorophore of a bound probe and the increase in fluorescence is due to the secondary fluorophore fluorescing after being dark. To investigate this further other static foci can be examined.

Figure 4.13 shows that for the first 30 frames of another acquisition of a single cell a static focus was observed with a steady fluorescence intensity. At frame 35 the focus diffused away from the original position and the fluorescence intensity recorded at the localisation was now the background fluorescence detected in the cell. In the frame before, frame 34, a second dimmer focus appeared in close proximity to the original focus with the first focus dimmed and slightly shifted in location. Degradation of the doubly labelled probes in vivo was not observed with the FRET probes, therefore, the secondary focus can be attributed to a bound oligo. One can speculate that due to the proximity of the foci and the reduced fluorescence of the original foci that two probes were originally bound at the static focus possibly to two RNAs transcripts being transcribed in short succession. This

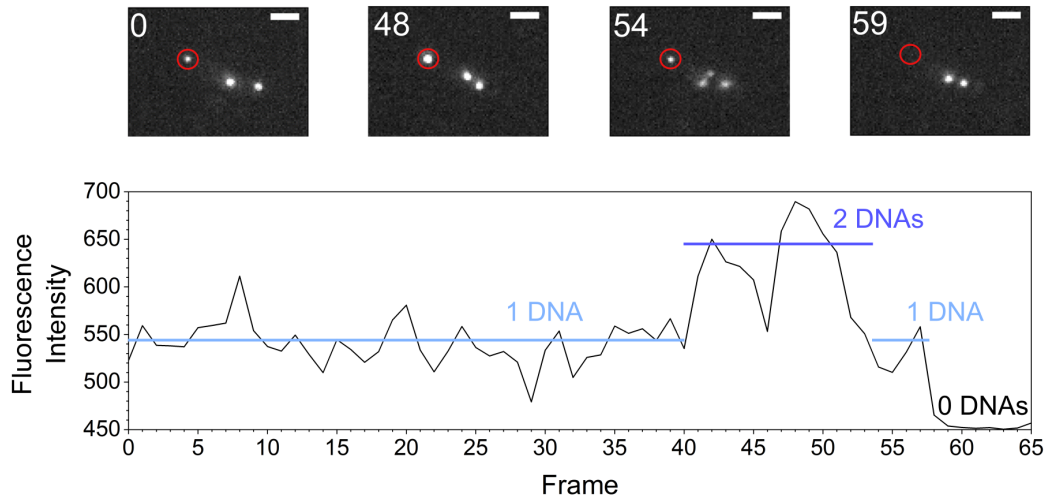


Figure 4.12: Fluorescence intensity of a single static focus (red circle) over the first 65 frames of acquisition. Top: individual frames of the cell, 50 ms exposure time. Scale bars $2\ \mu\text{m}$. Bottom: Fluorescence intensity vs. time trace for the focus over 65 frames with fluorescence intensity levels corresponding to multiple probes bound to the target RNA at the localisation.

is another example of a cell observed to show two probes bound at static focus suggesting these static foci with multiple levels of fluorescence intensity are due to multiple probes binding to target RNA, with the foci static due to the RNA being produced during active transcription.

A way to measure transcription duration using this method would be to observe the appearance of a static focus and measure the length of time it is stationary before the focus diffuses away. For a single cell a static focus was chosen to be investigated. In Figure 4.14 a static focus appeared at frame 79 in the acquisition and remained at the same location until frame 177, where a fluorescent focus appeared further towards the edge of the cell. The focus was static for 98 frames equal to 4.9 seconds. The expected time for full transcription of the gene is ≈ 71 seconds. The static time of the focus is only a fraction of the time for a full

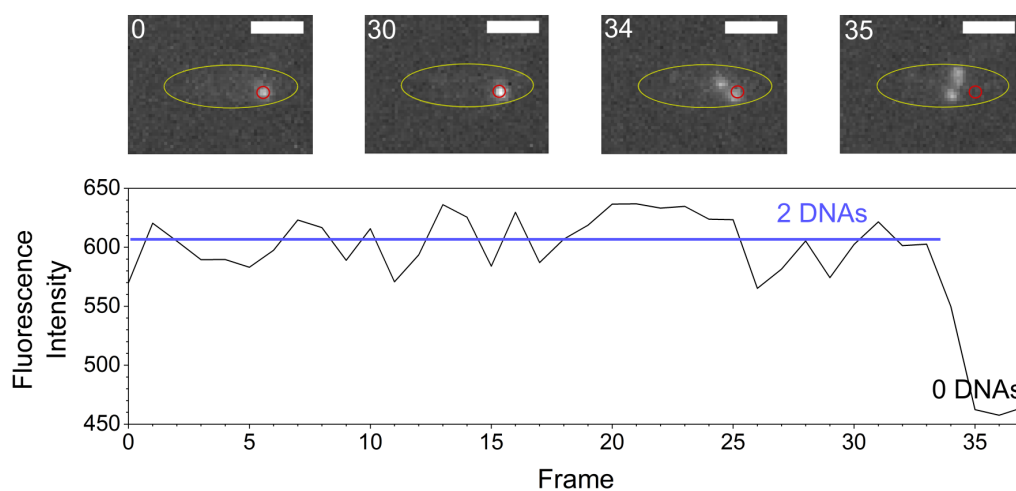


Figure 4.13: Fluorescence intensity of a single static focus (red circle) over the first 37 frames of an acquisition. Top: individual frames of the cell, 50 ms exposure time. The focus is seen to split at frame 34 and diffuse away from the original localisation. Scale bars $2 \mu\text{m}$. Bottom: Fluorescence intensity vs. time trace for the focus over 37 frames with a steady fluorescence intensity level corresponding to two probes bound to the target RNA being produced from active transcription at the localisation.

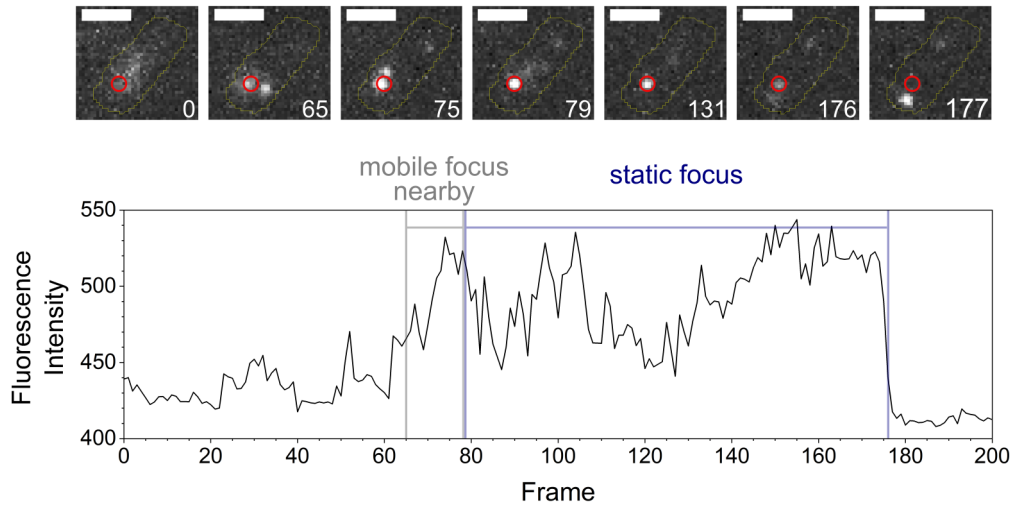


Figure 4.14: Fluorescence intensity of a single static focus (red circle) over the first 200 frames of acquisition. Top: individual frames of the cell, 50 ms exposure time. The static focus is seen to appear at frame 79 and remain until frame 176, preceded by a mobile focus in close proximity in frames 65-78. Scale bars $2\ \mu\text{m}$. Bottom: Fluorescence intensity vs. time trace for the area of the red circle over 200 frames.

transcription event under the above assumptions. On closer inspection of the movie and fluorescence intensity vs. time trace, a mobile focus was in the proximity of the location of the static focus in frames 65-78 before the static focus appeared in frame 79. One could hypothesise that the mobile focus becomes a static focus by capture of the target RNA by an immobile RNase. In frame 177, the focus moves away from the location and this could be presumed to be release of the RNA by the RNase after cleavage. The presence of the initial mobile focus in frames 65-78 suggests that this is not a transcription event being observed, however, this method of measuring static foci indicates how transcription and other kinetics related to RNA can be observed using *in vivo* smFISH. With a small amount of further optimisation, this assay has the potential to measure the full expected duration of transcription of the gene through a higher concentration of probes used during electroporation. Further changes to the method could include the use of stable organic fluorophores, as well as stroboscopy, which could yield an imaging time of several minutes, allowing the capture of full transcription events even for the longest genes.

4.3 Discussion

The results here show a proof-of-principle for *in vivo* single-molecule FISH (smFISH), which is a novel way to conduct FISH in live cells. It was shown that short fluorescently labelled probes can be used to bind to specific RNA targets inside live cells, without degradation of the probes. Target-specific binding was observed for both plasmid and chromosomal expression showing how to localise target RNA and, with some further optimisation, the ability to measure full tran-

scription events and transcription kinetics. In vivo smFISH is an advancement on previous experiments of FISH in live bacteria as single-molecules can be detected inside cells, not seen by Batani et al. [136] and the cells are still alive, unlike the the first attempt at FISH in bacteria by Silverman et al. [135].

Future short-term experiments include increasing the concentration of the probes in the cell in order to facilitate more parallel probe binding events allowing the observation of specific RNA over a longer time period. Also, reduction in the laser power would slow the speed of photobleaching of the probes, as long as a good signal to background ratio is maintained. Stroboscopy, with dark frames in between excitation frames, could be used to preserve the photon budget of the fluorophores and increase possible imaging times. With these slight adjustments, the in vivo smFISH method could be used to measure the time taken for transcription of selected genes and compare the transcription kinetics of different sequences.

In the longer-term, genes that have been modified to include a pause sequence could be compared to unmodified sequences, as longer transcription events could be identified by longer lived static foci within cells. Other possible experiments could include combining in vivo smFISH with chromosomal labels to look for co-localisation of a known site on the chromosome and the localisation of the target RNA. It may also be possible to use two different probes labelled with spectrally separated dyes, such as Cy3B and ATTO647N, to mark the start and end of transcription by placing binding sequences at the two ends of the RNA transcript allowing the time for transcription to be measured precisely.

To allow the in vivo smFISH method to be established as a key technique, the implementation of tracking software is essential allowing the velocity and location of many static and mobile fluorescent foci to be recorded. Once developed

further, in vivo smFISH can be used for, but not limited to, measuring sequence specific pausing of transcription, cell stress response, and cell response to antibiotic treatments, making it a powerful tool for studying live cell processes.

Chapter 5

Materials and Methods

5.1 Chapter 2

5.1.1 DNAs

A table of the gapped-DNA constructs sequences is included in the appendix to the chapter, see Figure 5.1. All DNA constructs were annealed in equimolar amounts in annealing buffer (200 mM Tris-HCL pH 8, 500 mM NaCl and 1 mM EDTA pH 8) by heating to 90 ° C and slow cooling to 4 ° C. A second table is included of the seals used in this chapter, see Figure 5.2. Explanation of the nomenclature - an example strand is GSeq(1-27)-T^{Cy3B, P18}. GSeq refers to the gap sequencing experiments, (1-27) documents the length of the strand and the position within the construct with respect to the 5' end, T represents top strand and ^{Cy3B, P18} notes the type and position of modifications.

The Biotinylated Hairpin, GSeq(1-27)-T^{Cy3B, P18}, GS12-compl-T^{AT647N, P28}, GS9-compl-T^{AT647N, P28}, GS6-compl-T^{AT647N, P28} and GS6-mis3G- T^{AT647N, P28} were pur-

chased from iba. All other DNAs were synthesised by Brown Group, Department of Chemistry, University of Oxford.

The biotinylated DNA constructs were immobilised on polyethylene glycol coated coverslips [145] at an approximate density of $0.23 \mu\text{m}^{-2}$.

5.1.2 Buffers

Standard imaging buffer: 0.5 M NaCl, 50 mM HEPES, 6 mM BSA, 1 mM Trolox and seals at specified concentration. The imaging buffer also contained an oxygen scavenging system: 1% glucose, $40 \mu\text{g ml}^{-1}$ catalase and 0.2 mg ml^{-1} glucose oxidase. For experiments with tetramethylammonium chloride (TMAC), the standard imaging buffer was adjusted to include 3 M TMAC. For experiments with seals containing 5-nitroindole universal bases the standard imaging buffer was adjusted slightly to contain 0.25 M NaCl.

5.2 Chapter 3

5.2.1 DNAs

The table in Figure 5.3 gives the DNA sequences for the DNA constructs and seals used in Chapter 3. All pre-annealed DNA constructs were annealed in equimolar amounts in annealing buffer (200 mM Tris-HCL pH 8, 500 mM NaCl and 1 mM EDTA pH 8) by heating to 90°C and slow cooling to 4°C . The nomenclature for the DNA names is the same as those in Chapter 2. GSeq-CAP- $\text{G}^{\text{P}25}, \text{C}^{\text{P}38}$ -(1-49)- $\text{B}^{\text{Bio}, \text{P}49}$ and GSeq-CAP- $\text{G}^{\text{P}38}$ -(30-49)-T were purchased from Sigma-Aldrich. GSeq-CAP- $\text{C}^{\text{P}25}$ -(1-49)- $\text{T}^{\text{Cy}5, \text{P}1}$ was purchased from Integrated DNA Technolo-

gies. GSeq-CAP-C^{P25},G^{P38}-(1-49)-T^{Cy3B,P13}, STD45-T^{Bio,P1,Cy3,P45} and STD45-B^{AT647N,P28} were purchased from iba. All other DNAs were synthesised by the Brown Group, Department of Chemistry, University of Oxford.

The biotinylated DNA constructs were immobilised on polyethylene glycol coated coverslips [145] at an approximate density of 0.23 μm^{-2} .

5.2.2 Protein Expression, Purification and Labelling

The doubly mutated catabolite activator protein (CAP) was expressed on the plasmid pHSCR_P C178SH17C created by Dr. El Sayyed, Kapanidis Group. The plasmid expresses CAP/CRP with the cysteine at position 178 changed to serine (to prevent labelling at this position) and the histidine at position 17 changed to cysteine (making cysteine labelling possible at this position). Position 17 was chosen as labelling with a fluorophore here should not interfere with DNA binding. The CAP gene was controlled by a T7 promoter downstream of LacI allowing transcription of the CAP gene in the presence of Isopropyl β -d-1-thiogalactopyranoside (IPTG). The plasmid was transformed into competent *E. coli* cells BL21 DE3 via heat shock. The cells were plated on LB agar plates with 100 $\mu\text{g}/\text{ml}$ Ampicillin overnight at 37 $^{\circ}\text{C}$, a colony picked then cultured in the presence of 100 $\mu\text{g}/\text{ml}$ Ampicillin to optical density (OD) of 0.6 OD units at 600 nm before 1 mM IPTG was added. After three hours shaking at 37 $^{\circ}\text{C}$, the cells were centrifuged at 4000 rpm for 30 minutes, collected and stored at -80 $^{\circ}\text{C}$.

The cells were later thawed and lysed using a French press in 50 ml lysis buffer (20 mM Tris-HCL pH8, 0.5 M NaCl, 5% glycerol) and protease inhibitor. Then centrifuged at 13000 rpm for 40 minutes and the supernatant collected. For pu-

rification of the protein a Ni-NTA column was prepared with 4 ml of Ni-NTA slurry and washed with 40 ml of equilibrium buffer (20 mM Tris-HCl pH8, 500 mM NaCl, 5% glycerol). The cell lysis supernatant was added to the column and the flow-through collected, twice. The column was then washed with washing buffer (20 mM Tris-HCl pH8, 500 mM NaCl, 5% glycerol, 500 mM imidazole). CAP was then exchanged into a storage buffer (40 mM Tris-HCl pH8, 100 mM NaCl, 5% glycerol, 1 mM TCEP) and stored in aliquots at -80°C .

Aliquots of CAP were then labelled with 10-fold excess Alexa647 maleimide by Dr. Mazumder, Kapanidis Group. Excess dye was removed from the labelled CAP by a G25 Sephadex column. The absorption spectra of fractions of CAP in CAP storage buffer after gel filtration were recorded to indicate fractions containing labelled protein. A final labelling efficiency of the CAP with Alexa647 was 52%.

5.2.3 Buffers

Single-molecule CAP binding assays were conducted in CAP binding buffer (40 mM Tris-HCl pH8.0, 100 mM KCl, 10 mM MgCl_2 , 5% glycerol, 0.2 mM cAMP). The CAP binding buffer also contained an oxygen scavenging system: 1% glucose, $40\text{ }\mu\text{g ml}^{-1}$ catalase and 0.2 mg ml^{-1} glucose oxidase.

Standard imaging buffer: 0.5 M NaCl, 50 mM HEPES, 6 mM BSA, 1 mM Trolox and seals at specified concentration. The imaging buffer also contained an oxygen scavenging system: 1% glucose, $40\text{ }\mu\text{g ml}^{-1}$ catalase and 0.2 mg ml^{-1} glucose oxidase.

5.2.4 Assay for Denaturing and Reannealing In Situ

Annealed biotinylated FRET standards were immobilised on the surface via polyethylene glycol (PEG) [145], washed with 1X PBS (Phosphate-Buffered Saline) thrice and imaged at 532 nm in 20 μ l standard imaging buffer. Then washed with 1X PBS thrice before incubation with 10 mM NaOH (20 μ l) for 1 minute. The sample was then washed again with 1X PBS thrice before imaging at 532 nm in 20 μ l standard imaging buffer. After washing three times with 1X PBS, the sample was incubated with 50 nM STD45-T^{AT647N,P28} with 0.5 M NaCl for 1 minute. The sample again was washed thrice with 1X PBS before 20 μ l of standard imaging buffer was added and the sample imaged at 532 nm.

The denaturing and in situ hybridisation of the CAP DNA constructs was conducted similarly. Annealed biotinylated CAP consensus binding sequence dsDNA construct was immobilised rather than the FRET standard. Also, 25 mM NaOH for 2 minutes was used instead of 10 mM for 1 minute. For gapped-DNA hybridisation in situ, gap hybridisation mix (10 nM of GSeq-CAP-(1-20)-T^{Cy3B,P11}, 10 nM GSeq-CAP-(30-49)-T, annealing buffer) was placed in the well for 5 minutes.

5.2.5 Assay of CAP Binding Followed by Gap-Seq

The whole assay for single-molecule CAP binding through to Gap-Seq was as follows. First, biotinylated dsDNA CAP binding constructs were immobilised on a PEG surface and washed thrice with 1X PBS. CAP was placed in the well in CAP binding buffer and imaged at 532 nm and 640 nm. The well was then washed three times with 1X PBS before a 2 minute incubation with 25 mM NaOH. The sample was washed thrice with 1X PBS and replaced with standard imaging buffer and

imaged at 532 nm and 640 nm. The sample was washed thrice with 1X PBS and the gap hybridisation mix was added for 5 minutes. The well was then washed with 1X PBS three times before being replaced with standard imaging buffer containing 100 nm seal DNA and imaged at 532 nm and 640 nm.

5.2.6 Mixed DNA on the Surface Followed by Gap-Seq

The assay used here was very similar to the previous assay with some slight changes. Two sequences CAP binding dsDNA constructs were immobilised, consensus binding sequence and 1 x mutant sequence instead of a single sequence. Gold nanoparticles were immobilised on the surface for 2 minutes with an average density of 6 per field of view in the green emission channel and 3 per field of view in the red emission channel. After imaging the dsDNA on the surface, 640 nm excitation was used in the presence of 1X PBS to bleach the Cy5 fluorophores on the 1 x mutant CAP binding sequence DNA constructs allowing localisation of the nanoparticles and Alexa647 CAP binding to visualised in future assays. No CAP in CAP binding buffer was added to the cell and the following washing step was skipped. Also, 25 mM NaOH was only used for 2 seconds, rather than two minutes, to denature dsDNA. The hybridisation mix was also contained 100 nM of the top flanking strands. The rest of the assay was the same as above.

5.3 Chapter 2 & 3 Data Collection

5.3.1 Microscopy and Data Analysis

Movies were collected using the Oxford Nanoimager microscope [146] at a maximum frequency of 100 Hz. The microscope was used as a widefield single-molecule fluorescence microscope with objective based total internal reflection illumination at 53 °C. Continuous wave excitation lasers (532 nm and 640 nm) were used to excite Cy3B and ATTO647N/Alexa647 respectively with the laser power adjusted to the experiment.

Drift Correction

Software used for x/y drift correction of the field of view was ‘Localize’ and ‘Render’ from the Picasso software [147]. Further drift correction across multiple movies by using the fiducial markers was performed by hand.

Time Traces and HMM Analysis

The FRET data was analysed using single-molecule FRET analysis software from Oxford Nanoimaging. The localisation of gapped-DNAs and the analysis of direct excitation of ATTO647N labelled seals of Alexa647 labelled CAP was conducted using custom built software built by Dr. Steuer, formally Kapanidis Group.

Data from experiments was collected in the form of .tiff or .fits movies where images of the field of view (FoV) of the sample were separated by the exposure time of the camera (10-100 ms). The custom-built software detected regions of high intensity fluorescence and determined them to be a point of fluorescence within the movie and identified the same region in the FoV in the other fluorescence

channel of the microscope. The intensity at these points was then graphed through the length of the movie to produce fluorescence intensity vs. time traces. The software automatically performed background fluorescence subtraction using the area surrounding the point of fluorescence. The fluorescence intensity vs. time traces were then sorted to identify traces that show fluorescent events which were analysed.

Hidden Markov Modelling (HMM) analysis was then be conducted on sorted fluorescence intensity vs. time traces [148]. HMM identified different intensity states within the traces and measured the number of frames spent in each state.

State Histograms

The collated state times were then appropriately binned and plotted as a histogram for each state. The histograms were then fitted with a decaying exponential fit in order to determine an average time spent in each state. A two-state model can be fitted with a single exponential fit, however, kinetics with more than two states are more accurately fitted with multiple exponentials [75]. The dwell time histograms were fitted with a decaying biexponential with off time histograms fitted with a single exponential. The times calculated from the fits of the histograms were be interpreted as the average time for that event. Dwell time histograms gave the average dwell time T_1 interpreted as the average time for partial/abortive binding and the average dwell time T_2 interpreted as the average time for full binding.

5.4 Chapter 4

5.4.1 DNAs

The table in Figure 5.4 gives the DNA sequences for probes used in Chapter 4.

5.4.2 Electroporation of Cells

Thawed MG1655WT cells were incubated with 1 μ M probes or Milli-Q water for 30 minutes. Cells were then placed in chilled 25 μ l electrocuvettes and electroporated with 1.4 kV. Cells recovered in S.O.C. media at 37 °C, 350 rpm for 30 minutes. Cells were centrifuged for 2.5 minutes at 3300 g at 4 °C, the supernatant removed and chilled 1X PBS added. The centrifuging and washing procedure was repeated three times. After the final wash, cells were suspended in 20 μ l of 1X PBS and placed on ice.

5.4.3 Imaging

For imaging, cells were placed on 1% agarose pads containing 1X M9 that were made between two burnt coverslips. Images of cells were collected using bright field and HILO microscopy (illumination at 532 nm and 640 nm) on a custom built widefield microscope and on the Oxford Nanoimager microscope [146] at a maximum frequency of 20 Hz. The data was viewed and analysed using FIJI [149].

5.4.4 Plasmid Creation

The plasmid pBEC1 was created to allow for expression of a designed target RNA. pBEC1 was generated by Gibson assembly with the plasmid backbone amplified by

PCR from the plasmid pACBSR [141] and the mini-gene insert was assembled by hybridising two long PCR primers. The mini-gene insert was put under the control of an arabinose-inducible promoter [142] and followed by a terminator creating an RNA of ≈ 100 -102 bases. The plasmid pBEC1 was transformed into MG1655WT cells via electroporation (2.5 kV), plated on a LB plate with 35 $\mu\text{g/ml}$ Chloramphenicol (CAM) and cultured in the presence of 0.1% arabinose. Cells were harvested in early exponential phase $< \text{OD}_{600} = 0.2$, were concentrated 600 times and stored at -80°C . Cells containing pBEC1 were made under the supervision of Dr. El Sayyed, Kapanidis Group.

5.4.5 Chromosome Expression

Cells modified to allow expression of target RNA on the chromosome were made by Dr. El Sayyed, Kapanidis Group. For chromosomal expression, the repeated binding array was inserted by Gibson assembly on a plasmid pTARA [144] carrying a T7 RNA polymerase gene under the control of the arabinose inducible promoter pBAD. The ribosome binding site was replaced by the 8 nt probe array. The plasmid was then used as a template to generate a lambda red substrate to insert the whole construct on the chromosome between *glnA* and *typA* genes in MG1655 E.coli. Experiments were made with electro competent cells harvested in early exponential phase $< \text{OD}_{600} = 0.2$, were concentrated 600 times and stored at -80°C .

5.5 DNA Sequences

Construct Name	Name of Strand	DNA sequence
Biotinylated Hairpin (15 nt gap)	Biotinylated Hairpin	5' CTG CTG CCT CAG GCT CAT --A AGA GCC TGA GGC AGC AGA TTA TCG ACG GTG GAT GTA TGG TAA TGG GAC GAA GAA TGA GG-bio 3'
	GSeq(1-27)-T ^{Cy3B,P18}	5' CCT CAT TCT TCG TCC CA <u>T</u> TAC CAT ACA 3'
8 nt gap	GSeq(1-65)-B ^{Bio,P1}	5' CAG TCA AGA GCC TGA GGC AGC AGA TTA TCG ACG GTG GAT GTA TGG TAA TGG GAC GAA GAA TGA GG-bio 3'
	GSeq(1-27)-T ^{Cy3B,P18}	5' CCT CAT TCT TCG TCC CA <u>T</u> TAC CAT ACA 3'
	GSeq(36-65)-T	5' CGA TAA TCT GCT GCC TCA GGC TCT TGA CTG 3'
6 nt gap	GSeq(1-65)-B ^{Bio,P1}	5' CAG TCA AGA GCC TGA GGC AGC AGA TTA TCG ACG GTG GAT GTA TGG TAA TGG GAC GAA GAA TGA GG-bio 3'
	GSeq(1-27)-T ^{Cy3B,P18}	5' CCT CAT TCT TCG TCC CA <u>T</u> TAC CAT ACA 3'
	GSeq(34-65)-T	5' GTC GAT AAT CTG CTG CCT CAG GCT CTT GAC TG 3'
7 nt gap (A-T exclusive)	GSeq-Atrich(1-65)-B ^{Bio,P1}	5' CAG TCA AGA GCC TGA GGC AGC AGA TTA TCG <u>AAT AAT ATT</u> GTA TGG TAA TGG GAC GAA GAA TGA GG-bio 3'
	GSeq(1-27)-T ^{Cy3B,P18}	5' CCT CAT TCT TCG TCC CA <u>T</u> TAC CAT ACA 3'
	GSeq(35-65)-T	5' TCG ATA ATC TGC TGC CTC AGG CTC TTG ACT G 3'
7 nt gap	GSeq(1-65)-B ^{Bio,P1}	5' CAG TCA AGA GCC TGA GGC AGC AGA TTA TCG ACG GTG GAT GTA TGG TAA TGG GAC GAA GAA TGA GG-bio 3'
	GSeq(1-27)-T ^{Cy3B,P18}	5' CCT CAT TCT TCG TCC CA <u>T</u> TAC CAT ACA 3'
	GSeq(35-65)-T	5'-TCG ATA ATC TGC TGC CTC AGG CTC TTG ACT G-3'
13 nt gap	GSeq(1-65)-A ^{P31} -B ^{Bio,P1}	5' CAG TCA AGA GCC TGA GGC AGC AGA TTA TCG ACG <u>GAG</u> GAT GTA TGG TAA TGG GAC GAA GAA TGA GG-bio 3'
	GSeq(1-27)-T ^{Cy3B,P18}	5' CCT CAT TCT TCG TCC CA <u>T</u> TAC CAT ACA 3'
	GSeq(41-65)-T	5' ATC TGC TGC CTC AGG CTC TTG ACT G 3'

Figure 5.1: Gapped-DNA constructs sequence's used in Chapter 2 - labelled bases are in bold and underlined.

Sequence Name	DNA Sequence
GS12-compl- $T^{AT647N,P28}$	5' <u>T</u> CC ACC GTC GAT 3'
GS9-compl- $T^{AT647N,P28}$	5' <u>T</u> CC ACC GTC 3'
GS6-compl- $T^{AT647N,P28}$	5' <u>T</u> CC ACC 3'
GS8-compl- $T^{ATTO647N,P28}$	5' <u>T</u> CC ACC GT 3'
GS8-mis5G- $T^{ATTO647N,P28}$	5' <u>T</u> CC <u>AG</u> C GT 3'
GS8-mis3G- $T^{ATTO647N,P28}$	5' <u>T</u> <u>C</u> G ACC GT 3'
GS6-mis3G- $T^{AT647N,P28,P33}$	5' <u>T</u> <u>C</u> G <u>ACC</u> 3'
GS8-compl- $T^{AT647N,P28,DABCYL,P35}$	5' <u>T</u> CC ACC G <u>T</u> 3'
GS8-compl- $T^{AT647N,P28,P35}$	5' <u>T</u> CC ACC G <u>T</u> 3'
GS7-ATrich- $T^{AT647N,P28,P34}$	5' <u>A</u> TA TTA <u>T</u> 3'
GS7-degen ^{P30,P32} - $T^{ATTO647N,P28,P34}$	5' <u>T</u> CN ANC <u>G</u> -3'
GS7-mis4T-degen ^{P30,P32} - $T^{ATTO647N,P28,P34}$	5' <u>T</u> CN <u>T</u> NC <u>G</u> -3'
GS13-N ^{10,11,12,13} - $T^{AT647N,P28,P35}$	5' <u>T</u> CC ACC G <u>T</u> C NNN N -3'
GS13-mis9G-N ^{10,11,12,13} - $T^{AT647N,P28,P35}$	5' <u>T</u> CC ACC G <u>T</u> G NNN N -3'

Figure 5.2: Seal DNA sequences used in Chapter 2 - labelled bases are in bold and underlined, mismatch bases are underlined only.

Construct Name	Name of Strand	DNA sequence
FRET Standard	STD45-B ^{Bio,P1,Cy3B, P45}	5' <u>T</u> AA ATC TAA AGT AAC ATA AGG TAA CAT AAC GTA AGC TCA TTC GCG bio 3'
	STD45-T ^{AT647N,P28}	5' CGC GAA TGA GCT TAC GTT ATG TTA CCT <u>T</u> AT GTT ACT TTA GAT TTA 3'
Consensus dsDNA	GSeq-CAP-(1-49)-B ^{Bio,P49}	5' bio- GTG CCT AAA ATG TGA TCT AGA TCA CAT TTA TTG CGT AGA GCT CAC TGC C 3'
	GSeq-CAP-(1-49)-T ^{Cy3B,P13}	5' GGC AGT GAG CTC <u>T</u> AC GCA ATA AAT GTG ATC TAG ATC ACA TTT TAG GCA C 3'
1 x mutant dsDNA	GSeq-CAP-G ^{P25} -(1-49)-B ^{Bio,P49}	5' bio- GTG CCT AAA ATG TGA TCT AGA TCA GAT TTA TTG CGT AGA GCT CAC TGC C 3'
	GSeq-CAP- C ^{P25} -(1-49)-T ^{Cy3B,P13}	5' GGC AGT GAG CTC <u>T</u> AC GCA ATA AAT <u>CTG</u> ATC TAG ATC ACA TTT TAG GCA C 3'
2 x mutant dsDNA	GSeq-CAP- C ^{P25} ,G ^{P38} -(1-49)-T ^{Cy3B,P13}	5' -GGC AGT GAG CTC <u>T</u> AC GCA ATA AAT CTG ATC TAG ATC AGA TTT TAG GCA C -3'
	GSeq-CAP-G ^{P25} ,C ^{P38} -(1-49)-B ^{Bio,P49}	5'bio- GTG CCT AAA ATC TGA TCT AGA TCA <u>GAT</u> TTA TTG CGT AGA GCT CAC TGC C -3'
1 x mutant dsDNA (Cy5)	GSeq-CAP-G ^{P25} -(1-49)-B ^{Bio,P49}	5' bio- GTG CCT AAA ATG TGA TCT AGA TCA <u>GAT</u> TTA TTG CGT AGA GCT CAC TGC C 3'
	GSeq-CAP- C ^{P25} -(1-49)-T ^{Cy5,P1}	5' - <u>G</u> GC AGT GAG CTC TAC GCA ATA AAT CTG ATC TAG ATC ACA TTT TAG GCA C -3'
Gapped DNA - Consensus	GSeq-CAP-(1-49)-B ^{Bio,P49}	5' bio- GTG CCT AAA ATG TGA TCT AGA TCA CAT TTA TTG CGT AGA GCT CAC TGC C 3'
	GSeq-CAP-(1-20)-T ^{Cy3B,P11}	5' GGC AGT GAG C <u>T</u> C TAC GCA AT 3'
	GSeq-CAP-(30-49)-T	5' CTA GAT CAC ATT TTA GGC AC 3'
Gapped DNA - 1 x mutant	GSeq-CAP-G ^{P25} -(1-49)-B ^{Bio,P49}	5' bio- GTG CCT AAA ATG TGA TCT AGA TCA <u>GAT</u> TTA TTG CGT AGA GCT CAC TGC C 3'
	GSeq-CAP-(1-20)-T ^{Cy3B,P11}	5' GGC AGT GAG C <u>T</u> C TAC GCA AT 3'
	GSeq-CAP-(30-49)-T	5' CTA GAT CAC ATT TTA GGC AC 3'
Seals	GS9-CAP-T ^{AT647N,P21,P29}	5' <u>A</u> AA TGT GAT <u>T</u> 3'
	GS9-CAP-mis5C-T ^{AT647N,P21,P29}	5' <u>A</u> AA TCT GAT <u>T</u> 3'

Figure 5.3: DNA sequences for the DNA constructs and the seals used in Chapter 3 - labelled bases are in bold and underlined.

Sequence Name	DNA Sequence
GS8-compl- T ^{ATTO647N,P28}	5' <u>T</u> CC ACC GT 3'
GS8-compl- T ^{AT647N,P28,P35}	5' <u>T</u> CC ACC G <u>T</u> 3'
GS7-mis4T- T ^{Cy3B,P28,AT647N,P34}	5' <u>T</u> CC TCC <u>G</u> 3'
EUB338 - P1, Cy3	5' <u>G</u> CT GCC TCC CGT AGG AGT 3'

Figure 5.4: DNA sequences of the probes used in Chapter 4 - labelled bases are in bold and underlined.

Chapter 6

Conclusion

In this thesis I present experiments in the development of two single-molecule methods, which utilise DNA hybridisation, one to link phenotype to DNA sequence for sequence dependent protein-DNA interactions and the other to view gene expression in live prokaryotic cells.

The first experiments presented were the proof-of-principle results for Gap-Seq - a high throughput single-molecule DNA/RNA sequencing method that uses the binding kinetics of labelled oligonucleotides (seals) binding to surface-immobilised gapped-DNAs. The difference in the binding kinetics of seals (8-6 nt), for a single base mismatch compared to full complementary binding, is a way to sequence a single base at a time. The ability to sequence unknown sequences using Gap-Seq was demonstrated through the use of degenerate or universal bases. After showing that single base changes in the binding sequence of the catabolite activator protein (CAP) affect its phenotype, an assay was developed to show how single-molecule measurements of catabolite activator protein (CAP) binding to surface-immobilised DNA can be followed by Gap-Seq to investigate the binding sequence.

The assay was made possible by the formation of gapped-DNAs in situ, from double stranded DNA, through chemical denaturing of the DNA and hybridisation of DNA strands in situ. Further experiments also indicated how the method could be used for multiple DNA sequences immobilised on the surface.

Gap-Seq has the potential to become a high-throughput short-read single-molecule sequencing method able to analyse many DNA sequences on a surface and link this to sequence dependent protein behaviour on both commercial and custom-built microscopes. The method does have some limiting points such as the need for drift correction, as is standard with super-resolution imaging techniques, and the necessity of fluorescently labelled seals, which can be costly. To improve the method, the implementation of fluidics for solution exchange is necessary and the use of artificial intelligence to classify fluorescence intensity vs. time traces will allow faster processing of results. Other related research that would provide more insights would be computational modelling of the kinetics of DNA hybridisation to gapped-DNA, to generate expected kinetics for specific sequences. Once developed, the Gap-Seq method would allow new insights into highly sequence dependent reactions such as those involving aptamers and sequence dependent processes such as transcription.

The other method under development in this thesis was ‘in vivo single-molecule FISH (smFISH)’. In vivo smFISH is a technique similar to fluorescence in situ hybridisation (FISH), however, in live prokaryotic cells rather than fixed cells. Short (8 nt) doubly fluorescently labelled oligonucleotides (probes) can be electroporated into live cells and bind to target RNAs producing highly fluorescent static or slowly diffusing spots (foci). Through analysis of foci, particularly the static foci, gene expression can be localised and the kinetics measured. In vivo

smFISH presents advancements on current attempts of FISH in live prokaryotic cells as single-molecules can be observed, giving access to the kinetics of cellular processes. With some further optimisation of the assay and implementation of tracking software, in vivo smFISH could be applied to study the effect of sequence dependent pausing during transcription for example, creating a powerful tool for understanding live bacterial processes.

The two biophysical methods developed in this thesis add to the growing evidence that interdisciplinary research is needed, now and in the future, to allow for innovation in science and to further humankind's understanding of the natural world.

Bibliography

- [1] Miescher, F. Über die chemische Zusammensetzung der Eiterzellen. *Medicinisch-chemische Untersuchungen* **4** 441–460 (1871).
- [2] Levene, P. A. & Jacobs, W. A. Über die Hefe-Nucleinsäure. *Berichte der deutschen chemischen Gesellschaft* **42**(2) 2474–2478 (1909).
- [3] OpenStax. https://cnx.org/contents/GFy_h8cu@9.87:yxeAKc4X@8/Nucleic-Acids. Accessed 30/03/20 .
- [4] Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6** 201–209 (1950).
- [5] Watson, J. & Crick, F. A Structure for Deoxyribose Nucleic Acid. *Nature* **171**(4356) 737–738 (1953).
- [6] Wilkins, M. H., Stokes, A. R. & Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **171**(4356) 738–740 (1953).
- [7] Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**(4356) 740–741 (1953).

- [8] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. *Molecular Biology of The Cell*. 4th edn. Garland Science (2002).
- [9] Voet, D. & Voet, J. *Biochemistry*. 2nd edn. Wiley (1995).
- [10] Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research* **34**(2) 564–574 (2006).
- [11] Porschke, D. *Chemical Relaxation in Molecular Biology*. 1st edn. Springer-Verlag Berlin Heidelberg (1977).
- [12] Devoe, H. & Tinoco, I. The stability of helical polynucleotides: Base contributions. *Journal of Molecular Biology* **4**(6) 500–517 (1962).
- [13] Crothers, D. M. & Zimm, B. H. Theory of the melting transition of synthetic polynucleotides: Evaluation of the stacking free energy. *Journal of Molecular Biology* **9**(1) 1–9 (1964).
- [14] Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences* **83**(11) 3746–3750 (1986).
- [15] Sugimoto, N., Nakano, S.-I., Yoneyama, M. & Honda, K.-I. Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes. *Nucleic Acids Research* **24**(22) 4501–4505 (1996).
- [16] Garrett, R. H. & Grisham, C. M. *Biochemistry*. 4th edn. Cengage Learning (2008).

- [17] Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry*. 5th edn. New York: W H Freeman (2002).
- [18] Ng, H.-L., Kopka, M. L. & Dickerson, R. E. The structure of a stable intermediate in the A left-right-arrow B DNA helix transition. *Proceedings of the National Academy of Sciences* **97**(5) 2035–2039 (2000).
- [19] Burtis, C. A., Ashwood, E. R. & Bruns, D. E. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 5th edn. Elsevier (2012).
- [20] Loakes, D. & Brown, D. M. 5-nitroindole as an universal base analogue. *Nucleic Acids Research* **22**(20) 4039–4043 (1994).
- [21] Loakes, D., Brown, D. M., Linde, S. & Hill, F. 3-nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR. *Nucleic Acids Research* **23**(13) 2361–2366 (1995).
- [22] Gallego, J. & Loakes, D. Solution structure and dynamics of DNA duplexes containing the universal base analogues 5-nitroindole and 5-nitroindole 3-carboxamide. *Nucleic Acids Research* **35**(9) 2904–2912 (2007).
- [23] Kunitsyn, A., Kochetkova, S., Kolganova, N., Tishchenko, E., Gottikh, B. & Florentiev, V. Stabilizing Effect of 5-Nitroindole (Universal Base) on DNA Duplexes Immobilized on Gel Matrix. *Journal of Biomolecular Structure and Dynamics* **15**(3) 597–603 (1997).
- [24] Lakowicz, J. *Principles of Fluorescence Spectroscopy*. 3rd edn. Springer (2006).

- [25] Nagashima, K. & Velan, S. S. Understanding the Singlet and Triplet States in Magnetic Resonance. *Concepts in Magnetic Resonance Part A* **42**(5) 165–181 (2013).
- [26] Stokes, G. G. On the Change of Refrangibility of Light. [Abstract]. *Abstracts of the Papers Communicated to the Royal Society of London* **6** 195–200 (1850).
- [27] Crisalli, P. & Kool, E. T. Multi-Path Quenchers: Efficient Quenching of Common Fluorophores. *Bioconjugate Chem.* **22**(11) 2345–2354 (2011).
- [28] Herschel, J. F. W. On a case of superficial colour presented by a homogeneous liquid internally colourless. *Abstracts of the Papers Communicated to the Royal Society of London* **5** 547 (1851).
- [29] Cooper, M., Ebner, A., Briggs, M., Burrows, M., Gardner, N., Richardson, R. & West, R. Cy3B: Improving the Performance of Cyanine Dyes. *Journal of Fluorescence* **14**(2) 145–150 (2004).
- [30] ATTO-TEC GmbH. Product Information : ATTO 647N. Tech. rep. (2019).
- [31] Umezawa, K., Matsui, A., Nakamura, Y., Citterio, D. & Suzuki, K. Bright, Color-Tunable Fluorescent Dyes in the Vis/NIR Region: Establishment of New “Tailor-Made” Multicolor Fluorophores Based on Borondipyrromethene. *Chemistry - A European Journal* **15**(5) 1096–1106 (2009).
- [32] ISS Fluorescence and Biomedical Instrumentation. Lifetime Data of Selected Fluorophores. Tech. rep. (2020).

- [33] Johnson, I. D. *The Molecular Probes Handbook—A Guide to Fluorescent Probes and Labeling Technologies*. 11th edn. Life Technologies Corporation (2010).
- [34] Fluorophores.org. <http://www.fluorophores.tugraz.at/substance/419>. Accessed 02/05/2020 .
- [35] ATTO-TEC GmbH. www.atto-tec.com/product_info.php?language=en&info=p114_atto-647n.html. Accessed 02/05/2020 .
- [36] Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J. & Waggoner, A. S. Cyanine dye labeling reagents: Sulfoindocyanine succinimidyl esters. *Bioconjugate Chemistry* **4**(2) 105–111 (1993).
- [37] Kapanidis, A. N., Lee, N. K., Laurence, T. A., Doose, S., Margeat, E. & Weiss, S. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proceedings of the National Academy of Sciences of the United States of America* **101**(24) 8936–41 (2004).
- [38] Zheng, Q. & Blanchard, S. C. *Single Fluorophore Blinking*, 2322–2323. Springer Berlin Heidelberg, Berlin, Heidelberg (2013).
- [39] Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods* **3**(11) 891–893 (2006).
- [40] Cordes, T., Vogelsang, J. & Tinnefeld, P. On the Mechanism of Trolox as Antiblinking and Antibleaching Reagent. *Journal of the American Chemical Society* **131**(14) 5018–5019 (2009).

- [41] Hecht, E. *Optics, Global Edition*. Pearson Education Limited (2016).
- [42] Kapanidis Laboratory. BPO4: Single-molecule fluorescence microscopy. Tech. rep., University of Oxford (2016).
- [43] Kapanidis, A. N., Laurence, T. A., Lee, N. K., Margeat, E., Kong, X. & Weiss, S. Alternating-Laser Excitation of Single Molecules. *Accounts of Chemical Research* **38**(7) 523–533 (2005).
- [44] Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods* **5**(2) 159–161 (2008).
- [45] Saunders, M. & Ross, P. D. A simple model of the reaction between polyadenylic acid and polyuridylic acid. *Biochemical and Biophysical Research Communications* **3**(3) 314–318 (1960).
- [46] Wetmur, J. G. & Davidson, N. Kinetics of renaturation of DNA. *Journal of Molecular Biology* **31**(3) 349–370 (1968).
- [47] Jung, J., Ihly, R., Scott, E., Yu, M. & Van Orden, A. Probing the complete folding trajectory of a DNA hairpin using dual beam fluorescence fluctuation spectroscopy. *Journal of Physical Chemistry B* **112**(1) 127–133 (2008).
- [48] Zhang, J. X., Fang, J. Z., Duan, W., Wu, L. R., Zhang, A. W., Dalchau, N., Yordanov, B., Petersen, R., Phillips, A. & Zhang, D. Y. Predicting DNA hybridization kinetics from sequence. *Nature Chemistry* **10**(1) 91–98 (2018).

- [49] Dupuis, N. F., Holmstrom, E. D. & Nesbitt, D. J. Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophysical Journal* **105**(3) 756–766 (2013).
- [50] Yang, H., Li, H. & Liu, T. Photobleaching statistics in single-molecule on-/off-time distributions. *The Journal of Chemical Physics* **151**(17) 174101 (2019).
- [51] Traeger, J. C. & Schwartz, D. K. Surface-Mediated DNA Hybridization: Effects of DNA Conformation, Surface Chemistry, and Electrostatics. *Langmuir* **33**(44) 12651–12659 (2017).
- [52] Gietl, A., Holzmeister, P., Grohmann, D. & Tinnefeld, P. DNA origami as biocompatible surface to match single-molecule and ensemble experiments. *Nucleic Acids Research* **40**(14) e110–e110 (2012).
- [53] Melnychuk, N., Egloff, S., Runser, A., Reisch, A. & Klymchenko, A. S. Light-Harvesting Nanoparticle Probes for FRET-Based Detection of Oligonucleotides with Single-Molecule Sensitivity. *Angewandte Chemie International Edition* **59**(17) 6811–6818 (2020).
- [54] Schickinger, M., Zacharias, M. & Dietz, H. Tethered multifluorophore motion reveals equilibrium transition kinetics of single DNA double helices. *Proceedings of the National Academy of the Sciences of the United States of America* **115**(32) E7512–E7521 (2018).
- [55] Peterson, A. W. The effect of surface probe density on DNA hybridization. *Nucleic Acids Research* **29**(24) 5163–5168 (2001).

- [56] Springer, T., Sipova, H., Vaisocherova, H., Stepanek, J. & Homola, J. Shielding effect of monovalent and divalent cations on solid-phase DNA hybridization: surface plasmon resonance biosensor study. *Nucleic Acids Research* **38**(20) 7343–7351 (2010).
- [57] Qiao, W., Chiang, H. C., Xie, H. & Levicky, R. Surface vs. solution hybridization: Effects of salt, temperature, and probe type. *Chemical Communications* **51**(97) 17245–17248 (2015).
- [58] Vainrub, A. & Pettitt, B. M. Coulomb blockage of hybridization in two-dimensional DNA arrays. *Physical Review E* **66**(4) (2002).
- [59] Johnson-Buck, A., Nangreave, J., Jiang, S., Yan, H. & Walter, N. G. Multifactorial Modulation of Binding and Dissociation Kinetics on Two-Dimensional DNA Nanostructures. *Nano Letters* **13**(6) 2754–2759 (2013).
- [60] Peterson, E. M., Manhart, M. W. & Harris, J. M. Single-Molecule Fluorescence Imaging of Interfacial DNA Hybridization Kinetics at Selective Capture Surfaces. *Analytical Chemistry* **88**(2) 1345–1354 (2016).
- [61] Monserud, J. H. & Schwartz, D. K. Mechanisms of surface-mediated DNA hybridization. *ACS Nano* **8**(5) 4488–4499 (2014).
- [62] Yazawa, K. & Furusawa, H. Probing Multiple Binding Modes of DNA Hybridization: A Comparison between Single-Molecule Observations and Ensemble Measurements. *ACS Omega* **3**(2) 2084–2092 (2018).

- [63] Boukobza, E., Sonnenfeld, A. & Haran, G. Immobilization in Surface-Tethered Lipid Vesicles as a New Tool for Single Biomolecule Spectroscopy. *The Journal of Physical Chemistry B* **105**(48) 12165–12170 (2001).
- [64] Cisse, I. I., Kim, H. & Ha, T. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nature Structural & Molecular Biology* **19**(6) 623–627 (2012).
- [65] Cisse, I., Okumus, B., Joo, C. & Ha, T. Fueling protein–DNA interactions inside porous nanocontainers. *Proceedings of the National Academy of Sciences* **104**(31) 12646–12650 (2007).
- [66] Okumus, B., Arslan, S., Fengler, S. M., Myong, S. & Ha, T. Single Molecule Nanocontainers Made Porous Using a Bacterial Toxin. *Journal of the American Chemical Society* **131**(41) 14844–14849 (2009).
- [67] Ishitsuka, Y., Okumus, B., Arslan, S., Chen, K. H. & Ha, T. Temperature-independent porous nanocontainers for single-molecule fluorescence studies. *Anal Chem* **82**(23) 9694–9701 (2010).
- [68] Jonkman, J. & Brown, C. M. Any Way You Slice It-A Comparison of Confocal Microscopy Techniques. *Journal of biomolecular techniques : JBT* **26**(2) 54–65 (2015).
- [69] Liu, C., Obliosca, J. M., Liu, Y.-L., Chen, Y.-A., Jiang, N. & Yeh, H.-C. 3D single-molecule tracking enables direct hybridization kinetics measurement in solution. *Nanoscale* **9**(17) 5664–5670 (2017).

- [70] Wang, Q. & Moerner, W. E. Single-molecule motions enable direct visualization of biomolecular interactions in solution. *Nature Methods* **11**(5) 555–558 (2014).
- [71] You, Y., Tataurov, A. V. & Owczarzy, R. Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers* **95**(7) 472–486 (2011).
- [72] Zimmers, Z. A., Nicholas, Gabella, W. E. & Haselton, F. R. Fluorophore-quencher interactions effect on hybridization characteristics of complementary oligonucleotides. *Analytical Methods* **11**(22) 2862–2867 (2019).
- [73] Moreira, B. G., You, Y. & Owczarzy, R. Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: Predictive thermodynamic model. *Biophysical Chemistry* **198** 36–44 (2015).
- [74] Hartmann, A., Krainer, G. & Schlierf, M. Different Fluorophore Labeling Strategies and Designs Affect Millisecond Kinetics of DNA Hairpins. *Molecules* **19**(9) 13735–13754 (2014).
- [75] Sobek, J., Rehrauer, H., Schauer, S., Fischer, D., Patrignani, A. & Landgraf, S. Single-molecule DNA hybridisation studied by using a modified DNA sequencer : a comparison with surface plasmon resonance data. *Methods and Applications in Fluorescence* **4** 15002 (2016).
- [76] Ouldrige, T. E., Sulc, P., Romano, F., Doye, J. P. K. & Louis, A. A. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Research* **41**(19) 8886–8895 (2013).

- [77] Strauss, S. & Jungmann, R. Up to 100-fold speed-up and multiplexing in optimized DNA-PAINT. *Nature Methods* **17** 789–791 (2020).
- [78] Tang, J., Sun, Y., Pang, S. & Han, K. Y. Spatially encoded fast single-molecule fluorescence spectroscopy with full field-of-view. *Scientific Reports* **7**(1) 10945 (2017).
- [79] Deußner-Helfmann, N. S., Auer, A., Strauss, M. T., Malkusch, S., Dietz, M. S., Barth, H.-D., Jungmann, R. & Heilemann, M. Correlative Single-Molecule FRET and DNA-PAINT Imaging. *Nano Letters* **18**(7) 4626–4630 (2018).
- [80] Jungmann, R., Steinhauer, C., Scheible, M., Kuzyk, A., Tinnefeld, P. & Simmel, F. C. Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Letters* **10**(11) 4756–4761 (2010).
- [81] Auer, A., Strauss, M. T., Schlichthaerle, T. & Jungmann, R. Fast, Background-Free DNA-PAINT Imaging Using FRET-Based Probes. *Nano Letters* **17**(10) 6428–6434 (2017).
- [82] Whitley, K. D., Comstock, M. J. & Chemla, Y. R. Elasticity of the transition state for oligonucleotide hybridization. *Nucleic Acids Research* **45**(2) 547–555 (2017).
- [83] Johnson-Buck, A., Su, X., Giraldez, M. D., Zhao, M., Tewari, M. & Walter, N. G. Kinetic fingerprinting to identify and count single nucleic acids. *Nature Biotechnology* **33**(June) 1–4 (2015).

- [84] Su, X., Li, L., Wang, S., Hao, D., Wang, L. & Yu, C. Single-Molecule Counting of Point Mutations by Transient DNA Binding. *Nature Publishing Group* (March) 1–9 (2017).
- [85] Li, Z., Zhou, X., Li, L., Liu, S., Wang, C., Li, L., Yu, C. & Su, X. Probing DNA Hybridization Equilibrium by Cationic Conjugated Polymer for Highly Selective Detection and Imaging of Single-Nucleotide Mutation. *Analytical Chemistry* **90**(11) 6804–6810 (2018).
- [86] Li, L., Yu, Y., Wang, C., Han, Q. & Su, X. Transient Hybridization Directed Nanoflare for Single-Molecule miRNA Imaging. *Analytical Chemistry* **91**(17) 11122–11128 (2019).
- [87] Hayward, S. L., Lund, P. E., Kang, Q., Johnson-Buck, A., Tewari, M. & Walter, N. G. Ultraspecific and Amplification-Free Quantification of Mutant DNA by Single-Molecule Kinetic Fingerprinting. *Journal of the American Chemical Society* **140**(37) 11755–11762 (2018).
- [88] Ratushna, V. G., Weller, J. W. & Gibas, C. J. Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. *BMC genomics* **6** 31 (2005).
- [89] Gao, Y., Wolf, L. K. & Georgiadis, R. M. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Research* **34**(11) 3370–3377 (2006).
- [90] Chen, C., Wang, W., Wang, Z., Wei, F. & Zhao, X. S. Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Research* **35**(9) 2875–2884 (2007).

- [91] Jo, J. J., Kim, M. J., Son, J. T., Kim, J. & Shin, J. S. Single-fluorophore monitoring of DNA hybridization for investigating the effect of secondary structure on the nucleation step. *Biochemical and Biophysical Research Communications* **385**(1) 88–93 (2009).
- [92] Blumhardt, P., Stein, J., Mücksch, J., Stehr, F., Bauer, J., Jungmann, R. & Schille, P. Photo-Induced Depletion of Binding Sites in DNA-PAINT Microscopy. *Molecules* **23**(12) 3165 (2018).
- [93] Lackey, H. H., Chen, Z., Harris, J. M., Peterson, E. M. & Heemstra, J. M. Single-Molecule Kinetics Show DNA Pyrimidine Content Strongly Affects RNA:DNA and TNA:DNA Heteroduplex Dissociation Rates. *ACS Synthetic Biology* **9**(2) 249–253 (2020).
- [94] Schueder, F., Stein, J., Stehr, F., Auer, A., Sperl, B., Strauss, M. T., Schille, P. & Jungmann, R. An order of magnitude faster DNA-PAINT imaging by optimized sequence design and buffer conditions. *Nature Methods* **16**(11) 1101–1104 (2019).
- [95] Nakano, S. Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Research* **27**(14) 2957–2965 (1999).
- [96] Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A. & Walder, J. A. Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations. *Biochemistry* **47**(19) 5336–5353 (2008).
- [97] Mitchell, M. L., Leveille, M. P., Solecki, R. S., Tran, T. & Cannon, B. Sequence-Dependent Effects of Monovalent Cations on the Structural Dy-

- namics of Trinucleotide-Repeat DNA Hairpins. *The Journal of Physical Chemistry B* **122**(50) 11841–11851 (2018).
- [98] Lee, J., Park, S., Kang, W. & Hohng, S. Accelerated super-resolution imaging with FRET-PAINT. *Molecular Brain* **10**(1) 63 (2017).
- [99] Holmstrom, E. D., Dupuis, N. F. & Nesbitt, D. J. Pulsed IR Heating Studies of Single-Molecule DNA Duplex Dissociation Kinetics and Thermodynamics. *Biophysical Journal* **106**(1) 220–231 (2014).
- [100] Pyshnyi, D. V., Pyshnaya, I. A., Levina, A. S., Goldberg, E. L., Zarytova, V. F., Knorre, D. G. & Ivanova, E. M. Thermodynamic Analysis of Stacking Hybridization of Oligonucleotides with DNA Template. *Journal of Biomolecular Structure and Dynamics* **19**(3) 555–570 (2001).
- [101] Kotler, L. E., Zevin-Sonkin, D., Sobolev, I. A., Beskin, A. D. & Ulanovsky, L. E. DNA sequencing: modular primers assembled from a library of hexamers or pentamers. *Proceedings of the National Academy of Sciences* **90**(9) 4241 – 4245 (1993).
- [102] Lane, M. J., Paner, T., Kashin, I., Faldasz, B. D., Li, B., Gallo, F. J. & Benight, A. S. The thermodynamic advantage of DNA oligonucleotide ‘stacking hybridization’ reactions: energetics of a DNA nick. *Nucleic acids research* **25**(3) 611–617 (1997).
- [103] Chen, J., Bremauntz, A., Kisley, L., Shuang, B. & Landes, C. F. Super-resolution mbPAINT for optical localization of single-stranded DNA. *ACS Applied Materials and Interfaces* **5**(19) 9338–9343 (2013).

- [104] Wang, Y., Woehrstein, J. B., Donoghue, N., Dai, M., Avendaño, M. S., Schackmann, R. C. J., Zoeller, J. J., Wang, S. S. H., Tillberg, P. W., Park, D., Lapan, S. W., Boyden, E. S., Brugge, J. S., Kaeser, P. S., Church, G. M., Agasti, S. S., Jungmann, R. & Yin, P. Rapid Sequential in Situ Multiplexing with DNA Exchange Imaging in Neuronal Cells and Tissues. *Nano Letters* **17**(10) 6131–6139 (2017).
- [105] Spahn, C., Hurter, F., Glaesmann, M., Karathanasis, C., Lampe, M. & Heilemann, M. Protein-Specific, Multicolor and 3D STED Imaging in Cells with DNA-Labeled Antibodies. *Angewandte Chemie International Edition* **58**(52) 18835–18838 (2019).
- [106] Schueder, F., Strauss, M. T., Hoerl, D., Schnitzbauer, J., Schlichthaerle, T., Strauss, S., Yin, P., Harz, H., Leonhardt, H. & Jungmann, R. Universal Super-Resolution Multiplexing by DNA Exchange. *Angewandte Chemie International Edition* **56**(14) 4052–4055 (2017).
- [107] Lin, D., Gagnon, L. A., Howard, M. D., Halpern, A. R. & Vaughan, J. C. Extended-Depth 3D Super-Resolution Imaging Using Probe-Refresh STORM. *Biophysical Journal* **114**(8) 1980–1987 (2018).
- [108] Schueder, F., Lara-Gutiérrez, J., Beliveau, B. J., Saka, S. K., Sasaki, H. M., Woehrstein, J. B., Strauss, M. T., Grabmayr, H., Yin, P. & Jungmann, R. Multiplexed 3D super-resolution imaging of whole cells using spinning disk confocal microscopy and DNA-PAINT. *Nature Communications* **8**(1) (2017).

- [109] Böger, C., Hafner, A.-S., Schlichthärle, T., Strauss, M. T., Malkusch, S., Endesfelder, U., Jungmann, R., Schuman, E. M. & Heilemann, M. Super-resolution imaging and estimation of protein copy numbers at single synapses with DNA-point accumulation for imaging in nanoscale topography. *Neurophotonics* **6**(03) 1 (2019).
- [110] Brockman, J. M., Su, H., Blanchard, A. T., Duan, Y., Meyer, T., Quach, M. E., Glazier, R., Bazrafshan, A., Bender, R. L., Kellner, A. V., Ogasawara, H., Ma, R., Schueder, F., Petrich, B. G., Jungmann, R., Li, R., Mattheyses, A. L., Ke, Y. & Salaita, K. Live-cell super-resolved PAINT imaging of piconewton cellular traction forces. *Nature Methods* (2020).
- [111] Jungmann, R., Avendaño, M. S., Woehrstein, J. B., Dai, M., Shih, W. M. & Yin, P. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature Methods* **11**(3) 313–318 (2014).
- [112] Sanger, F. & Nicklen, S. DNA sequencing with chain-terminating. *Proceedings of the National Academy of Sciences of the United States of America* **74**(12) 5463–5467 (1977).
- [113] Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**(2) 560–564 (1977).
- [114] Drmanac, R., Brukner, I., Labat, I., Brukner, I. & Crkvenjakov, R. Sequencing of megabase plus DNA by hybridization : Theory of the Method. *Genomics* **7543** 114–128 (1989).

- [115] Bains, W. & Smith, G. A Novel Method for Nucleic Acid Sequence Determination. *Journal of Theoretical Biology* **135** 303–307 (1988).
- [116] Lysov, V., Floretiev, V., Khorlyn, A., Khrapko, K., Shick, V. & Mirzabekov, A. DNasequencing by hybridisation with oligonucleotides. *Dokl. Acad. Sci. USSR* **303** 1508–1511 (1988).
- [117] Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P. & Rank, D. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**(January) 133–139 (2009).
- [118] Clarke, J., Wu, H.-c., Jayasinghe, L., Patel, A., Reid, S. & Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* **4**(April) 265–270 (2009).
- [119] Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13** 278–289 (2015).
- [120] Oxford Nanopore Technologies. nanoporetech.com/how-it-works. Accessed 25/03/2020 .
- [121] International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431** 931 (2004).
- [122] Ebright, R. H., Ebright, Y. W., Gunasekera, A. & Brunswick, N. Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E.coli lac DNA site. *Nucleic Acids Research* **17**(23) 16–18 (1989).

- [123] Melchior, W. B. & Hippel, P. Alteration of the Relative Stability of dA-dT and dG-dC Base Pairs in DNA. *Proceedings of the National Academy of Sciences of the United States of America* **70**(2) 298–302 (1973).
- [124] Ball, S., Reeve, M. a., Robinson, P. S., Hill, F., Brown, D. M. & Loakes, D. The use of tailed octamer primers for cycle sequencing. *Nucleic acids research* **26**(22) 5225–7 (1998).
- [125] Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Helen, M. & Ebright, R. H. Catabolite activator protein (CAP): DNA binding and transcription activation. *Current Opinion in Structural Biology* **14**(1) 10–20 (2004).
- [126] Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nature Protocols* **12**(6) 1198–1228 (2017).
- [127] Rudkin, G. T. & Stollar, B. D. High resolution detection of DNA–RNA hybrids in situ by indirect immunofluorescence. *Nature* **265**(5593) 472–473 (1977).
- [128] Weng, X., Bohrer, C. H., Bettridge, K., Lagda, A. C., Cagliero, C., Jin, D. J. & Xiao, J. Spatial organization of RNA polymerase and its relationship with transcription in Escherichia coli. *Proceedings of the National Academy of Sciences* **116**(40) 20115 LP – 20123 (2019).
- [129] Skinner, S. O., Sepúlveda, L. A., Xu, H. & Golding, I. Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nature protocols* **8**(6) 1100–1113 (2013).

- [130] Speicher, M. R., Ballard, S. G. & Ward, D. C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genetics* **12**(4) 368–375 (1996).
- [131] Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R. & Stahl, D. A. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and environmental microbiology* **56**(6) 1919–1925 (1990).
- [132] Amann, R. & Fuchs, B. M. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nature Reviews Microbiology* **6**(5) 339–348 (2008).
- [133] Vargas, D. Y., Raj, A., Marras, S. A. E., Kramer, F. R. & Tyagi, S. Mechanism of mRNA transport in the nucleus. *Proceedings of the National Academy of Sciences* **102**(47) 17008–17013 (2005).
- [134] Zhang, X., Song, Y., Shah, A. Y., Lekova, V., Raj, A., Huang, L., Behlke, M. A. & Tsourkas, A. Quantitative assessment of ratiometric bimolecular beacons as a tool for imaging single engineered RNA transcripts and measuring gene expression in living cells. *Nucleic Acids Research* **41**(15) e152–e152 (2013).
- [135] Silverman, A. P. & Kool, E. T. Quenched autoligation probes allow discrimination of live bacterial species by single nucleotide differences in rRNA. *Nucleic Acids Research* **33**(15) 4978–4986 (2005).

- [136] Batani, G., Bayer, K., Böge, J., Hentschel, U. & Thomas, T. Fluorescence in situ hybridization (FISH) and cell sorting of living bacteria. *Scientific Reports* **9**(1) (2019).
- [137] Crawford, R., Torella, J. P., Aigrain, L., Plochowietz, A., Gryte, K., Uphoff, S. & Kapanidis, A. N. Long-Lived Intracellular Single-Molecule Fluorescence Using Electroporated Molecules. *Biophysical Journal* **105**(11) 2439–2450 (2013).
- [138] Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* **85**(1) 21–38 (2012).
- [139] Sanamrad, A., Persson, F., Lundius, E. G., Fange, D., Gynna, A. H. & Elf, J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. *Proceedings of the National Academy of Sciences* **111**(31) 11413–11418 (2014).
- [140] Mohapatra, S., Choi, H., Ge, X., Sanyal, S. & Weisshaar, J. C. Spatial Distribution and Ribosome-Binding Dynamics of EF-P in Live Escherichia coli. *mBio* **8**(3) e00300–17 (2017).
- [141] Herring, C. D., Glasner, J. D. & Blattner, F. R. Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli. *Gene* **311** 153–163 (2003).
- [142] Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology* **177**(14) 4121 LP – 4130 (1995).

- [143] Gotta, S. L., Miller Jr, O. L. & French, S. L. rRNA transcription rate in *Escherichia coli*. *Journal of bacteriology* **173**(20) 6647–6649 (1991).
- [144] Wycuff, D. R. & Matthews, K. S. Generation of an AraC-araBAD Promoter-Regulated T7 Expression System. *Analytical Biochemistry* **277**(1) 67–73 (2000).
- [145] Joo, C. & Ha, T. Single-Molecule FRET with Total Internal Reflection Microscopy. In *Single-Molecule Techniques: A Laboratory Manual*, 3–36 (2008).
- [146] Kapanidis, A., Jing, B. & Crawford, R. Compact Microscope. *WO/2015/059682* (2015).
- [147] Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat Protoc* **12**(6) 1198–1228 (2017).
- [148] Mckinney, S. A., Joo, C. & Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophysical Journal* **91**(5) 1941–1951 (2006).
- [149] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**(7) 676–682 (2012).