Single-molecule DNA Sensors and Cages For Transcription Factors *In Vitro* and *In Vivo*

Robert Crawford

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

St John's College

University of Oxford

28^{th} February 2011



Abstract

Gene regulation is vital to the success of all living organisms. Understanding this complex process is crucial to our knowledge of how cells function and how in some cases they can lead to debilitating or even fatal disease. In this thesis I focus on a set of DNA-binding proteins known as transcription factors (TFs), proteins fundamental to the process of gene regulation at the level of transcription. I develop assays and techniques for the detection and quantitation of TFs *in vitro* and *in vivo* as well as a method for TF encapsulation and release.

The advantages of the TF detection assays in this thesis are made possible through the use of single-molecule (sm) fluorescence. This methodology enables detection of individually labeled molecules allowing discrimination of sample heterogeneities inaccessible with ensemble techniques. Here I present two different TF assays based on two sm observables: relative probe stoichiometry and Förster resonance energy transfer (FRET).

The first assay design, based on stoichiometry, detects TFs using TF-dependent coincidence of two distinctly labelled DNA 'half-sites'. I demonstrate sensitive detection (~ pM) in solution and on surfaces, multiplexed detection of multiple TFs, and detection in cell lysates. A kinetic model of the system is also developed, verified experimentally and used to quantify TF concentrations without the need for a calibration curve.

The second assay design, based on FRET, is a novel approach to TF detection using TF-mediated DNA bending. TFs are detected by bending the sensor and monitored with FRET at the single-molecule or ensemble level. I demonstrate TF detection in purifed form and expressed in cell lysates. As this sensor was designed for use *in vivo*, methods to hinder nuclease degradation are explored.

For TF detection *in vivo*, I describe a successful strategy to internalise fluorescently labeled molecules into live *E.coli*. Viability and internalisation efficiency are characterised and ensemble measurements with FRET standards are demonstrated. Importantly, sm FRET measurements *in vivo* are achieved opening many exciting possibilities. The FRET based TF sensor is then internalised as a step towards real-time *in vivo* monitoring of TF concentrations.

Finally a system based on DNA nanotechnology is presented for the non-covalent encapsulation and release of TFs. Such a system could be delivered into a cell to alter levels of gene expression using external stimuli as inputs.

We believe these tools will generate valuable information in the study of prokaryotic gene expression as well as providing a potential commercial avenue towards diagnostics.

Acknowledgements

First and foremost I would like to thank all the members of the Kapanidis group for a very enjoyable 3 years of research. Whilst at times 'it was the CAP!' , the overall experience has been thoroughly rewarding.

Thanks also go to the reviewers of my chapters: Louise Aigrain, Christoph Erben, Seamus Holden, Johannes Hohlbein, Achillefs Kapanidis, Joseph Torella and Andrew Turberfield. Contributions and additional thanks for individual chapters are included at the end of each.

A special mention certainly goes to Joe for his support in the lab, teaching this misplaced Electronic Engineer more than a thing or two Biochemical...

To my friends and family. I thank my parents for their continued support of my academic ventures. My sister Elizabeth, my brother-in-law, James, and my girlfriend Emily have also supported me greatly over this period.

Finally I would like to express my gratitude to Achilles for accepting me into his lab, his guidance over the years and his ever enthusiastic support and optimism even when it was the CAP.

Contents

1 Introduction **7**

2 Detecting Single Molecules **11**

2.1 Introduction 11

2.2 Principles of Fluorescence 11

2.3 Förster Resonance Energy Transfer (FRET) 13

2.4 Single-Molecule Fluorescence 16

2.4.1 Single-molecule Confocal Spectroscopy 17

2.4.2 Total Internal Reflection Fluorescence (TIRF) Microscopy 18

2.5 Alternating Laser Excitation Spectroscopy (ALEX) 19

2.6 Experimental Setup 23

3 Transcription Factor Detection: Stoichiometry Approach **27**

3.1 Introduction 27

3.2 Assay Design 29

3.3 Results 31

3.3.1 Transcription Factor Detection 31

3.3.2 Transcription Factor Quantitation 34

3.3.3 Multiplexed Transcription Factor Detection 36

3.3.4 Transcription Factor Detection In Cell Lysates 39

3.4 Discussion 40

3.5 Contributions 41

4 Transcription Factor Detection: FRET Approach **42**

4.1 Introduction . . . 42

4.2 Assay Design . . . 43

4.3 Results . . . 45

4.3.1 FRET-based TF Detection In Solution . . . 45

4.3.2 FRET-based TF Detection On A Surface . . . 47

4.3.3 Nuclease resistance . . . 49

4.4 Discussion . . . 55

4.5 Contributions . . . 57

5 Toward Transcription Factor Detection *in vivo* **58**

5.1 Introduction . . . 58

5.2 Internalisation Method: Electroporation . . . 61

5.3 Results . . . 64

5.3.1 Internalisation Of Fluorescent Species . . . 64

5.3.1.1 Single-stranded DNA . . . 64

5.3.1.2 Double-stranded High FRET DNA . . . 66

5.3.1.3 Fluorescent NTPs . . . 68

5.3.2 Internalisation Efficiency And Viability . . . 70

5.3.3 Single-molecule Fluorescence Measurements *in vivo* . . . 73

5.3.4 Towards Transcription Factor Detection *in vivo* . . . 77

5.4 Summary . . . 79

5.5 Contributions . . . 81

6 Transcription Factor Encapsulation And Release Within A DNA Cage **82**

6.1 Introduction . . . 82

6.2 Protein-Encapsulating DNA Cage . . . 88

6.2.1 Design . . . 88

6.2.2 Results . . . 94

6.2.2.1 Protein-cage Binding and Encapsulation . . . 94

6.2.2.2 Confirmation of Protein Encapsulation Using Single-Molecule Fluorescence . . . 98
6.2.2.3 Protein Retention Within Cage . . . 101
6.2.3 Discussion . . . 104
6.3 Reversibly Opening DNA Cage . . . 106
6.3.1 Design . . . 106
6.3.2 Results . . . 107
6.3.2.1 Double Nicked Edge Stability . . . 107
6.3.2.2 Correct Formation Of Reversibly Opening Cage . . . 110
6.3.2.3 Cage Opening Experiments Using PAGE . . . 112
6.3.2.4 Cage Opening Experiments Using Ensemble Fluorescence . . . 115
6.3.3 Discussion . . . 122
6.4 Overall Discussion . . . 123
6.5 Contributions . . . 124

7 Conclusion and outlook **125**
7.1 Summary . . . 125
7.2 Future Work . . . 127

A Supplementary Material **130**
A.1 DNA Used . . . 130
A.1.1 Transcription Factor Detection: Stoichiometry Approach . . . 130
A.1.2 Transcription Factor Detection: FRET Approach . . . 131
A.1.3 Towards Transcription Factor Detection In-vivo . . . 132
A.1.4 Transcription Factor Encapsulation And Release Within A DNA Cage . . . 132
A.1.4.1 Protein-Encapsulating DNA Cage . . . 132
A.1.4.2 Reversibly Opening Tetrahedron . . . 133
A.2 Experimental Protocols & Supplementary Data . . . 135
A.2.1 Transcription Factor Detection: Stoichiometry Approach . . . 135
A.2.2 Transcription Factor Detection: FRET Approach . . . 140
A.2.3 Towards Transcription Factor Detection In vivo . . . 141

A.2.3.1 Preparation Of Electrocompetent *E.coli* 141
A.2.3.2 Electroporation Of *E.coli* With Fluorescent Molecules 142
A.2.3.3 Widefield Imaging Of *E.coli* With Internalised Fluorescent Molecules 143
A.2.3.4 Data Analysis Of Widefield Fluorescent Cell Images 144
A.2.3.5 Single-molecule Confocal Spectroscopy Of Live *E.coli* With Internalised Fluorescent Molecules 144
A.2.4 Transcription Factor Encapsulation and Release 145
A.2.4.1 Formation Of Cages and Complexes 145
A.2.4.2 Encapsulation Of CAP Within A DNA Cage 145
A.2.4.3 Removal Of cAMP From CAP-Cage Complexes 146
A.2.4.4 CAP Off-rate Competition Assay 146
A.2.4.5 Opening Cage Ligation And Exonuclease Procedure 146
A.2.4.6 Opening Cage Bulk Fluorescence Measurements 147
A.3 Scientific Contributions Arising From This Thesis 148

Bibliography **149**

List Of Abbreviations Used

ALEX - Alternating laser excitation

APD - Avalanche photodiode

bp - base pairs

CAP - Catabolite activator protein a.k.a cyclic AMP receptor protein (CRP)

ds - double stranded

dNTP - Deoxyribonucleoside tri-phosphate

DNA - deoxyribonucleic acid

E.coli - *Escherichia coli*

EP - Electroporation

FCS - Fluorescence correlation spectroscopy

FOV - Field of view

FP - Fluorescent protein

FRET - Förster resonance energy transfer

LacR - Lactose repressor

mRNA - Messenger RNA

NTP - Nucleoside tri-phosphate

PAGE - Polyacrylamide gel electrophoresis

PCR - Polymerase chain reaction

POI - Protein of interest

RNA - Ribonucleic acid

RT - Room temperature

s.d. - Standard deviation

sm - single-molecule

ss - single stranded

TF - Transcription factor

TIRF - Total internal reflection microscopy

Chapter 1

Introduction

Gene regulation is fundamental to all life. Multicellular organisms, consisting of several trillion cells, of several hundred types, each must develop from a single cell through a complex program of cell division and differentiation governed by tightly regulated gene expression. All cells contain exactly the same genetic information (DNA) yet assume very different roles to carry out very different tasks [1]. Cells must appropriately respond to environmental stimuli, communicate with each other and die when required, all for the success of the larger organism and species as a whole [1]. Organisms are an example of an amazingly complex, yet perfectly engineered control system. While malfunctions can occur, such as uncontrollable cell division (cancer), given the complexity and sheer numbers involved, it is a wonder there are not many more problems that arise. Evolution has refined highly efficient systems that can even intervene with their own natural fate to correct malfunctions e.g. through the use of gene therapy [2].

In this thesis, I explore a fundamental set of gene regulatory molecules known as transcription factors (TFs). TFs are sequence-specific DNA binding proteins that act to directly modulate gene expression by aiding or hindering the attachment of RNA polymerase to promoter sites [3]. Thus intracellular TF concentrations must be precisely controlled for each cell to carry out its particular role. Loss of genetic control can be catastrophic, leading to a number of debilitating or even fatal diseases; mutations in the TF p53 are linked to a large fraction of human tumours [4]. In contrast, alterations in TF levels can be used in a beneficial medical context. Intracellular delivery or expression of only four TFs can revert a fully differentiated cell into an embryonic stem cell state [5, 6]; a state that could be further re-differentiated as a proposed treatment of a number of

diseases such as leukaemia or Parkinson's disease [7, 8, 9]. The ability to accurately detect and quantify active TF concentrations both *in vitro* and *in vivo* would not only greatly improve our understanding of gene regulation, but would also serve as important markers indicative of disease [10]. As discussed in the following chapters, current TF assays less than optimal. Aspects such as signal amplification, taking several hours, can hinder their use for rapid diagnostics and prohibit their use for *in vivo* studies. In this thesis I present two novel approaches for sensitive TF detection and quantitation *in vitro* and *in vivo* using some of the latest methods in Biophysics.

The method predominately used in this thesis is single-molecule fluorescence. Fluorescence, the emission of light from a molecule given excitation at a shorter wavelength, is employed here to observe individually labelled molecules either freely diffusing in solution or attached to a surface. First observed in 1989 by Moerner et al. [11], single-molecule fluorescence has developed into a mature field incorporating a wide range of tools to study populations of individual molecules [12, 13]. This approach to molecular characterisation has led to a number of advancements in our knowledge of enzyme kinetics [14, 15] and conformational dynamics [16, 17] that would otherwise be inaccessible with ensemble experiments due to sample averaging. As a modality for biosensing, the single-molecule approach is ideally suited; sensitivity, specificity, time-to-result, multiplexing capability and identification of separate molecular subpopulations provide a platform on which many novel and exciting assays can and have been developed [18, 19, 20]. Single molecule methods themselves encompass a wide variety of techniques, not all of which are fluorescence based [21]. Techniques such as atomic force microscopy (AFM) [22, 23] or tethered particle motion (TPM) [24] have allowed new insight into biological processes such as nucleoid formation, TF-mediated DNA looping or protein folding. Other well developed fields such as those using optical [25] or magnetic tweezers [26] can infer protein-DNA interactions from positional distance changes of attached beads. In the work of this thesis I focus on single-molecule fluorescence methods; the principles behind which are outlined in Chapter 2.

The single-molecule experiments I perform in this thesis use a technique known as alternating laser excitation (ALEX) [27, 28]. Briefly, ALEX allows virtual 'sorting' of labelled molecules based on two ratiometric parameters: relative fluorophore stoichiometry and Förster resonance energy transfer (FRET). Relative fluorophore stoichiometry is measure of relative fluorophore brightness, and FRET is the measure of the distance between a 'donor' and 'acceptor' fluorophore. These

parameters enabled the development of the two versatile single-molecule assays to detect TFs, one using stoichiometry as the observable (Chapter 3) and the other, FRET (Chapter 4). The assays were developed to detect bacterial TFs from *Escherichia coli* (*E.coli*). As gene regulations systems have been studied extensively in this organism [1], it served as an excellent model system that could be used as a platform before extending to more complex eukaryotic systems.

While *in vitro* studies provide a great deal of information and precise control over experimental variables, they may not fully represent molecular behaviour *in vivo*. Macromolecular crowding, unknown cofactors, or concentrations of competitor or inhibitor molecules may all play a role in affecting enzyme activity in the cellular environment [29, 30]. One of the difficulties in performing fluorescence measurements *in vivo* is tagging molecules of interest. Green fluorescent protein (GFP) and its derivatives [31] fused to the N and/or C terminus of the protein of interest (POI) are one successful methodology that have contributed significantly to single-molecule efforts *in vivo* [32, 33, 34]. However limited photostability and brightness, requirement to genetically modify the POI, inability to label internal protein sites and often comparable size to the POI may hinder GFP use in certain scenarios. The use of smaller, brighter and more photostable organic dyes offer an attractive alternative. Such dyes can be covalently conjugated to small tags, again placed at the N- and/or C- terminus of the POI [35] but suffer similar drawbacks in terms of genetic modification and inability to label internal sites. Direct internalisation of fluorescently labelled DNA or proteins would open up many possibilities for *in vivo* fluorescence measurements. Whilst this has been achieved in mammalian cells via microinjection [36], unfortunately the dimensions of bacteria preclude its use here; bacterial cell walls also pose challenges to the introduction of foreign material.

In Chapter 5, I describe a general method for the internalisation of fluorescently labelled molecules into live *E.coli*. I demonstrate successful internalisation of labelled single- and double-stranded DNA and nucleoside tri-phosphates characterising internalisation efficiency and effects on cell viability. As a step towards real-time TF detection *in vivo*, we internalise the FRET-based TF sensor developed in Chapter 4. An important result achieved in this chapter was the first observation of single-molecule FRET in bacterial cells. This ability will allow the advantages of single-molecule techniques such as ALEX to be brought into an *in vivo* context in live bacterial cells. Single-molecule detection is especially important for gene regulation *in vivo* due to the in-

herent stochasticity of gene expression [37]; a single TF dissociation event has even been shown to change physiological outcome [38]. We believe the methods developed in this chapter significantly add to the single-molecule toolbox and open the possibility of a wide variety of interesting experiments in live bacterial cells at the sm level.

Finally in Chapter 6, as a proof-of-principle, I describe the use of DNA nanotechnology to develop a novel TF encapsulation and release system. DNA, normally known as the molecule storing genetic information, is a remarkably versatile material with which to rationally design and build nanodevices from the bottom up. As such, many structures [39, 40], motors [41, 42] and 'logic gates' [43, 44, 45] have been produced. In this work I adapt the design of a DNA tetrahedron [46] to produce two nanodevices; one design is used for TF encapsulation and the other is designed to have the ability to reversibly open and close. It may be possible in the future to combine the designs of this chapter with the internalisation methods of Chapter 5 to produce a system for controlled TF delivery into bacterial cells as a means to modulate gene expression.

Chapter 7 summarises the main findings of the thesis and suggests avenues for future research. The Appendix contains supplementary information such as DNA sequences, protocols and some supplemental data.

Chapter 2

Detecting Single Molecules

2.1 Introduction

As discussed in Chapter 1, the modality of single-molecule fluorescence is ideally suited to biosensing applications. In this chapter I discuss the physical phenomena of fluorescence and Förster Resonance Energy Transfer (FRET)[47, 48], and methods employed in this thesis that utilize these phenomena to characterise individual molecules, such as Alternating Laser Excitation (ALEX) [27, 28]. In Chapters 3 and 4, I describe the use of ALEX to detect and quantify single DNA-binding proteins known as transcription factors (TFs) using two different single-molecule observables. Similar methods are used in Chapter 6 to confirm encapsulation of a TF inside a DNA cage. Experiments in these chapters involve the observation of molecules *in vitro*, either freely diffusing in solution or attached to a surface. In Chapter 5 I describe the extension of the methods to the detection of individual molecules in live bacterial cells.

2.2 Principles of Fluorescence

Fluorescence is the process of photon emission caused by relaxation of an orbital electron from an excited state back to the ground state [49]. Fluorescence is preceded by the absorption of a photon of a shorter wavelength (higher energy) than that of the emitted photon and is normally illustrated using a Jablonski diagram as shown in Figure 2.2.1.

Following light absorption, a ground state electron of the fluorophore is excited from S_0 to

an electronically excited state, which then decays rapidly to the vibration level, S_1 [49]. The fluorophore can then return to the ground state through a variety of routes. The most important mechanism for discussions in this thesis is that of fluorescence whereby a photon is emitted which occurs with a lifetime of ~10 ns (rate: k_{FL}). Other mechanisms by which the fluorophore returns to the ground state are that of quenching and phosphorescence. Quenching can occur through collisions with other molecules or through formation of non-fluorescent complexes with quenchers. Quenching is dependent on quencher concentration ([Q]) and is accounted for with rate k_Q [49]. Phosphorescence occurs when excited molecules undergo spin conversion to the first triplet state, called intersystem crossing, which then return to the ground state emitting a photon (of lower energy, longer wavelength than fluorescence). As crossing to the first triplet state is forbidden, this process occurs several of orders of magnitude more slowly than fluorescence (micro- to milli-seconds) and is denoted with a rate k_{PH}. Phosphorescence however is not the dominant decay process of most fluorophores and for the purposes of this work we will focus on fluorescence. All other non-radiative paths to the ground state are collectively accounted for with the rate k_{NR}. Bright fluorophores with a high quantum yield (ratio of emitted/absorbed photons) will have $k_{NR} < k_{FL}$ [49].

Figure 2.2.1 b) shows the absorption and emission spectra for a typical fluorophore (Cy3). As photons emitted from the fluorophore are of lower energy (longer wavelength) than those absorbed, the maximum point of the emission peak is at higher wavelength than that of the absorption peak. This difference is known as the Stokes shift, first observed by Sir G G Stokes with a solution of quinine in 1852 [50]. A large Stokes shift is an advantage as more of the emission light can be collected and separated from the excitation light.

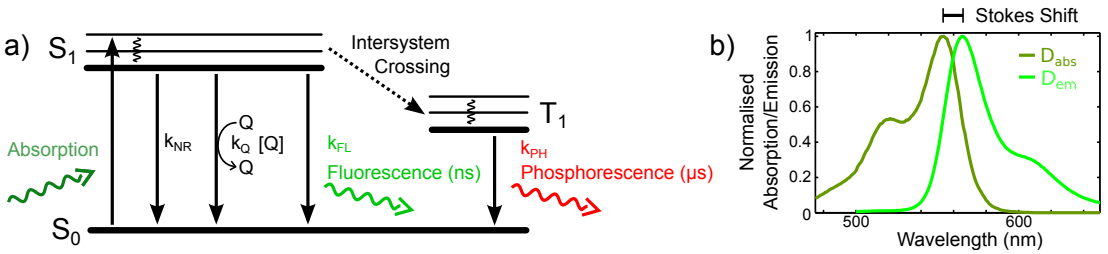


Figure 2.2.1: Principles of fluorescence. a) Jablonski diagram illustrating the electronic transitions occurring between the absorption and emission of light during fluorescence. Absorption of a photon causes fluorophore transition from S_0 to S_1 (ground to first excited singlet state). Return to ground state occurs either through: fluorescence by emission of a photon - timescale of ns (rate: k_{FL}), quenching - collisional or static (rate: k_Q, dependent on quencher concentration [Q]), or by phosphorescence through intersystem crossing to T_1 (first triplet state) resulting in the emission of a photon of lower energy than fluorescence - timescale of μs to ms (rate: k_{PH}) . All other non-radiative paths to the ground state are represented by k_{NR}. b) Normalised absorption and emission spectra of a typical fluorophore (Cy3) illustrating the Stokes shift.

2.3 Förster Resonance Energy Transfer (FRET)

A useful fluorescence phenomena for biological observations is Förster Resonance Energy Transfer, commonly known as FRET, which describes the process of non-radiative energy transfer between a 'donor' fluorophore and 'acceptor' molecule with overlapping emission/absorption spectra. If the acceptor molecule is a fluorophore, the energy transfer may result in photon emission. First discovered by Theodore Förster in 1948 [47], the donor and acceptor are coupled by a strongly distance dependent dipole-dipole interaction operating effectively in the range of ~2-10 nm. FRET is therefore ideally suited to study biological molecules of interest, such as DNA or proteins that have dimensions on a similar scale, far below the resolution of a light microscope (~200 nm). As such, FRET is sometimes known as a 'spectroscopic ruler' [48].

We illustrate FRET using again a Jablonski diagram in Figure 2.3.1 a). Here, on absorption of a photon of an appropriate wavelength, a ground state electron of the donor fluorophore undergoes a transition from $S_{0,D}$, to an electronically excited state, which then decays rapidly to the vibration level, $S_{1,D}$. The excited donor fluorophore can then return to the ground state either by emitting a photon as with normal fluorescence, or transfer its energy to a nearby acceptor molecule at a rate k_T. This energy transfer causes the transition of an acceptor ground state electron, at $S_{0,A}$ to its first vibration level, $S_{1,A}$, which then decays back to the ground state, releasing a photon

(via fluorescence) from the acceptor fluorophore at a much longer wavelength than the original donor excitation photon. Whilst other non-radiative paths to the ground state described for the fluorescence case (Figure 2.2.1) exist, they are omitted here for simplicity. The rate of energy transfer, k_T, is given by:

$$k_T(r) = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6, \tag{2.3.1}$$

where r is the distance between the donor and acceptor, τ_D is the lifetime of the donor in the absence of the acceptor and R_0 is the Förster radius at which the energy transfer efficiency between the donor and acceptor is 50% [49]. R_0 is specific to the particular donor-acceptor pair and is given by:

$$R_0^6 = \frac{9\ln(10)Q_D\kappa^2}{128\pi^5 n^4 N_A}J(\lambda), \tag{2.3.2}$$

where Q_D is the quantum yield (ratio of absorbed/emitted photons) of the donor in the absence of the acceptor. κ^2 is the relative dipole orientation factor for the transition dipoles of the donor and acceptor. κ^2 can vary between 0 and 4 depending on the rotational freedom of the fluorophores. For all cases considered in this thesis, it is assumed the dipoles have unrestricted rotational freedom, implying a κ^2 value of 2/3. n is the refractive index of the medium (typically 1.4 for molecules in aqueous solution). N_A is Avogadro's number. $J(\lambda)$ is the integral quantifying the overlap between donor emission and acceptor absorption spectra. $J(\lambda)$ is given by:

$$J(\lambda) = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda, \tag{2.3.3}$$

where F_D is the normalised donor emission spectrum (dimensionless). ε_A is the molar extinction coefficient (in $M^{-1}cm^{-1}$). λ is the wavelength (in nm). The spectral overlap for the common donor-acceptor pair Cy3B-ATTO647N is shown in Figure 2.3.1 b).

If k_T, the rate of energy transfer is much faster than the rate of decay of the donor fluorophore, the energy transfer will be efficient. Conversely if k_T is much slower than the donor decay rate, little energy will be transferred during the donor excited state lifetime, and hence energy transfer will be inefficient [49]. Energy transfer efficiency (or FRET efficiency), E, is defined as the fraction

of photons absorbed by the donor that are transferred to the acceptor. E is given by the ratio of the transfer rate to total decay rate of the donor in the presence of the acceptor:

$$E = \frac{k_T(r)}{\tau_D^{-1} + k_T(r)}. \tag{2.3.4}$$

Substituting equation 2.3.1 into this equation and rearranging yields:

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}. \tag{2.3.5}$$

We see from this equation the very strong (sixth power) distance dependence of FRET efficiency. The distance dependence is most sensitive at R_0, the Förster radius for the particular donor-acceptor pair at which the E is 50%. R_0 dictates the dynamic range of distances observable, and hence care should be taken in the selection of the right donor-acceptor pair for a particular experiment. The distance dependence of FRET efficiency on donor acceptor distance for a variety of R_0 values is shown in Figure 2.3.1 c). In a typical experiment, FRET efficiency values between 10-90% are considered most reliable as they lie within an approximately linear region of each curve.

FRET efficiency can be measured in a variety of ways including fluorescence intensities, fluorescence lifetime, donor quantum yield, and fluorescence anisotropy measurements [51, 49]. For the work performed in this thesis, for a single donor-acceptor pair, FRET efficiency is calculated as the ratio of acceptor fluorescence to total fluorescence (donor + acceptor) given donor excitation light:

$$E = \frac{F_D^A}{F_D^D + F_D^A}, \tag{2.3.6}$$

where F_X^Y is the fluorescence measured in the Y channel, given X excitation.

Thus, by tagging molecules of interest with donor and acceptor fluorophores, distances and distance changes can be monitored, allowing the measurement of molecular interactions, dynamics, and associated kinetic rates as illustrated in Figure 2.3.1 d).

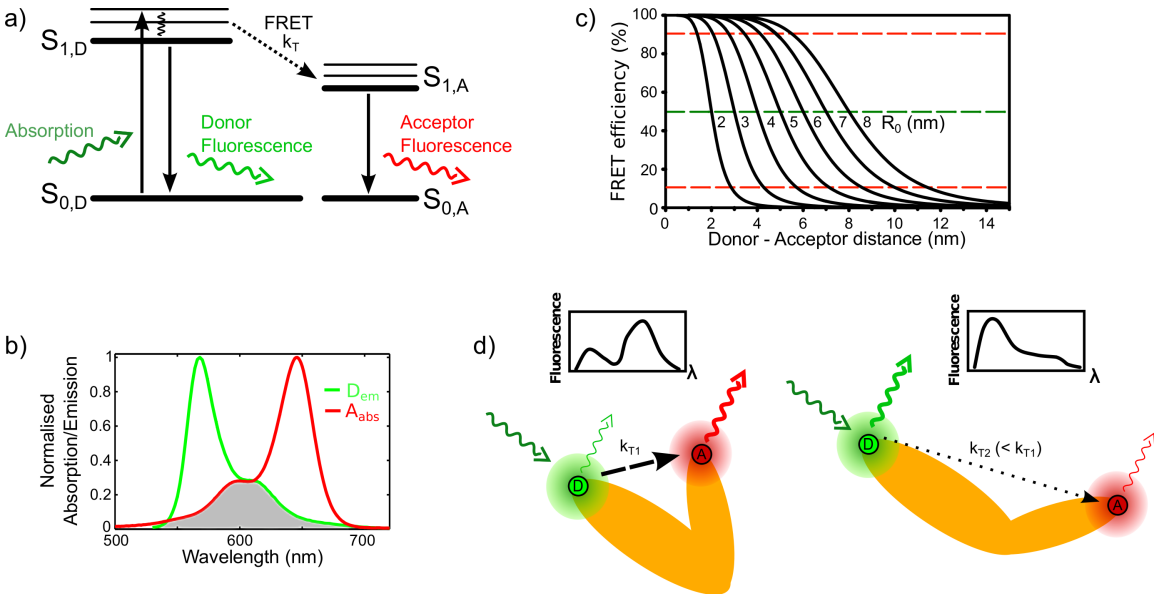


Figure 2.3.1: Förster Resonance Energy Transfer (FRET). a) Simplified Jablonski diagram illustrating the process of FRET. Here, a photon is absorbed by the 'donor' causing a transition of a ground state electron from $S_{0,D}$ to $S_{1,D}$. The excited electron can then return to the ground state either by releasing a photon (as in fluorescence), or by transferring its energy to a nearby 'acceptor', which if a fluorophore can lead to the subsequent release of a photon through fluorescence (at a longer wavelength). b) The normalised absorption and emission spectra of a common FRET pair Cy3B-ATTO647N. The overlapping region, indicated in grey, is one of the determining factors of the Förster radius, R_0, at which FRET efficiency is 50%. c) Graph depicting the sixth power dependence of donor-acceptor distance on FRET efficiency for a range of R_0 values between 2-8 nm (indicated by the green dashed line). The most reliable FRET data is between 10-90%, in an approximately linear region of the curve (bounded by the red dashed lines). d) Schematic of a biomolecule of interest labelled with a donor-acceptor pair undergoing a conformational change from 'closed' (left, high FRET), to 'open' (right, low FRET). Above each molecule is a schematic of spectral data expected for each case.

2.4 Single-Molecule Fluorescence

Whilst ensemble fluorescence and FRET measurements yield a lot of information on biological systems [52, 53], one of the main drawbacks is due to the unavoidable averaging that occurs when observing many molecules at one time. Consider the molecules illustrated in Figure 2.3.1 d). If there existed a mixture of such molecules at a 1:1 ratio, observed at the same time, neither a high or a low FRET state would emerge. Instead, a single averaged FRET value would result, intermediate between the two extremes. The ability to observe individual molecules would allow the discrimination of such sample 'heterogeneity'.

Sample heterogeneity can come in two forms: static and dynamic. Static heterogeneity refers to collections of molecules fixed in certain states e.g. double-stranded (ds) DNA molecules labelled in with donor-acceptor pairs in different positions [54]. Dynamic heterogeneity refers to molecules that each can interconvert between certain states, e.g. the opening and closing of a DNA polymerase around template DNA[16] which at the ensemble level would require synchronisation of all molecules. By considering molecules on an individual basis, single-molecule fluorescence and FRET are ideally suited to study such molecular subpopulations.

Since the first experiments to observe single-molecules over 20 years ago [11, 55], the number of publications using single-molecule (sm) fluorescence based techniques has increased exponentially, producing a vast body of information inaccessible with ensemble methods [12]. SM fluorescence can study molecules that are: freely diffusing, attached to surfaces, confined in channels or wells, encapsulated within vesicles, or attached to optically or magnetically trapped beads [12]. The experiments in this thesis concern the study of freely diffusing or surface attached molecules, using confocal spectroscopy and total internal reflection fluorescence microscopy respectively. These methods are discussed in the following sections.

2.4.1 Single-molecule Confocal Spectroscopy

For freely diffusing experiments we use sm confocal spectroscopy. Here the excitation laser light is focused to a femtolitre-sized diffraction-limited spot through which fluorescently labelled molecules diluted to picomolar (pM) concentrations can diffuse via Brownian motion. As labelled molecules diffuse through the spot, they give rise to 'bursts' of fluorescence [56], which are subsequently detected by avalanche photodiodes (APDs) with single photon detection sensitivity. A schematic of the process is illustrated in Figure 2.4.1 The main advantages of single-molecule confocal spectroscopy are that the experiments have very high time resolution and do not suffer from unwanted side effects involved with surface based approaches e.g. non-specific absorption. The main disadvantage is that molecules can only be observed for the time they take to diffuse through the confocal spot, on the order of ~1 ms [57].

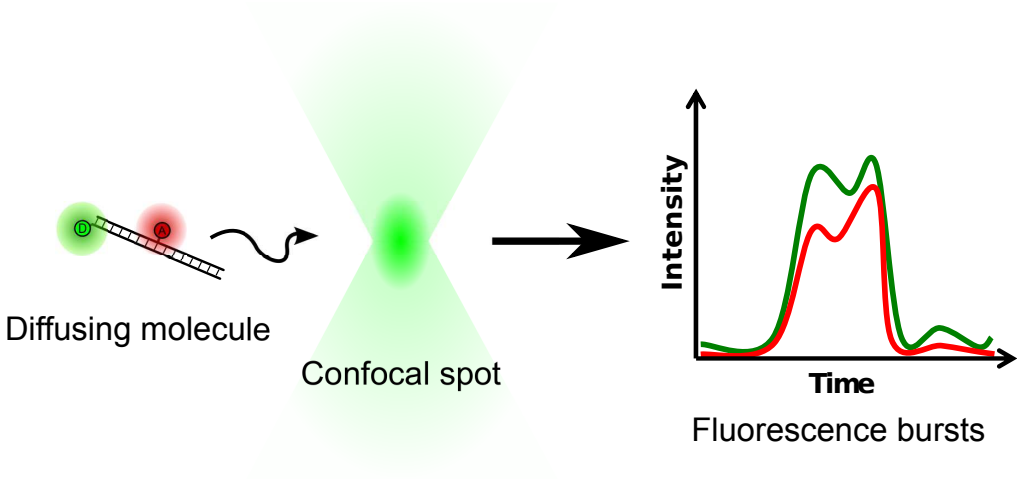


Figure 2.4.1: Single-molecule confocal spectroscopy. Fluorescently labelled molecules of interest diluted to pM levels diffuse via Brownian motion through a femtolitre-sized confocal spot formed from focused excitation laser light. The molecules give rise to fluorescent 'bursts', which are then detected using avalanche photodiodes.

2.4.2 Total Internal Reflection Fluorescence (TIRF) Microscopy

An alternative to confocal spectroscopy is to probe molecules attached to the surface using total internal reflection fluorescence microscopy [58, 59]. Attachment to the surface is mediated through neutravidin-biotin linkage; molecules are excited using a thin evanescent wave just above the coverslip. This evanescent wave is formed by focusing the incoming laser light at the edge of the back focal plane of the objective (illustrated in Figure 2.4.2, left) such that the light hits the coverslip just beyond the critical angle. The result being exponentially decreasing laser intensity from the coverslip on the order of ~100 nm. This evanescent wave has the advantage of only illuminating those molecules attached to the surface and not those freely diffusing above, leading to a high signal-to-noise ratio (SNR). The resulting fluorescence from the attached molecules is split into green and red emission channels and detected using an electron multiplying charge couple device (emCCD) (Figure 2.4.2, right).

The advantages of single-molecule fluorescence using TIRF are the ability to observe the same molecule for long periods of time and its high parallelism; hundreds of molecules can be imaged at once. The disadvantage being that the time resolution achievable is limited to the frame rate of the emCCD camera (typically ~10ms). Another disadvantage of TIRF is the surface effects on molecules, such as non-specific absorption, resulting in a higher background signal and/or potential

inactivation of protein activity [60].

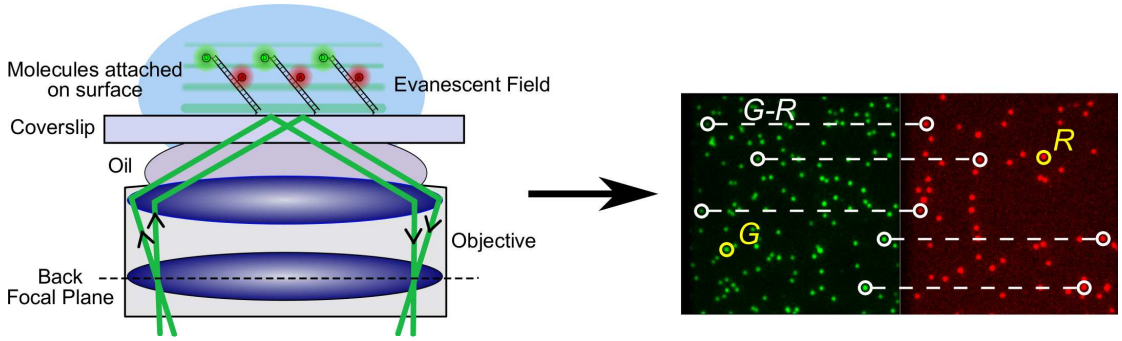


Figure 2.4.2: Single-molecule total internal reflection microscopy (TIRF). Left: schematic illustrating the formation of the evanescent wave above the coverslip. Excitation light (green lines) is focused at the edge of the back focal plane of the objective, such that it hits the coverslip just above the critical angle. This produces an exponentially decaying field of depth ~100 nm illuminating only those molecules attached to the surface. Right: Typical fluorescence image obtained using TIRF microscopy. Green and red channels are split onto the left and right halves of the emCCD chip. Individual molecules are then detected and quantified as described in the main text.

2.5 Alternating Laser Excitation Spectroscopy (ALEX)

Method

A significant improvement to single-molecule FRET measurements employed in this thesis is the use of Alternating Laser Excitation Spectroscopy (ALEX) [27, 28]. ALEX adds to the information available in smFRET experiments by probing the presence of the acceptor. Knowledge of the presence of the acceptor in a FRET pair allows discrimination of molecules containing single or multiple distinct fluorophores and to discriminate photophysics from real events.

To implement ALEX spectroscopy, two laser excitation sources are used, that alternately excite the donor and acceptor fluorophores on the microsecond (confocal) or millisecond timescale (TIRF). During donor excitation, donor and acceptor emission, F_D^D and F_D^A respectively, are recorded and FRET efficiency calculated using equation 2.3.6. During acceptor excitation, the acceptor emission, F_A^A , is recorded allowing the calculation of a new parameter, relative fluorophore stoichiometry, S, given by:

$$S = \frac{F_D^D + F_D^A}{F_D^D + F_D^A + F_A^A}, \tag{2.5.1}$$

where F_X^Y is the fluorescence measured in the Y channel, given X excitation. With ALEX, molecules can then be sorted virtually using two parameters: FRET efficiency (E) and relative fluorophore stoichiometry (S).

Relative fluorophore stoichiometry can be thought of as the ratio of total fluorescence under donor excitation to total fluorescence under acceptor excitation. The value of S thus varies from 0 to 1: ~1 for donor-only molecules, ~0 for acceptor-only and ~0.5 for a 1:1 ratio of donor:acceptor molecules (assuming the laser powers are adjusted correctly). Using both E and S, a 2-dimensional ES histogram can be plotted, where each point represents the number of molecules with those specific values. By observing many molecules, subpopulations on the ES histogram become apparent which identify particular biomolecular states or interactions.

A schematic representation of an ES histogram is shown in Figure 2.5.1 a). Donor- and acceptor-only molecules are shown at the top left and bottom right of the histogram at either end of the S-axis. In these cases, the FRET efficiency parameter is meaningless, but due to the relative fluorescence values, donor only molecules appear low FRET and acceptor only as mid to high FRET. Double labelled molecules appear in the middle of the histogram (S ~ 0.5), with a FRET efficiency dependent on the proximity between the donor and acceptor. The sorting afforded by the ALEX approach allows discrimination of the molecules of interest, e.g. doubly labelled proteins, from molecules which would contaminate the signal in a normal smFRET experiment e.g. mis-labelled / photobleached proteins or free dye.

Data Acquisition And Analysis

The fluorescence measured from each molecule will depend on the implementation of ALEX.

In confocal (diffusion based) experiments, photons and their arrival times are recorded by the avalanche photodiodes. Bursts of fluorescence, manifested by a large number of photons in a short period of time, indicate the transit of a fluorescent molecule through the confocal spot. Bursts are quantified using a modified detection algorithm described by the Seidel group [61]. Briefly a burst is recorded if, for every photon, there exist N neighbouring photons within a time window, T, of that photon. Typical parameters are N = 7, and T = 500μs. E and S are then calculated from the number of photons in each channel for each burst.

For TIRF (surface based) experiments, fluorescence is measured by analysing images obtained

from the emCCD camera. Images of individual molecules attached on the surface are split into the green and red channels and projected onto the left and right halves of the camera chip itself (Figure 2.4.2). Briefly, the green and red channel data are aligned using a transformation matrix, and molecules identified in each channel. For each molecule, green and red fluorescence values are then recorded allowing calculation of E and S values. All data analysis was performed using custom written LabVIEW and MATLAB programs.

One caveat is noted here for real experimental data. Factors such as detector efficiency differences (confocal), uneven illumination (TIRF), spectral crosstalk (leakage), direct excitation or quantum yield difference result in E and S values that are offset from their 'true' value. To measure accurate values, for example to calculate distances from E, one can measure and compensate for these factors [62]. However for all cases in this thesis, this is not a problem, as we consider *relative* changes in E and S which are still valid when considering interactions or kinetics. For real experimental data, doubly labelled molecules are considered at an S value between ~0.35-0.8, with FRET values ranging from ~0.1 to ~0.9.

Applications

Some applications of ALEX are shown in Figure 2.5.1 b) - d). Schematics of the molecules involved and the expected ES histogram are shown top and bottom in each panel. b) shows single-stranded (ss) donor and acceptor labelled DNA molecules (left) displaying high E values from secondary structure bringing the donor and acceptor into close proximity. The addition of the complementary strand (middle) results in the formation of a linear dsDNA segment, increasing the distance between the donor and acceptor, producing low E values. Further, the addition of nucleases (right) to the sample, results in degradation of the dsDNA molecules such that coincident detection (mid S values) is lost. c) shows an acceptor labelled dsDNA (left) with a single stranded overhang corresponding to a low S, acceptor-only population. The addition of donor labelled DNA polymerase (middle) causes a high E population to emerge from polymerase binding to the DNA. Addition of a subset of nucleotide tri-phosphates (NTPs) (right) allows DNA synthesis to occur and translocation of the polymerase along the DNA up to a certain point. The ES histogram now shows a low E population. d) shows donor and acceptor labelled DNA 'half-sites' (left) each containing half the binding site of a particular protein and complementary single-stranded overhang regions

freely diffuse in solution. Donor-only and acceptor-only populations are present in the histogram (high and low *S*). The presence of the particular protein (right) results in association of the two half-sites generating a low *E*, mid *S* population. This stoichiometric approach to detection would not be possible with single excitation source smFRET experiments. Such an assay was developed for the detection of particular proteins and is discussed in the next chapter.

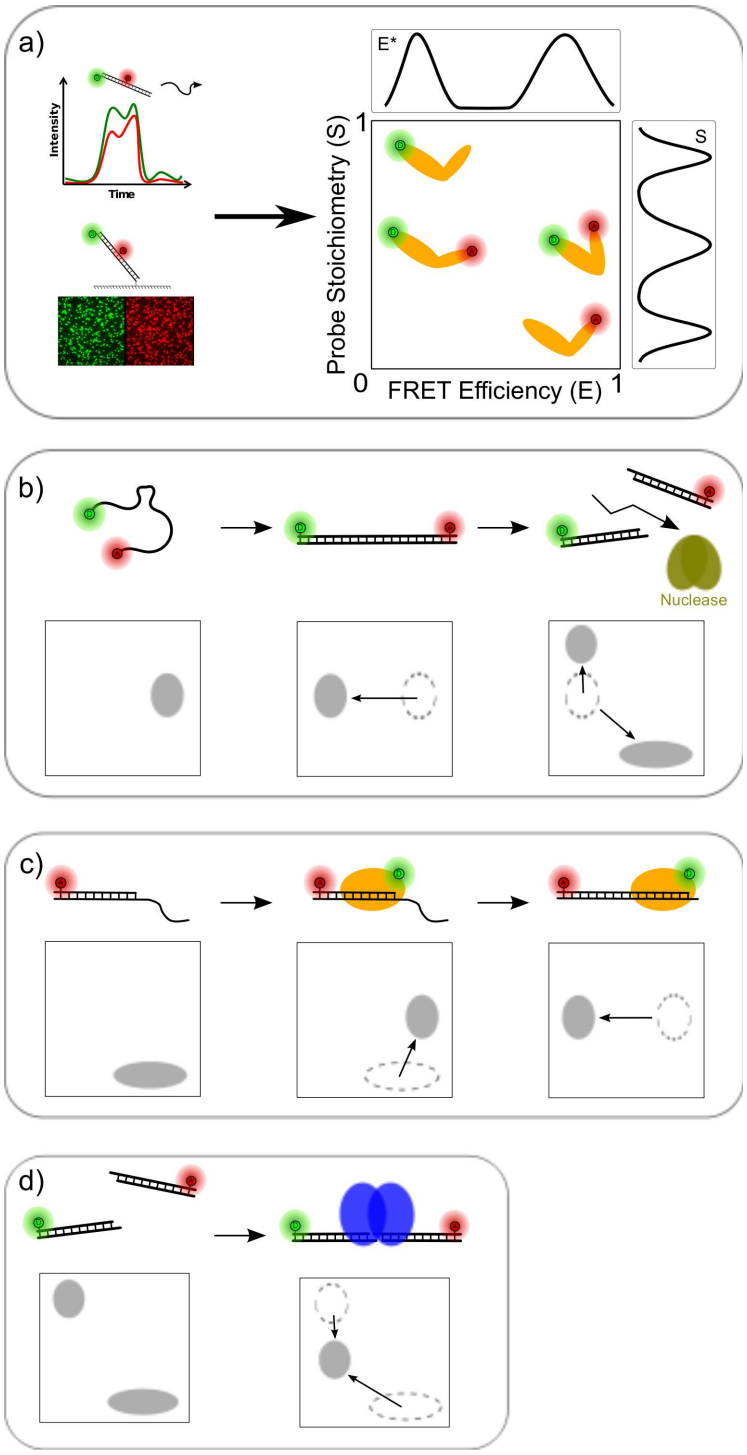


Figure 2.5.1: Alternating Laser Excitation Spectroscopy (ALEX). a) Fluorescence from individual molecules in the form of bursts (confocal) or images (TIRF) is analysed, and quantified in terms of relative probe stoichiometry (*S*) and FRET efficiency (*E*). Observing many molecules allows clusters to form on an ES histogram clearly identifying molecular subpopulations. Shown are the approximate positions of donor only (high *S*), acceptor only (low *S*), and donor-acceptor (mid *S*, low and high *E*) molecular populations. b)-d) Applications of ALEX showing molecule schematic (top) and ES histogram predictions (bottom). b) Left: Donor-acceptor labelled ssDNA resulting in a high *E* population due to secondary structure. Middle: Addition of complementary DNA strand extends DNA, shifting the population from high to low *E*. Right: Addition of a nuclease degrades DNA creating donor-only and acceptor-only molecules removing the low *E*, mid *S* population. c) Left: Acceptor labelled dsDNA containing a single-stranded overhang shows an acceptor-only population (low *S*). Middle: Donor labelled DNA polymerase binding results in high *E* values. Right: translocation of the polymerase along DNA (given NTPs), causes a change from high to low *E*. d) Left: Freely diffusing donor and acceptor labelled DNA 'half-sites' containing the binding site for a particular protein and complementary overhang regions. Donor-only and acceptor-only bursts (low and high *S*) should be observed. Right: the addition of the particular protein causes association of the two half-sites, observable through a low *E*, mid *S* population.

2.6 Experimental Setup

Single-molecule fluorescence spectroscopy was performed on a customised inverted Olympus IX-71 microscope, the schematic layout of which is shown in Figure 2.6.1.

For green / red fluorescence measurements, the microscope was equipped with two lasers, a 635 nm diode laser (Cube; Coherent, Santa Clara, CA) and a 532 nm Nd:YAG laser (Samba; Cobolt AB, Solna, Sweden). For ALEX experiments, the individual laser beams were modulated either directly through a TTL pulse, or by using an acousto-optical modulator (AOM; Isomet, Springfield, VA), and combined into a single-mode optical fibre (Thorlabs, Newton, NJ) using dichroic DC1 (500 DRLP, Chroma). Alternatively widefield illumination could be performed using a mercury lamp for fluorescence at shorter wavelengths or with a white light lamp for brightfield imaging.

Confocal Spectroscopy

To perform confocal measurements, green and red alternating laser excitation (ALEX) was used. In this case, the collimated excitation light enters the oil immersion objective (100x, N.A 1.4; Olympus) (Figure 2.6.1) . Light is focused by the objective to a diffraction-limited spot in the solution above the coverslip. Fluorescence emission is collected by the same objective and separated from excitation light using dichroic DC2 (FF545/650-Di01; Semrock). Mirror MR2 is moved into

the emission path using the translation stage (MS-2000, XYZ Microscope stage; Thorlabs) such that the emission light is directed towards the APDs (SPCM-AQR14; Perkin Elmer). Emission is spatially filtered by focusing onto a 100μm pinhole to reject out of focus light in the axial direction. The light is then split into green and red using dichroic DC4 (630DRLP; Omega Optical) and filters FT2 & 3 (HQ585/F70 and HQ650LP, respectively; Chroma) and focused to each APD to detect incoming photons and their arrival times. Photon arrival times were recorded with a PC counting board (PCI-6602; National Instruments), and data processing was performed using custom-written software written in LabVIEW (National Instruments). Typical excitation powers (continuous wave) for confocal spectroscopy were 180μW for green and 60μW for red.

Total Internal Reflection Microscopy (TIRF)

For TIRF, green and red ALEX is also used. In this case, the TIRF lens was placed in the excitation path (Figure 2.6.1) such that excitation light was focused to the edge of the back focal plane of the oil immersion objective (the same as used for confocal measurements). Aligning the incoming light such that it approached the coverslip slightly beyond the critical angle resulted in an exponentially decaying excitation field with a depth of ~100 nm above the coverslip. Fluorescence emission for molecules attached to the surface were then collected through the same objective and separated from excitation light using dichroic DC2. Mirror MR2 is moved out of the emission path such that light is directed towards the emCCD camera (iXon+, BI-887; ANDOR). Emission light is passed through an aperture to shape the light appropriately for the emCCD sensor. Light is spectrally filtered to remove remaining excitation light through a long-pass filter, FT4 (HQ545LP; Chroma) and a notch filter, FT5 (NF02-633S; Semrock). Green and red emission were separated by dichroic DC5 (630DRLP, Omega Optical), and directed to the two halves of the chip of an emCCD camera. Movies were recorded using manufacturer's software and were further processed using custom-written MATLAB (Mathworks) software for image alignment and analysis. Typical excitation powers (continuous wave) for TIRF were 2mW for green and 1mW for red.

Widefield Imaging

For widefield imaging e.g. of bacterial cells, green or red excitation from the lasers could also be used. Fluorescence images for different wavelengths e.g. in the blue region could be obtained

using a mercury lamp (U-LH100HG; Olympus), filter FT1 (D440/7; Chroma) and dichroic DC3 (470dcxru; Chroma). Alternatively brightfield images could be acquired using a white light lamp (IX2-ILL100; Olympus) and condenser (IX2-LWUCD; Olympus) attached to the microscope. Widefield imaging was performed without the confocal lens and MR2 moved out of the emission path (Figure 2.6.1) to project light onto the emCCD camera. Movies or images were recorded using the manufacturer's software and then analysed using custom-written MATLAB software.

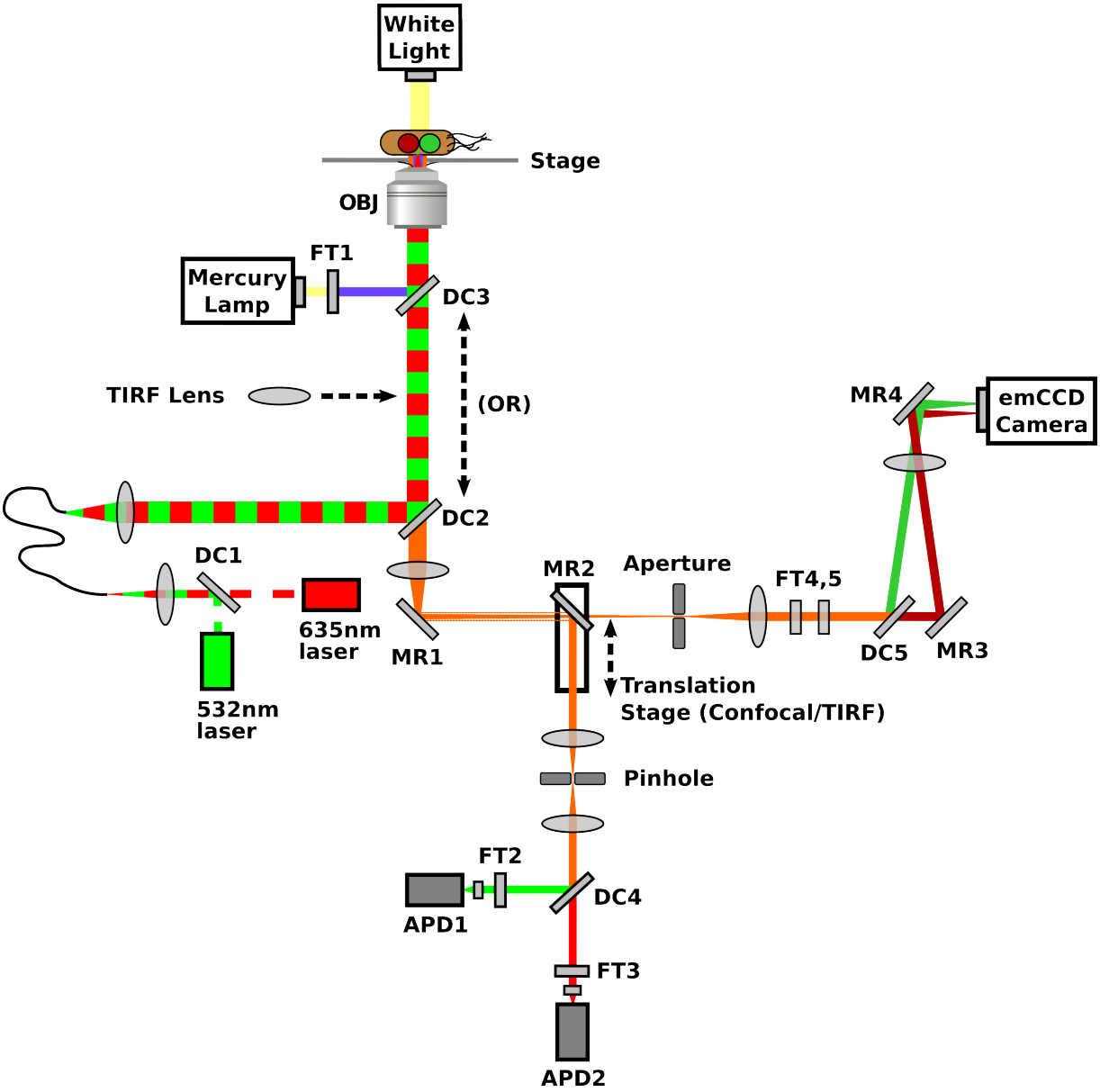


Figure 2.6.1: Schematic of complete ALEX setup incorporating confocal, TIRF and widefield paths. Guide: FT - Filter, APD - Avalanche Photodiode, DC - Dichroic, MR - Mirror, OBJ - Objective. Alternating excitation from the green (532 nm) and red (635 nm) laser sources are coupled, using DC1, into a single mode optical fibre and collimated via a lens into the oil immersion objective via DC2. *Confocal imaging path*: Fluorescence emission collected from OBJ is then coupled into the APDs by moving MR2 into the emission path using the translation stage. Emission is focused on a 100μm pinhole to reject out of focus light. Emission is then split into red/green light for each APD by DC4 and FT2 & 3. *TIRF/Widefield path*: The TIRF lens is placed into the excitation path to focus excitation light at the edge of the back focal plane of OBJ creating an evanescent field above the coverslip. Emission light is then collected by OBJ and directed towards the emCCD camera - MR2 is moved out of the emission path by the translation stage. The light is passed through an aperture to shape the light that will finally reach the camera. Light is then spectrally separated using FT4 & 5 and DC5 into red and green fluorescence and directed towards each half of the emCCD camera sensor. *Widefield illumination*: Without the confocal lens and MR2 moved out of the emission path, widefield illumination using the green/red lasers, the mercury lamp (for different wavelengths), or white light lamp (brightfield) could be performed, imaging on the emCCD camera.

Chapter 3

Transcription Factor Detection: Stoichiometry Approach

3.1 Introduction

In this chapter I explore the use of single-molecule detection to probe the presence and concentration of proteins known as Transcription Factors (TFs) using the observable of probe stoichiometry (S).

Transcription factors (TFs) are sequence-specific DNA-binding proteins [3] that control much of gene expression. TFs are natural biosensors and switches, translating chemical and physical signals (temperature shifts, light exposure, chemical concentrations, redox status) into transcriptional changes by modulating the binding of RNA polymerase to promoter DNA. Since changes in TF levels underlie fundamental biological processes such as DNA repair and cell-cycle progression, alterations in the levels of active TFs both lead to and indicate disease; e.g., mutations in transcription factor p53 contribute to the rapid growth of cancer cells and, due to their prevalence (p53 is mutated in $\approx$50% of all human tumors), have served as cancer biomarkers [10]. Thus, methods for sensitive detection and quantitation of TFs provide both fundamental information about gene regulation and a platform for diagnostics.

TF detection often involves gel-based assays and Western blotting; although helpful in characterizing TF-DNA interactions, these assays are tedious, expensive, qualitative, and consume large quantities of sample. Enzyme-linked immunosorbent assays (ELISAs) are more sensitive and offer

higher throughput, but they require many preparation and signal-amplification steps to detect low-abundance TFs. Amplification is also required in the proximity-based ligation assay [63], making it incompatible with TF detection in living cells and diagnostic settings that demand results in minutes.

An additional TF detection assay is based on Förster resonance energy transfer (FRET) between two double-stranded DNA (dsDNA) fragments containing fluorescently labeled single-stranded complementary overhangs ("molecular beacons") [64, 65]. In the presence of TF, the DNAs associate, resulting in donor fluorophore quenching due to FRET. This assay still requires significant amounts of sample and cannot detect low-abundance TFs; and due to the short dynamic range of FRET (1-10 nm), it also requires close proximity among the fluorophore, the quencher, and the protein-DNA interface, increasing the likelihood of steric interference with protein-DNA binding and complicating sensor design. Moreover, placing the fluorophore and the quencher on either side of the protein-binding site (usually 15-30 bp long) on DNA results in very low FRET signals for most TFs.

Here, alternating-laser excitation (ALEX) spectroscopy [27, 66] is used to detect TFs and small molecules via TF-dependent coincidence of fluorescent DNA. As in the molecular-beacon assay, our method is based on TF-driven DNA association, is rapid and requires no amplification. However, the assay can detect pM levels of TFs in small amounts of sample, and is FRET-independent, bypassing the need to optimize fluorophore position or know the structural details of TF-DNA binding; this flexibility in labelling ensures unperturbed TF-DNA binding. Using ALEX, TF and small-molecule detection is demonstrated, as well as assay multiplexing and compatibility with complex biological samples.

Transcription Factors Used

The proteins used to test our assay were two well known TFs, namely the catabolite activator protein (CAP) and the lactose repressor (LacR). Both proteins are most famously known for their roles in catabolite repression of lactose through the *lac* operon, a textbook example of gene regulation [1]. CAP however is not only involved with the lac operon, and is known as a global regulator involved with over 100 genes [67]. CAP aids recruitment of RNA polymerase (RNAP) via protein-protein interactions between CAP and the α subunit C-terminal domain of RNAP. CAP-dependent

promoters are classed according to the relative locations of their binding site(s): Class I where the binding site is located upstream of the RNAP promoter (e.g. the *lac* promoter [1]), Class II where the binding site overlaps the RNAP promoter e.g. in the *galP1* promoter [68]) , and Class III where multiple activator molecules are required, involving two or more CAP molecules or one or more CAP molecules and one or more other activators that cooperatively recruit RNAP (e.g. *ansB* promoter [69]). As CAP requires the allosteric effector cyclic AMP (cAMP) to bind to its specific DNA recognition sequence, it is also known as the Cyclic AMP Receptor Protein (CRP). Global regulation by CAP is therefore essentially enabled via the intracellular level of cAMP which is turn is synthesised by adenylate kinase encoded by the *cya* gene [70]. CAP exists as a ~45 kDa homodimer which in the presence of cAMP binds to a 22 bp inverted repeat sequence, which for the consensus sequence binding site (used in this chapter) has very high affinity with a dissciation constant, K_D, of $\approx$20 pM [71].

Lac repressor (LacR), best known as a genetic switch in the lac operon blocks transcription by RNAP of lactose utilisation genes under conditions where the preferred carbon source glucose is present [72]. If glucose is not available and lactose abundant, the cell changes to lactose metabolism which is conferred by a conformational change in LacR due to the binding of allolactose, dramatically reducing its affinity for its recognition site. LacR exists as a ~154 kDa homotetramer that under repression conditions binds tightly to its recognition site of ~21 bp (distinct from CAP) with an affinity of ~10 pM [73]. In the work here we use a symmetrical variant of the wild-type sequence, shown to have ~10 fold higher affnity for LacR [73].

As both proteins have been so well studied and characterised, they served as excellent model proteins with which to develop our assay before moving to more complex TFs. CAP is also used throughout the thesis in Chapters 4, 5 and 6.

3.2 Assay Design

The design of the TF detection assay is shown in Figure 3.2.1 (a). Here the full DNA-binding site for a TF is split in two (as in [64]): the left half-site (H1) and the right half-site (H2). Each site contains half of the TF-binding determinants and short, complementary 3'-overhangs. H1 is labeled with a "green" fluorophore ("G") to give half-site $H1^G$, whereas H2 is labeled with a spec-

trally distinct "red" fluorophore ("R") to give $H2^R$. In the absence of TF, at DNA concentrations of ≈10-100 pM, H1 and H2 diffuse independently, and associate only transiently. In contrast, in the presence of a TF that binds to the fully assembled DNA site, H1 and H2 diffuse as a complex ($H1^G$-TF-$H2^R$; Figure 3.2.1(a), bottom). The assay is based on the assumption that the affinity of the TF for either half-site alone will be far lower than that of the full site. In the case of CAP, a titration by Heyduk et al. [65] with similar half-sites showed no CAP binding under conditions that allowed binding to a full site. It may also be possible to rationally split the half-sites to minimize any single half-site interactions with the TF of interest. Careful splitting of the binding sequence must also be taken to minimise pertubations of TF binding to the fully formed site. On assembly of the full site, there will exist two nicks in the DNA, one on each strand. For the assays used here, the nicks were placed between the main nucleotides making contact with the DNA binding domains of the TFs [74, 73]. For the CAP half-sites used by Heyduk et al., they found similar binding affinity for CAP to the half-sites as to a fully formed site [65].

TF-dependent DNA coincidence is detected using ALEX spectroscopy [27, 62] (described in section 2.5). A search for all R-labeled molecules (i.e. G-R molecules plus R-only molecules) when the two half-sites diffuse independently yields a single population on the S histogram (H2R, a R-only species; Figure 3.2.1 (b), top). In contrast, when the two half-sites diffuse as a TF-bound complex, a species with higher S appears, reporting on TF presence (Figure 3.2.1 (b), bottom).

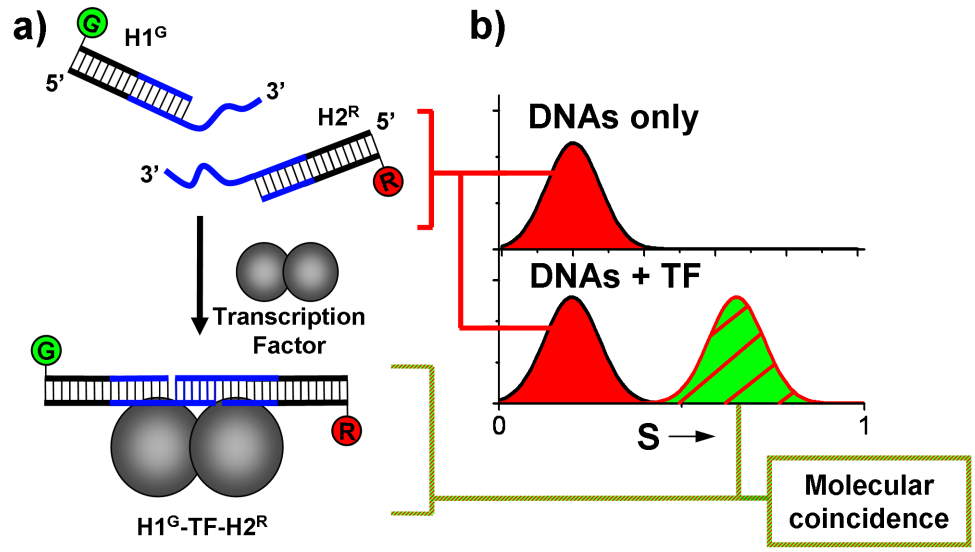


Figure 3.2.1: Stoichiometry-based TF detection. (a) Biosensor concept. In the absence of TF, two DNA half-sites, each with half of the site for TF binding and complementary 3'-overhangs, diffuse independently (top); the full site assembles transiently. When a TF binds to a full site, half-sites and TF diffuse as a complex ($H1^G$-TF-$H2^R$, bottom). (b) Schematic of an ALEX-based stoichiometry histogram. The $H1^G$-TF-$H2^R$ complex is detected as a "green-red" coincident species on ALEX histograms, distinct from free "red-only" half-sites. Reproduced with permission from [19].

3.3 Results

3.3.1 Transcription Factor Detection

ALEX was first used to detect the presence of CAP and LacR in solution (Figure 3.3.1).

Upon incubation of 100 nM of CAP-specific half-sites $H1^G$ and $H2^R$ with 200 nM CAP (including 0.2 mM cAMP) and dilution to ≈10 pM DNA, the ES histogram (Figure3.3.1 (a)) shows a high-S species with G-R stoichiometry (S ≈0.63, due to $H1^G$-CAP-$H2^R$), and a low-S species with a R-only stoichiometry (S ≈0.2, due to free $H2^R$). In contrast, without CAP, no G-R molecules are detected (Figure 3.3.1 (b)). Similar results were obtained for the same assay designed to detect lactose repressor (LacR a.ka. LacI; Figure 3.3.1 (c)-(d)).

The fraction of the observed bound complexes can be maximised by reducing the excitation laser power to minimize photobleaching; by increasing the ratio of G-labelled to R-labelled DNA half-sites from 1:1 to 3:1; and by minimising the dissociation rate of the TF-DNA complex by using KG7-LS (see A.2) for sample observation. Using half-site concentrations of 1.5 nM for $H1^G$

and 500 pM for $H2^R$ in KG7-LS, we achieved clear detection of LacR down to pM levels (Figure 3.3.1 (e)).

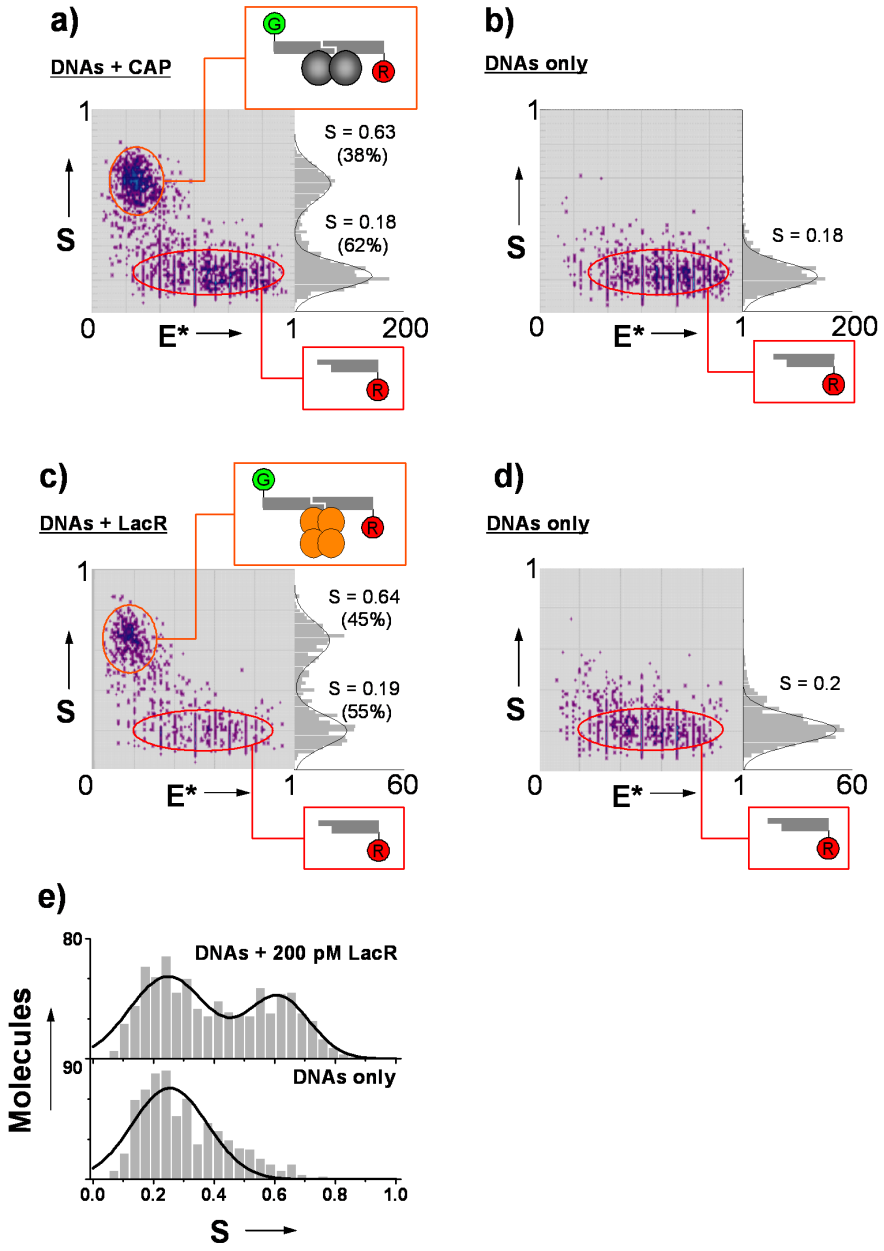


Figure 3.3.1: Use of the ALEX-based biosensing assay to detect transcription factors CAP and LacR. (a) Incubation of 200 nM CAP with 100 nM of each CAP half-site and 0.2mM cAMP gives rise to a species with S ≈0.63, corresponding to the complex of DNA half-sites with CAP, $H1^{G}$-CAP-$H2^{R}$ (orange rectangle); the second species is the free R-only half-site, $H2^{R}$ (red rectangle). (b) In the absence of CAP, only few counts due to random coincidence are present in the area previously occupied by the complex. (c-d) Similar results to a-b were obtained using 400 nM LacR with 100 nM of each LacR half-site. (e) TF detection at sub-nanomolar concentrations without amplification. Incubation of 200 pM LacR with 1.5 nM H1 and 500 pM H2 in low salt buffer (see A.2) gives rise to a new population with S ≈0.6 (top panel) when compared to the S histogram in the absence of LacR (bottom panel). Reproduced with permission from [19].

3.3.2 Transcription Factor Quantitation

To quantify the levels of transcription factor present, for the case of the CAP assay, a kinetic model was developed that describes the system using two coupled equilibria: one for half-site association, and a second for TF binding to $H1^G - H2^R$ with dissociation constants K_{D1} and K_{D2}, respectively:

$$H1^G + H2^R \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} H1^G - H2^R$$

$$H1^G - H2^R + CAP \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} H1^G - CAP - H2^R$$

By solving the related equilibrium equations, an expression is generated that relates CAP concentration ([CAP]) to the fraction F_B of half-site bound to TF (as determined via ALEX), to K_{D1}, K_{D2}, and to the total concentration H_{tot} of each half-site:

$$[CAP] = \left[H1^G\right] F_B + \frac{F_B K_{D2}\left(2\left[H1^G\right](1-F_B) + K_{D1} + \sqrt{K_{D1}}\sqrt{4\left[H1^G\right](1-F_B) + K_{D1}}\right)}{2\left[H1^G\right](1-F_B)^2}$$

Model derived by J.Torella. For a full description of the derivation see A.2.1). As mentioned in section 3.2 it is not expected that there will be any significant CAP binding to either half-site alone affecting the signal, hence the model does not account for this. If TF binding to individual half-sites was of concern, titration experiments using PAGE (as performed in [65]) or fluorescence anisotropy could be performed to determine the affinity. If significant, this could then be factored in as additional equilibria between CAP and $H1^G$ and $H2^R$ individually . However the basic premise of the assay is that TF binding to individual half-sites is weak and binding to the fully associated site, strong. Hence efforts to achieve this would be preferable to adding additional complexity to the assay. As we see later, our model does agree well with experimental results suggesting that it does describe reasonably well the relative interactions involved.

The model suggests simple ways of optimizing assay sensitivity and dynamic range. When H_{tot} is on the order of K_{D1}, half-site hybridization is efficient and K_{D2} controls CAP detection: for TFs with a low K_{D2}, the bound fraction is a linear function of [CAP], allowing CAP detection between 0 and H_{tot} (Figure 3.3.2(b), top). Conversely, by setting $H_{tot} << K_{D1}$, the bound fraction

increases hyperbolically, shifting the dynamic range and increasing sensitivity (Figure 3.3.2(a)) for lower [CAP]. Assay sensitivity and dynamic range can thus be adjusted by altering H_{tot}, K_{D1} and K_{D2} (Figure A1); altering H_{tot} is trivial, and rational changes in K_{D1} (e.g., by modifying the length of the 3'-overhangs [73, 75]) or K_{D2} (by using mutant TF-binding sites) can easily be made.

To test the ability to quantify TFs, we titrated either 1 nM or 100 nM half-site DNA with 0-400 nM CAP, diluted to ≈10 pM, and the resulting normalised DNA coincidence, F_B, was measured (Figure 3.3.2(a-b), bottom). To compare the 1 nM data with predictions (Figure 3.3.2(a), bottom), the model (Eq. A6) was supplied with H_{tot}=1 nM and K_{D1}=300 nM (from ensemble measurements; see Methods), and fitted for K_{D2}; the best-fit value for K_{D2} was ≈10 pM, in good agreement with the published value (≈20 pM [71]). For the 100 nM half-sites (Figure 3.3.2(b) bottom), the expected linear increase of F_B with CAP concentration is observed, until saturation at [CAP]≈H_{tot}, and gives an excellent agreement between our experiments and the model predictions for K_{D1} = 300 nM and K_{D2}=10 pM (red line, Figure 3.3.2(d)). Thus it should be possible, given K_{D1}, K_{D2} and H_{tot}, to quantify [CAP] without the requirement of a calibration curve.

Small molecule detection, like TF detection, can also form a useful bioassay. Since TFs act as natural sensors for small molecules (e.g., sugars, nucleotides, metals, amino acids), the assay was tested to probe the presence small molecules. As a proof of principle, the LacR sensor was used to detect the lactose analog IPTG (isopropyl β-D-1-thiogalactopyranoside) (Figure 3.3.2(c)). IPTG binds to LacR and induces a conformational change that reduces its DNA-binding affinity by 1000-fold [76, 72]. Fixed concentrations of the half-sites and LacR were incubated with 0-100μM IPTG. As expected, at low IPTG concentration, LacR leads to high DNA coincidence (Figure 3.3.2(d)). Increasing the IPTG concentration reduced DNA coincidence to the levels observed without LacR (Figure 3.3.2(d)). Analysis of the binding-inhibition curve yields a half-maximum binding-inhibition value of ≈5 μM, in excellent agreement with published values of ≈3 μM in similar buffers [77].

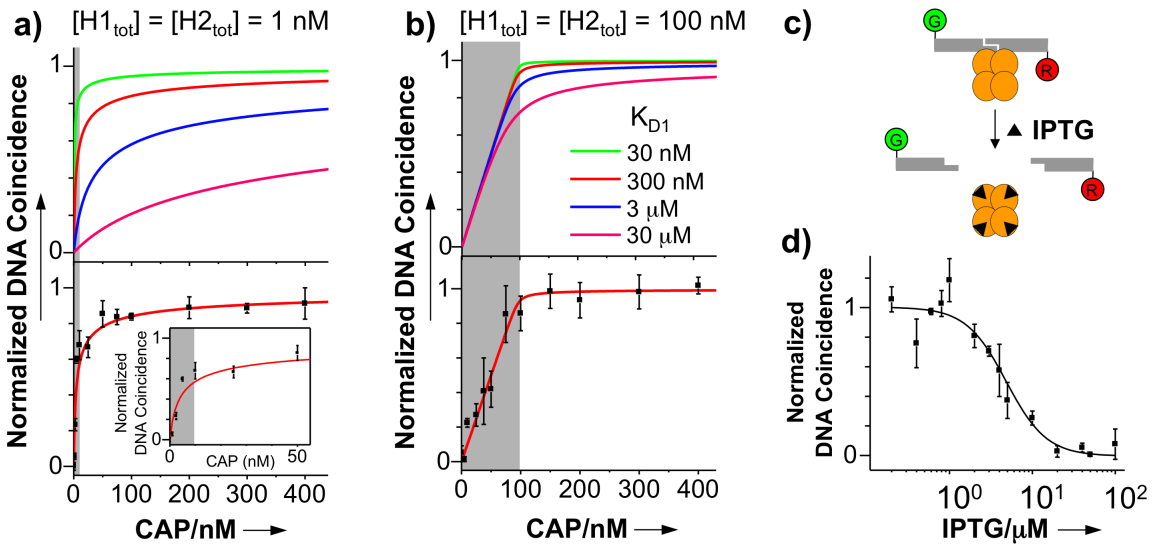


Figure 3.3.2: Quantifying TFs and their ligands. (a-b) Normalised DNA coincidence as a function of [CAP] using 1 nM half-site (a) and 100 nM half-site (b). Top: Model predictions as a function of K_{D1}. Bottom: fit of model (red curve) to ALEX data. The inset in panel (a) shows a close-up of the assay's dynamic range. Grey area: dynamic range of the measurement; errors bars: standard error of the mean. (c). Small molecule detection: sensing IPTG (triangle). At low IPTG, presence of LacR results in DNA coincidence (top). At high IPTG, LacR dissociates from DNA, reducing DNA coincidence (bottom). (d) Titration of the $H1^G$-LacR-$H2^R$ complex with IPTG shows loss of DNA coincidence at increased IPTG levels. The IPTG concentration at 50% binding inhibition is $\approx$ 5 μM. Small molecule detection experiment (c & d) performed by K. Lymperopoulos. Reproduced with permission from [19].

3.3.3 Multiplexed Transcription Factor Detection

The ability to detect multiple analytes simultaneously (multiplexing) leverages diagnostic assays. To determine whether assay multiplexing in solution was possible, CAP-specific half-sites were designed that yielded a high-S species when bound to CAP, and LacR-specific half-sites that yielded a lower-S species when bound to LacR; this strategy led to simultaneous detection of both TFs (Figure A2). The results provide support for developing assays with increased resolution between TF-bound species (e.g., S-based coding using arrays of G or R fluorophores).

We also explored use of a simple logical function (an AND gate) to detect the presence of both CAP and LacR in the same solution. Logic evaluators based on nucleic acids has been used extensively [45, 78], but protein-based evaluators (apart from naturally occurring transcriptional networks [79] have not been explored. The key DNA design element in this assay is an unlabeled DNA fragment ('bridge') that contains the H2 unlabelled CAP half-site at one end and the H1

unlabelled LacR half-site at the other (Figure 3.3.3 a)); the remaining DNAs are the complementary half-sites for CAP and LacR ($H1^G$ for CAP and $H2^R$ for LacR). In the presence of CAP and LacR, half-site coincidence is expected and indeed it is observed experimentally (Figure 3.3.3 b) bottom); no coincidence is seen in the absence of TFs (Figure 3.3.3 b) top). An 'OR' gate can also be constructed from simple rearrangement of fluorophore positions. By placing a red fluorophore in the middle of the bridge and a green fluorophore on H2 LacR (Figure 3.3.3 c)) a positive output should be observed in the presence of CAP OR LacR (or both).

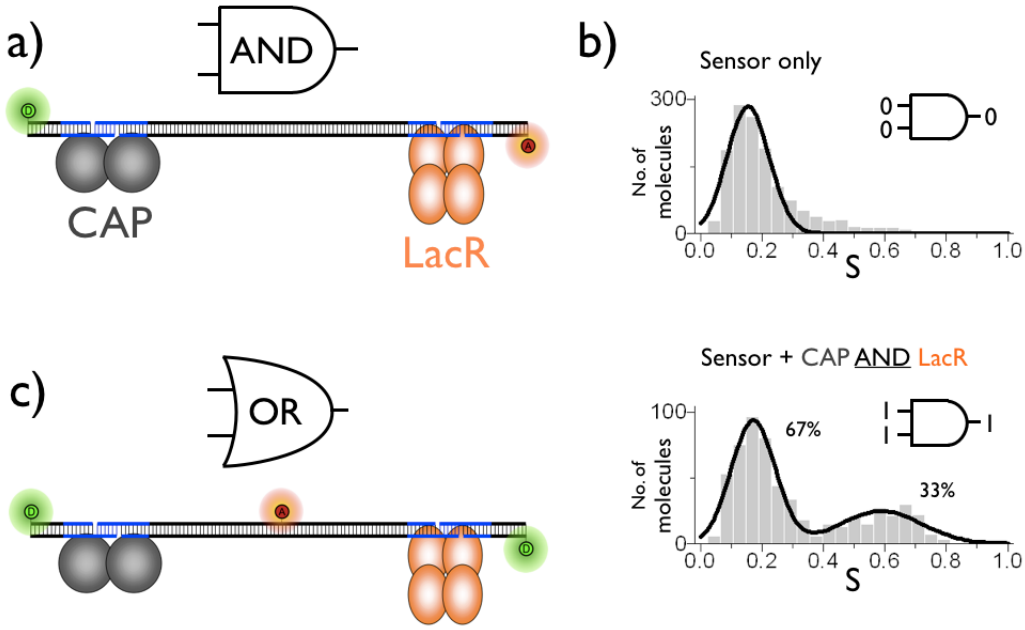


Figure 3.3.3: Multiplexed detection using 'molecular logic gates'. a) The design for an 'AND' logic gate; a positive signal (mid-high S population) will be observed only if CAP AND LacR are present. b) Experimental confirmation of AND gate operation. Without either analyte, a single low S population is observed; in the presence of CAP and LacR, a secondary higher S population is observed (c) The design for an 'OR' logic gate; a positive signal will be observed if either CAP OR LacR (or both) are present.

While solution-based multiplexing may allow simultaneous detection of multiple TFs, surface-based techniques can enable truly high-throughput assays (i.e. by using microarrays [80]) capable of detecting hundreds of TFs and small molecules. To test whether the assay was compatible with surface immobilization, we attached LacR-designed biotinylated half-site $H1^G$ on a glass surface, incubated it with LacR and $H2^R$, and imaged the surface on an emCCD camera (Figure 3.3.4(a)) using total internal reflection (TIRF) microscopy (described in 2.4.2). In the presence of LacR, the

aligned images show co-localization of G and R fluorophores (Figure 3.3.4(a)), whereas without LacR, no co-localised spots appear (Figure 3.3.4(b)). Using this method, LacR was detected down to the pM level (Fig. 3.3.4(c)). Similar results were obtained for CAP (Figure A3).

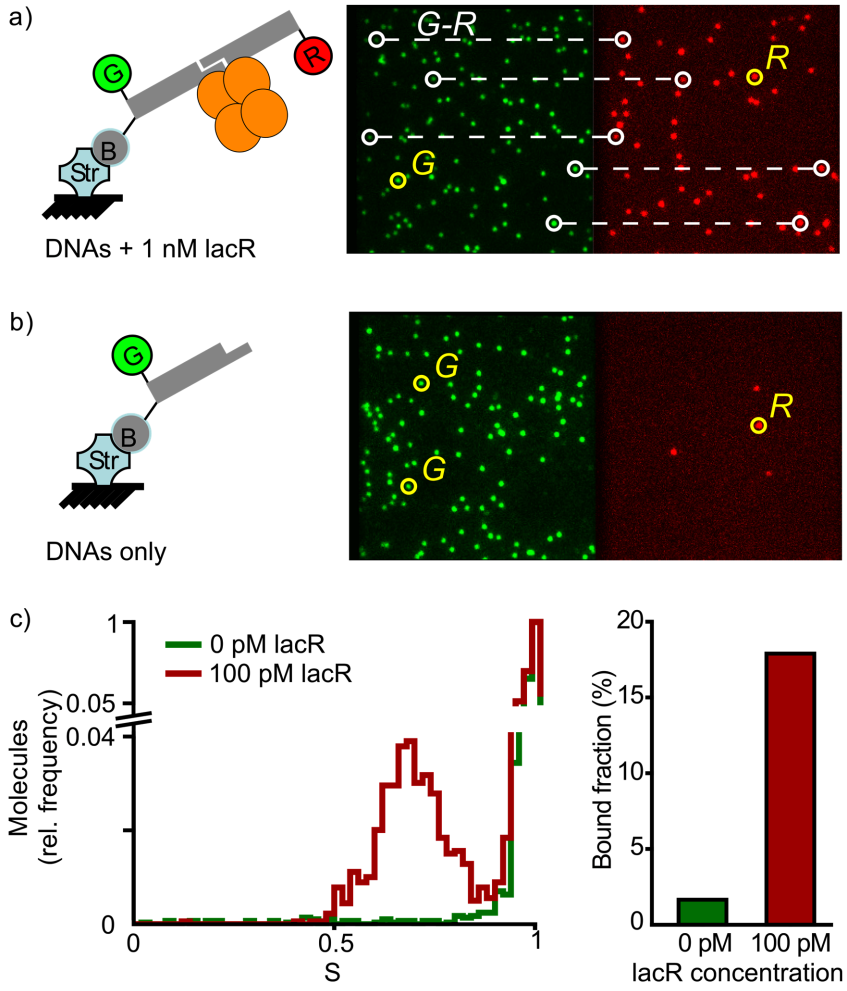


Figure 3.3.4: Enabling multiplexing. (a-b) TF detection on surfaces. (a) A biotinylated $H1^G$ was attached on streptavidin-coated glass, incubated with 1 nM LacR and 1 nM $H2^R$; after washing, the surface was imaged. In the presence of LacR, the images show diffraction-limited spots where G fluorophores (green spots) and R fluorophores (red spots) co-localize. G-only spots are due to $H1^G$-only, plus $H1^G - lacR - H2^R$ complexes with bleached R fluorophore; R-only spots are due to $H1^G - lacR - H2^R$ complexes with bleached G fluorophore, plus non-specifically bound $H2^R$ molecules. (b) Without LacR, few R-spots appear (due to non-specific surface binding of $H2^R$). (c) TF detection on the surface with high sensitivity. Left: Relative molecular frequency vs. stoichiometry in the presence and absence of 100 pM LacR. Right: Bound fraction (the fraction of G fluorophores that co-localize with R fluorophores), calculated by summing the molecular frequency in the S histogram, for S<0.87 (co-localized G and R) and S>0.87 (G-only). TIRF Experiments co-performed and analysed with S. Holden. Reproduced with permission from [19].

3.3.4 Transcription Factor Detection In Cell Lysates

For either gene-expression analysis or point-of-care diagnostics, biosensors must perform in biological fluids that contain nucleases, proteases and other molecules that may interfere with detection, producing false-positive or false-negative results. To test the robustness of the assay, we first detected TFs in human cell extracts; we added CAP-specific half-sites with or without CAP to HeLa cell nuclear extracts, incubated for 10 min, diluted to ≈50 pM DNA and performed ALEX experiments. Without CAP (Figure 3.3.5(a), top), no DNA coincidence is observed. In the presence of CAP, however, strong DNA coincidence is detected (Figure 3.3.5(a), bottom). Reaction conditions are described in A.2.

We then tested the ability of the sensors to detect changes in gene expression in *E.coli* whole-cell lysates for increasing CAP synthesis, which was induced in cells containing a plasmid carrying the CAP gene; cells were then analyzed at various points just before and up to 3 hours after induction. Using ALEX, an increase in active CAP concentration is observed from a basal level before induction, to a saturated level within ≈1 hr (Figure 3.3.5(b)). Accounting for dilutions during lysate preparation, the *active* CAP concentration at saturation corresponds to ≈300μM (≈3x10^5 molecules of active CAP per cell, or 20-30% of the total cellular protein [81]), consistent with direct quantification of total CAP via SDS-PAGE (≈40% of total protein).

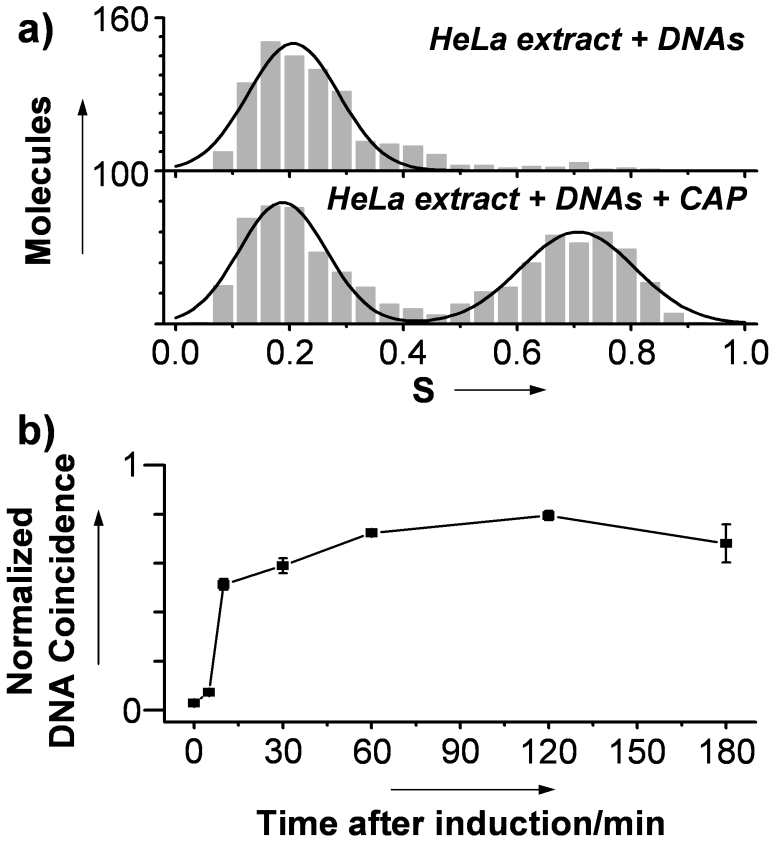


Figure 3.3.5: TF detection in biological samples. (a) Detection of 10 nM CAP in HeLa nuclear extracts. Without CAP (top), little DNA coincidence is seen; in the presence of CAP (bottom), substantial DNA coincidence is seen (species with S≈0.7). (b) Detection of gene expression in bacterial lysates. CAP expression was induced at 0 hrs, and DNA coincidence was measured over time using 100 nM half-sites; here, DNA coincidence of 0.8 corresponds to ≈80 nM active CAP. DNA coincidence was normalised relative to the maximum observable coincidence using the CAP assay. Errors bars: standard error of the mean. Reproduced with permission from [19].

3.4 Discussion

In conclusion, a versatile assay has been developed for detecting TFs and small molecules. We show rapid detection of the TFs CAP and LacR in solution, on surfaces and in cell lysates. For CAP detection we develop a model describing the kinetics of the assay allowing us to calculate CAP concentration from the bound fractions observed in our experiments. This assay should be useful for determining TF concentrations given different environmental stimuli, detecting particular TFs associated with disease, or to screen drugs that may affect TF binding (as demonstrated with IPTG detection). Detection of CAP and LacR (or other TFs) using this assay is a significant improvement

in term speed and cost compared to other techniques such as Western blots, ELISAs or PCR based methodologies such as the proximity dependent ligation assay [63].

The compatibility with surface-based detection opens exciting prospects for multiplexing, e.g., using arrays of spatially addressable half-sites. The assay is compatible with important eukaryotic TFs that possess high-to-moderate DNA affinity, such as p53 (Kd $\approx$1.6 nM [82]), NF-κB (Kd $\approx$8 pM [83]) and estrogen receptor (Kd$\approx$1.8 nM [84]. Since some of these TFs dissociate from the DNA in the 1-5 min timescale, it will be best to detect them using a TIRF format that allows rapid, parallel sampling of >1,000 molecules within seconds and accumulates enough statistics to establish the concentration and dissociation rate of complexes. The ability to detect 10 nM of a TF in nuclear extracts (corresponding to $\approx$10,000 TF molecules in a "typical" mammalian cell) compares favourably to the copy number of many disease-associated TFs (e.g. p53 ranges from 30,000-200,000 copies per cell in many cancer cell lines [85]). The robustness of our assay in cell lysates provides a starting point for TF-detection in living cells.

Since single fluorophores have been detected in bacteria [32], and as 1 nM corresponds to a single molecule in a single *E.coli* cell, our ability to detect pM levels of TFs may allow detection down to a few protein copies in single cells without amplification. Methods to introduce labelled probes as a step towards real time TF detection in bacterial cells is discussed in Chapter 5; such techniques may permit detection of low-abundance biomarkers in populations of cells where only few cells signify diseased states. Finally, combining our assay with compact or portable instruments will aid rapid, on-site diagnostics.

3.5 Contributions

Kinetic model for the half-site assay was derived by Joseph Torella. Small molecule detection and S-based multiplexing experiments (section A.2.1) performed by K Lymperopoulos. I thank Seamus Holden for introducing me to TIRF and helping with the measurements and data analysis. Support and feedback provided by Joseph Torella and Achillefs Kapanidis.

Chapter 4

Transcription Factor Detection: FRET Approach

4.1 Introduction

In Chapter 3 we presented an assay for the detection of transcription factors (TFs) using the ALEX observable: relative probe stoichiometry (S) [19]. This versatile assay demonstrated rapid, sensitive and robust detection of one or multiple analytes. In this chapter, we explore the use of the other ALEX observable, FRET (E) to detect the presence of TFs.

Binding of TFs to DNA is known to directly affect gene expression by modulating promoter binding by RNA polymerase [3]. One of the mechanisms this is achieved is through DNA bending to bring distal DNA elements into close proximity [86] for example to mediate interactions between other TFs. A large number of TFs in both prokaryotic and eukaryotic systems are known to bend DNA [86] and hence methods to detect DNA bending could serve as a sensitive observable to detect the presence of specific TFs.

DNA bending can be observed in a number of ways. Electrophoretic shift assays such as circular permutation analysis or cyclization analysis [86] can infer bend angles from shifts caused by protein binding site variations or degree of DNA cyclization by a ligase. Such assays are extremely time consuming, require large sample concentrations, and have low sensitivity. Alternatives are force measurement approaches whereby DNA bending is inferred from a paramagnetic bead attached to surface-immobilised DNA using magnetic tweezers [26]. This method requires long

DNA fragments (λ DNA) containing multiple binding sites for the proteins of interest. Thus the ability to distinguish individual DNA bending events, and bending events caused by different proteins would be technically challenging with this approach. FRET, known as the 'spectroscopic ruler' between the range of 1-10 nm [48] is an attractive solution for the detection of DNA bending and has been employed at the ensemble level [87, 88]. However, for the detection of DNA bending, FRET requires close proximity between the donor and acceptor fluorophores in either the bound or unbound case; DNA bend angles would either have to be extreme [89, 88] or would require fluorophore placement close to the protein binding site [64] or on the protein itself [87].

Here we describe the design of a novel DNA probe that can detect the presence of a particular TF via FRET from DNA bending induced at its recognition site. The sensor consists of a double-labelled dsDNA molecule with fluorophores placed far from the protein binding site. The assay can be used for ensemble and single-molecule measurements on surfaces as well as in solution. The probe was designed with the aim of TF detection in live bacterial cells (discussed in Chapter 5). Unlike the stoichiometric half site approach described in Chapter 3, this design does not require single-molecule sensitivity for TF detection; ensemble FRET *in vivo* could be probed initially before moving towards single-molecule detection. Secondly without single-stranded overhang regions, the design should be less susceptible to nucleases. No DNA probe will be completely resistant to nuclease degradation, hence methods to improve resistance are also explored.

4.2 Assay Design

The assay design is shown in Figure 4.2.1 a). Following a similar design by Stümeier et al. [90], the TF FRET based assay is a single dsDNA sensor, containing three five-adenine (A5) bulge loops and a binding site for the TF of interest. By incorporating A5 bulge loops within a dsDNA structure, it is possible to kink the DNA by its intrinsic sequence. These loops contain five unpaired adenine bases; the crystal structure of these loops indicate that the kink induced in the DNA is around 74° (+/- 11°) [91]. By incorporating several A5 loops within a single dsDNA, the two ends of the DNA are brought into close proximity, within the range of FRET. The 5' end of the 'top' strand was labelled with a donor (Cy3B) and the 3' end with a biotin tag. The 5' end of the 'bottom' strand was labelled with an acceptor (ATTO647N). Sequences used are included in section A.1.2.

In the absence of the TF, the two labelled ends of the sensor are in close proximity approximately at the level of the characteristic Förster distance (R_0) for the donor-acceptor pair used[1] resulting in an intermediate FRET efficiency (Fig 4.2.1 a) of ~ 50%. In the presence of the TF, the protein binds the DNA along one edge and induces DNA bending pulling the ends of the sensor apart by ~ 15Å [92], decreasing the observable FRET efficiency (Figure 4.2.1 b)) to ~ 24 %. The A5 bulges incorporated into the probe enables amplification of bend angles produced by the TF that would be otherwise unobservable using linear dsDNA. The design also holds promise for multiplexed detection of several DNA-bending TFs. For example, the same sensor design incorporating different recognition sequences for TFs with different bend angles would produce distinct FRET populations. Alternatively, the edge lengths of the sensor could be altered to create distinct FRET populations for TFs with similar bend angles.

The TF target selected for this assay was catabolite activator protein (CAP, as used in section 3.3.1). As CAP is well known to significantly bend DNA [87, 93] by approximately 80˚, it served as an ideal candidate for TF sensing via DNA bending. In this case the 'bending sensor' was designed to include the wild-type recognition sequence for CAP as compared with the consensus site used in Chapter 3. CAP has an approximately 450 fold *lower* affinity for the wild-type binding site than for the consensus site [71] with a dissociation constant of ~11 nM compared with ~24 pM for the consensus. This was used with a view to later internalisation within *E.coli* cells; the consensus site, due to native CAP concentrations [94], would be permanently bound making changes in CAP concentration impossible to detect.

[1] ~62Å , Ludovic Le Reste, personal communcation

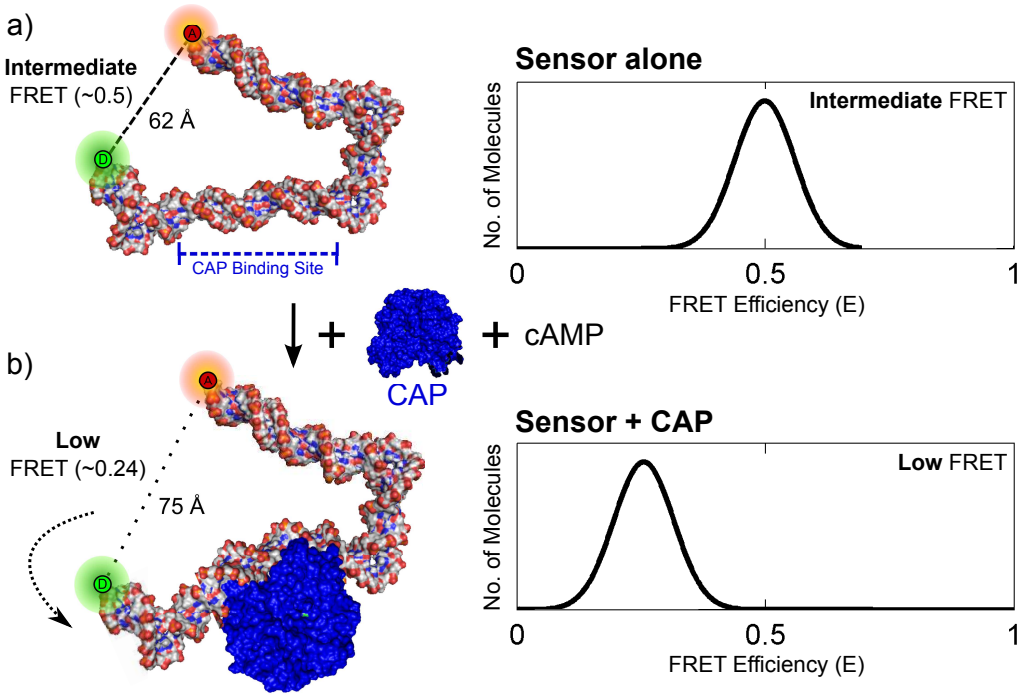


Figure 4.2.1: FRET-based TF detection using a 'bending' sensor. (a) Left - the design of the sensor is a single dsDNA containing three A5 bulge loops with the TF binding site (in this case CAP) placed along one edge. A donor (Cy3B) is placed at one end of the sensor and an acceptor (A647N) at the other. Right - the proximity of the fluorophores in the unbound state at approximately the Förster distance (R_0) results in an intermediate FRET population. (b) Left - in the presence of the CAP (and cAMP, required for CAP-DNA binding), the protein binds to and bends the sensor at its binding site thus pulling the two fluorophores apart. Right - the increased distance between the donor and acceptor results in low FRET population. Reproduced with permission from [95].

4.3 Results

4.3.1 FRET-based TF Detection In Solution

We initially confirmed qualitatively successful CAP-sensor binding using gel-shift assays and bending using ensemble FRET (data not shown). To quantitate binding affinity we used ensemble FRET (measuring the decrease in ensemble FRET) and fluorescence anisotropy (measuring the decrease in rotational freedom [49]). In this experiment we titrated 0-3 μM of CAP in the presence of 0.2 mM cAMP (required by CAP to bind its target sequence) with 10 nM of sensor in binding buffer (KG7-LS; section A.2.1). Using FRET, the dissociation constant (K_D) of the interaction was ~790 nM and with anisotropy, the K_D was ~600 nM (section A.2.2). These values were in close

agreement with the value derived using FRET by Stühmeier [90] of the construct with the same binding sequence and kinks (~600 nM). However this value differs from the K_D for CAP binding to a 40 bp wild-type sequence on linear DNA which resulted in a value of ~10 nM [71]. In our case the length of the edge used for CAP binding is 24 bp. It has previously been found that CAP has maximum affinity for consensus sequences of a length at least 28-30 bp [96]. Gunasekera et al. [96] found that a 24 bp segment resulted in ~100 fold lower affinity than that of a 30 bp segment for the consensus site. Hence the lower affinity in our case seems likely due to sequence length rather than any influence from the DNA kinks.

We then probed detection of CAP using confocal single-molecule alternating laser excitation (ALEX, section 2.5). In this case we observed either 300 pM bending sensor alone or incubated with 3.8 μM of CAP (to maintain a high bound fraction) for 10 min at 14°C. Complexes were then diluted to ~10 pM sensor and placed on a coverslip on the setup. Using a red-only burst search and filtering for double labelled molecules ($S > 0.45$), the unbound sensor shows a dominant FRET population at ~ 0.47. Incubation of CAP with the sensor causes a lower FRET population ~ 0.32 demonstrating the protein dependent change in proximity between the donor and acceptor fluorophore due to DNA bending. The FRET values and relative FRET change are in agreement with those predicted from previous work (section 4.2). Discrepancies are not entirely unexpected due to errors associated with Förster radius (R_0, equation 2.3.5), measurement errors in both cases and the fact that the results reported here are uncorrected for factors such as leakage and direct excitation (section 2.5).

Protein binding induced changes in fluorescence of the dyes on the construct may also play a small role in the FRET values observed. This effect is expected to be minimal due to the placement of the dyes far from the protein binding site. Such changes would be reflected by different stoichiometry values of the populations observed, which were found to lie at very similar values (~0.6, data not shown) for both the bound and unbound case. Control experiments either at the single-molecule or ensemble molecule using singly-labelled constructs could determine relative levels of fluorescence enhancement or quenching upon CAP binding if deemed necessary. However the clear FRET shift indicates its utility as the observable for CAP detection.

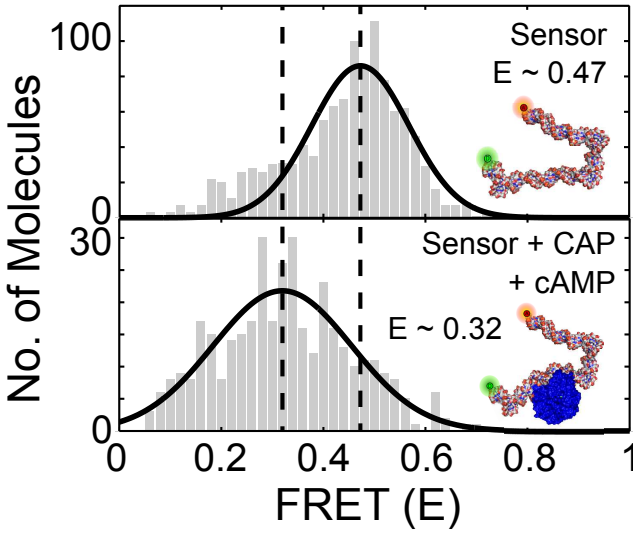


Figure 4.3.1: Single-molecule confocal detection of CAP using DNA bending. Data was collapsed onto a 1-dimensional histogram using a red burst search and filter for double labelled molecules (S > 0.45). Top: The bending sensor alone (~10 pM) shows an FRET efficiency of ~0.47. Bottom: Incubation of 3.8 μM active CAP with 300 pM bending sensor and 0.2 mM cAMP and dilution to ~10 pM sensor results in a lower FRET efficiency of ~0.32 consistent with CAP binding and bending of the sensor increasing the distance between the donor and acceptor. Experiment performed by Douglas Kelly.

4.3.2 FRET-based TF Detection On A Surface

Surface based approaches for TF detection allows the parallel detection of many molecules rapidly enabling multiplexing; sensors for different targets could either be separated spatially or report different FRET values. As the bending sensor design is uni-molecular, it could also be re-used by washing off the previous analyte allowing sufficient time for dissociation of the target. The ability to directly monitor association and dissociation kinetics could be used to measure K_D on the surface to compare with solution measurements. This measurement will be included for future work.

To investigate whether the assay was compatible with surface-based approaches, the sensor was immoblized on a glass coverslide using a biotin/neutravidin/biotin linkage (section A.2.2). Using total internal reflection microscopy (TIRF; section 2.4.2), the sensors were imaged on an emCCD camera; individual molecules were then sorted on the basis of FRET and stoichiometry.

Figure 4.3.2 (a) shows TIRF results for the detection of purified CAP using the bending sensor, filtered for double-labelled molecules. Without CAP, the sensor itself results in a FRET efficiency

of ~0.45 (Fig. 4.3.2 (a) top). After addition of 540 nM active CAP to the immobilised sensors and incubation for 10 min at room temperature (RT), a significant drop in FRET is observed to ~ 0.32 in good agreement with solution based experiments. Recoverable FRET efficiency of ~0.45 was achieved by washing the sensor 3x with binding buffer (KG7-LS; section A.2.1) to allow full dissociation of CAP. The sensor could then be re-used to detect CAP (data not shown).

As the sensor was designed to detect CAP *in-vivo*, an initial step towards this goal was to test the sensor response in complex biological samples such as cell lysates. For this experiment, *E.coli* cell lysates containing highly expressed CAP were prepared (section A.2.2). We measure again the sensor itself, resulting in a FRET efficiency of ~0.44 (Figure 4.3.2 b) top). Addition of 0.17 mg/ml cell lysate to the sensor notably without additional cAMP results in a FRET efficiency of ~0.42, slightly lower but within error of the case in buffer alone (Figure 4.3.2 b) middle). It is also noted that there exists a small but non-negligible population at a higher FRET value of ~0.7. As the lysate contains a cocktail of many different proteins, there will likely be a certain degree of non-specific protein binding to the sensor which in some case may affect the distance between the donor and acceptor.

As cAMP is required by CAP to bind to its recognition sequence [70] it proved a useful tool in identifying the performance of the sensor in the lysate. Addition of 0.2mM cAMP and incubation at RT for 10 min results in a FRET decrease to ~0.32 (Figure 4.3.2 b) bottom) consistent with the purified case, highlighting the sensor's specific response to CAP. The small higher FRET population (~0.7) noted in the previous case (without additional cAMP) has now disappeared. With CAP now in its active binding form, it would seem to have outcompeted any non-specific protein interaction.

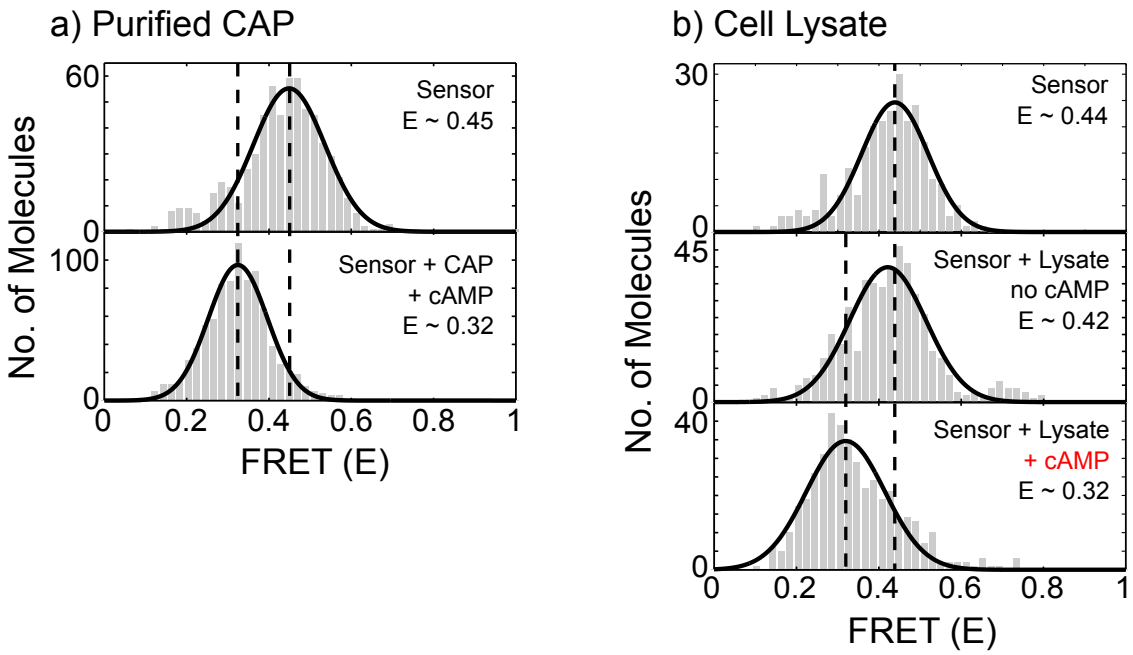


Figure 4.3.2: Single-molecule TIRF detection of CAP using DNA bending in purified form in KG7-LS buffer (section A.2.1) a) or expressed in a cell lysate. Data was collapsed onto a 1-dimensional histogram having been filtered for double labelled molecules. b). a) Top - the sensor itself in KG7-LS buffer exhibits a FRET efficiency of ~0.45. Bottom - addition of 540nM CAP, 0.2 mM cAMP and incubation for 10 min RT, the FRET decreases to ~ 0.32 consistent with confocal measurements. b) Top - the sensor itself measured again (new experiment) in KG7-LS buffer exhibits a FRET efficiency of ~ 0.44. Middle - in the presence of 0.17 mg/ml *E.coli* lysate containing overexpressed CAP but without additional cAMP produces a FRET efficiency of ~0.42. Bottom - addition of 0.2mM cAMP and incubation for 10 min at RT allows CAP to adopt a DNA-binding conformation resulting in a FRET decrease to ~0.32, demonstrating specific detection of CAP.

4.3.3 Nuclease resistance

Monitoring DNA Degradation

One of the major difficulties of using DNA probes inside live bacterial cells is nuclease mediated degradation. Due to an evolutionary 'arms race' between microbes and parasites [97], bacteria have developed a range of efficient enzymes to break down foreign DNA. Combined with nucleases intended for DNA repair [98], survival of DNA probes long enough for observation would be critical to the success of the assay *in vivo*.

Thus far, to detect TFs in cell lysates in this Chapter and in Chapter 3, the lysates were supplemented with ethylenediaminetetraacetic acid (EDTA). EDTA acts as a chelating agent for metal ions, notably Mg^{2+}, a cofactor required by cellular degradation machinery such as nucleases to

degrade exogenous DNA. Whilst possible for detection *in-vitro*, this step would be unfortunately impossible *in vivo*, as EDTA would prevent normal cell processes taking place, resulting in cell death. An alternative might be to flood the cell with other nucleic acids to sequester nuclease activity. Here, we try to limit the *activity* of nucleases. Limited nuclease activity could then be further benefitted by such sequestering methods.

To improve the lifetime of DNA sensors inside cells, genetic modifications to the *E.coli* as well as chemical modifications to the DNA itself were investigated. To probe nuclease-mediated degradation rate, we monitored loss of FRET efficiency from labelled DNA constructs over time; as the DNA is degraded, the donor and acceptor become separated, resulting in a loss of FRET. To demonstrate that FRET could be used as an observable for degradation rate, we incubated a high FRET standard (sequence given in section A.2.2) in 5 mg/ml cell lysate (*E.coli* strain DH5α) in the presence of 5 mM Mg^{2+} or 10 mM EDTA. Samples incubated with Mg^{2+} were then quenched after a certain time interval by the addition of EDTA. Samples were observed using ensemble FRET, and samples incubated with Mg^{2+} were analysed using denaturing PAGE (Figure 4.3.3). From the ensemble FRET analysis (Figure 4.3.3 top), we see complete loss of FRET over the course of 60 min for the Mg^{2+} case (solid line). Incubation with EDTA shows no FRET loss over the course of the experiment (dashed line). Analysis using denaturing PAGE of the Mg^{2+} incubated samples shows that even after 5 minutes, the distinct band at time 0 corresponding to intact dsDNA had degraded partially, giving rise to smaller DNA fragments with faster mobility rates. Over the course of the experiment this distinct band is lost and replaced primarily with a fuzzy band which is likely due to accumulation of free dye or dye conjugated to a single nucleotide. Such molecules, whilst much smaller than the original dsDNA construct also have a weaker electrostatic charge resulting in slower migration and increased diffusion within the gel, leading to the fuzzy appearance. Complete transition to free dye (or single nucleotide conjugated) occurs after 60 min, on the same timescale as the FRET experiment supporting the use of FRET as an observable for DNA degradation and highlighting the need for DNA protection from nucleases.

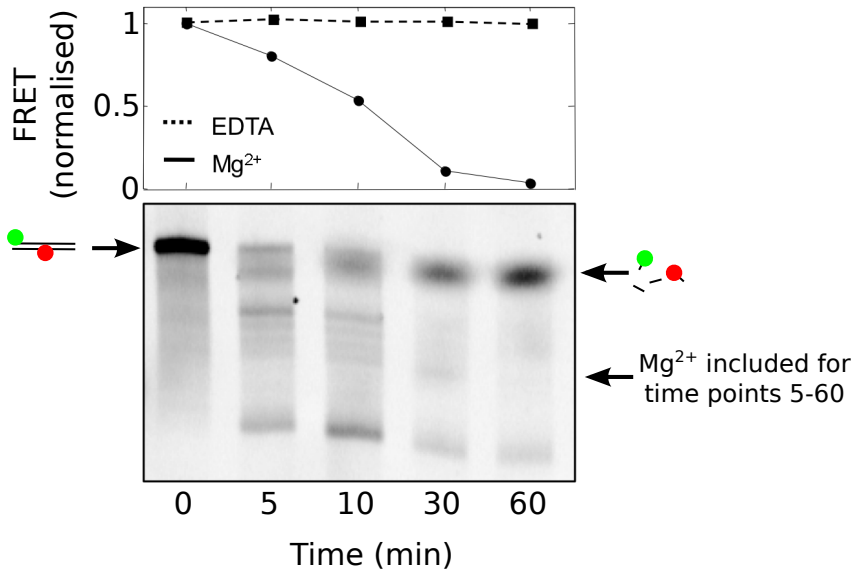


Figure 4.3.3: Nuclease mediated dsDNA degradation analysed via ensemble FRET (top) and denaturing PAGE (bottom). Top: A high-FRET dsDNA was mixed with 5 mg/ml cell lysate from *E.coli* strain DH5α and loss of FRET was monitored over time in a fluorimeter in the presence of either 10mM EDTA (dashed line) or 5mM Mg^{2+} (solid line). With Mg^{2+}, a cofactor necessary for nuclease activity, FRET is entirely lost over the course of 60 min. With EDTA, a chelator of Mg^{2+}, no FRET loss is observed. FRET was normalised to 1 for the first time point, and 0 for a time point 2 hrs after the start of the experiment at which FRET stopped decreasing. Bottom: Samples from the Mg^{2+} case were analysed using denaturing PAGE, visualised using acceptor (635nm) excitation. The 0 time point contained 10mM EDTA and shows a single band, whereas by 5 min, clear degradation is observed. By 60 min accumulation of a new band is present, which is likely due to acceptor-labelled nucleotides. Experiment performed by J.Torella

Genetic and DNA Modifications To Prevent DNA Degradation

To investigate the effect of genetic modifications to prevent DNA degradation, three different *E.coli* strains were used:

1. DH5α: Commonly used stain derived from K12 to clone plasmids. Contains a loss of function mutation in endA1, to eliminate non-specific digestion by endonuclease I. Also lacks functional recA, to avoid unwanted recombination with cloned DNA.

2. MG1655 / pKD46: Wild-type K-12 strain containing pKD46 plasmid which expresses the Gam protein from lambda phage. Gam inhibits the host RecBCD nuclease (Exonuclease V), responsible for recombinatorial repair, involving helicase, nicking and exonuclease activities.

3. Sure2 (Stratagene): Lacks restriction systems: McrA$^-$, McrCB$^-$, McrF$^-$, Mrr$^-$, HsdR$^-$. Contain loss of function mutations to endA1 (endonuclease I) and exonucleases recB and recJ (exonucleases involved in recombination)

The effect of chemical DNA modifications was investigated using the following constructs:

1. Linear unmodified high FRET dsDNA standard

2. Incorporations of ligated hairpins at either end of the dsDNA segment producing a 'dumb-bell' DNA with no exposed 5' or 3' ends. The circular ends of the dumbbell constructs suppresses degradation by exo- and endonucleases [99].

3. Incorporation of four phosphorothioate (PTO) linkages at either end of the dsDNA. Phosphorothioate linkages are a modification of a normal DNA backbone whereby one of the nonbridging oxygens is replaced by a sulphur slowing degradation by nucleases whilst not disrupting the structure of the DNA [100].

All sequences are given in section A.1.2.

Cell lysates for each strain were prepared (section A.2.2) and a degradation experiment was performed for each construct. Each degradation experiment was carried out at RT in 5mg/ml cell lysate, containing 10nM of fluorescently labelled dsDNA construct. Fluorescence spectra between 555-700 nm, given excitation at 525 nm, were recorded at regular intervals over the course of ~ 6 hrs. Fluorescence was background-subtracted and detector-efficiency corrected (section A.2.2); FRET calculated according to equation 2.3.6.

Figure 4.3.4 a) shows a schematic of each contruct and below each in b), shows the loss of FRET over time in each of the three lysates as indicated on the left. As the degradation behaviour varied in each experiment, it was difficult to apply a single model to the rate of FRET loss for all cases. For example, in DH5α lysate, the linear DNA construct could be approximated as a single exponential decay, whereas the linear PTO construct may be bi-exponential. This is not entirely surprising as the cell lysate contains a complex mixture of DNA processing enzymes. In order to compare degradation results we find the half-life for each construct. We define the half-life as the time point at which the construct degraded to half its starting value above donor-leakage (donor emission into acceptor channel) . This leakage contribution was measured by calculating the FRET

observed under the same conditions using a Cy3B-only labelled ssDNA. This value should be the FRET efficiency observed if the sensor has completely degraded. In practice there will also be a small contribution to FRET from direct excitation of the acceptor at the donor wavelength, however we omit this in our analysis. The half-lives are shown in each degradation graph and indicated with magenta lines in Figure 4.3.4 b).

From the data in Figure 4.3.4 b), it can be seen that both genetic and DNA alterations help to decrease degradation. An interesting result was that despite the large number of deletions and mutations in the Sure2 strain, it did not perform significantly better than the strain expressing Gam protein (inhibiting RecBCD). Sure2 cells are also $RecB^-$ which may suggest RecBCD as the dominant enzyme in DNA degradation. The other factor to consider here is that both of these strains contain active recA, which by binding to the sensors may protect them from degradation. The lac of recB systems and active recA in these strains may in part explain the longer survival times.

While genetic alterations to the *E.coli* strain decreased DNA degradation, such modifications can be detrimental to overall cell fitness; cells may not entirely represent wild-type behaviour. To obtain a general protection strategy for DNA probes for use in a range of cell strains, protection would have to be mediated through modifications on the probes themselves. Compared to the linear dsDNA standard in DH5α lysate, both dumbbell and PTO incorporating constructs had ~2-10x longer half-lives. Both modifications hinder access by exonucleases, suggesting exonuclease activity may play the larger role in degradation than endonucleases. We also noticed, mainly with the dumbbell construct, a potential synergistic effect of genetic and DNA modifications in hindering degradation. In Gam protein expressing compared to DH5α lysate the linear construct had ~2x longer half-life whereas the dumbbell had ~7x longer half-life.

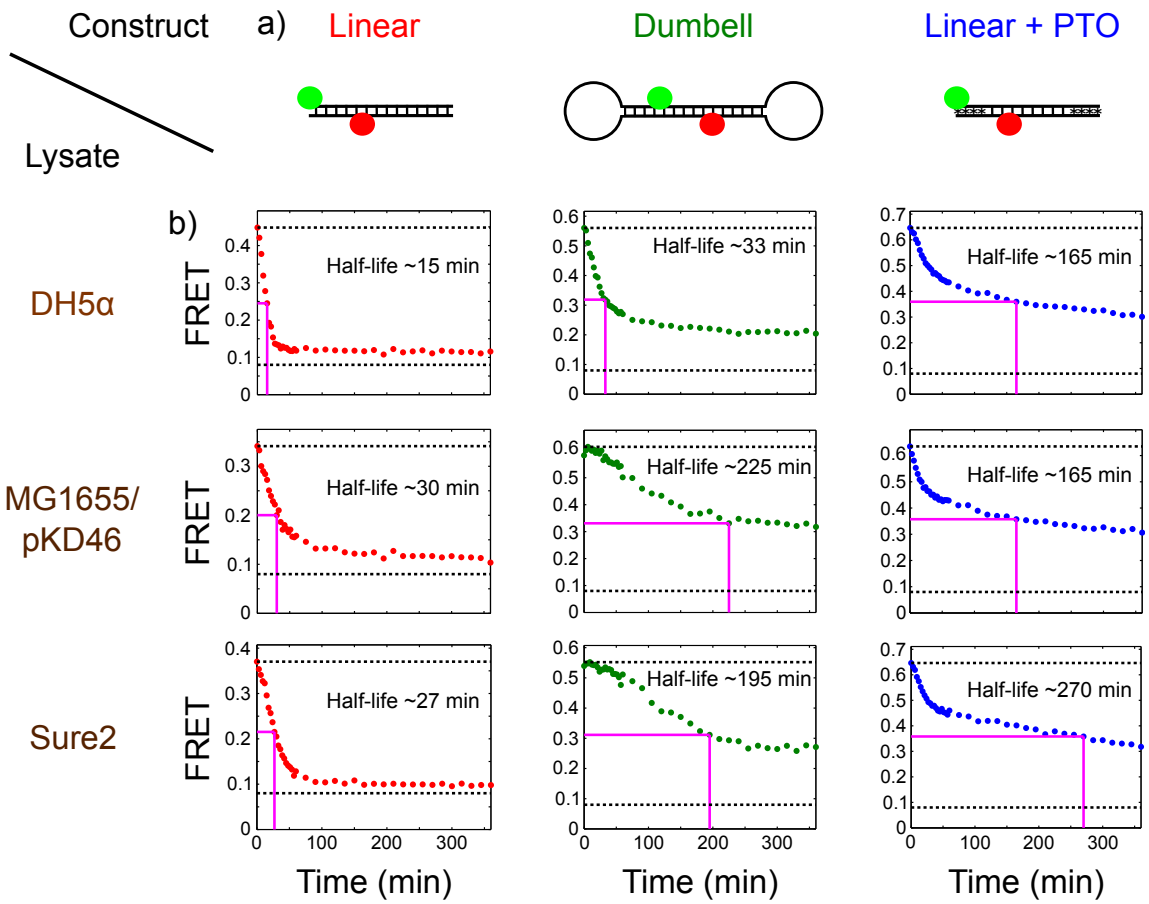


Figure 4.3.4: Degradation of 10 nM linear, dumbbell and phosphorothioate (PTO) high FRET dsDNA constructs in 5 mg/ml DH5α, MG1655/pKD46 and Sure2 strains of *E.coli* measured by ensemble FRET loss. a) Schematic of each construct and under each in: b) plot of FRET loss vs. time (min) for each construct in each lysate (indicated on the left). Each plot indicates with magenta lines the half-life of each construct. The half-life is defined as the time at which the construct degrades to half its initial FRET value above leakage. Dotted lines indicate initial FRET value (top) and leakage-based FRET (bottom) calculated from the spectra of a Cy3B-only labelled ssDNA.

Bending Sensor Protection

The DH5α lysate, containing active RecBCD and lacking active recA was considered a more difficult environment for foreign DNA to survive. Hence as the PTO linked construct in the previous experiment performed best in this environment we incorporated four PTO linkages at either end of the bending sensor and tested its degradation in DH5α lysate.

Figure 4.3.5 shows the degradation experiment, as monitored by loss of FRET of bending sensor with or without four PTO linkages in DH5α lysate under the same conditions as above. FRET was normalised to 1 for the first time point in each case and 0 for the contribution to FRET

from donor leakage. Interestingly, the unmodified bending sensor degraded with a similar half-life to the dumbbell construct in the previous experiment. The incorporation of three A5 bulges and a binding site for CAP (a fairly abundant protein [94]) may prevent nuclease processivity along the sensor. Incorporation of the PTO linkages significantly improved the survival of sensor constructs to the point that a half-life was not achieved over the course of the ~6 hr experiment. This construct was used in the next chapter for *in vivo* measurements.

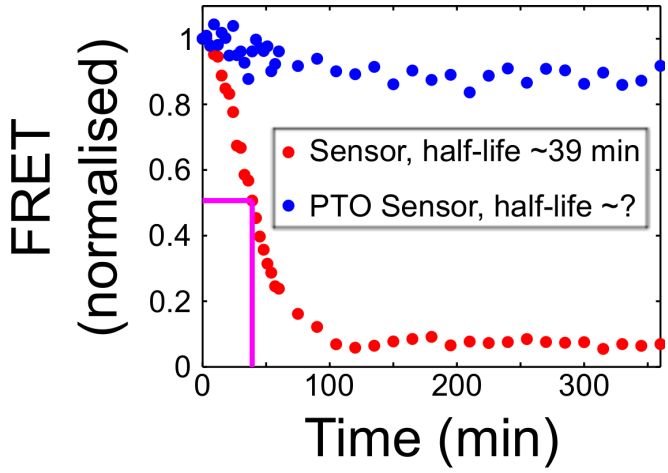


Figure 4.3.5: Degradation of 10 nM bending sensor with (blue) and without (red) PTO linkages in 5 mg/ml DH5α lysate monitored by ensemble FRET loss. FRET was normalised to 1 for the first time point of each case and 0 for the contribution to FRET from donor leakage. The half-life for the unmodified bending sensor is indicated with magenta lines. Inclusion of PTO linkages significantly improves percentage of sensors surviving degradation.

4.4 Discussion

We have designed and validated the operation of a TF sensor operating on the basis of TF-mediated DNA bending. As an increasing number of TFs have been found to induce DNA bending such as NF-κB (involved in immune response to infection) [86], we believe our assay provides a novel approach to TF-detection for gene regulation research or as a diagnostic sensor.

In our assay, a single ds DNA probe contains the binding site for the TF of interest and 3 A5 kinks bringing the donor and acceptor labelled ends into close proximity (~60 Å). The presence of the TF binding and bending is observed by a drop in FRET efficiency which can be monitored either at the ensemble or single-molecule level. We demonstrated specific detection of CAP in solution and on surfaces with either purified protein or expressed in a cell lysate. Our assay can

be performed rapidly, using very little sample volume, and without the requirement of fluorophore labelling near the binding site or on the TF itself, improving significantly on previous assays [86, 87, 26]. The design of the assay allows amplification of moderate TF bend angles improving the generality of the approach away from only TFs that induce extreme ($> 160^{\circ}$) bend angles [88]. As the kink angles induced by the A5 bulges are known from crystallographic and FRET analysis [91, 90], the use of corrected FRET [62] to calculate donor-acceptor distances would also allow accurate determination of DNA-bend angles induced by the TF.

In Chapter 3, we presented a TF sensor based on the single-molecule ALEX observable of relative probe stoichiometry (S). The ability to observe TFs using single-molecule FRET (E) extends our tools such that detection of several TFs could be identified simultaneously using specific S and E values. While the stoichiometry design did not require TF DNA-bending, it required single-molecule sensitivity for detection. With one of the goals being detection TFs in live *E.coli* cells, the bending sensor had the advantage that it could be used at the ensemble level before progressing to single-molecule experiments.

One of the hurdles of working *in vivo* was bypassing the cell's natural defence mechanism of degrading foreign DNA by nucleases. To optimise probe survival in the presence of nucleases, we investigated different genetic *E.coli* strains and chemical modifications to the DNA probes themselves. Both interventions proved successful at limiting DNA degradation. However, as a general approach to improve survival, we focused on chemical modifications to the DNA probes. It was found that incorporation of phosphorothioate linkages (PTOs) at either end of a DNA probe was simple but effective; incorporation of PTO linkages at either end of the TF bending sensor significantly improved the percentage of constructs surviving degradation. Other chemical DNA modifications not investigated here such as peptide or locked nucleic acids [101, 102] may yield similar benefits.

Towards the goal of real-time TF sensing *in vivo*, we now discuss in the next chapter the delivery of such DNA probes into live *E.coli* cells.

4.5 Contributions

Douglas Kelly performed initial experiments to check correct operation of the bending sensor. I thank Joseph Torella for contributions towards the nuclease-resistant oligonucleotide designs. Support and feedback provided by Achillefs Kapanidis.

Chapter 5

Toward Transcription Factor Detection *in vivo*

5.1 Introduction

In this chapter I describe methods to internalise and observe fluorescent molecules in live *E.coli* cells at the ensemble and single-molecule (sm) levels. I then describe the use of such methods to internalise the bending sensor developed in Chapter 4 as a step towards transcription factor detection in-vivo.

I discussed in Chapters 3 and 4 the crucial role transcription factors (TFs) play in gene regulation. As a significant amount of gene regulation takes place at the level of transcription, changes in active cellular TF concentrations (those in a state for DNA binding) can control fundamental biological processes such as differentiation and response to environmental stimuli; indeed, the delivery or expression of certain TFs have recently been used to revert fully differentiated cells back into stem cells through 'induced pluripotency' [5, 6]. TF concentrations and their activity *in vivo* must therefore be accurately controlled for each cell to perform its various functions. If control is lost in a cell, it can lead to a number of debilitating or even fatal diseases; loss of the correct function of the TF p53 is observed in the majority of human tumours [4]. Quantitation of the level and activity of TFs can therefore not only provide essential information on regulatory process occurring within the cell but also serve as important markers indicative of disease [10]. While *in vitro* TF studies provide valuable information and precise control over experimental variables, aspects

such as macromolecular crowding, ionic strength, pH, and cellular concentrations of particular molecules (e.g. inhibitors, unknown activators or cofactors) may dramatically change the activity of TFs (or any other protein) *in-vivo* [76, 103, 104, 105]. The question we wanted to answer was: can we monitor in real-time, down to the level of single-molecules, active TF levels *in vivo* and observe their changes over time under different conditions or environmental stimuli? Such data is crucial to complement *in vitro* studies for research into gene regulation and drug discovery.

Biochemical techniques such as ELISAs[1] provide one avenue for TF quantitation. In such assays TFs from cell lysates bind to immobilised consensus oligonucleotides and are subsequently detected by antibody binding coupled to a colorimetric readout. This *in vitro* approach provides a useful 'snapshot' of TF concentrations but would not be suitable for dynamic TF detection in cells over time. Similarly, newer technologies such as ChIPSeq [106], which allow genome-wide mapping of protein-DNA interactions using chromatin precipitation followed by sequencing, provide a picture of TF binding at a given point, and not variations in active TFs. ChIPSeq could be used to probe different TF binding patterns given different stimuli, but would require individual experiments for each new condition, clearly increasing the time and cost associated with the various isolation and sequencing steps. Another important point is that techniques such as these also suffer from ensemble averaging of many cells at once such that important variations between individual cells would be hidden.

A major breakthrough in observing and quantifying proteins *in vivo*, came with the discovery and development of the green fluorescent protein (GFP) [107, 34]. By genetically fusing GFP to a protein of interest (POI), this protein could be observed within the cell. Since the discovery of GFP, many fluorescent protein (FP) variants with different excitation/emission spectra, photostability and quantum yield have been developed [31]. GFP fusions have been used to measure protein diffusion coefficients, concentrations, localization patterns both at the ensemble [108] and sm level [109, 32, 34, 33].

However, as a method to quantify intracellular TF concentrations GFP tagging would record total TF levels, not levels of *active* TFs in the cell. As TF activity is often affected by binding an effector molecule resulting in a conformational change to a DNA-binding form (e.g. cyclic AMP

[1] TransAM™ Transcription Factor ELISAs, Active-motif (http://www.activemotif.com/catalog/180.html?gclid=CLK7sOg3aQCFQHS4wodg200JQ)

to catabolite activator protein [70]), it would be ideal to measure not only total concentrations but activities. While two FPs placed at either end of a protein can be used to measure conformational changes using FRET [110], this may not be a general approach to detect TF activity. Potential activity issues of the tagged POI due to the large size of FPs (~4 nm) can be a problem, as well as the inability to label internal sites.

In comparison to much smaller (<1 nm) organic fluorophores, FPs generally exhibit lower brightness and suffer from much poorer photostability [111, 31] reducing the signal-to-noise ratio and available observation time. As such sm FRET *in vivo* using FPs has yet to be achieved. Organic fluorophores can be covalently attached to polypeptide tags attached the the POI [35] but as with FPs, require genetic modification and are limited to N or C terminal labelling.

By internalising into cells fluorescent DNA 'probes' such as those described in Chapters 3 or 4, one may be able to observe directly the binding of that TF to the probe(s) either through relative probe stoichiometry, S, or FRET efficiency, E, by continuously monitoring whole cells either at the ensemble or single-molecule level. In this sense, one detects only those TFs that are active i.e. that are in conformations active for DNA binding. Thus a probe designed with the binding site for a particular TF could report on the dynamic changes in active concentration of that TF in the cell given different environmental stimuli. Such a system should also be less perturbative to the cell's real physiological state as no protein modification would be required. The main perturbation to the cell in our case would be the sequestering of TF binding from the chromosome due to probe binding. The question is whether this effect is significant enough to affect the cell's natural gene regulation mechanisms. As we will see later in the chapter, for the target CAP used, the natural number of molecules of CAP/cell outweigh the number of internalised probes by an order of magnitude. However, it remains to be determined how much this does actually affect the cell's physiological state. A good handle on internalisation efficiency may even allow investigation of how exogeneous DNA could affect regulatory mechanisms.

As with *in vitro* experiments, similar arguments of the advantages of sm over ensemble fluorescence (section 2.4) approach apply to *in vivo* studies. The ability to discriminate sample heterogeneity within one and between individual cells is particularly important in terms of gene regulation due to the inherent stochasticity of gene expresion within and between cells [112, 109, 37]; indeed a single TF dissociation event has even been shown to change the phenotype of a cell[38].

For mammalian cells, internalisation of organic fluorophore labelled molecules for sm fluorescence studies has been demonstrated [113, 36]. In this case molecules of interest were introduced into the cytoplasm using microinjection, a technique unfortunately not possible for bacteria due to their much smaller size. Other complications arise with internalising molecules into bacteria due to differences in their physiology compared to mammalian cells. In addition to the cytoplasmic membrane, prokaryotes contain a tough cell wall to protect it from high internal pressure [114] as well as in some cases a secondary outer membrane. These extra components act as barriers to the environment making prokaryotes more difficult to transfect with DNA [115].

Here, we develop a general method for the internalisation of fluorescent species into bacteria. I show the internalisation into *E.coli* of different fluorescent species at the ensemble level, characterising viability and internalisation efficiency and then demonstrate sm detection *in vivo*. The FRET-based TF sensor (Chapter 4) is then introduced into live *E.coli* cells as a step towards real-time active TF concentration monitoring *in vivo*. It is important to point out that the methods developed in this chapter need not be limited to active TF detection. The summary at the end of the chapter highlights some other exciting avenues of research involving the use of internalised fluorescent molecules.

5.2 Internalisation Method: Electroporation

We investigated several methods to internalise the fluorescent species including the use of detergents [116, 99], sonication [117, 118] and electroporation [119, 120]. In terms of the trade off between high internalisation efficiency and high viability, electroporation proved to be the most promising method. Thus, all experiments shown in this chapter use electroporation mediated transfection**.**

Electroporation or electropermeabilization, as the name suggests, is the process of increasing the permeability of the cell membrane by exposure to an electric field. Electroporation (EP) is a commonly used technique for high-efficiency transformation of cells [119].

EP is typically performed by discharging an RC circuit across a high resistance solution of cells. Figure 5.2.1 a) shows a general circuit diagram for an electroporator. Before electroporation, 'Switch 1' is closed and the capacitor is charged. To electroporate cells, a cuvette containing

a solution of electrocompetent cells and the molecules to be internalised (in a low ionic, high resistance medium) is placed in the electroporator (right half of the circuit diagram). 'Switch 1' is then opened and 'Switch 2' closed discharging the capacitor across the cell suspension. The capacitor discharge results in an exponential decay of electric field across the cell suspension (Fig. 5.2.1 b) with a time constant that is governed by the capacitance of the capacitor and resistance of the circuit (internal circuit resistance and cell suspension). The discharge of the electric field causes pores to form in the cell membranes, through which molecules can pass [120] (Fig. 5.2.1 c). Immediately after electroporation, the cells are transferred to a recovery medium, which allows the pores to reseal. If the pores do not reseal and grow, the cell membrane will lyse, killing the cell.

The magnitude of the applied electric field affects the efficiency of transformation, but has an inverse impact of the viability of the cells. Careful parameter selection is therefore required to ensure cells have probes internalised but also maintain viability. For the 'Bio-Rad MicroPulser' used in these experiments and *E.coli* in a 0.1 cm cuvette, we characterised internalisation efficiency and viability on the standard setting for *E.coli* ('EC1'), using a voltage of 1.8 kV with a discharge time constant of ~5 ms.

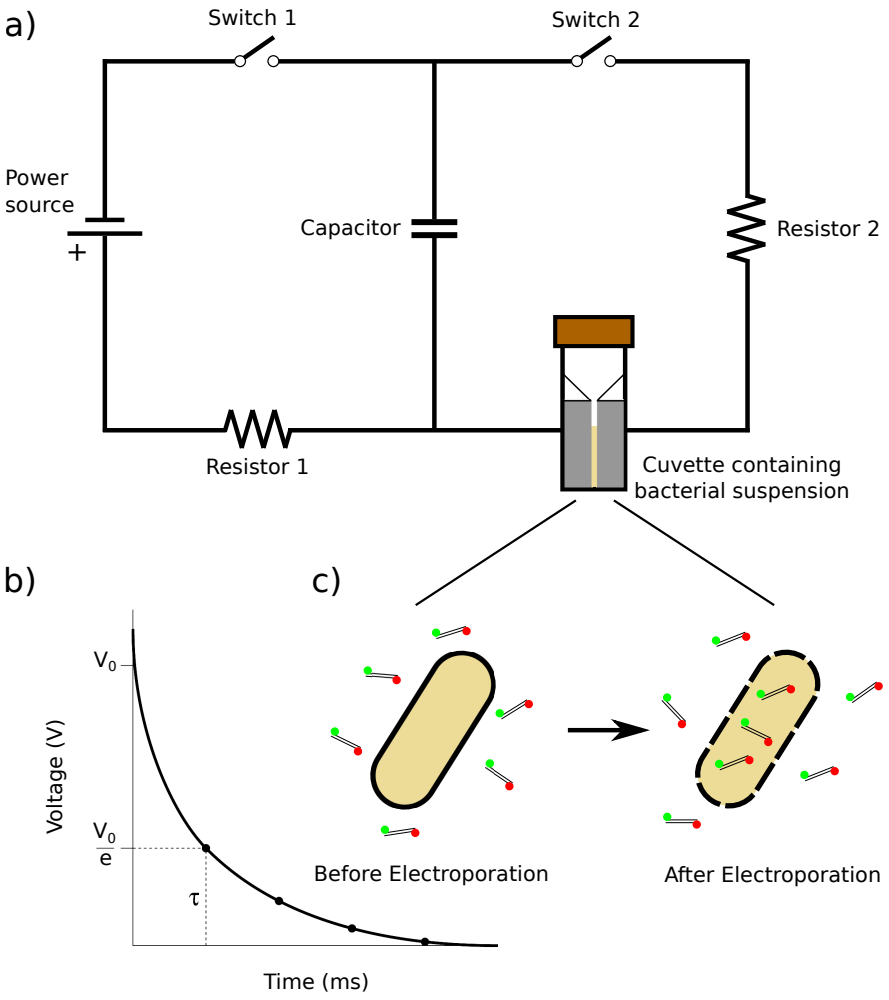


Figure 5.2.1: Electroporation theory. a) Resistor-capacitor (RC) circuit diagram illustrating the basic components of an electroporator. Initially 'Switch 1' is closed to charge the capacitor, while 'Switch 2' is left open. A cuvette containing cells prepared for electroporation (section A.2.3.1) in a high resistance medium is then placed into the right half of the circuit. On electroporation (EP), 'Switch 1' is opened and 'Switch 2' closed and the capacitor is discharged across the cell suspension. b) A typical exponential decay trace of the voltage across the cell suspension with time. τ indicates the time constant at which the voltage is 1/e times the initial voltage (V_0). For the experiments performed in this chapter, the voltage used was 1.8kV resulting in a discharge time constant of ~ 5ms. c) Schematic illustrating the state of the bacterial membrane before and after EP. Before EP the membrane(s) are intact, and no fluorescently labelled molecules can enter the cell. After EP, pores are formed in the membrane allowing internalisation of fluorescent molecules.

5.3 Results

5.3.1 Internalisation Of Fluorescent Species

5.3.1.1 Single-stranded DNA

Initial experiments to demonstrate whether fluorescent molecules can be internalised into *E.coli* cells using electroporation used single-stranded singly labelled DNA strands. In this case, 40μL of electrocompetent cells were mixed with 10 pmoles of Cy3B (green) labelled single-stranded DNA ('FRET Standard top ', section A.1.3), incubated for 10min on ice and then electroporated in a 1mm cuvette with a field strength of 18kV/cm (setting 'EC1' on a BioRad Micropulser). Cells were then recovered for 20 min at 37˚C in SOC (Super Optimal broth with Catabolite repression) media, washed 5 times in PBS (phosphate buffered saline) and immobilised for 5 min on a glass coverslip using polyethyleneimine (PEI); the cells were washed twice on the coverslip and then the coverslip was placed on the microscope. Images were then acquired using either white light (brightfield) or using green laser light (532nm) in widefield mode at a power of 200μW - a power at which cellular autofluorescence was negligible. Fluorescence images were acquired from an accumulation of 50x 0.04s exposure frames (2s overall exposure). As a control the same procedure was performed without electroporation. The results are shown in Figure 5.3.1. Detailed protocols are described in the Appendix - section A.2.

As seen from Figure 5.3.1, no fluorescence can be seen when the sample is not electroporated. In contrast, after electroporation under the same conditions, many bright fluorescent cells can be observed with fluorescence dominating within the cell body and not localised at the edges as would be the case for membrane localisation. It was also often noticed that the fluorescence was not entirely uniform. While care was taken that the sequence used for internalisation shared no major homology to the *E.coli* chromosome using BLAST[2], it is possible that the internalised DNA interacts with various processing enzymes such as nucleases; other factors such as dye properties (hydrophobicity, charge), or imaging cells in the process of division may result in non-uniform fluorescence. Experiments performed by J. Torella [121] on the same project demonstrated using Fluorescence Correlation Spectroscopy (FCS) the presence of three fluorescent species in the cell with different diffusion times. In this work it was assumed that two of the species corresponded to

[2]Basic Local Alignment Search Tool (BLAST). http://blast.ncbi.nlm.nih.gov/Blast.cgi

free Cy3B (either from a free dye component, or from strand degradation by nucleases) and intact Cy3B labelled DNA with intracellular diffusion times of ~3 ms and ~ 20 ms respectively. The third species had a much slower diffusion time of ~200 ms which could be attributed to DNA-binding proteins in the cytosol. Using FCS the average number of fluorescent molecules internalised into an *E.coli* cell using the current protocol was 106 ± 21 (for an average of 4 cells). This number includes all fluorescent species: free Cy3B, Cy3B labelled DNA and slowly diffusing Cy3B labelled DNA. To obtain a more accurate picture of intracellular molecule numbers more statistics will be needed.

Not all cells take up fluorescent DNA. EP settings and the concentration of the fluorescent DNA play a role in the internalisation efficiency (see section 5.3.2). As the DNA was labelled with Cy3B (green, donor), the excitation scheme used was continuous wave excitation at 532 nm. Leakage of Cy3B fluorescence in the acceptor channel can be seen with an average of 17.1 ± 0.9% (average of 3 cells ± 1 s.d) which is at the level expected for this dye and the filters used in the setup. Cells were picked if they had a sufficient signal-to-noise ratio (see 'Internalisation Efficiency and Viability', section 5.3.2) and lay completely in the plane of focus. Unfortunately attachment with PEI was not optimal; a certain proportion of cells would only be partially adhered to the slide, partly in and out of focus. In future cells will be immobilised using agarose pads, found (later) to produce much flatter distribution of cells.

Similar results were produced using ATTO647N labelled single stranded DNA (data not shown). In this case widefield red light (635nm) at a power of 100μW was used to excite the sample - a level at which no autofluorescence occured. Direct excitation (acceptor fluorescence given donor excitation) of these samples using 200μW green light resulted in a Dex contribution of 11.7 ± 4.3% (average of 3 cells ± 1 s.d) again within the bounds expected for this dye and laser powers used. All experiments in this chapter use a green laser power of 200μW and red of 100μW unless specified. Fluorescence (and later, FRET) values were obtained by analysing images in MATLAB by subtracting the background from a region of interest (cell) to quantify the background subtracted fluorescence of that cell in each channel (see section A.2.3).

It was noticed that after electroporation, there did seem to be a proportion of cells much longer than the expected 2μm length for *E.coli*. This may be normal, as cells may still be in the process of recovery, and have yet to split into individual cells. Extremely long cells however (>15μm), were

often found to have uptaken fluorescent molecules, but were often also found to be dead using viability assessment (section 5.3.2).

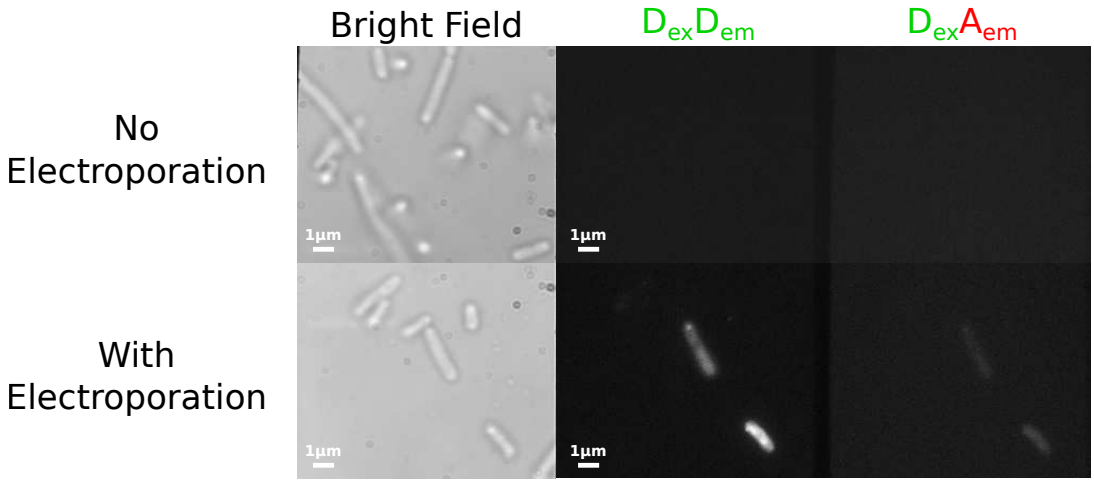


Figure 5.3.1: Electroporation of *E.coli* with fluorescent DNA. 40μL electrocompetent *E.coli* were mixed with 10 pmoles Cy3B labelled single stranded DNA without (top) or with electroporation (bottom), recovered for 20 min at 37˚C in SOC media, washed 5x and then immobilised on a coverslip. $D_{ex}D_{em}$ - Donor excitation, Donor emission, $D_{ex}A_{em}$ - Donor excitation, Acceptor emission. -EP: without electroporation, no fluorescence is observed in either channel. +EP: with electroporation, bright fluorescent cells are observed in the $D_{exc}D_{em}$ channel with some leakage occuring in the $D_{exc}A_{em}$. Donor excitation power: 200μW in widefield mode.

5.3.1.2 Double-stranded High FRET DNA

To confirm that the species internalised was in fact DNA (and not free dye), we used double stranded (ds) DNA FRET standards and measured widefield FRET efficiency in single cells. High FRET efficiency due to physical proximity of two dyes would confirm that the DNA standards were intact.

The DNA FRET standards were nuclease protected using phosphorothioate linkers (PTO) at either end of each strand to hinder degradation in the cell (see section 4.3.3). It had previously been noticed that FRET standards without nuclease protection would drop in FRET rapidly, within ~ 5min, consistent with experiments in section 4.3.3. To further minimize degradation for this experiment the same standard internalisation protocol (section A.2.3.2) was used but without recovery in SOC media. Using these steps, high FRET PTO protected DNA standards could be observed for over an hour.

DNA standards were labelled with Cy3B as the donor and ATTO647N as the acceptor. A low,

intermediate, and high FRET standard were used with donor-acceptor spacing of 33, 18, and 8 bp respectively and theoretical FRET efficiencies of ~ 3%, 52% and 99% respectively. In practice the uncorrected FRET efficiencies obtained for the three standards were ~19%, 46% and 91 % as determined using sm confocal measurements (data not shown); the difference between these values reflects the absence of correcting for experimental factors such as leakage and direct excitation [62].

Figure 5.3.2 shows brightfield and fluorescence images of cells internalised with low, intermediate and high FRET standards, using continuous wave donor excitation. From the images, the standards produce the donor/acceptor fluorescence pattern expected: the low FRET standard shows fluorescence mainly in the donor channel (Figure 5.3.2 a)), whereas the high FRET standard shows fluorescence mainly in the acceptor channel (Figure 5.3.2 c)). The intermediate FRET standard shows fluoresence in both donor and acceptor channels (Figure 5.3.2 b)). The uncorrected FRET efficiencies for the low, intermediate and high FRET standards were 21.1 ± 2%, 40.4 ± 1.2%, and 72.2 ± 4.8% (average of 11, 8 and 6 cells ± 1 s.d) respectively (measured as described in section A.2.3). The individual measurement error for the low, intermediate and high FRET samples was 0.6%, 1% and 0.7% which was calculated by re-sampling each measurement (for each cell) three times, then taking the mean of the standard deviations.

The intermediate and high FRET standards had slightly lower FRET efficiencies than expected which is likely due to ensemble averaging of mislabelled, photobleached and a certain proportion of degraded DNA standards. The FRET efficiency for the low FRET standard is around the level expected for a G-only labelled DNA, i.e. dominated by leakage and direct excitation, hence any degradation of this standard would be difficult to see.

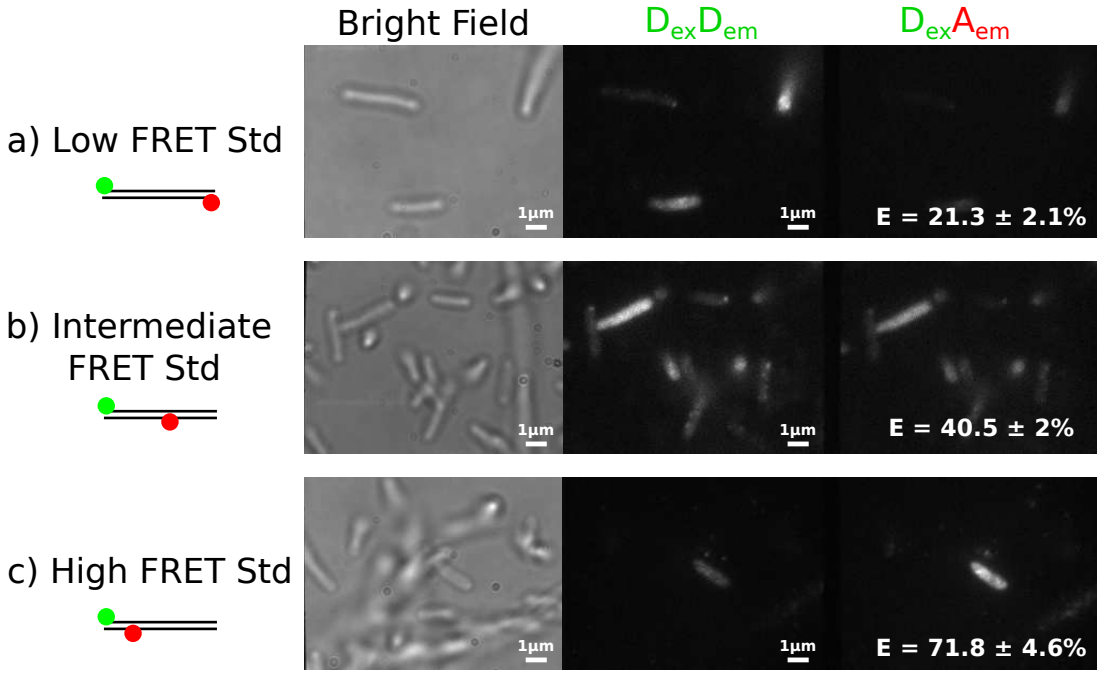


Figure 5.3.2: Internalisation of dsDNA FRET standards observed under brightfield and donor excitation. 40μL electrocompetent *E.coli* were mixed with 10 pmoles phosphorothiate linked low/intermediate/high dsDNA FRET standards, electroporated, washed 5x and then immobilised (without recovery) on a coverslip. a) The low FRET standard shows fluorescence mainly in the donor channel. Average E* = 21.3 ± 2.1%. b) The intermediate FRET standard shows similar levels of fluorescence in both donor and acceptor channels. Average E* = 40.5 ± 2% c) The high FRET standard shows fluorescence mainly in the acceptor channel. Average E* = 71.8 ± 4.6%. $D_{ex}D_{em}$ - Donor excitation, Donor emission, $D_{ex}A_{em}$ - Donor excitation, Acceptor emission. Donor excitation power: 200μW in widefield mode. All uncorrected FRET efficiencies quoted (E*) are from an average of 3 cells ± 1 standard deviation.

5.3.1.3 Fluorescent NTPs

In this experiment we internalised fluorescently labelled nucleoside tri-phosphates (NTPs) and deoxyribonucleoside tri-phosphates (dNTPs). These NTPs and dNTPs, if incorporated into newly synthesised RNA and DNA (respectively) would be of great interest, due to their ability to highlight cellular structural features such as mRNA localisation or the structure of the bacterial chromosome. Certain dyes especially are known for their ability to photoswitch, and therefore may allow super resolution imaging of such structures using new techniques such as stochastic optical reconstruction microscopy (STORM) [122].

Figure 5.3.3 shows the results of internalising different fluorescently labelled NTPs and dNTPs. The standard internalisation protocol (section A.2.3.2) is used except with 1 nmole of fluorescent

species (100 fold higher than previously) and a longer (1 hr) recovery time. It was hoped the longer incubation would decrease the ratio of a cytosolic immobile fraction of fluorescent molecules to those potentially incorporated into nucleic acid polymers in the cell. Fluorescent dNTPs incorporated into the chromosome should be essentially immobile even in living cells and therefore produce a localised spot that is much brighter than those diffusing throughout the cell, which would act to increase the background noise.

Figure 5.3.3 a) and b) shows the internalisation of Cy3 and Cy5 ATP (NU-805-CY3 & NU-805-CY5; Jena Biosciences) excited with green and red light respectively. Both NTPs were N6-(6-Amino)hexyl-ATP modified, with the dye label attached by an amine group on a linker from carbon 6 of the base. N6-(6-Amino)hexyl modified NTPs have previously been shown to incorporate into RNA [123, 124]. At the concentrations used, it was noticed that a small fraction (~ 5%) of cells took up the fluorescent NTPs without EP. Similar results were achieved using ATTO647N ATP (NU-805-647N; Jena Biosciences) - data not shown.

Fluorescence non-uniformity throughout the cells is present, but with few specific loci. ATP, often known as the unit of energy currency within the cell, is a ubiquitous molecule involved in many cellular processes including metabolic processes, signalling pathways and as a constituent of messenger RNA [1]. In combination with the cytosolic fraction of unbound fluorescent ATPs, it was unclear whether any dominant structure should be obvious. A more promising approach was to internalise Cy5 dUTP (PA55022; GE Life Sciences) shown in Figure 5.3.3 c). This dUTP is 5-Amino-propargyl-2'-deoxyuridine 5'-triphosphate with the Cy5 attached by an amine group from a linker on carbon 5 on the base. dUTP analogues have previously been shown to incorporate into the bacterial chromosome of *B.subitilis* [125].

Efforts towards attaining super-resolution images were attempted. For super-resolution images such as STORM [122], the ability to photoswitch the fluorophore was required. For the Cy5 nucleotides, preliminary results showed that a certain amount of limited photoswitching in live cells was possible before bleaching using much higher laser powers (~3.5mW). Cells could be also be fixed and permeabilised to allow the use of photoswitching buffer enabling much longer photoswitching times.

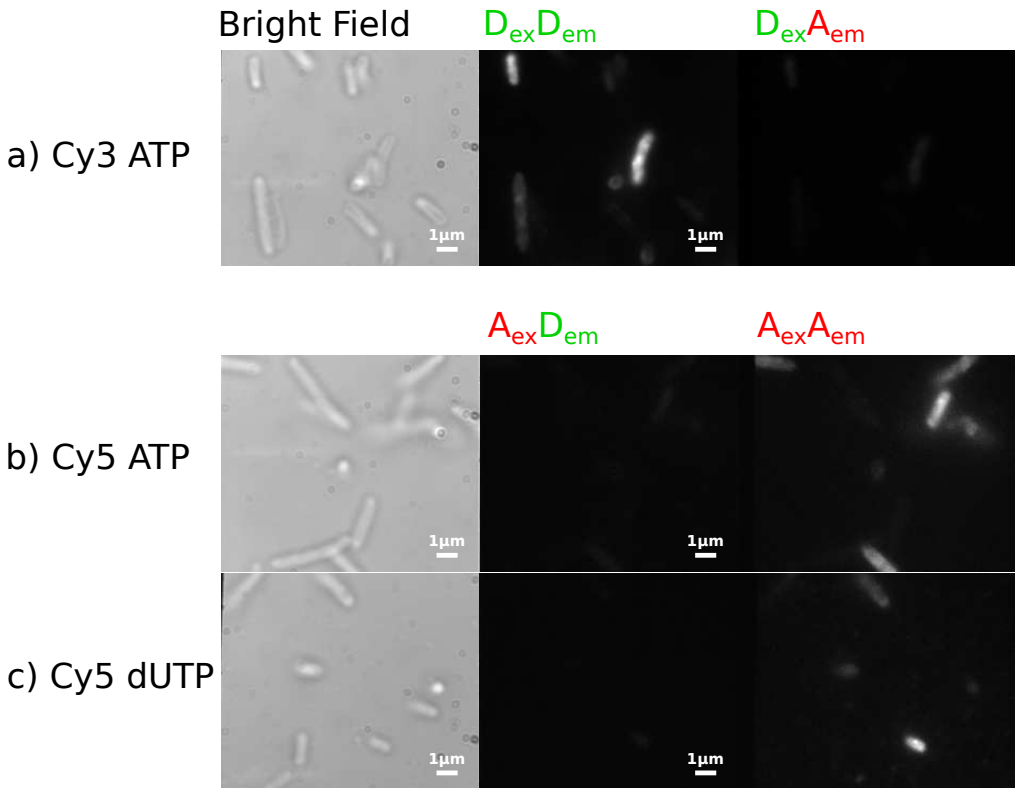


Figure 5.3.3: Internalisation of fluorescently labelled nucleotide tri-phosphates. In each case 40μL electrocompetent *E.coli* were mixed with 1nmol fluorescently labelled NTPs, electroporated, recovered for 1 hr at 37˚C in SOC media, washed 5x and then immobilised on a coverslip. a) Cy3 ATP imaged using donor excitation. b) Cy5 ATP imaged with acceptor excitation. c) Cy5 dUTP imaged with acceptor excitation. $D_{ex}D_{em}$ - Donor excitation, Donor emission, $D_{ex}A_{em}$ - Donor excitation, Acceptor emission, $A_{ex}A_{em}$ - Acceptor excitation, Acceptor emission. Donor excitation power: 200μW, acceptor excitation power: 200μW, both in widefield mode.

5.3.2 Internalisation Efficiency And Viability

Having demonstrated the ability to internalise fluorescent molecules, the next question was how many cells in the population had taken up significant numbers of fluorescent molecules and what was the effect of electroporation on cell viability.

To identify cells that had taken up a significant number of fluorescent molecules, a simple brightness threshold was used. Internalisation efficiency is defined as the ratio of cells above the threshold to the total number of cells visible in the bright field image. At 200μW widefield illumination, most cells that had not undergone EP with a fluorescent sample, or had undergone EP *without* a fluorescent sample, showed negligible autofluorescence. Occasionally, there existed a small fraction of cells that had slight autofluorescence. Hence the brightness threshold used

was set higher than this autofluorescence value to ensure that fluorescence came from internalised molecules and not from the cell itself. Therefore the internalisation efficiency numbers stated represent only bright cells with high numbers of fluorescent molecules inside which may underestimate the total number of cells that have at least some internalised molecules. Using a ssDNA labelled with Cy3B and the standard EP protocol (section A.2.3.2), we observe an internalisation efficiency of 9.8 ± 3.8% (3 separate experiments, counting an average of ~160 cells each).

To measure cell viability after electroporation, a bacterial viability assay called the '*LIVE/DEAD BacLight Bacterial Viability Kit*' (Invitrogen) was used. This viability assay contains two nucleic acid stains (SYTO 9 and propidium iodide) that differ in their ability to penetrate the membranes of healthy cells. SYTO 9 (green) can stain all cells, but propidium iodide (red) would only stain those cells with damaged membranes i.e. dead cells. Both stains were excited using the same source: a mercury lamp with a blue (440nm) filter. A live cell would show only green fluorescence whereas a dead cell would show both green and red fluorescence. Control experiments have shown that the assay provided very clear discrimination between live and dead cells. Example pictures are shown in Figure 5.3.4. In a), the sample contains only live cells (green) and in b) the same cells killed with a 1 hr incubation in 70% isopropyl alcohol (green and red). The fluorescence could be false coloured and overlaid for easy counting (right).

Figure 5.3.4 c) shows a mixture of live and dead cells that had gone through electroporation. In this particular case, 71.1% of the cells were determined to be alive (69 live, 28 dead). Interestingly, if an arc occurred during electroporation, the viability of the cells dropped to ~8% (data not shown). Without EP, but with the 20-min recovery in SOC media, from two separate experiments the cell viability was determined to be 91.2 ± 3.8% (from ~100 cells). With electroporation at a field strength of 18kV/cm and then recovery, the viability dropped to 73.1 ± 2.7% (from ~100 cells), such that relative to the non-electroporated control the cell viability was approximately 80%. This number was in agreement with independent colony counting assays performed by J.Torella, with a viability of 70 ± 6%.

To conclude, using the protocols developed in this chapter, the internalisation efficiency for a standard DNA sample is ~10% with a cell viability of ~80%. Internalisation efficiency could be improved using higher field strengths and possibly higher concentrations of fluorescent molecules. However because higher field strengths also reduce cell viability, care would have to be taken to

optimise the trade-off between these two parameters.

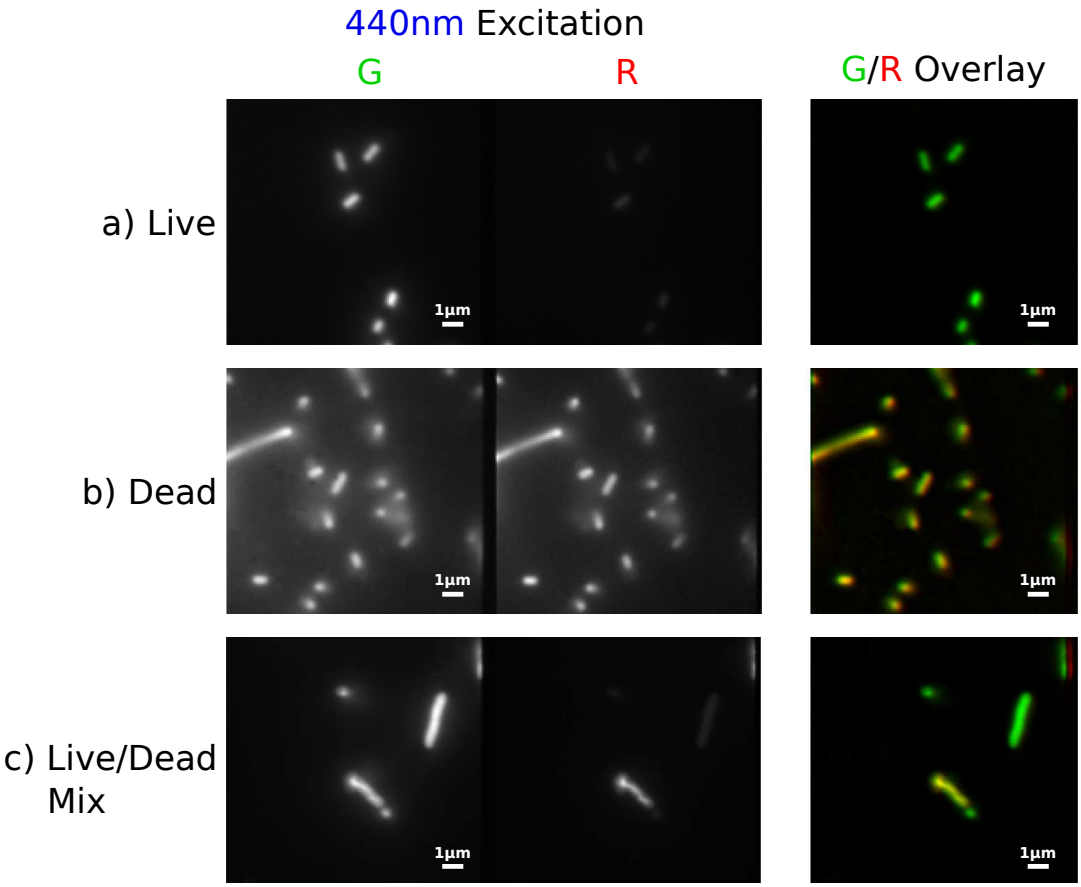


Figure 5.3.4: Example cell fluorescence using the '*LIVE/DEAD BacLight Bacterial Viability Kit*'. Fluorescence images are obtained by exciting the sample with a mercury lamp supplied with a 440nm filter. Images on the left show the raw fluorescence image, with fluorescence split into green (G) on the left and red (R) on the right. Images on the right show a false-colour overlay of the green/red images shown on the left allowing easy distinction of live (green) and dead (yellow) cells. a) Live cells. Cells were grown to late log phase, and are predominately live (green). b) Dead cells. The same cells, incubated in 70% isopropyl alcohol for 1hr, are predominately dead (yellow). c) A mixture of live and dead cells. Cells here have gone through electroporation and a 20min recovery at 37˚C in SOC media. Counting live/dead cells from several images, this experiment resulted in 69 live and 28 dead (71.1% alive).

For future work, it would be useful to know the overlap between cells that had significant numbers of internalised molecules and cells that are alive i.e. were the fluorescent cells all live, all dead, or live/dead in a ratio corresponding to or different from the population whole. To achieve this, several brightfield and fluorescent fields-of-view (FOV) of cells electroporated with a fluorescent sample could be observed with the positions recorded using an automated microscope stage. The live/dead cell viability assay could then be applied to the sample, with the assay fluorescence of

the same FOVs (from the position recordings) of the same cells recorded using blue illumination from the mercury lamp and no gain on the emCCD camera. Initial exploratory experiments have shown this approach to work, and suggest that the cells loaded with fluorescent molecules have a high average viability. This may be because the fluorescent cells are those whose pores had re-sealed successfully after electroporation, both preserving cell viability and sealing the fluorescent molecules inside. Further experiments are required to quantify this.

5.3.3 Single-molecule Fluorescence Measurements *in vivo*

A major objective of this work was to observe single-molecule (sm) fluorescence *in vivo*. As discussed previously, the ability to perform measurements at the sm level *in vivo* would provide far more information in terms of identifying molecular subpopulations, providing quantitative information on kinetic rates or conformational dynamics of individual molecules.

To achieve this, we first electroporated cells with fluorescently labelled DNA using the standard protocol used in the previous experiments (section A.2.3.2), located a fluorescent cell using widefield illumination, switched to confocal mode (section 2.4.1) and positioned the confocal spot on the cell, illuminating a femtolitre-sized spot inside the cell (Figure 5.3.5 a)). 'Switching to confocal mode' involved removing the TIRF lens in the excitation path, and moving the emission mirror (via a translation stage) to direct the light to the APDs instead the emCCD camera (section 2.6, Figure 2.6.1). We then recorded the resulting fluorescence time traces using green/red avalanche photodiodes (APDs) as in a normal sm confocal experiment.

Figure 5.3.5 b-d) show schematic illustrations (left) and sm confocal time traces using donor-only excitation (right) from 3 cells with different molecules internalised. Cells b) and c) come from the same experiment, whereby a sample was electroporated with green only (Cy3B) labelled dsDNA; b) shows the time trace of a cell that had no detectable fluorescence under widefield illumination, and c), the time trace from a cell with visible widefield illuminated green fluorescence. Cell d) has undergone electroporation with a high FRET standard (all sequences are given in section A.1.3).

The time trace of the 'empty' cell (Figure 5.3.5 b) shows a higher than normal background in the red channel (~10 kHz, compared with ~5 kHz) which remains constant throughout the experiment and is likely due to scattering from the cell. No bursts above this background are observed.

The fluorescent cell in panel c) however shows clear bursts above the background (highlighted in the grey shaded regions). Burst sizes ranged from 5-25ms, expected from the diffusion time for Cy3B labelled DNA in the high viscosity environment of the cell [121]. Longer bursts (> 100ms) were also present, which is expected to be due to a combination of anomalous diffusion inside the cell, and the labelled DNA being bound non-specifically or specifically by certain DNA binding proteins. Figure 5.3.5 d) shows the time trace from a cell internalised with a high FRET DNA standard. In this case, using donor-only excitation, clear bursts above background are observed in the acceptor channel consistent with intact high FRET DNA. Interestingly, green-only bursts also occur, which could arise from DNA that was mislabelled, partially photobleached or nuclease degraded.

One of the difficulties with sm detection in *E.coli* is because of their comparable size to the confocal spot. A typical *E.coli* cell has a volume between 0.3-1fL, and a confocal spot, using green (~532nm) light, illuminates a region of ~0.1fL. Thus ~10-30% of the cell will be illuminated in a typical sm confocal experiment. To maintain sm sensitivity, one therefore needs on the order of 10 molecules or less per cell (~60 nM, given a cell volume of $0.65\mu m^3$ [126]). Such a small number of molecules will be repeatedly observed, and quickly photobleached dramatically limiting observation times.

To increase sm observation times *in vivo*, one either has to reduce the size of the excitation volume or increase the size of the cell. While efforts to achieve the former have been successful using nanofabricated devices such as zero-mode waveguides for solution and surface based *in vitro* experiments [127, 128], and even in lipid membranes [129], it would be very difficult to employ them for the detection of molecules in a bacterial cytosol. However for the latter, it is possible to increase cell size by using of certain antibiotics that affect cell wall formation such as cephalexin[130, 131]. The use of cephalexin causes elongation of *E.coli* by up to 100x their normal length, producing a continuous cytoplasmic space, enabling much longer observation times for sm measurements *in vivo*. This approach was successfully implemented by J.Torella [121] where an observation time of ~ 4min allowed the detection of ~ 1000 bursts, enough to generate the first ES histogram from *in vivo* data. However, preliminary viability experiments using cephalexin-elongated cells showed that few survive electroporation. Much useful information should still be obtainable using the elongated cells, but future work will be directed towards long sm observation

times in *live* cells.

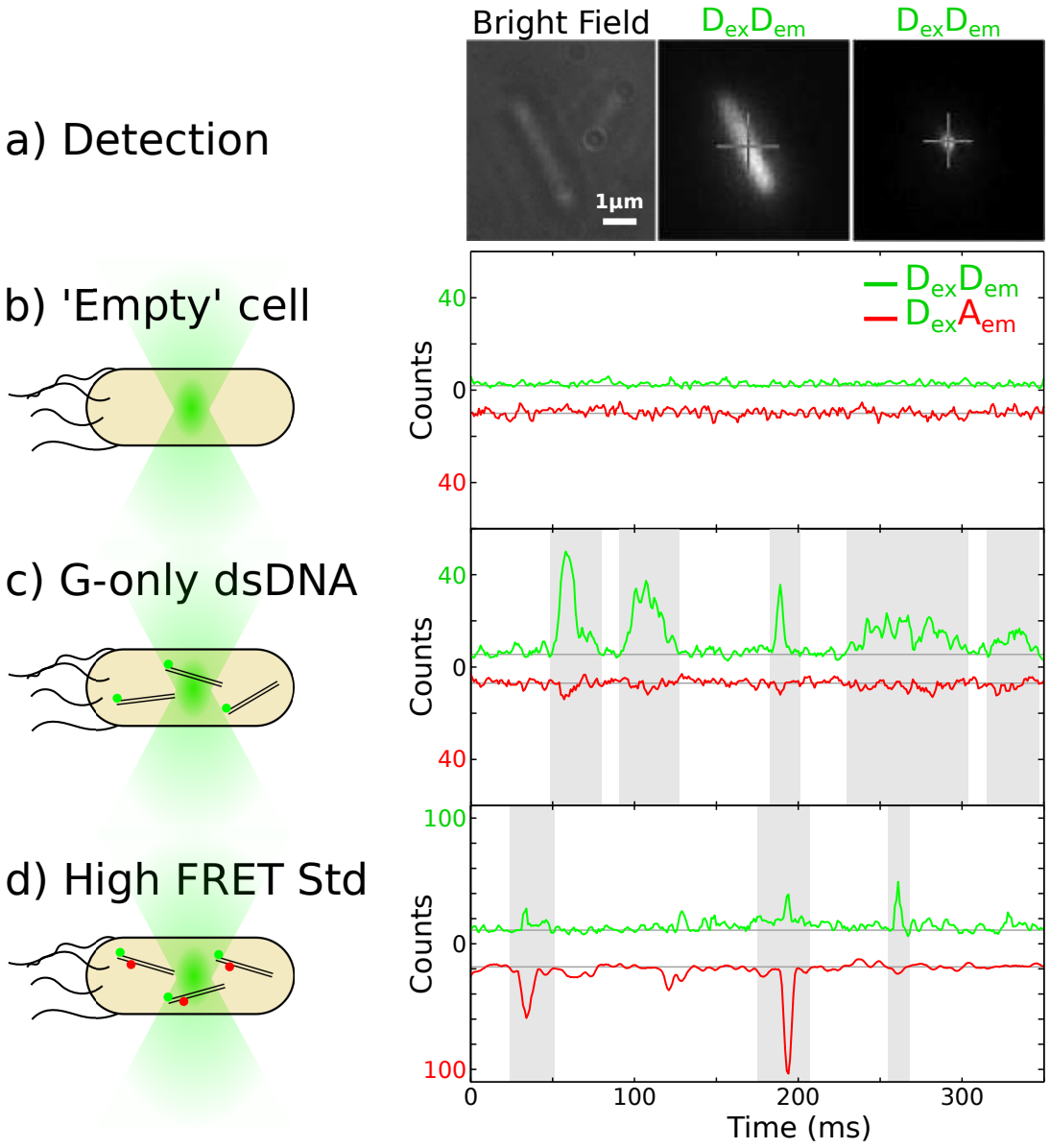


Figure 5.3.5: Single-molecule fluorescence *in vivo*. a) Detection method. To detect single molecules in live cells, first a suitable fluorescent cell is picked using widefield illumination - see brightfield (left) and widefield (middle) images. The setup is then switched to confocal mode, producing a diffraction limited detection volume inside the cell, observable as a bright spot inside the cell (right). b-d) Schematic (left) and 1ms binned with 3-point smoothing confocal time trace (right) for each sm experiment *in vivo*. Grey lines indicate mean background, shaded areas show bursts. b) An 'empty' cell, where no detectable fluorescence under widefield is observed, produces a sm confocal time trace with higher than normal background, but no apparent bursts. c) A cell containing G-only dsDNA produces clear bursts in the donor channel under donor excitation. d) A cell containing high FRET dsDNA produces clear bursts in the acceptor channel under donor excitation demonstrating intact DNA. Also visible are donor-only bursts due to mislabelled, partially photobleached or degraded high FRET dsDNA. $D_{ex}D_{em}$ - Donor excitation, Donor emission, $D_{ex}A_{em}$ - Donor excitation, Acceptor emission. Donor excitation power: 150μW in all cases. Experiment performed by J.Torella. Reproduced with permission from [132] Copyright © 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

5.3.4 Towards Transcription Factor Detection *in vivo*

Finally as a step towards real-time dynamic TF detection, the bending-based TF sensor (for CAP detection) described in Chapter 4 was successfully internalised into *E.coli* cells. Figure 5.3.6 shows fluorescence from the FRET sensor in live cells using widefield illumination. In an attempt to modulate CAP binding activity, the cells were incubated with different concentration of the small molecule 'cyclic AMP' (cAMP), the co-factor required by CAP to alter its conformation into the DNA binding form. cAMP has previously been reported as cell permeable for *Staphylococcus aureus* [133], and has been added exogenously in the growth medium to *E.coli* to affect catabolite repression and the synthesis of β-galactosidase [70, 94].

Given an average cell volume of $0.65\mu m^3$ [126], and the number of CAP dimers in an *E.coli* cell [94] grown in LB to the optical density used to prepare electrocompetent cells (section A.2.3.1), the intracellular CAP concentration should be ~14.5μM (or ~2400 molecules / cell). With the measured dissociation constant of the bending sensor of ~800 nM (using FRET, section A.2.2), the bound fraction of bending sensors in-vivo should be ~94 %, at the appropriate concentration of cAMP. Even at half this intracellular CAP concentration, 90% of the sensors should still be bound.

At this point we must take note of the biphasic influence of cAMP concentration on CAP binding affinity. Without cAMP, CAP exists in a conformation with low affinity for its recognition site [134, 135]. For DNA binding, at cAMP concentrations of ~0.2 mM, one molecule of cAMP binds to each monomer of the CAP dimer resulting in a conformational change allowing tight binding of CAP to its recognition site [134, 135]. At higher millimolar concentrations of cAMP, CAP-DNA affinity dramatically decreases [136, 137, 138]; indeed at these concentrations, it was found that two cAMP molecules could bind to each CAP monomer [93]. As this secondary binding site lay in the helix-turn-helix DNA binding domain of CAP, it may explain the decrease in affinity for its recognition site [93]. In terms of our experiment, if we could change the intracellular concentration of cAMP, then the change in CAP binding affinity should be observable using the bending sensor.

Estimates of the natural intracellular cAMP concentration of *E.coli* range from ~0.5-10μM, depending on the strain used [70]. At the higher end of these natural cAMP concentrations, CAP binding activity should still be non optimal. Thus by varying the concentration of cAMP from 0 to

0.2 to 2 mM we should observe weak, strong then weak binding again respectively. As the FRET efficiency observed previously for CAP bending sensor *in vitro* experiments in Chapter 4 changed from around 45 % in the unbound state to 32 % when fully saturated by CAP binding we might expect shifts between 45 % and 32 % for weak and strong binding respectively.

Figure 5.3.6 shows the brightfield and fluorescence images of cells internalised with the CAP bending sensor incubated in the presence of a) 0, b) 0.2 mM, and c) 2 mM cAMP. Cells were internalised with the standard protocol (section A.2.3.2) using 25 pmoles of CAP bending sensor, given no recovery and then immobilised on a coverslip in PBS initially with no cAMP. Separate 5 minute incubations of 0.2 mM and then 2 mM cAMP were used with the same immobilised cell sample. FRET efficiencies were calculated from the background-subtracted intensity of the average of the number of cells indicated ± 1 standard deviation (section A.2.3.4). Cells were picked if they were of sufficient signal to noise (section 5.3.2) and completely within a single plane of focus.

Without any cAMP (Figure 5.3.6 a)), we observe a FRET efficiency of 39.3 ± 1.8% (9 cells analysed), midway between the two extremes (45% to 32%) expected for unbound to fully bound. Thus it is possible that even without additional cAMP, intracellular CAP and cAMP concentrations may be at a level for intermediate binding to the bending sensor. On application of 0.2mM cAMP (Figure 5.3.6 b)), we see a FRET efficiency of 36.1 ± 1.8% (6 cells analysed) and then with 2 mM cAMP (Figure 5.3.6 c)) a rise to 37.8 ± 2.4% (11 cells analysed). The individual measurement error for 0mM, 0.2mM and 2mM cAMP experiments was 1.2%, 1.1% and 1.1% respectively (calculated in the same way as the FRET standards, section 5.3.1.2).

(Figure 2.6.1)

The drop and then rise in FRET, on addition of 0.2 mM and then 2 mM cAMP seems at first glance consistent with the biphasic predictions of CAP-DNA affinity on cAMP concentrations. However it is quite possible that this small FRET change is attributable just to experimental noise. Secondly as these experiments are performed *in vivo*, they should be considered in light of several factors. For example, the dissociation constant of the sensor may have changed inside the cellular milieu due to different ionic strength, pH or contributions from macromolecular crowding. Non-specific protein binding to the introduced sensor by for example nucleases may to some extent degrade the sensor and/or block access to the sensor by CAP. Along with potential photophysical effects on the fluorophores used, it is difficult to exactly predict the FRET change expected inside

the cell. Nevertheless, this initial result is encouraging and further work will be aimed at better control over CAP binding activity inside cells to thoroughly characterise the sensor *in vivo*. This data should be compared to similar experiments using a control sensor with no CAP binding site. It will also be necessary to increase the statistics of observed cells. Efforts to automate image analysis for cell counting and quantification are currently underway.

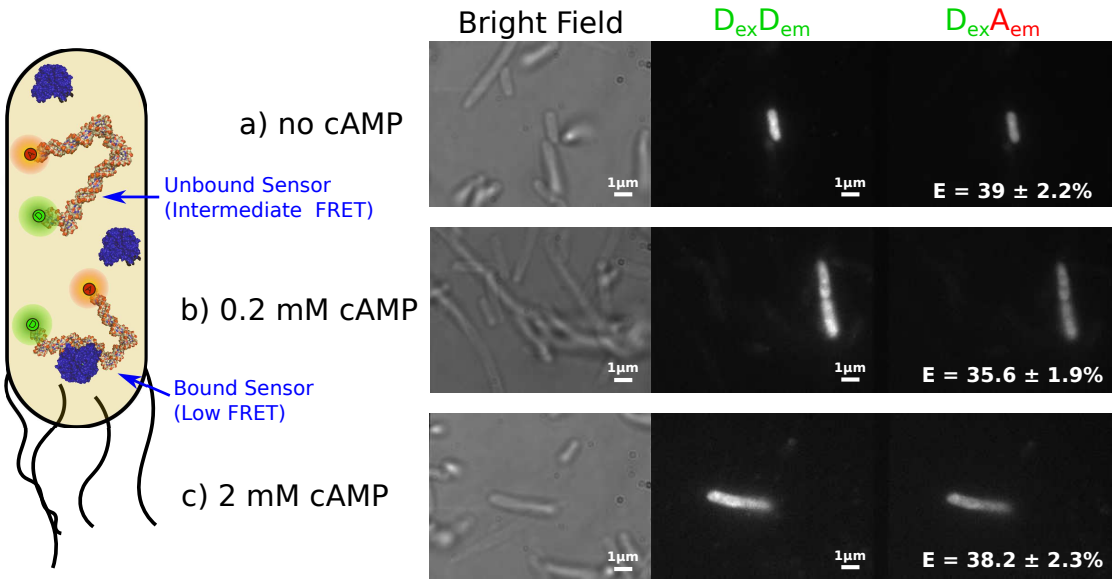


Figure 5.3.6: Internalisation of FRET-based Transcription Factor sensor. 40μL electrocompetent *E.coli* were incubated for 10 min on ice with 25 pmoles nuclease protected FRET bending CAP sensor (as described in Chapter 4), electroporated, washed 5x and then immobilised (without recovery) on a coverslip. The dependence on FRET of the sensor by CAP binding due to different cAMP concentrations was probed by incubating cells with different cAMP concentrations on the coverslip for 5 min each. Left: schematic illustrating *E.coli* with internalised bound and unbound CAP FRET sensors in the presence of CAP molecules (blue). a) Without cAMP: E* = 39 ± 2.2%. b) At 0.2mM cAMP: E* = 35.6 ± 1.9%. c) At 2mM cAMP: E* = 38.2 ± 2.3%. BF - Bright field, D_{exc},D_{em} - Donor excitation, Donor emission, $D_{exc}A_{em}$ - Donor excitation, Acceptor emission. Donor excitation power: 200μW in widefield mode. All uncorrected FRET efficiencies quoted (E*) are from an average of 3 cells ± 1 s.d.

5.4 Summary

In this chapter we have demonstrated a successful strategy for the internalisation into and observation of fluorescent molecules in live *E.coli* cells as a route towards real time TF detection *in vivo*. Using electroporation, cells were loaded with fluorescent single- and double-stranded oligonucleotides and fluorescent nucleoside tri-phosphates (NTPs) . Doubly labelled double-stranded oli-

gos that contained specific FRET signatures were used to confirm that intact DNA existed inside cells.

The internalisation of fluorescent molecules into bacterial cells would not be limited to detection of active transcription factor concentrations. Internalisation of fluorescent nucleotides may allow incorporation into the replicating bacterial chromosome, as demonstrated recently at the single-cell level in *Bacillus subtilis* [125]. Combining this approach with sub-diffraction imaging techniques such as STORM (Stochastic Optical Reconstruction Microscopy) [122, 139] could allow novel views of the organisation of the bacterial chromosome. Doubly labelled DNA promoter sites could also be used to infer RNA polymerase dynamics using assays already performed *in vitro* [140, 141]. Further work will incorporate attempts to internalise other molecules such as short polypeptides (e.g. GFP fragments [142]) for protein labelling methodologies, or even doubly labelled proteins for example DNA polymerase [16] to directly observe protein conformational dynamics *in vivo*.

The protocols developed in this chapter demonstrate that using 10 pmoles of labelled DNA, and an electroporation field strength of 18kV/cm, we achieved ~10 % internalisation efficiency and ~80 % viability relative to a non-electroporated control. These protocols were originally developed for the internalisation of plasmids into cells and not short labelled oligos. Thus there may well be room for optimisation of the protocols, by changing parameters such as: electroporation field strength, cuvette size, cell volume, concentration of molecule to be internalised, or even through the use of molecules that have properties aiding pre-electroporation membrane adhesion. Ideally, to help optimise the protocols, some automation of the cell image analysis would be an advantage. This would help to separate and count cells, experiment with different counting metrics, quantify fluorescence / FRET and also determine live/dead cell viability ratios.

We then demonstrated single-molecule (sm) FRET detection in bacterial cells which is to our knowledge the first report of such an observation. Single-molecule FRET *in vivo* is an important result for accurately quantifying protein-nucleic acid interactions and conformational dynamics in the natural cellular environment. We show clear bursts using a green only and high FRET DNA standard. Single-molecule observation times are limited in normal cells due to the similar dimensions of the confocal spot and the cells themselves. Thus alterations of cell size, through the use of certain antibiotics such as cephalexin, are a route to much longer sm observation times.

Finally we demonstrate successful internalisation of the TF FRET-based sensor developed in Chapter 4 to detect catabolite activator protein (CAP). In normal *E.coli* cells, without additional cAMP, the bending sensor displays a FRET efficiency of ~39%, approximately intermediate between values for completely bound and unbound sensors observed in *in vitro* experiments (45% to 32%). While a small FRET change was observed on addition of different concentrations of the co-factor cAMP, it is difficult to attribute this to real causes, or to measurement noise. Thus future work is aimed at more accurately controlling the level of active CAP molecules inside cells, to fully characterise the sensor in a cellular context. To modulate active CAP levels, either CAP numbers or cAMP numbers could be altered. CAP expression could be modulated by the expression of a plasmid encoding CAP whereas cAMP levels could be affected by including different carbon sources in the growth medium such as galactose, glycerol, glucose or succinate [70]. An important control for these experiments will be the construction of a sensor containing a different sequence to the CAP binding site. Another interesting route may be to use a mutant E.coli incapable of producing cAMP such that cAMP could be added exogenously to modulate CAP binding activity. We are also aiming to improve our detection statistics by automating the fluorescent cell image analysis to count and quantify in the range of 50-100 cells for each experiment. This should allow us to prepare histograms of the FRET values from single-cells highlighting potential sample heterogeneity perhaps even with FRET standards alone.

We believe the methods developed in this chapter significantly add to the single-molecule toolbox and act as a general approach to the possibility of a wide variety of interesting *in vivo* experiments at the sm level. Not least of these is the ability to monitor TF activity in real-time, in living cells.

5.5 Contributions

I thank Joseph Torella for all his help and guidance in this project. Single-molecule measurements *in vivo* shown peformed by Joseph Torella. Many thanks to Ludovic Le Reste for helping with setup alignment and modifications. Support and feedback provided byAchillefs Kapanidis, Javier Periz and Joseph Torella.

Chapter 6

Transcription Factor Encapsulation And Release Within A DNA Cage

6.1 Introduction

In this chapter I explore the use of DNA nanotechnology in collaboration with Prof Andrew Turberfield's lab, to develop systems for transcription factor encapsulation and release.

In chapters 3-5 I developed methods to detect the *presence* and *concentration* of transcription factors (TFs) both *in vitro* and *in vivo*. From another perspective, it was thought that it might be possible to *deliver* TFs into a bacterial cell, to change its behaviour due to the large role in gene regulation that TFs play. Intracellular delivery of TFs into mammalian cells has already been demonstrated to induce pluripotency [6]. Even small numbers of internalised TFs may play a role in determining cell fate as it has been shown that the stochastic single-molecule event of the TF lactose repressor dissociating from its binding site caused a switch in cell phenotype [38]. Through the use of DNA nanotechnology, it would be theoretically possible to encapsulate a transcription factor within a rationally designed DNA 'cage' with the ability to release (and activate) the protein given required inputs. While encapsulated within the cage, the active site of the TF would be blocked thus rendering the protein 'inactive'. Such a device could then be transferred into a cell and opened under the stimulus of an external trigger e.g. from a particular mRNA transcribed under the control of an inducible promoter.

Another interesting application of single protein encapsulation within a DNA cage would be

the measurement via smFRET of potential protein conformational dynamics. Whilst trapped within the cage, the protein and cage could be labelled in such a way to probe the distance between labels, thus directly reporting on dynamics using the cage as a reference point. Such information may complement analogous work of protein folding in chaperonin cages [143]. New techniques such as switchable FRET [144] or multi-colour FRET [145] may allow the measurement of several FRET pairs, for multiple accurate distance measurements that could be used to help elucidate protein structure through triangulation [146].

DNA nanotechnology is a field that has been rapidly growing over the last 10-15 years. First suggested in 1982 [147], the specificity of DNA base-pairing interactions could be used as the building blocks for rationally engineered nanofabrication from the bottom-up. A hugely diverse array of DNA-based nanodevices have been produced, from 2 and 3-dimensional structures [39, 148, 40] to molecular motors [41, 42] and molecular 'logic gates' [43, 44, 45]. Formation of structures are based on the fact that they have the highest number of Watson-Crick base-pairs (bps). However it is possible that other structures with only marginally less stable conformations can form if the component strands are simply mixed in one step. Hence a lot of the approaches to form such structures involve specific strategies such as intermediate purification steps which can result in low yields.

In this work I focus on adapting the design of a DNA tetrahedron [46] which is a robust structure that can be self-assemble in high yield in a single annealing step, in under 1 minute, in one stereoisomer, using equimolar concentrations of four component strands. Each strand runs around one face of the tetrahedron, hybridizing to its neighbour and forming an edge. Unpaired nucleotide spacers at each vertex allows freedom for the strand to bend to the next edge. Figure 6.1.1 shows a space-filling representation of a tetrahedron with all six edges 20 bp long.

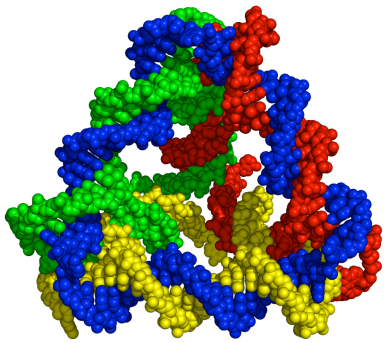


Figure 6.1.1: Space-filling representation of a regular 20 bp edged DNA tetrahedron [46].

Several methods to encapsulate atoms, molecules and proteins have been proposed and confirmed experimentally. Methods vary greatly in size and encapsulation target, from the smallest - fullerenes (diameter 0.4 - 1nm), shown to encapsulate a single lanthanum atom [149, 150], to mid-sized hollow metal nanoparticles (diameter ~100nm), shown to encapsulate the enzyme horseradish peroxidase [151], to larger lipid bilayer structures such as liposomes and vesicles (diameter 100-800nm) that have been demonstrated in pharmaceutical applications encapsulating drugs [152] or more recently for single-molecule fluorescence studies of individual molecules [153, 154].

Constructing a molecular cage out of DNA has several advantages. Due to the specificity of Watson-Crick base pairing, cages of dimensions similar to protein targets could be rationally designed, and easily self-assembled rapidly in a single step. Edges could be extended or broken through the addition of complementary strands to outcompete those forming the structure in order to open such cages. Also as a simple modification for protein attachment, DNA sequences with high affinity for protein targets e.g. aptamers or sequence specific recognition sites could also be adopted into a DNA cage design. In this work we chose to adapt the design of the DNA tetrahedron in [46] due to its ease of assembly, robust structure and dimensions comparable a protein of approximately 60kDa [155].

For a fully 20 bp edged DNA tetrahedron it was shown that such a structure did have the capability of encapsulating a protein [155]. In this study cytochrome-c was conjugated to the 5' end of one of the DNA strands used to form the tetrahedron such that the protein would be attached to one edge. By altering the position of where the 5' and 3' ends of the strand met along that edge, it was possible to place the protein facing into or out from the tetrahedron, as demonstrated by

migration differences using PAGE (Fig 6.1.2 c)).

However, in order to attach the protein to the DNA, both the DNA and protein required several chemical modification and purification steps, resulting in conjugation through bifunctional crosslinker molecules. As the protein modification involved attaching a crosslinker to primary amine sites (of which there are 19 on cytochrome-c), the linker position, and therefore protein orientation relative to the tetrahedron would be a distribution of 19 different positions, as well as likely affecting the activity of the protein itself.

The protein encapsulated, cytochrome-c, has a molecular weight of ~12 kDa and approximate radius of 1.75nm. For the geometry of a regular 20 bp tetrahedron, the largest sphere that can fit through one face is ~2 nm (calculating the incircle - see section 6.2.1) while the largest sphere that can fit inside would be 2.4 nm (calculating the insphere). Thus, encapsulation is produced from the combined effect of orienting the protein attachment site into the tetrahedron, and from physical attachment to the protein, as opposed to encapsulation through non-covalent means, via physical confinement or protein affinity.

To release the protein from such a cage, chemical breaking of the linker molecule would be required, whereas if the protein was encapsulated through non-covalent means, could be released through a change to the DNA topology, such as extending or breaking an edge from the addition of an input oligonucleotide.

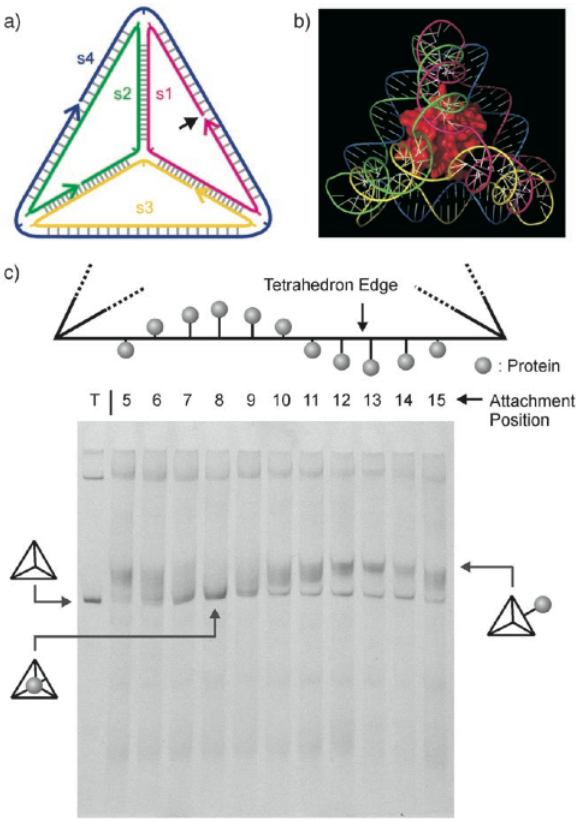


Figure 6.1.2: Protein encapsulation within a DNA tetrahedron. a) The design of the tetrahedron consists of four oligonucleotides, each running around one face, producing a nick on each edge where the 5' and 3' ends meet. Arrow indicates attachment point for cytochrome-c at the 5' end of strand s1. b) Scale schematic illustrating cytochrome-c can fit within the dimensions of the tetrahedron. c) Native PAGE gel demonstrating mobility shifts dependent upon attachment position of cytochrome-c. T indicates unmodified tetrahedron (no protein attachment). Reproduced with permission from [155].

A DNA tetrahedron whose topology could be changed has already been demonstrated [156]. This design consists of a tetrahedron containing 5x20 bp edges and 1x10 bp edge that contained a ssDNA hairpin (Fig 6.1.3). To this hairpin was added a complementary 'fuel' stand which on hybridisation caused extension of that edge from 10 to 30 bp. The 'fuel' strand could further be removed by the addition of an 'antifuel' such that the tetrahedron could be reversibly opened and closed as demonstrated using PAGE, ensemble and single-molecule FRET [156].

Protein encapsulation and release using a reversibly opening tetrahedron of this design may prove difficult. In the closed state, the tetrahedron would only have a depth at its widest point of 10 bp (~3.4 nm) and so would be limited to small proteins. Larger proteins may be accommodated by encapsulation in the open state, but would not be released on closing, as it would only be confined

further. Some optimisation of the design may yield a workable solution for this task which the design was not originally intended for.

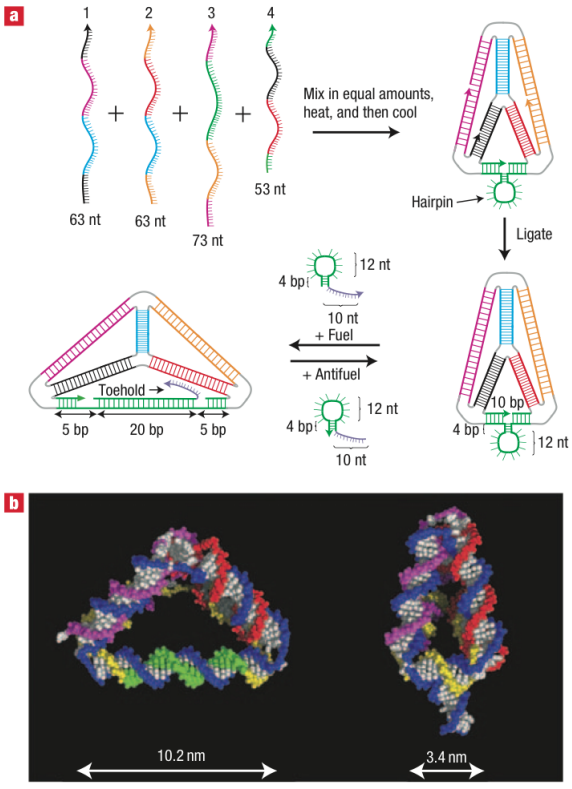


Figure 6.1.3: Reconfigurable tetrahedron. (a) Synthesis scheme. Four oligonucleotides with complementary regions are annealed forming a tetrahedron with a hairpin on one edge. Nicks, where the 5' and 3' end of each strand met, were ligated shut. On the addition of a 'fuel' strand, the 10 bp edge extends to 30 bp. This can be reversed through the addition of an 'anti-fuel' strand that displaces the fuel strand. (b) Schematic illustrating the open (left) and closed (right) forms of the reconfigurable tetrahedron. Reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology [156], copyright 2008.

Both the protein-encapsulating [155] and reconfigurable [156] tetrahedron designs serve as important examples demonstrating the flexibility of the tetrahedron design to adopt novel functionality. In this chapter I will describe modification of the original tetrahedron design [46] to produce two different DNA nanostructures; one for non-covalent protein encapsulation without the need for any DNA or protein modifications and another to achieve a reversible opening of a similar nanostructure (through complete breaking of one edge).

6.2 Protein-Encapsulating DNA Cage

6.2.1 Design

To encapsulate a protein non-covalently within a DNA cage a method of producing affinity between the DNA and the protein was required. Transcription factors, due to their inherent affinity for specific recognition sequences, were the ideal choice. By placing the recognition sequence for a specific transcription factor along one edge of the tetrahedron, the transcription factor could bind to this edge, resulting in encapsulation either through the formation of the cage around the TF-binding site, or by binding directly to a pre-formed cage if the cage dimensions were favourable. This approach would avoid the need for any further linker molecule between the cage and the target e.g. aptamer or antibody as the affinity would be provided by the cage sequence alone.

The target transcription factor for encapsulation was catabolite activator protein (CAP) as used in all previous chapters. The dimensions of CAP compare favourably with the DNA tetrahedron produced in [46] which was used as a basis for the design of the protein-encapsulating DNA cage.

Figure 6.2.1 shows a schematic representation of the protein-encapsulating DNA cage with and without CAP. The cage consists of four complementary oligonucleotides that each run around one face of the cage forming double-stranded edges with their neighbouring strands. On one edge the binding site for CAP was placed.

The consensus binding site for CAP is 22 bp long [71], whereas the regular DNA tetrahedron [46] has an edge length of 20 bp. To accommodate the binding site, one edge length was increased. The regular DNA tetrahedron structure has been verified from a number of methods, cryo-EM, AFM and PAGE [46, 157]. As it was known that the major groove of the regular DNA tetrahedron faces inwards at the vertices , to keep them facing inwards with a longer edge, the edge was increased by one full turn of DNA i.e. to 30 bp (Figure 6.2.1 left). Knowledge of the orientation of this edge was crucial to position the CAP binding site to face into the cage and also would not hinder the stereoselectivity of formation [46]. In fact it was shown in [46] that a 5x20bp and 1x30bp tetrahedron did form in one stereoisomer with the major groove facing inwards at the vertices. To orient the protein into the cage, we observed from the crystal structure of CAP bound to a 30 bp fragment, 1CGP [74], the major groove of the DNA fragment at either end was oriented approximately in the same plane as CAP (on the side of binding). Therefore placing the CAP

binding site directly in the centre of the 30 bp edge, bound CAP would lie in the same plane as the major grooves at each end of the edge which faced into the cage. As a control we formed an outward facing site by shifting the binding site one half turn (5 bp) towards vertex 1.

We saw in Chapter 4 that CAP is known to induce a significant bend in the DNA of its binding site of approximately 80-100˚ [74, 87, 158]. Therefore on encapsulation within the cage, the binding of CAP to the 30 bp edge would cause such a bend, decreasing the distance between vertex 1 and 2 (v1 and v2 - Figure 6.2.1 Right) changing the shape of the cage. Calculation of the cage geometry in the encapsulated and non-encapsulated states reveals why the cage is so effective at trapping CAP.

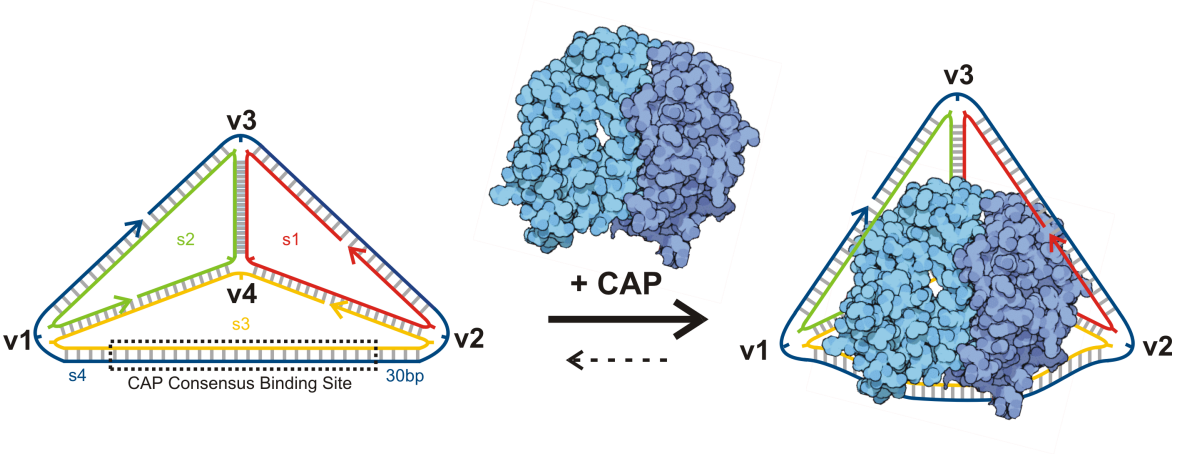


Figure 6.2.1: Schematic design of the protein-encapsulating DNA cage designed for the encapsulation of catabolite activator protein (CAP). The cage is shown as a combination of four coloured strands (s1-s4: red, green, yellow, blue) and CAP is shown in light/dark blue. Left: Without CAP the DNA cage dimensions (in bp) are given. v1-v4 denote vertices 1-4. Right: In the presence of CAP, the protein induces a bend on the binding edge decreasing the distance between v1 and v2. Illustration of CAP provided by David Goodsell, RCSB. Reproduced with permission from [159].

Cage Geometry

Two dimensions of the cage are relevant to encapsulation, namely the radius of the largest sphere that can fit through one face of the cage, and the radius of the largest sphere that can fit inside the cage. In the case of a regular tetrahedron, these two dimensions are given from the size of the incircle of one face (triangular window) and from the size of the midsphere (the largest sphere tangent to the *edges*) of a regular tetrahedron respectively. The incircle radius, r, of a triangle with sides of length a, b and c is given by:

$r = \frac{2A}{P}$ where P is the perimeter of the triangle and A is the area of the triangle given by Heron's formula: $A = \sqrt{s(s-a)(s-b)(s-c)}$, where s is the semiperimeter.

For a regular tetrahedron of edge length a, the radius of the midsphere is given by:

$r = \frac{a}{\sqrt{8}}$ [160]

For a regular tetrahedron of edge length 20 bp (6.8 nm) the incircle radius of one face is ~1.96 nm, and the radius of the midsphere is ~2.4nm. However this calculation of the midsphere does not account for the width of dsDNA. Accounting for this difference, using a model of the structure, it was found that the radius of the largest sphere that fitted inside was ~2.6nm [155, 161], which was approximated to correspond to a protein size of ~60kDa. CAP, a ~45kDa homodimer, was thus a suitable candidate for a cage of similar geometry.

For the CAP encapsulating cage, a regular tetrahedron where one edge is extended to 30 bp, we now have two triangular faces with edge lengths of 30 bp, 20 bp and 20 bp and an incircle radius of ~1.93 nm and two facces with all 20 bp edge lengths and the same incircle radius as that of the regular 20 bp tetrahedron face, ~1.96 nm as illustrated in Figure 6.2.2 a) left, middle. CAP, of approximate width, length, and depth of 5.5 nm, 5.9 nm and 2.9 nm (Figure 6.2.2c)) should not be able to pass through any of these faces.

The calculation of the midsphere of this now irregular tetrahedron is unfortunately not as trivial as in the regular case. However, we know the position of CAP binding within the cage and the bend that CAP induces to this edge. Thus it is possible to calculate the smallest distance between the middle of the binding edge and the edge opposite the binding edge. Without CAP bound inside the cage, this distance is calculated from straightforward geometry as ~2.95 nm, far smaller than the length of CAP itself, ~5.9 nm (Figure 6.2.2c)). Using the same approximations of protein densities as [155, 161], sphere of this diameter would equate to a protein size of ~11 kDa.

It would appear that CAP should not fit inside the cage. However, due to the bend in the DNA induced by CAP, the dimensions of the cage change quite drastically. As has been well documented [74, 87, 158], the bend in DNA caused by CAP occurs at 2-3 symmetrical kink sites. The largest bend occurs at the primary kink around position 6/7 on either side of the middle of the binding site. As an approximation, I consider the overall bend to occur from the primary kinks alone with a kink angle of 45˚ (overall DNA bend of 90˚).

In the presence of the bend induced by CAP, the radius of the largest sphere that can fit through

the the faces now affected by the bend (v1-v2-v3 and v1-v2-v4, Figure 6.2.1) is the radius of the circumcircle of the triangle connecting the midpoints of each edge (therefore tangent to each edge - Figure 6.2.2 b) left). For a triangle of side lengths a, b and c, the radius of the circumcircle is given by:

$r = \frac{abc}{4A}$, where A is the area of the triangle given by Heron's formula (above).

In this case, the radius is approximately 2.75nm, potentially large enough to allow CAP to pass through those faces. Furthermore, the distance between the middle of the binding site to the opposite edge has now increased to ~6.6 nm (~122 kDa [155, 161]). The bending CAP induces, changes the cage geometry allowing it to fit inside. With CAP inside the cage, if it tries to dissociate, the cage would try to snap back to its original size where the inner dimensions were insufficient for the accommodation of CAP. At the same time, CAP would be pushed back to its binding site and thus experience an extremely high local concentration that would result in a high likelihood of rebinding. It is proposed that these mechanisms help trap CAP inside the cage.

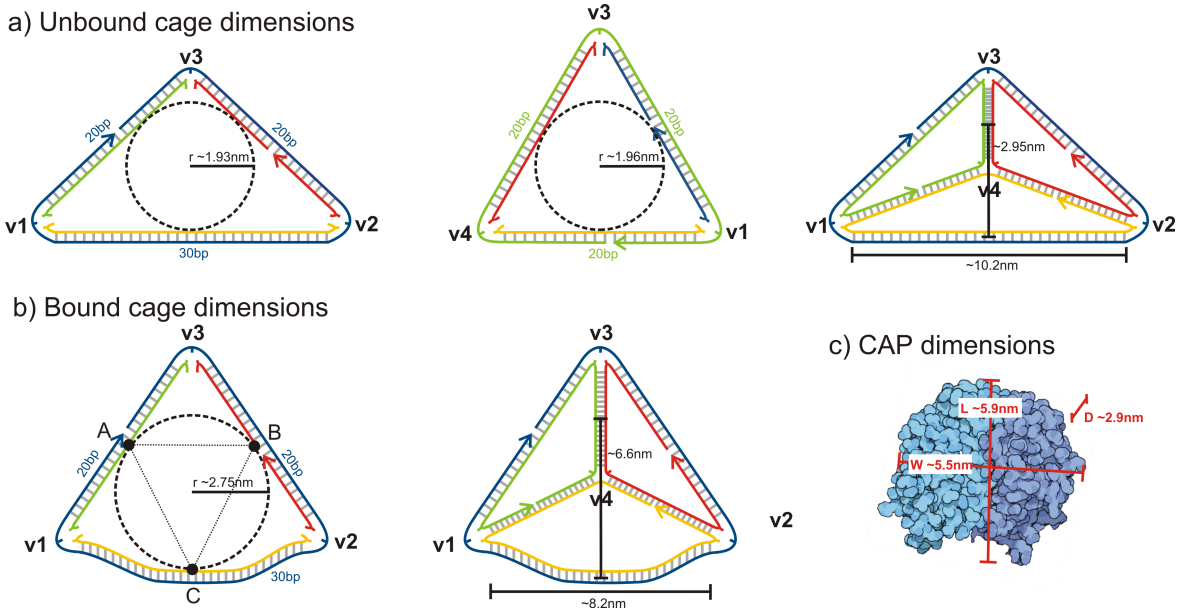


Figure 6.2.2: Protein-Encapsulating DNA cage geometry calculations. a) Left, middle: Unbound by CAP, the cage has two face sizes, with incircle radii of ~1.93 and 1.96 nm. Right: the shortest distance between the middle of the binding site to opposite edge is ~2.95 nm. b) Left: in the bound state, the binding edge is bent by CAP causing a change in one face size. The largest circle that now fits inside that face has a radius of ~2.75 nm. Right: Due to the bend in the DNA, the distance now between the middle of the binding site to the opposite edge has increased to ~6.6 nm. c) Approximate dimensions of CAP in terms of width, length and depth - 5.5 nm, 5.9 nm and 2.9 nm respectively.

Figure 6.2.3a) and b) show a series of images from different perspectives of a 3-dimensional scale model of the cage with and without CAP bound. The model was made using DNA strands of accurate dimensions and the CAP-DNA crystal structure from the protein data bank (structure 1O3T from [162]) where CAP is bound to a 30 bp dsDNA segment (the same length as the CAP binding edge in the cage). Figure 6.2.3a) demonstrates the cage dimensions without CAP bending; CAP is shown on the right for comparison. It can be seen from both the front and side perspectives that CAP would not fit inside a cage with this geometry. Comparing to 6.2.3b) where CAP is bound, bending the DNA on the 30 bp edge, we see now from all perspectives how CAP fits inside the cage. A similar model was fabricated using a 3d printer, a picture is shown in Figure 6.2.3c).

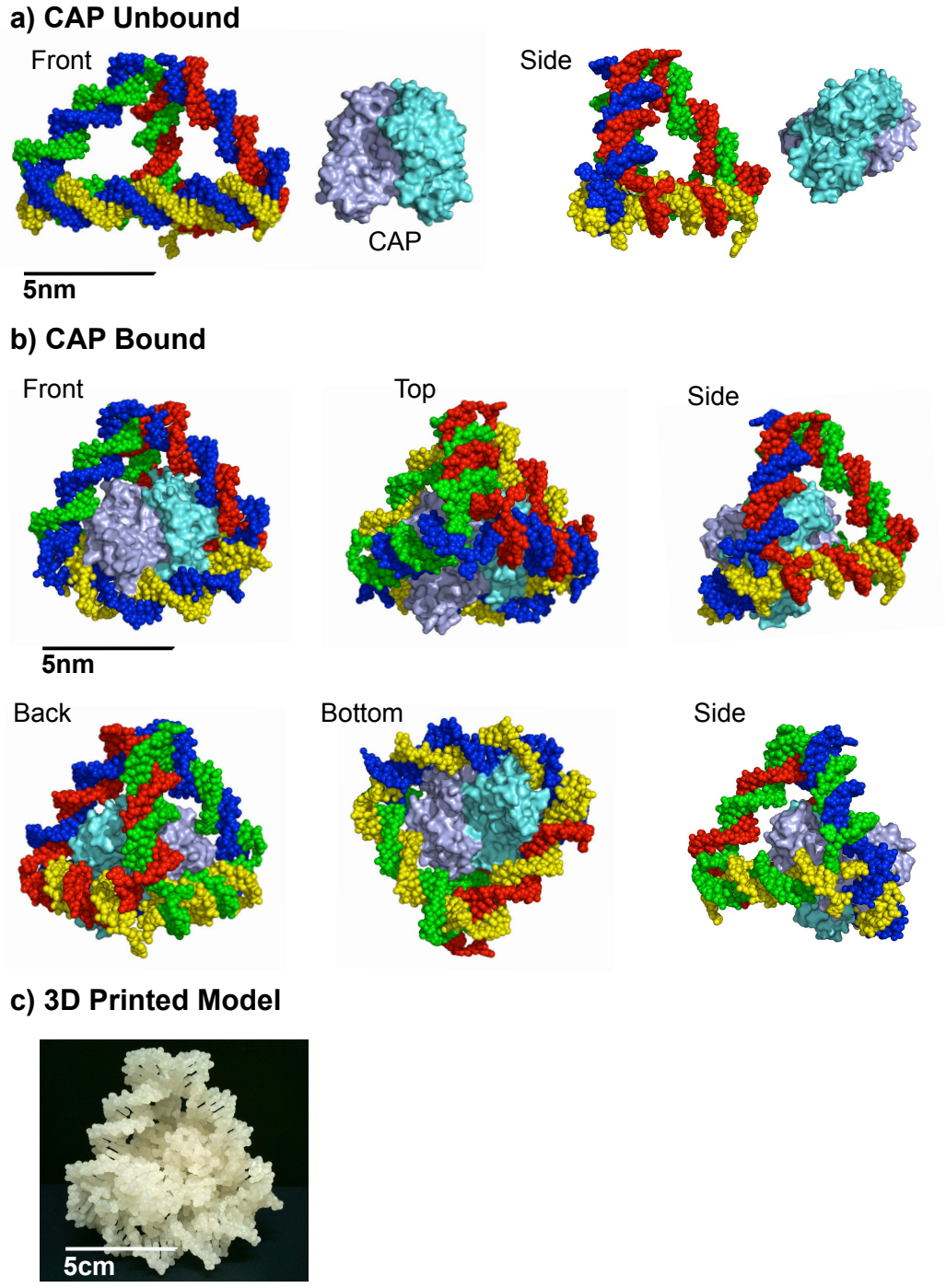


Figure 6.2.3: 3-dimensional scale model of the protein DNA cage from different perspectives with and without CAP bound. a) The unbound cage is shown next to CAP for comparison from the front and side perspectives illustrating its inability to fit in the cage with such dimensions. b) Due to CAP induced DNA bending in the bound state, the cage geometry changes to allow encapsulation. The cage with CAP encapsulated inside is shown from all perspectives. c) Picture showing a similar 3-dimensional scale model (approximately 7 orders of magnitude larger) that was fabricated using a 3D printer.

6.2.2 Results

6.2.2.1 Protein-cage Binding and Encapsulation

Conditions for cage formation and CAP binding are given in section A.2.4. Briefly, to form the cage, equal amounts of the component strands were mixed and annealed at 90˚C in annealing buffer for 3 min and cooled to 4˚C over 30s. CAP binding was performed in KG7 buffer at 14˚C for 10 min. PAGE gels were stained with Sybr gold for imaging.

Formation of the cage and CAP binding was initially observed using PAGE. Figure 6.2.4 shows a titration of CAP with cages containing inward and outward facing CAP binding sites. Lane 1 contains a 100 bp DNA ladder and lane 2, 'normal' tetrahedra with all 20bp edges from [46] as controls for comparison. Lanes 3-6 contain the titration with the inward facing site and lanes 7-10 with the outward facing site.

Lane 3 contains the cage with the inward facing binding site alone. A shift is observed compared with the normal tetrahedron of lane 2, as is expected for the larger molecule due to the DNA bending induced by CAP. Lane 7 contains the cage with the outward facing site alone, and appears at a similar position to that with the inward site of lane 3. The surprising result is that on application of a 0.3x, 3.3x and 6.7x excess of *active* CAP to both inward and outward binding site cages (lanes 4-6, and 8-10 respectively), strong binding is observed for the inward but not the outward case. For the inwards facing site, we see about a 50% shift in lane 4, moving to a complete shift in lanes 5 and 6, whereas we only see a faint double band occurring in lanes 8-10 for the outwards site suggesting much weaker affinity. The hypothesis was that CAP would easily bind to its binding site if presented facing outwards from the cage but would have more difficulty in binding the site facing inwards as it would have to travel through the cage faces. We saw from the cage geometry calculated in Figure 6.2.2 b) that it is possible CAP could fit through the largest window if that edge was bent. If CAP had the conformational flexibility to bind partially or fully to the 30bp edge of the cage, it may explain its ability to move inside.

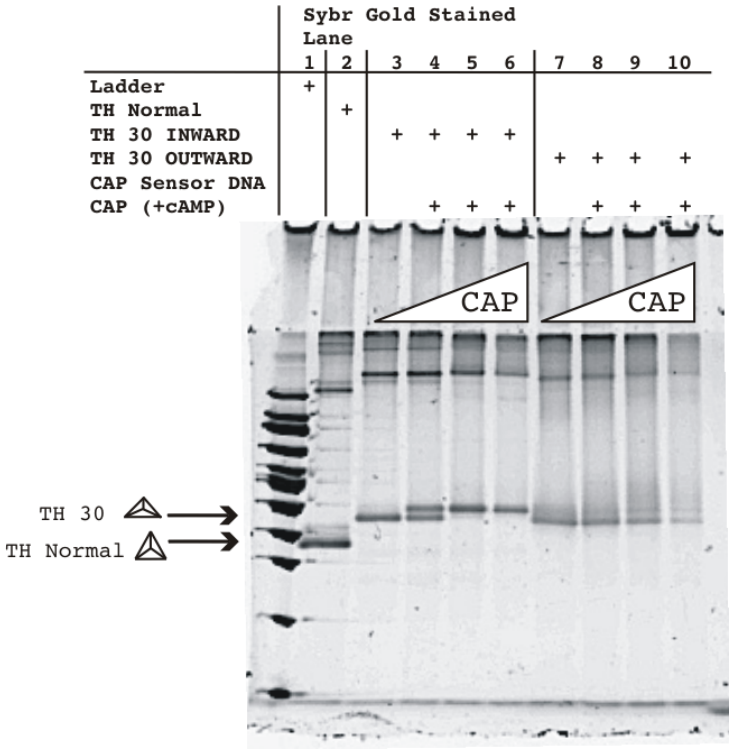


Figure 6.2.4: Titration of CAP with protein cages containing an inwards and outwards facing CAP binding site observed using PAGE. Each lane contains ~1 pmole DNA. Lanes 4-6 and 8-10 contain a 0.3x, 3.3x and 6.7x excess of *active* CAP respectively. A 100bp ladder is used in lane 1, and normal tetrahedra (from [46]) in lane 2 as controls. Gel conditions: 7.5% Tris-HCl (pH 8.8) separating, 4% Tris-HCl (pH 6.8) stacking, run in 1x Tris-Glycine at 4˚C.

CAP, also known as cyclic AMP receptor protein (CRP) requires the presence of cyclic AMP (cAMP) to bind to its recognition sequence. Binding of one cAMP molecule to each monomer of CAP is known to induce a large conformational change within the protein such that the DNA binding domain is oriented in such a way to facilitate site specific binding [135, 134]. It has been proposed that in the cAMP bound state, CRP from *Mycobacterium tuberculosis* (a close homolog to *E.coli* CAP), exhibits much enhanced flexibilty in the DNA binding domain [163] and connected helix due to release from the cAMP binding domain, resulting in large conformational dynamics ('swings') of the DNA binding domain itself. It is proposed that due to the conformational flexibility of CAP bound to cyclic AMP, it is possible the DNA binding domains of CAP occasionally make contact with the 30bp edge containing the consensus CAP binding site. If CAP partly bound this edge through one monomer, generating one kink and increasing the window size, this may aid binding of the second monomer in a cooperative fashion. Even if this interaction was rare, mani-

fested by a slower on-rate compared to binding linear DNA, once the CAP molecule was inside the cage it would have much more difficulty leaving due to the force of the cage trying to snap back to its original shape and the subsequent high local concentration of the binding site (put forward in section 6.2.1). It is expected that this interaction would result in a much slower off-rate compared to binding linear DNA containing the same binding site.

The lower affinity of CAP for the cage containing the outward facing binding site may be explained by two main factors. Firstly, similar to the inward binding site case, the bending induced by CAP will alter the conformation of the cage from its preferred unbent shape. If CAP binds and bends the DNA from outside the cage, the cage will try to resume its original shape, providing a force acting against DNA bending (and binding) pushing CAP off the cage. Secondly, as the binding site was shifted 5 bp to face outwards, the binding site was placed right next to one of the vertices of the cage. Such structures have been shown to demonstrate high structural rigidity [46] and thus DNA bending near the cage vertex would be more difficult. The off-rate of CAP from the outward binding site cage would therefore be expected to be faster than that from linear DNA with the same binding sequence.

It was also found using PAGE that CAP could bind to the inward facing site cage without cAMP although with a weaker affinity (data not shown). Without cAMP or at very low concentrations, CAP still retains some affinity for it's binding site [137], although it is much reduced. Hence in this case it seems that CAP could occasionally bind and enter the cage and once inside take longer to escape. As we see later in Figure 6.2.8, the affinity of CAP for linear DNA with the same binding site and without cAMP was not enough to cause a shift.

Lanes in Figure 6.2.4 also contain much higher molecular weight components are also formed in most lanes, which are the result of dimers and larger aggregates forming due to the inherent complementarity of the strands. Such larger molecules are common [155][161], and molecules of interest can be PAGE purified if required.

Having established CAP binding to fully formed cages, I explored whether it was possible to form the cage around the protein. Figure 6.2.5 shows the result of such an experiment where the complexes formed around the protein are compared with the fully formed cage alone and fully formed cages bound to CAP (as in Figure 6.2.4).

Lane 3 contains the protein cage with the outward facing binding site. On addition of CAP

to this cage (lane 4) we again see a double band occurring from incomplete binding due to the lower affinity of CAP for this construct. Lanes 5, 6, and 7 then show the outward facing binding site edge alone (strands 3 & 4), with the addition of CAP, and then with the addition of the other strands (1 & 2) required to complete cage formation respectively. Lane 5 shows the dsDNA alone; in lane 6 on the addition of CAP we see a clear shift, indicating strong binding of CAP to that edge. The further addition of strands 1 & 2 in lane 7 produces predominantly a higher weight band, whilst leaving a similar band to lane 6 due to incomplete binding[1]. This new band in lane 7, where the cage has been formed after CAP binding to one edge, appears slower than CAP bound to a pre-formed cage (the faint band in lane 4). It is likely that the presence of CAP has affected the correct formation of the cage, and strands 1 and 2 have annealed around the complex in the most favourable conformation which does not correspond to the design of the cage.

In lanes 8-12 I repeat the same experiments for the inward facing binding site. Lane 8 shows the protein cage with the inward binding site alone and in lane 9 the same sample with the addition of CAP. We see the shift caused from binding of CAP consistent with Figure 6.2.4. Lanes 10-12 show the inward facing binding site edge alone (strands 3 & 4), with the addition of CAP and then with the addition of the other strands (1 & 2) to complete cage formation respectively. Analogous to the outward site case, the inward site edge + CAP (lane 11) produces a clear shift compared to the DNA alone (lane 10) demonstrating strong binding of CAP. The further addition of strands 1 & 2 to this complex to form the full cage produces a higher band at an intermediate position between the unbound and bound pre-formed cages in lanes 8 & 9 respectively. A similar band to lane 11 is left behind due to incomplete binding. Again it would seem that the presence of CAP has affected correct cage formation. The band is similar in position to the bound pre-formed cage shown in lane 8, but further experiments would be required to check correct cage formation. As encapsulation seemed to not be a problem with pre-formed cages, simple binding reactions to the inward site cage were used for the rest of the work.

[1] Strands 1 & 2 were added at 37˚C, and not annealed using 90˚C, melting to avoid heat inactivating CAP.

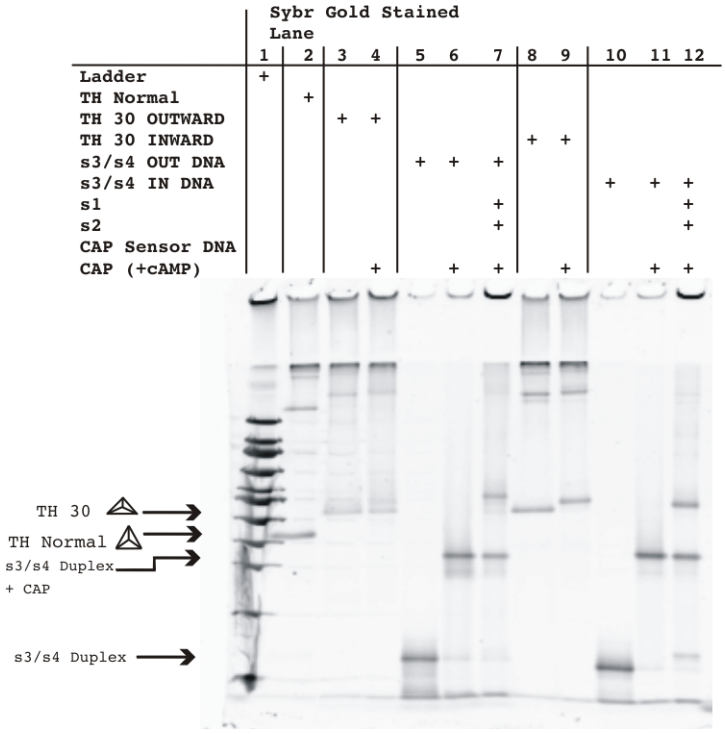


Figure 6.2.5: Binding of CAP to the protein cage with inward and outward facing binding sites either to a fully formed cage or with the cage formed around CAP bound to one edge as analysed using PAGE. Lane contents are shown above each lane. Each lane contains ~1 pmole DNA, and lanes with CAP used a ~3.3x excess of CAP. A 100bp ladder is used in lane 1, and normal tetrahedra (from [46]) in lane 2 as controls. Gel conditions: 7.5% Tris-HCl (pH 8.8) separating, 4% Tris-HCl (pH 6.8) stacking, run in 1x Tris-Glycine at 4˚C.

6.2.2.2 Confirmation of Protein Encapsulation Using Single-Molecule Fluorescence

To provide further evidence of encapsulation, fluorescence methods were used. Both the protein and cage were fluorescently labelled such that encapsulation could be detected using FRET. The cage was labelled green on the hinge base of strand 1 at vertex 3 with Cy3 (Figure 6.2.6 a)) and CAP red, by labelling site-specifically on each monomer a single cysteine with Alexa 647 maleimide (Figure 6.2.6 b)). As labelling would never be 100% efficient, the CAP dimer would be produced in four different labelled versions: no labels, single label (each monomer) or with both monomers labelled. CAP was labelled with a sub stoichiometric ratio of dye:CAP, hence predominately single labelled or unlabelled versions were expected. The approximate distance between the Cy3 donor label on the cage and the Alexa 647 acceptor position on each monomer is shown in Figure 6.2.6 c) as ~ 34Å for the proximal and ~50Å for the distal monomers. Given an R_0 of 52.3Å for the Cy3-Alexa647 pair (calculated using equation 2.3.5, Chapter 2, using κ^2 =

2/3), the expected FRET efficiencies for these two positions would be 93% for the proximal and 57% for the distal labelling position.

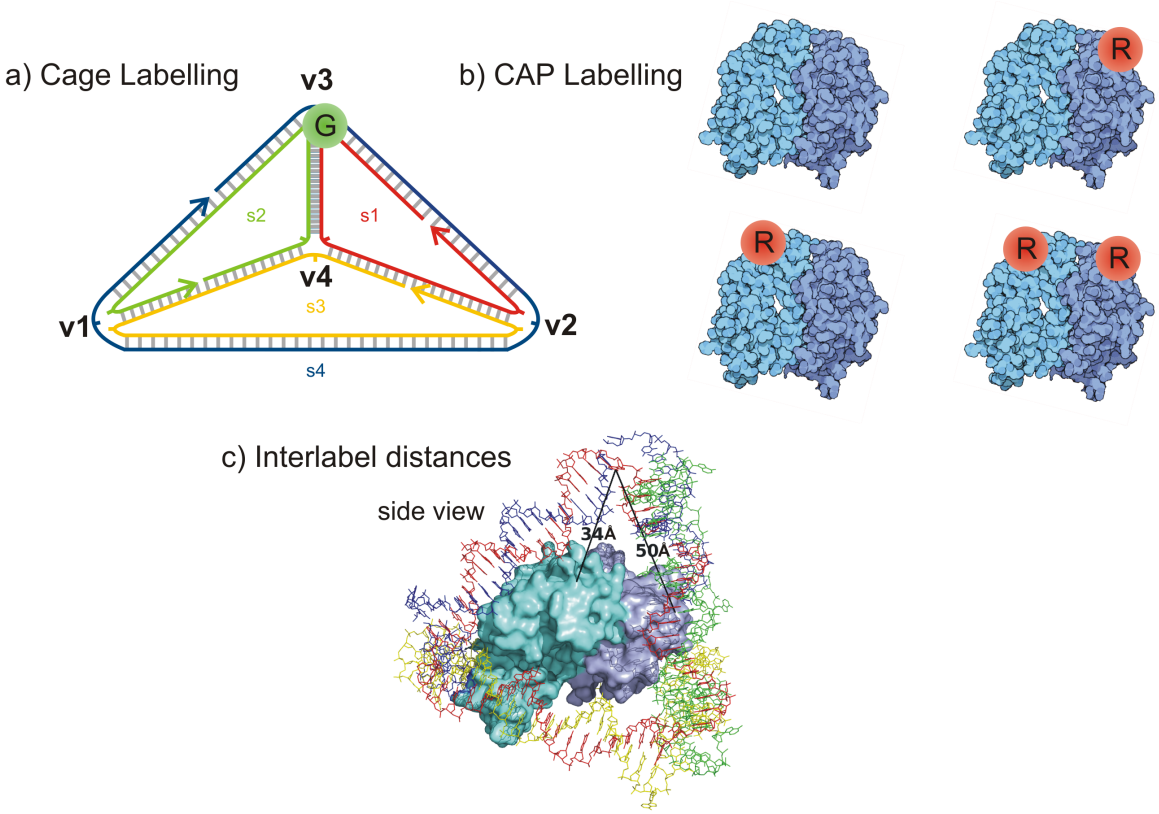


Figure 6.2.6: Cage and CAP fluorescent labelling scheme. a) The cage is labelled at the hinge base of strand 1 at vertex 3 with Cy3. b) One cysteine of each monomer of CAP was reacted with Alexa 647 maleimide resulting in the four labelling options. As the protein was sub stoichiometrically labelled, singly labelled or unlabelled dimers are expected to dominate the population. c) A model of CAP encapsulated within the cage showing the approximate interlabel distances from the single Cy3 site on the cage to the Alexa 647 site on each monomer.

Binding of the labelled CAP to the labelled cage was initially confirmed using PAGE (see section A.2) using the same conditions as before (section A.2.4). The bound complexes were then analysed using single-molecule confocal fluorescence. Figure 6.2.7 a) shows the ES histogram of labelled CAP encapsulated inside the labelled cage. We see two main populations, acceptor-only free CAP molecules with low S values, and CAP encapsulated within the cage with high-FRET (high E) mid-S values. The one-dimensional E histogram shows only the double-labelled molecules (with $S > 0.35$) indicated by the red box. Within the E distribution we see predominately a broad high FRET population with a mean around 74%. As CAP should be facing inwards to the

cage due to the positioning of the binding site (section 6.2.1), the donor on the cage and acceptor on CAP should be in close proximity. This high FRET population does indeed confirm CAP is positioned inside the cage. If CAP were bound on the outside of the cage the FRET expected would be much lower if not negligible.

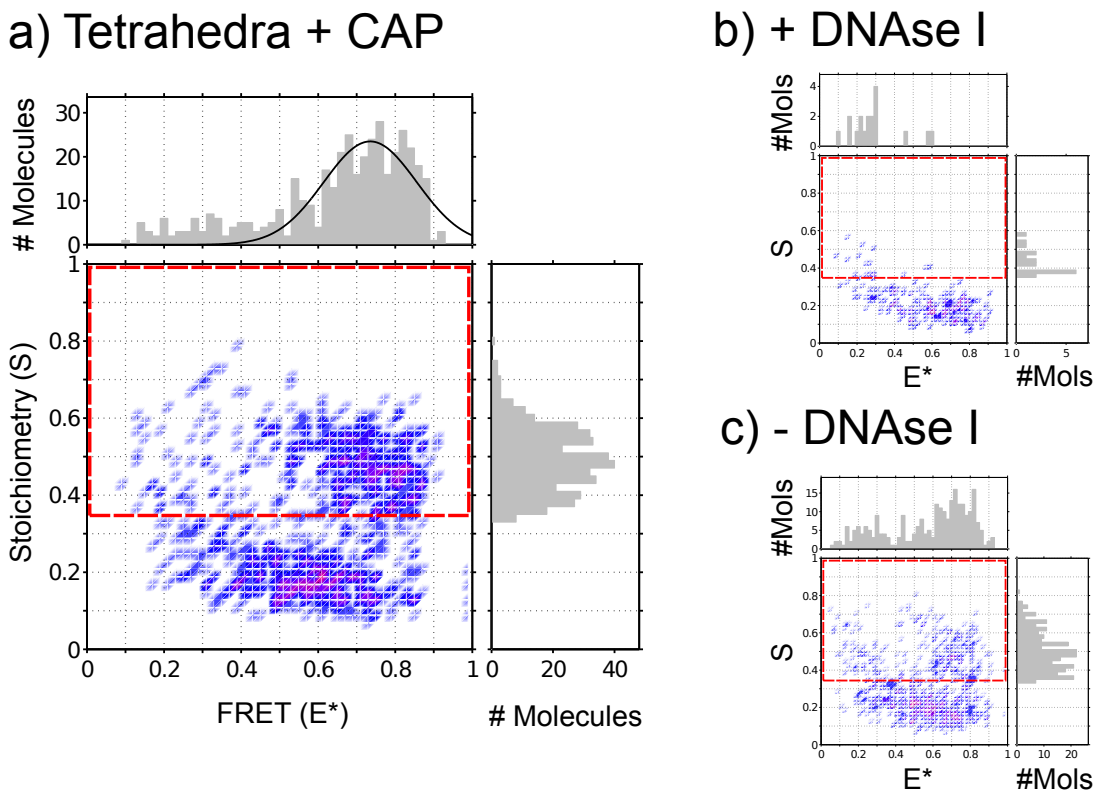


Figure 6.2.7: Single-molecule confocal data for protein encapsulation. a) ES Histogram of Alexa 647 labelled CAP encapsulated inside a Cy3 labelled cage. Two main populations are observed: an acceptor only population with low S and broad E corresponding to free CAP, and a high-FRET, mid-S population corresponding to CAP encapsulated inside the cage. b) The addition of DNAseI to a similar construct as in a) results in only a single free CAP population due to cage degradation. c) The same sample as b) without the addition of DNAseI shows high-FRET encapsulated CAP molecules. One-dimensional E* histograms (top) show double-labelled molecules with S > 0.35 as indicated by the red box.

Such a broad distribution is indicative of the two labelling positions on CAP, which as we saw earlier, for corrected FRET values should lie approximately at 57% and 93%. Longer acquisition times, and higher photon counts would allow for better distinction between these two assumed static populations. There are also a minority of lower FRET events in this distribution also. It is likely these arise mainly from random coincidence rather than protein dynamics within the cage.

Figure 6.2.7 b) and c) show an interesting control with the addition of DNAse I to cages encap-

sulating CAP. Figure 6.2.7 b) shows the result of forming a similar complex to that shown in a) but with the addition of the enzyme DNAse I. DNAse I is an endonuclease that non-specifically cleaves DNA, acting on single and double-strand DNA. Thus the addition of DNAse I should degrade the DNA cage releasing the encapsulated CAP. As we see from the ES histogram, there exists only a single population, corresponding to free CAP. The same sample, to which DNAse I was not added (Figure 6.2.7 c)) shows again both populations observed in a), the free CAP (low S, broad E) and the encapsulated CAP (mid S, high E*). It would also be interesting to probe the complexes in the presence of a protease such as proteinase K, to investigate whether the cage conferred any protection to the protein from such degradation machinery.

6.2.2.3 Protein Retention Within Cage

An important test was whether CAP remained encapsulated within the cage under conditions where it should dissociate. Removal of cyclic AMP was used to probe dissociation of CAP from the cage (inward site), the results are shown in the gel in Figure 6.2.8 a). A control where cyclic AMP was removed from CAP bound to linear DNA is shown in Figure 6.2.8 b).

Looking at Figure 6.2.8 a) left (sybr gold nucleic acid stain), we see in lane 3 the formation of the inward site cage, with the expected shift caused by binding of CAP in lane 4. In lane 5, cyclic AMP is then removed from the same sample used in lane 4, using a buffer exchange column (see section A.2). The interesting result is that the band in lane 5 appears at the same position as that in lane 4, showing that removal of cAMP from the CAP-cage complex does not cause dissociation of CAP from the cage. It is also confirmed from staining the same gel with sypro ruby protein stain (6.2.8 a) right), that protein is present in both lane 4 and 5, before and after cAMP removal. Lane 5 shows fainter staining due to sample dilution through the process of cAMP using buffer exchange.

To verify that removal of cAMP does indeed cause dissociation of CAP from its binding site, a similar experiment is shown in 6.2.8 b) whereby CAP is initially bound to a linear dsDNA and then cAMP is removed. Lane 1 shows the dsDNA alone that contains the same CAP consensus binding site as that of the cage. On the addition of CAP to this sample in lane 2, we see a complete shift from CAP binding. After buffer exchange to remove cAMP in lane 3, we see complete removal of the shifted complex and a band returning to a similar position of the unbound dsDNA.

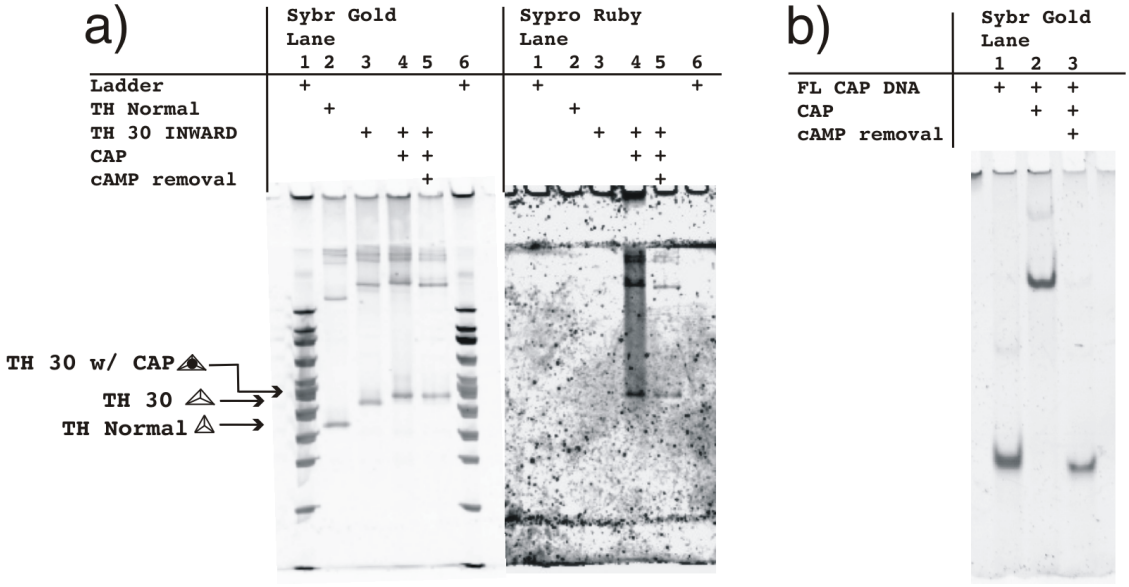


Figure 6.2.8: Removal of cAMP from CAP bound to either the inward site cage a) or to linear dsDNA containing the same binding sequence b). Lane contents are shown above each lane. a) Each lane contains 1 pmole DNA, lane 4 contains a 6.7x excess of CAP. cAMP was removed in lane 5 by buffer exchange (see section A.2). A 100bp ladder is used in lane 1 and 6, and normal tetrahedra (from [46]) in lane 2 as controls. The scan on the left shows nucleic acid stain Sybr gold and on the right, the protein stain Sypro Ruby. CAP is retained within the cage after cAMP removal. Gel conditions: 10% Tris-HCl (pH 8.8) separating, 4% Tris-HCl (pH 6.8) stacking, run in 1x Tris-Glycine at 4˚C. b) Each lane contains ~ 0.5 pmoles DNA, lane 2 contains a 2x excess of CAP. cAMP was removed in lane 3 by buffer exchange as in a). Removal of cAMP causes dissociation of CAP from linear DNA. Gel conditions: 10% 1x Tris-Glycine, no stacker, run in 1x Tris-Glycine RT at 4˚C.

Protein Off-rate Comparison: Cage and Linear DNA

To quantitate the increase in retention of CAP by the cage in comparison to linear duplex DNA, a competition experiment was performed to determine the off-rate from each. In this case ensemble fluorescence was used to measure the rate of donor dequenching of a complex between labelled CAP (acceptor; Alexa647) and either the labelled cage or linear DNA (donor; Cy3B) on the addition of a 50-fold excess of unlabelled competitor DNA. The inverse of the donor fluorescence (normalised) for each case is shown in Figure 6.2.9 with exponential fits to the data. For the cage data (green data points, red fit line) , we see clear biexponential dissociation behaviour containing a fast ($k_{d1} \approx 7.8 * 10^{-3} s^{-1}$) and slow component ($k_{d1} \approx 5.4 * 10^{-4} s^{-1}$) . This biexponential behaviour has previously been observed for CAP-DNA complexes [27] and may be attributed to a mixture of stable and unstable CAP-DNA complexes whereby stable complexes convert slowly to

unstable complexes which dissociate quickly.

Observing the data of CAP with duplex DNA (purple data points, blue line fit), we observe much faster dissociation. In this case the data was fitted to a single exponential. While fitting to a double exponential did achieve a marginally better fit, the rate of the slow component was so slow suggesting this was an experimental artefact such as photobleaching. The fast and slow components of dissociation in this data may be masked due to the relatively low time resolution of the experiment. A single exponential fit to the data results in an off-rate of $k_d \approx 4.2 * 10^{-2} s^{-1}$ which is approximately 17 fold slower when compared to a single exponential fit for the cage data ($k_d \approx 2.5 * 10^{-3} s^{-1}$). Thus the effect of caging CAP inside the cage does seem to have affected the kinetics of dissociation. It is interesting to note that while single-molecule complexes of CAP bound to the cage were observable (as shown in Figure 6.2.7), similar complexes using linear duplex DNA with the same binding site could not be observed (data not shown) which may be due to the faster rate of dissociation. A comparison of dissociation rate constants of the two cases, by obtaining the on-rates (using stopped-flow or similar) would be useful here.

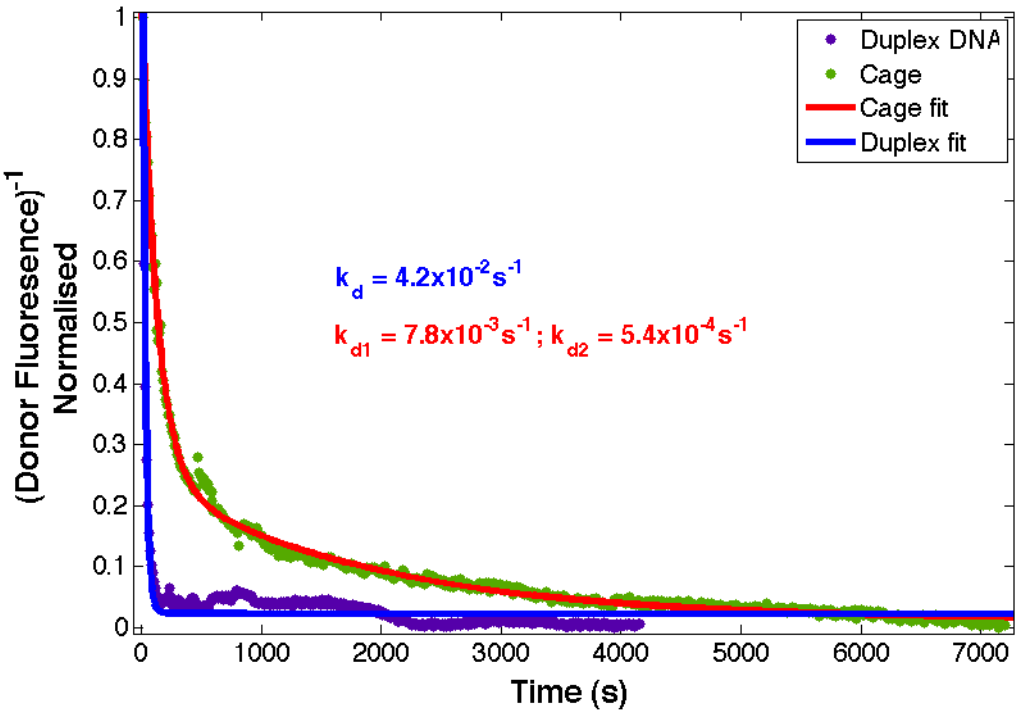


Figure 6.2.9: Comparison of off-rates of CAP from linear duplex DNA and the protein cage. In both cases 30nM of acceptor labelled CAP was added to 10nM of donor labelled DNA and left to equilibrate until a maximum FRET change was reached. A 50-fold excess of unlabelled competitor DNA containing the same CAP binding site was then added and donor fluorescence unquenching observed. The plots show the normalised inverse of donor fluorescence. Green and purple circles - cage and duplex data respectively. Red and blue lines - double exponential fit to cage data and single exponential fit to duplex data respectively.

6.2.3 Discussion

In this section I have demonstrated strong evidence of the encapsulation and entrapment of a protein (a transcription factor, CAP) inside a fully pre-formed DNA cage. The encapsulation was demonstrated initially using gel shift assays (PAGE) and further shown using single-molecule (sm) fluorescence. In sm experiments, an expected high FRET population indicated close proximity between the donor on the cage and acceptor on the protein observable only in the case that the protein was inside.

The dimensions of the cage without CAP-induced edge bending suggest that without CAP, the cage exists in a conformation in which CAP cannot fit. However, when bound (to an inward facing site) and bent by CAP, forms a window size that CAP can fit through. It is possible that

the conformational dynamics of CAP and the cage allow it to occasionally partially or fully bend that edge of the cage and move inside. Once inside, the ability of CAP to leave the cage would be impaired due to the force of the cage trying to adopt its original conformation and the subsequent high local concentration of the binding site presented to CAP. Similarly, binding to the cage with an outward facing site is expected to be weaker as the cage would not surround CAP as it tried to adopt its original configuration. This configuration change would try to pull the DNA out of its bent form, causing a shift towards CAP unbinding and subsequent diffusion away from the cage.

The entrapment of CAP was demonstrated with PAGE under conditions that would normally promote dissociation (the removal of cyclic AMP). A binding competition assay comparing CAP bound inside the cage and to linear duplex DNA with the same binding site suggested at least an order of magnitude slowing in off-rate for the case of the cage. Thus the dissociation kinetics of CAP from its binding site contained within a cage have been dramatically affected, which would likely also affect the dissociation constant (K_D). It is also possible that the *association* kinetics (on-rate) may have been affected. The geometry of the cage windows is such that it would be difficult for CAP to fit through without DNA bending and therefore the on-rate may also have been slowed. Further work for this project would include obtaining this rate, to quantitate better the shift in equilibrium this cage has introduced to CAP binding.

Having demonstrated the original goal of non-covalent single protein encapsulation within a DNA cage either as a potential drug delivery vehicle or to measure multiple protein-DNA distances, the project developed into an interesting study in terms of protein-DNA interactions. The unexpected result being the shift in the binding equilibrium of CAP to its consensus binding site introduced by the cage. CAP could bind to a fully pre-formed cage readily, but would leave the cage with a dramatically slowed rate compared to linear duplex DNA. The mechanism of CAP entering the cage also raises interesting questions. As the dimensions of the cage would seem to preclude CAP from entering, there is assumed to be significant conformational flexibility of CAP to allow its entry, as supported by recent literature [163, 135, 134] which suggest that, in the presence of cAMP, there is enhanced flexibility of the DNA-binding domains. These questions lie outside the scope for this chapter but I aim to address them in a publication for this work.

6.3 Reversibly Opening DNA Cage

6.3.1 Design

The second half of this chapter is dedicated to the design of a reversibly opening cage.

To fully open the tetrahedron for protein release, instead of extending one edge as in [156] the cage was designed such that at least one edge could be reversibly broken. Designs involving edges that were held together chemically or by a protein were considered but a design involving just DNA was chosen as the simplest.

The design for the reversibly opening cage is shown in Fig 6.3.1 (left). Following a similar design of the DNA tetrahedron [46], the opening cage consists of DNA strands that run around each face of the shape of a tetrahedron. Three strands (s1, s2, and s4(-4) Overhang) run around one complete face and two strands together run around one face (s2(-4) 3'Cy3 & s2(-4) 5'Cy5). Each strand contains complementary sections such that it binds to its neighbouring strands forming an edge. The 5' and 3' ends of each strand would meet to form a nick, the placement of which could be altered. Two nicks were placed on one edge to form an overlap holding that edge together. By adding an overhang onto the 3' end of the s4 strand (s4(-4) Overhang), it should be possible to bind a complementary strand to this overhang to displace the connection between s2 and s4, therefore breaking that edge (Fig. 6.3.1 right). Furthermore, by altering this complementary strand to contain itself an overhang, this 'fuel' strand could be removed by a complementary 'anti-fuel' strand allowing re-closing of the tetrahedron.

The face that comprises two strands (s2(-4) 3'Cy3 & s2(-4) 5'Cy5) is to allow the addition of a Cy3 (donor) and Cy5 (acceptor) fluorophore on opposite sides of the opening edge such that opening could be monitored via FRET. This was originally one strand (strand 2). At the time it was not possible to make an oligo of the length required that with a double modification. All nicks on this tetrahedron design (except for the two on the opening edge) contained 5' Phosphate groups and so could be ligated closed if desired to increase the rigidity of the structure or to check correct formation of the cage.

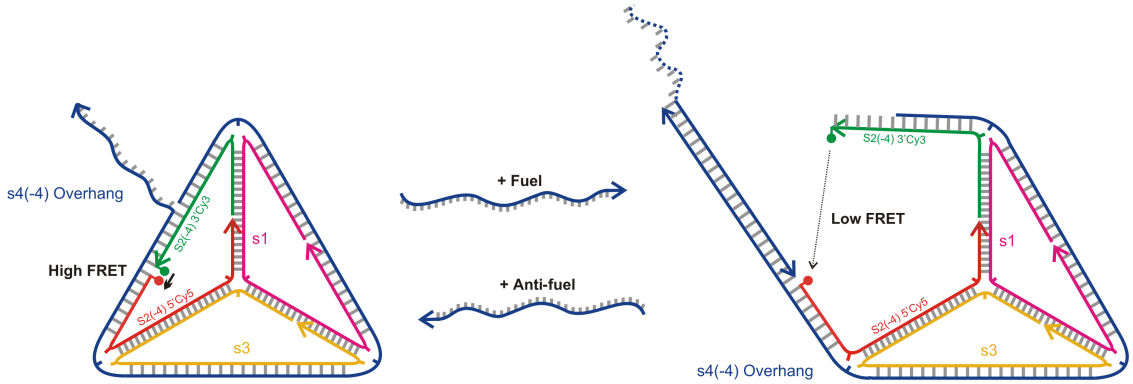


Figure 6.3.1: Schematic illustrating the operation of the reversibly opening cage. (Left) The cage is formed from five DNA strands that form the shape of a tetrahedron. Each strand runs around one or part of one face hybridising to their neighbour forming an edge. The cage is designed to open on the double nicked edge, containing an overlap between strand 4 (s4(-4) overhang) and strand 2 (split into two s2(-4) 3'Cy3 and s2(-4) 5'Cy5). An overhang is incorporated at the 3' end of strand 4 such that a 'fuel' strand could be added to outcompete the overlap on the opening edge and open the tetrahedron (right). The fuel strand contains an overhang that can be removed by a complementary 'anti-fuel' strand. Opening is monitored by FRET between the Cy3 (donor) and Cy5 (acceptor) fluorophores placed at opposite sides of the opening edge.

6.3.2 Results

6.3.2.1 Double Nicked Edge Stability

As this design (Fig. 6.3.1) used two nicks on one edge, it was a concern whether this edge would be stable under the conditions previously used for the regular 20bp edged tetrahedron [46] on which the design was based. Hence initial experiments involved checking the edge stability of the cage. The gel in Fig. 6.3.2 shows a comparison between cages made from combinations of different strands:

- Sample: Double nicked edge regular DNA tetrahedron with no ligation. Strands 2 and 4 contain no 5' phosphates (i.e. non ligatable).
- Control 1: Regular DNA tetrahedron: formed from the 4 original strands used in [46], all strands phosphorylated.
- Control 2: Equivalent to control 1, but with no 5' phosphates on strands 2 and 4.
- Control 3: Equivalent to 'sample' but with 5' phosphates on all strands. All nicks ligated closed.

All cages were annealed in equimolar concentrations in annealing buffer and then PAGE purified (see section A.2). Ligation was performed with NEB T4 DNA ligase (detailed in A.2).

From the gel (Fig. 6.3.2), it does appear that the double nicked edge cage does have some inherent instability as it appears as a smear (lane 2) compared to the regular DNA tetrahedron (lane 3), or to the same sample that has been ligated closed (lane 5). It is unclear whether there exist tetrahedra in this lane that are interconverting between states or exist in multiple distinct conformations, however the migration difference to the other controls seems to certainly indicate undesirable structures. This may be due to the fact the overlap between the two nicks formed only a 6bp overlap. Despite this, a 6bp overlap was chosen to avoid placing the nicks too close to each vertex, which would further decrease the stability of the whole structure.

To counter the instability on this edge, it was decided that the salt concentration ($MgCl_2$) used in the annealing buffer could be varied. The annealing buffer used in Fig. 6.3.2 was 10mM Tris and 5mM $MgCl_2$, whereas the work performed in [156] used also 10mM Tris but 15mM $MgCl_2$ in the annealing buffer and 16 mM $MgCl_2$ in the gel mixture itself.

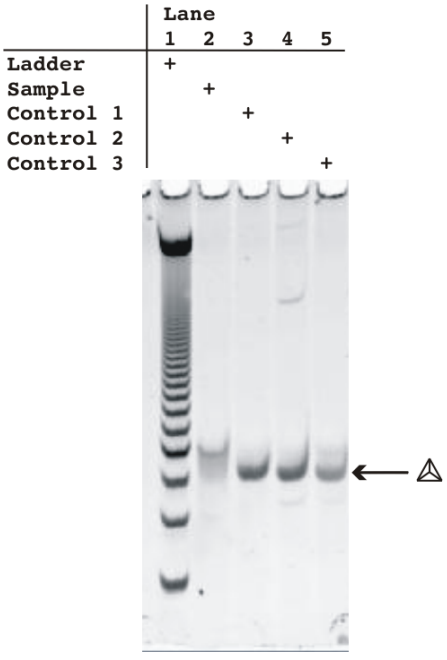


Figure 6.3.2: Edge stability test gel for double nicked edge. Lane contents shown above each lane. All cages used were initially gel purified. 'Sample' shows clear shift from other bands due to potential instability of the double-nicked edge. Indicated point - regular DNA tetrahedron migration point. Gel conditions: 6% 1x TAE (Tris-acetate-EDTA) run in 1x TAE at RT.

Figure 6.3.3 shows the results of annealing the cage design 'sample' in 10mM Tris and 5, 10, 15 and 20mM $MgCl_2$ (all pH 8) annealing buffer and running these designs in native gels, one containing no $MgCl_2$ and the other containing 16mM $MgCl_2$. In the gel containing salt (16mM $MgCl_2$), the samples run much closer to control 1 (the regular DNA tetrahedron). The band that is considered closest to the control is deemed the sample containing 15mM $MgCl_2$. Hence for the rest of the experiments in this work, the annealing buffer used contained 10mM Tris and 15mM $MgCl_2$, and the gels used contained contain 16mM $MgCl_2$. Also to further aid the overall structure stability, the nicks not on the opening edge could be ligated closed.

Still however the bands in Figure 6.3.3 are less well defined than that of the regular DNA tetrahedron suggesting there remained some instability. It was found that running the gels at 4˚C would result in much more defined bands (demonstrated in the following gels) and hence this approach was used for all experiments from then on.

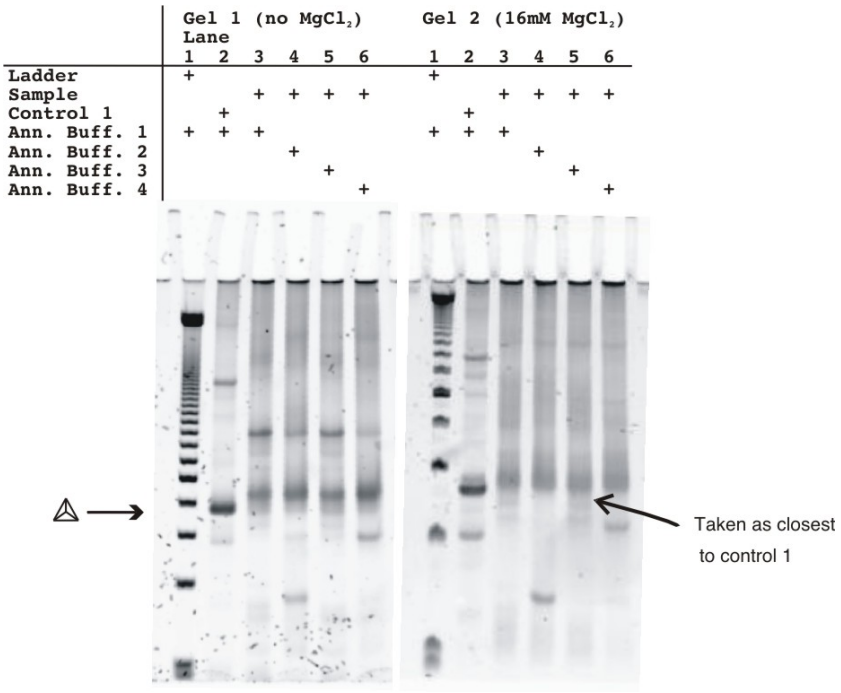


Figure 6.3.3: Salt variation gel experiment. Lane contents shown above each lane. Annealing buffers all contain 10mM Tris and different concentrations of $MgCl_2$. Ann. Buff 1, 2, 3, and 4 contain 5, 10, 15 and 20 mM $MgCl_2$ respectively. Gel 2 (right) which contains 16mM $MgCl_2$ shows all 'sample' designs to be much closer to the control 1 band (regular DNA tetrahedron). Of these, the sample containing 15mM $MgCl_2$ is considered closest (as noted with arrow). Gel conditions: 6% 1x TAE (Tris-acetate-EDTA) run in 1x TAE at RT with or without 16mM $MgCl_2$ at RT.

6.3.2.2 Correct Formation Of Reversibly Opening Cage

As the conditions for forming and running the samples had been established, an opening cage design with a double nicked edge and overhang on strand 4 was constructed (as portrayed in Figure 6.3.1). The 14bp overhang was designed such that it did not bind any other section of the tetrahedron.

Figure 6.3.4 demonstrates correct formation of the opening cage via native (left) and denaturing PAGE (right). From the native gel, we see a double band occur in lane 3 for the opening cage around the level of the normal tetrahedron ('Control 1') in lane 2. The upper band contains the opening cage design as it has slower mobility than the regular DNA tetrahedron (due to the extra overhang) and also displays fluorescence in both the Cy3 and Cy5 emission channels (green and red becoming yellow using a false colour representation from the combined emission scans). The lower green only band is likely due to using an imperfect 1:1 stoichiometry of all the required strands. Higher molecular weight bands are also present in both samples and are assumed to contain aggregates of many strands due to their inherent high degree of complementarity. Hence for later experiments PAGE purification of the correct structures was necessary.

To further confirm correction formation, cage samples could be ligated and then analysed via denaturing PAGE. By ligating nicks in DNA cages of tetrahedron shape that contain 5' phosphate groups, closed DNA circles (catenanes) will form, and, if two or more strands had been hybridized together on the same edge then a double or higher catenane will form. For example if a regular DNA tetrahedron comprising four 5' phosphorylated strands were to be ligated and run on a denaturing gel, a catenane consisting of four interlocked circles would be expected. However due to imperfect ligation, one would expect contributions of all other combinations of ligation to further include three and two interlocked catenanes, a single DNA catenane and single linear DNA strands. Such a technique has been used previously to confirm the correct formation of the regular DNA tetrahedron[46].

Shown on the right in Figure 6.3.4 is a denaturing PAGE gel of such a ligation experiment. The samples used were those from in the previous section (6.3.2.1) and also includes a single-stranded 63-mer and the opening cage design (named 'TetOH'). All structures have been treated with ligase, and in the adjacent lane, with ligase and then ExoIII. ExoIII nuclease is used to test for

successful ligation as it catalyses the removal of mononucleotides from the 3' termini of duplex DNA (including nicks in dsDNA).

After treatment with Exo III, each tetrahedra performs as expected on the gel:

Lane	Sample name	Sample Description	Catenanes observed
3	Sample	Double nicked edge regular DNA tetrahedron, two phosphorylated strands	single and double
5	Control 1	Regular DNA tetrahedron, all strands phosphorylated	single, double, triple and quadruple
7	Control 2	Regular DNA tetrahedron, two strands phosphorylated	single and double
9	Control 3	Double nicked edge regular DNA tetrahedron, all strands phosphorylated	single, double, triple and quadruple
13	TetOH	Opening cage design (double nicked edge & overhang), non opening edge strands phosphorylated. Differs from 'Sample' by inclusion of overhang.	single and double

Table 6.1: Summary of results from Figure 6.3.4.

Before treatment with Exo III, single strands are still present, including some with unexpected molecular weights (e.g. lane 2 top band). It is thought that this may be due to incompletely denatured complexes that are still hybridised to some degree. These bands are all fully removed after treatment with ExoIII demonstrating that they are not catenanes of unexpected size.

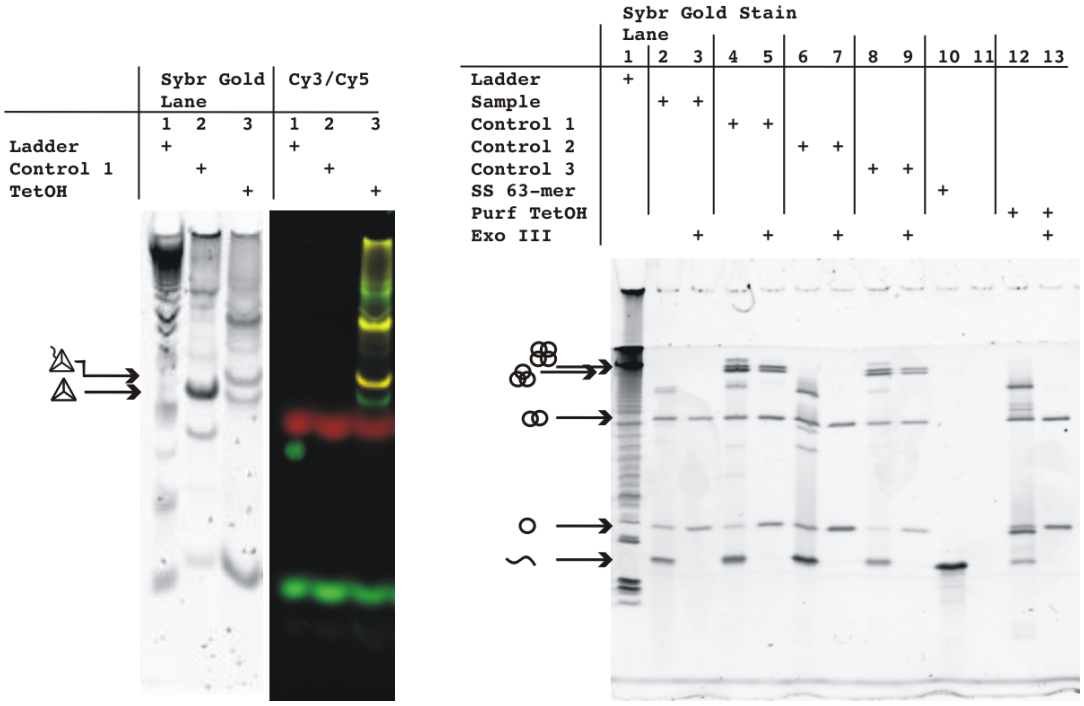


Figure 6.3.4: Confirmation of correct formation of opening cage design using native (left) and denaturing (right) PAGE. (Left) Native PAGE comparing the relative mobilities of the regular DNA tetrahedron ('Control 1' - lane 2) to the opening tetrahedron (lane 3). A 50bp DNA ladder is included in lane 1. Gels were stained with sybr gold with the sybr gold scan shown on the left and a false colour combined Cy3 and Cy5 emission scan shown on the right (Cy3 - green, Cy5 - red). Gel conditions: 6% 1x TAE containing 16mM $MgCl_2$ run at 4°C. Right - Denaturing PAGE of the following structures (described in the text): sample, control 1, control 2, control 3 (as given in section 6.3.2.1), single strand 63-mer, and the PAGE purified opening cage design ('TetOH') stained with Sybr gold. A 50bp DNA ladder is included in lane 1. All structures are treated with ligase and shown before and after application of Exo III. Lane contents shown above each lane. The migration positions of single strand, single circle, double/triple/quadruple catenanes are indicated on the left of the gel scans. Gel conditions: 5% stacking layer in Tris-HCl (pH 6.8), 13% separating layer in Tris-HCl (pH 8.8) both containing 7M UREA.

6.3.2.3 Cage Opening Experiments Using PAGE

Seven different input strands (2 negative controls) were designed to alter (or not alter) the structure of the opening cage design. Strand names and targets are given in Table 6.2.

Input Strand Name	Target	Length (nt)
Strip20	14nt overhang and 6bp of double-nicked edge overlap	20
StripWrong	14nt overhang, no complementarity to double-nicked edge overlap	20
StripWrong2	14nt overhang only	14
Strip Fuel	14nt overhang and 6bp of double-nicked edge overlap. Contains its own 10bp toe-hold	30
Strip Anti-fuel	Strip fuel via 10bp toe-hold	30

Table 6.2: Input strand designs for the opening cage.

Figure 6.3.5 tests the effects using PAGE of the all the stripping strands on equal amounts of opening cage design ('TetOH' - **Tet**rahedron **O**ver**H**ang). For the input strands, a large excess was used (~30x).

Lane 1 contains a 50bp ladder and lane 2, a regular DNA tetrahedron to be used as reference points. The opening cage is shown in lanes 4 and 5 before and after PAGE purification. The results of applying the different input strands are as follows:

- Lanes 6: Strip20 - designed to open the cage. A clearly slower band than the regular TetOH is observed consistent with the larger volume of an opened tetrahedron. The sharp band also suggests that the opened structure has some stability.

- Lanes 7 & 8: Strip Wrong and Strip Wrong2 respectively, used as negative controls. Both cause shifts with respect to unmodified TetOH but not by as much as Strip20 suggesting binding to the overhang but not opening the cage. StripWrong 2, designed to bind to the 14 bases of the overhang only, also caused less of a shift than Strip Wrong - a longer strand of 20 bases, with 6 mismatched bases to the double-nicked edge overlap.

- Lane 9: Fuel strand: to enable reversible opening. A double band is observed around the level of the shift caused by Strip20. In this case the fuel strand is assumed to either bind and open or just bind the overhang of TetOH i.e. remaining closed. If the fuel strand bound and opened the cage then the shift should be slightly slower than that of strip20 due to its own 10 bp toehold. If the fuel strand bound the overhang alone but did not open the cage a slightly faster band should be observed. The relative shifts in this lane would seem to confirm this. Overhang and overlap length could be optimised to generate all open or closed structures to avoid binding but no opening.

- Lane 10: Strip Anti-Fuel, another negative control. The target for this strand is strip fuel and so should not interfere with TetOH. No shift is observed compared to the unmodified TetOH.

- Lane 11: Strip-Fuel then Strip Anti-Fuel (2x excess to fuel), to open and then close TetOH. Two bands are visible in this lane. The lower band is expected, a closed cage migrating to the level of the purified TetOH. The upper band appears at the level of the lower band of Strip Fuel in lane 9 (i.e. Strip Fuel bound to overhang only and TetOH closed). It would seem that the Strip Anti-Fuel had mostly removed the Strip Fuel strands but not completely. A higher excess (to strip fuel) would be required to completely remove the strip fuel strands. Again, some optimisation fuel toehold length might be necessary to improve the ability of the Anti-Fuel stand to remove the Fuel strand. It is also interesting to note that in the sybr gold scan (Figure 6.3.5 left), there exists a wide dark band migrating near the bottom of the gel. This corresponds to the waste product of hybridized fuel and anti-fuel strands formed after opening and then re-closing the cage.

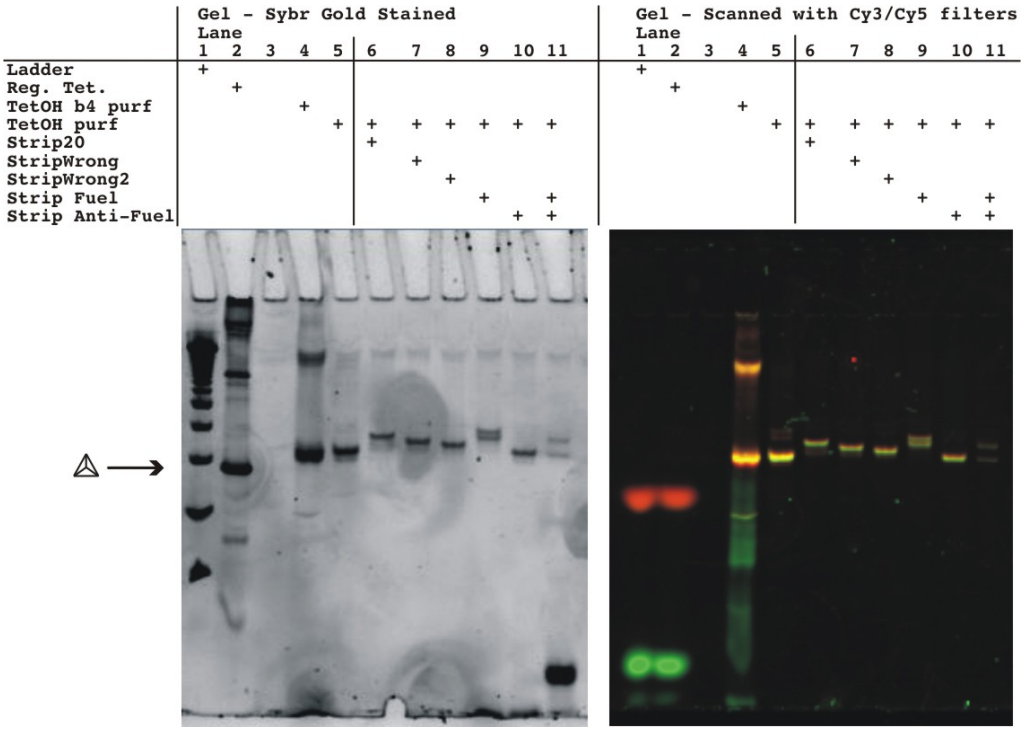


Figure 6.3.5: Native PAGE testing effect of all input strands (given in Table 6.2) on equal amounts of TetOH. Lane 1 includes a 50 bp ladder and lane 2, a regular DNA tetrahedron (indicated) as reference points. Lanes 4 and 5 show TetOH before and after purification. Lanes 6-11 contains TetOH + different input strands, the contents are shown above each lane. Lane 11 uses Strip Fuel first and then Strip Anti-Fuel second (2x excess, after 30 min). The scan on the left shows the gel stained with sybr gold, and the scan on the right shows false-colour representation of Cy3 (green) and Cy5 (red emission). Gel composition: 6% 1x TAE containing 16mM $MgCl_2$ run at 4˚C.

6.3.2.4 Cage Opening Experiments Using Ensemble Fluorescence

Demonstration Of Opening

To further investigate cage opening, ensemble fluorescence was used to probe the relative FRET efficiency between the Cy3 and Cy5 fluorophores placed at either side of the opening edge (see Fig 6.3.1). Using FRET as a reporter would unambiguously identify distance changes between the fluorophores resulting from tetrahedron opening.

Figure 6.3.6 (top) shows the changes in Cy3, Cy5 and FRET fluorescence signals over time from PAGE purified opening tetrahedra (TetOH) on application of 'Strip20' (designed for non-reversible opening). Cy3 excitation/emission is shown in green, Cy5 excitation/emission in red and Cy3 excitation/Cy5 emission in blue (FRET signal).

Initially (from t = 1500s onwards) only PAGE purified TetOH is contained in the cuvette.

At t = 2820s, Strip20 is added to the mixture and it can clearly be seen that both the Cy3 and Cy5 intensities immediately rise, eventually reaching saturation. The behaviour in these graphs is consistent with the hypothesis that the fluorophores are subject to contact-mediated or static quenching due to their close proximity (< 1 nm) whereby most of the absorbed energy is released as heat instead of light [164, 165]. Contact-mediated quenching has recently been used as an assay to detect transcription bubble expansion during abortive initiation by bacterial RNA polymerase [141]. Here, on application of the stripping strand, strip20, it is assumed that the opening edge has been broken; the stripping strand has outcompeted the 6 bp overlap holding edge together resulting in unquenching of the fluorophores.

In contrast to the large change in fluorescence signal of Cy3 and Cy5, the FRET signal does not vary as greatly. It is observed that after application of the stripping strand, the FRET signal does seem to gradually decrease. However it was unclear whether this was a reliable FRET signal due to the extreme proximity of the fluorophores (< 1nm) which resulted in high quenching. Because of quenching, the FRET signal in the closed state would also have a lower signal to noise ratio, increasing the error in the measurement. In the open state the fluorophores should be far apart, resulting in much lower FRET, a state less sensitive to distance changes. Thus in both open and closed states, the fluorophore positions are sub optimal for observation of opening via FRET. Whilst quenched fluorescence can be a useful tool for monitoring biomolecular interactions [141], in this case, to monitor opening and closing via FRET, the fluorophore positions would have to be altered to avoid quenching in the closed state. Having an unambiguous FRET signal would also allow calculation of accurate distance changes between the open and closed states using single-molecule fluorescence if required.

To test whether reliable FRET signal changes between the open and closed states could be observed, the placement of the Cy5 acceptor fluorophore on strand s2(-4) 5'Cy5 was moved further from the Cy3 donor, to an internally labelled position on the hinge base forming the vertex between itself and strands s3 and s4(-4)Overhang (see Figure 6.3.1). The spacing between the Cy3 and Cy5 fluorophores in the closed state should therefore be ~8 bp (~2.7 nm), within the observable range of FRET. Cages made using this modified strand are denoted 'internally labelled TetOH'.

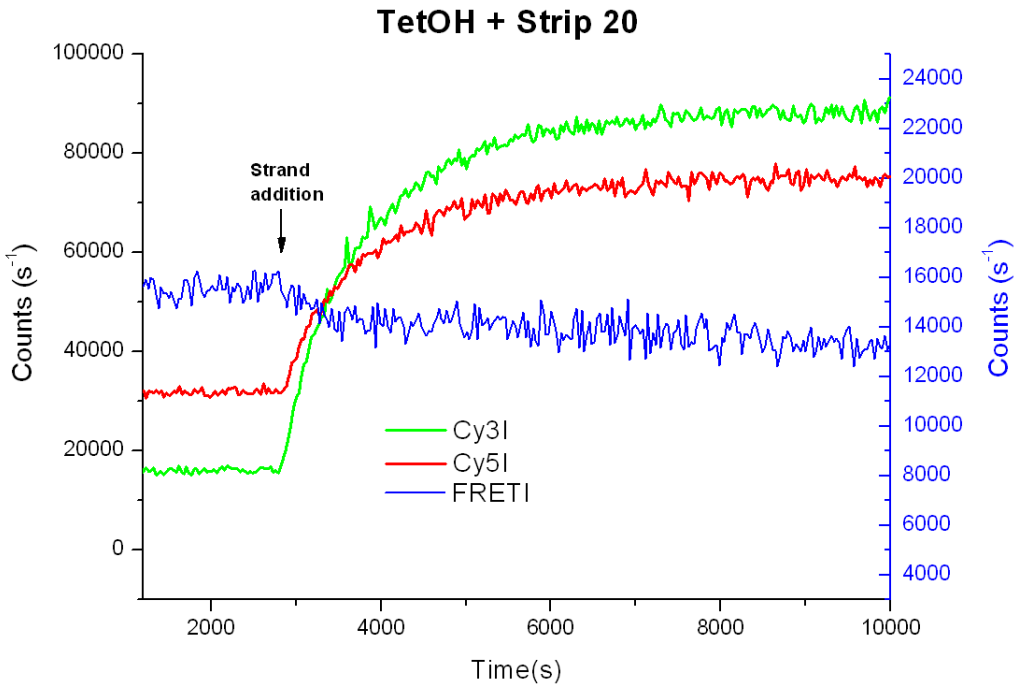


Figure 6.3.6: Bulk fluorescence vs. time of ~7 pmoles PAGE purified TetOH before and after the addtion of ~10 pmoles 'Strip 20' (for irreversible opening). Cy3 excitation/emission is shown in green (Cy3I), Cy5 excitation/emission is shown in red (Cy5I). FRETI (Blue) is the intensity of the FRET signal i.e. Cy3 excitation/Cy5 emission. After the addition of 'Strip20' at t = 1750s (indicated), Cy3 and Cy5 emission increases due to unquenching and FRET intensity decreases slightly.

Figure 6.3.7 shows the changes in Cy3, Cy5 and FRET fluorescence signal, and also FRET efficiency, using PAGE purified internally labelled TetOH on application of Strip20. The same colours are used as in Figure 6.3.6 with the addition of FRET efficiency shown in purple. The signal to noise in this experiment is higher than in the previous because of a higher concentration of sample being used.

Near the start of the experiment (from t =1200s onwards), only PAGE purified internally labelled TetOH is contained in the cuvette. At t = 1750s Strip20 is added to the mixture, resulting in a sharp increase in the Cy3 channel, a slight increase in the Cy5 channel and a sharp decrease in the FRET signal (and efficiency) as expected. Before the addition of the stripping strand, Cy3 is being quenched by its proximity to the Cy5 dye via FRET. Upon addition of the stripping strand, the cage opens, increasing the distance between Cy3 and Cy5 resulting in a large increase in Cy3 fluorescence due to less energy transfer to Cy5. Correspondingly, the FRET signal (FRET I) decreases

sharply on the addition of strip20. This much greater signal change than that using the quenched fluorophores in Figure 6.3.6 confirms the validity of moving the fluorophores further apart, giving an unambiguous picture of opening via FRET.

Observing Cy5 fluorescence, we see that it slightly increases on addition of Strip20. As any direct excitation of the now unquenched Cy3 fluorophore from Cy5 excitation is assumed to be negligible, it appears that the Cy5 in the closed state is still *marginally* quenched by proximity to the Cy3 dye, however not nearly to the same extent as in Figure 6.3.6 still allowing for the strong FRET signal to be observed.

In terms of a change in calculated FRET *efficiency*, there is sharp decrease from ~43% to saturation at ~11%, upon addition of the stripping strand, again confirming the large increase in distance between the two fluorophores, demonstrating strong evidence for tetrahedron opening.

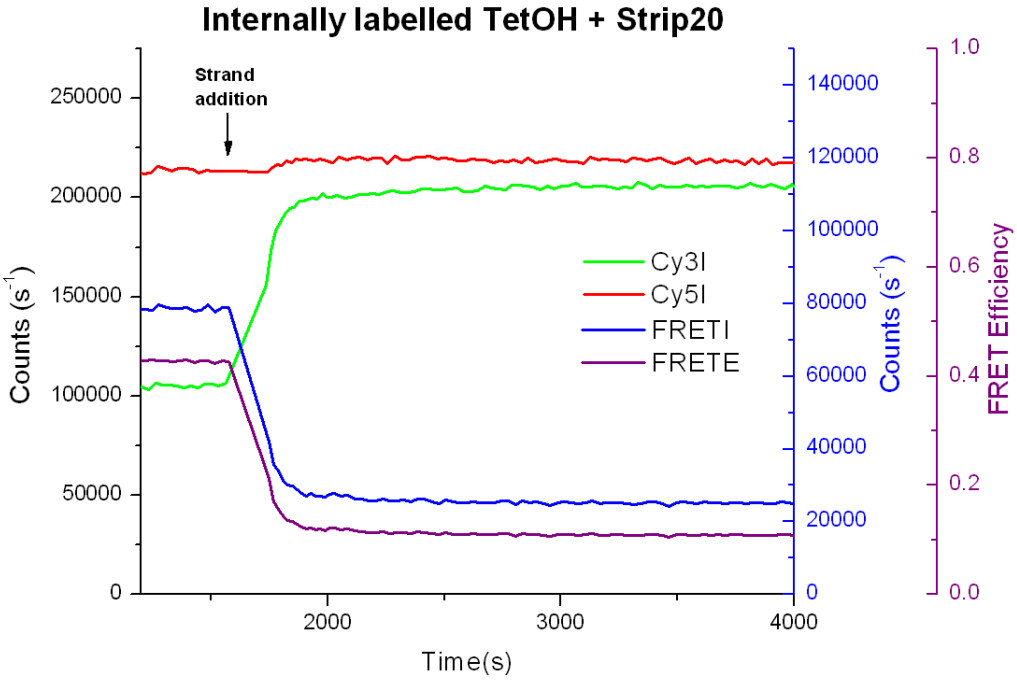


Figure 6.3.7: Bulk fluorescence vs. time of ~44 pmoles PAGE purified internally labelled TetOH before and after the addition of ~120 pmoles 'Strip 20' (for irreversible opening). Cy3 excitation/emission is shown in green (Cy3I), Cy5 excitation/emission is shown in red (Cy5I). FRETI (Blue) is the intensity of the FRET signal i.e. Cy3 excitation/Cy5 emission. The calculated FRET efficiency (FRET E) is shown in purple. After the addition of 'Strip 20' at t = 1750s (indicated), Cy3 emission increases, FRET intensity and efficiency decreases, and Cy5 emission increases slightly demonstrating strong evidence of cage opening.

Negative Control

The remainder of the experiments in this section use the TetOH quenched construct.

As a negative control, the same experiment as that in Figure 6.3.6 was carried out with addition of 'StripWrong' instead of 'Strip20'. StripWrong contains complementarity to the overhang region of TetOH but has 6 mismatch bases for the overlap on the double nicked edge. Therefore this strand should not cause tetrahedron opening and fluorescence from Cy3 & Cy5 should remain quenched throughout. Figure 6.3.8 shows the Cy3 and Cy5 fluorescence signals over time with the same colours as in the previous two experiments. The FRET signal, despite high quenching, is also shown for completeness.

Again near the start of the experiment (from t=1200s onwards) only PAGE purified TetOH is contained in the cuvette. At t = 2050s, strip wrong is added to the mixture. No major changes in the fluorescence signals are observed, and the general trend shows that the fluorescence of Cy3, Cy5 and FRET gradually decrease. It is possible that the decrease in fluorescence may be due to photobleaching or non-specific absorption of the fluorescent species onto the cuvette wall. This effect was not noticed for the experiment in Figure 6.3.6 although the observation may have been masked by the large fluorescence change upon opening. However, as the Cy3 and Cy5 fluorescence remained quenched throughout the experiment the distance between the fluorophores has not increased and it is assumed that the tetrahedron has remained closed.

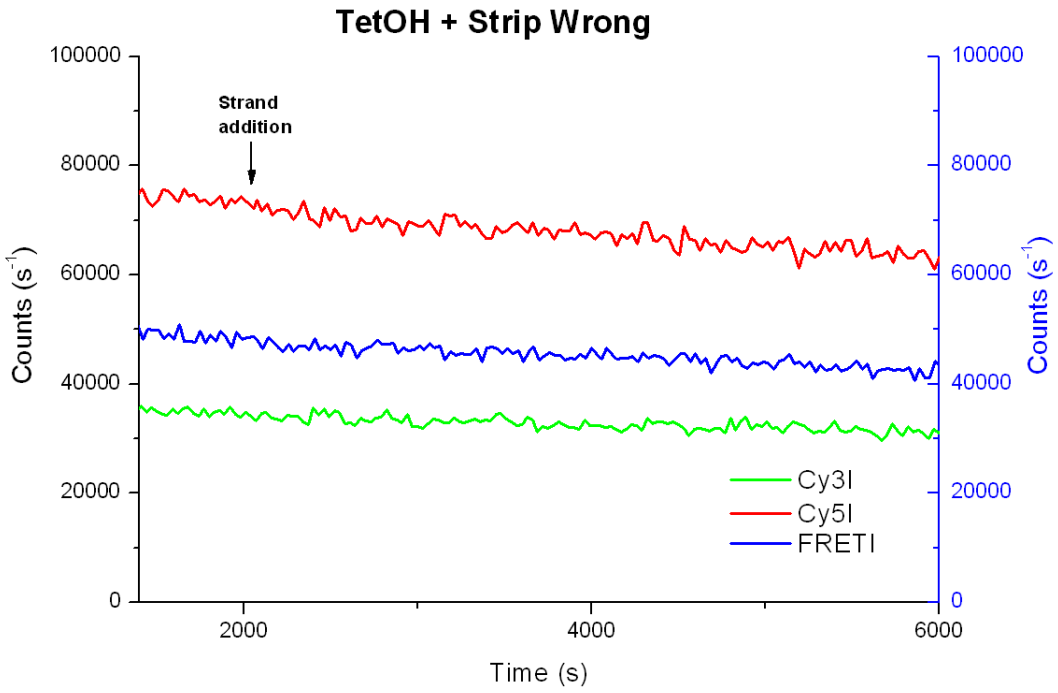


Figure 6.3.8: Bulk fluoresence vs. time of ~13 pmoles PAGE purified TetOH before and after the addition of ~40 pmoles 'StripWrong' (designed to bind to but not open TetOH). Cy3 excitation/emission is shown in green (Cy3I), Cy5 excitation/emission is shown in red (Cy5I). FRETI (Blue) is the intensity of the FRET signal i.e. Cy3 excitation/Cy5 emission. No major change in fluorescence in any channel is observed after the addition of 'StripWrong'.

Reversible Opening and Closing

Finally, the most interesting experiment involved the sequential addition of the 'Strip Fuel' and 'Strip Anti-Fuel' strand to TetOH to test reversible opening and closing. Figure 6.3.9 shows the result of such an experiment showing Cy3, Cy5 and FRET fluorescence signals over time with the same colours as in the previous graphs. Also indicated are the points at which 'Strip Fuel' (F) and 'Strip Anti-Fuel' (AF) strands were added to the cuvette.

TetOH alone is contained in the cuvette at the start of the experiment (from t=1000s onwards). On addition of the fuel strand at t = 1730s, at a 3x excess to TetOH, Cy3 and Cy5 fluorescence becomes unquenched and rises rapidly. These trends agree with Figure 6.3.6, indicating that the tetrahedra had opened. It is noted that the FRET signal, despite being treated with caution in the quenched state also decreases, in agreement with cage opening. All signals in this experiment have a much higher signal to noise than in Figure 6.3.6 as a higher concentration of TetOH was used. The resulting FRET signal changes in this case also appear larger which may due to the reduction

in measurement noise.

At t=9340s a 3x excess (to the previous fuel strand) of anti-fuel strands is added to the mixture. Here we observe the Cy3 and Cy5 fluorescence signal decrease sharply with a corresponding rise in the FRET signal consistent with the fluorophores becoming re-quenched and the tetrahedra re-closing. The baseline fluorescence for the closed state did not reach that of the original baseline due to incomplete removal of the fuel strands.

From here, two further opening and closing operations were performed by adding fuel at t = 14230s, anti-fuel at t = 19890s, fuel at t = 22080s and then anti-fuel again at t = 23940s (all using a 1:1 ratio with the previous strand). Opening and closing is again observed from changes in the Cy3 and Cy5 fluorescence and from the FRET signal. The amplitude of the change in fluorescence on sequential opening and closing decreases over time due to the lack of a sufficient excess of the previous strand. This insufficient excess results in a mixture of open and closed cages by the end of the experiment, as can be seen by the Cy3 and Cy5 fluorescence intensities which are midway between the initial open and closed states. The FRET signal also decreases overall over the course of the experiment, consistent with having a mixture of open and closed cages, as opposed to all closed (higher FRET, although more quenched) tetrahedra at the start of the experiment.

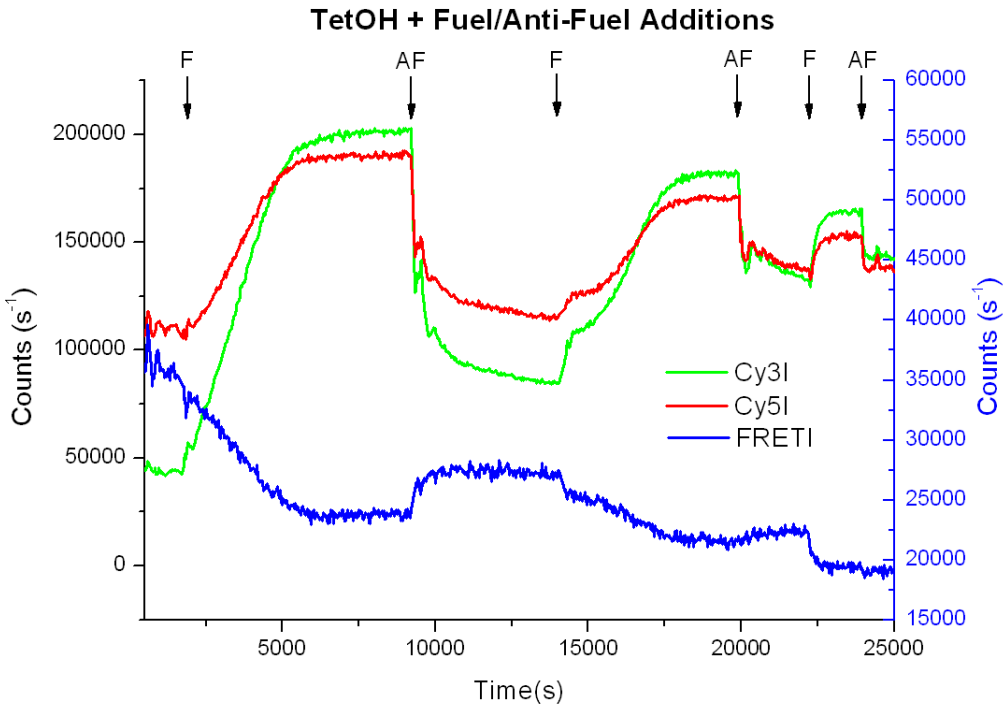


Figure 6.3.9: Bulk fluorescence vs. time of ~20 pmoles purified TetOH after sequential additions of 'Strip Fuel' and 'Strip Anti-Fuel' strands (initial 3x excess fuel, then 3x excess anti-fuel, then all 1x excesses to previous strand) designed to open and then re-close the tetrahedron respectively. Cy3 excitation/emission is shown in green (Cy3I), Cy5 excitation/emission is shown in red (Cy5I). FRETI (Blue) is the intensity of the FRET signal i.e. Cy3 excitation/Cy5 emission. F and AF indicated with arrow above the graph indicate the points at which fuel and anti-fuel strands were added.

6.3.3 Discussion

In this section I have demonstrated strong evidence for reversible opening and closing of a DNA cage initially using PAGE and then bulk fluorescence. Opening is achieved via the complete break of one of the cage edges by competitive binding from an input strand to the overhang and overlap region on that edge. A variety of input strands were designed to probe opening, reversible opening and no opening (as a negative control) and all performed as expected. An unintended consequence of the labelling positions was such that in the closed state, the donor and acceptor were subject to quenching due to extreme (< 1nm) proximity resulting in a FRET signal that was difficult to interpret. Changing the label positions removed the quenching problem and clear signals in the donor, acceptor and FRET channels could be observed.

By completely breaking one edge of the cage through the use of sequence-specific input strands,

such a device may find applications in controllable drug delivery. For example input strands could be particular messenger RNA transcribed under the control of an inducible promoter using external triggers as stimuli.

6.4 Overall Discussion

In this chapter I have shown the versatility with which DNA nanotechnology can be used. My original goal of developing a system for protein encapsulation and release was split into two sections, namely protein encapsulation within a DNA cage and then a design for DNA cage opening. It would certainly seem feasible to combine the two approaches to develop a complete protein encapsulation and release system. For such a system, it would be very interesting to measure the difference in protein dissociation rate observed in the open state and closed states where the opening edge is placed opposite the protein binding edge.

To demonstrate protein entrapment within the cage, I showed an approximately order of magnitude slowing in off-rate of CAP from the cage compared to linear dsDNA (section 6.2) due to the geometry of the cage. It is possible that a return to the off-rate of linear dsDNA may be obtained when the cage is opened.

The design of both cages was based on that of the regular DNA tetrahedron [46]. To accommodate the size of the encapsulation target, catabolite activator protein (CAP), one edge was extended from 20 to 30 bp. For proteins of different sizes, it would seem possible to rationally alter the geometry of the tetrahedron to generate an inside volume that would accommodate the target.

In order to act as a drug delivery vehicle, combining such a system with the internalisation methods developed in Chapter 5 would seem attractive. If pore sizes large enough could be formed using electroporation, whilst maintaining viability, such devices could be delivered directly into the cytoplasm of bacteria. Cages could then be opened in a controlled manner using an external stimulus (e.g. inducible promoter) to introduce active TFs into the cell to make changes in gene regulation and vice-versa: re-closed to sequester TFs from the cytoplasm to prevent their activity.

If multiple fluorescent labels are incorporated onto the protein and cage, the system may also find uses to investigate protein conformational dynamics using multiple distance measurements and used to collect data *in vitro* and *in vivo* for direct comparison. This may be considered as

an engineered analogy to molecular chaperonin cages which have been recently been investigated using similar single-molecule fluorescence methods [143].

In conclusion, this chapter provides important proof of principle experiments demonstrating the ability to non-covalently encapsulate a protein within a DNA cage and a separate system for protein release through cage opening. Future work would include combining the two systems developed in this chapter and internalising it into bacterial cells as well as demonstrating encapsulation of other targets to investigate conformational dynamics.

6.5 Contributions

I thank Christoph Erben for teaching me the ways of the tetrahedron and guiding me in the early days. Labelled CAP was provided by Javier Periz. I also thank Andrew Turberfield for agreeing for me to continue with the project and Achillefs Kapanidis for his enthusiastic support for a project outside his main focus. Support and feedback provided by Christoph Erben, Achillefs Kapanidis, and Andrew Turberfield.

Chapter 7

Conclusion and outlook

7.1 Summary

Control of gene expression is of fundamental importance to the success of all organisms [1]. In this post-genomic era, where genome sequencing can be performed at ever increasing speed and lower cost [166], we must focus our attention on how this genetic information is processed and regulated to produce fully functioning systems. Understanding the complex processes involved in gene regulation will not only help us understand how systems operate, but how they go wrong, and what we could do to intervene.

In this thesis I focus on a set of DNA-binding proteins known as transcription factors (TFs), crucial elements in the regulation of gene expression at the transcriptional level [3]. Intracellular TF concentrations must be tightly controlled for cells to function correctly; loss of control can lead to a range of debilitating or fatal diseases [167, 4]. The ability to rapidly detect and monitor TF concentrations *in vitro* and *in vivo* will be essential to understanding complex regulation pathways, complementing current gene expression profiling [168], and could also be used as diagnostic markers. To contribute towards this goal, we develop and validate two novel TF detection assays based on fluorescently labelled DNA probes.

Current TF assays can take several hours to produce a final result, require large sample volumes and can only be multiplexed spatially [169, 63, 64]. The TF assays developed in this work have many advantages mainly due to the methodology by which they are applied - single-molecule fluorescence. Single-molecule fluorescence techniques, discussed in Chapter 2, consider fluores-

cently labelled molecules on an individual basis allowing discrimination of sample heterogeneity. In this work, we use an implementation of single-molecule fluorescence known as alternating laser excitation (ALEX, section 2.5). ALEX allows the separation of labelled molecules on the basis of two ratiometric parameters: relative fluorophore stoichiometry (S) and FRET efficiency (E). S is a measure of relative fluorophore brightness whereas E is a measure of the proximity between the fluorophores. These two parameters were used to develop the two TF assays: TFs were detected using S in Chapter 3, and E in Chapter 4.

In Chapter 3 we demonstrated extremely sensitive detection of TFs down to pM levels, rapid detection time (~10 min), ability to multiplex detection of more than one TF and detection of TFs in complex biological samples such as cell lysates. We also developed and confirmed experimentally a kinetic model of the assay interactions such that TF concentrations could be quantified without the need of a calibration curve. In Chapter 4 we demonstrate a novel approach to TF sensing on the basis of TF-mediated DNA bending. By incorporating sequence-dependent bends in our DNA probe, moderate DNA bend angles induced by the protein of interest could be amplified and detected using FRET. We demonstrated detection of TFs in solution and on surfaces in purified form and expressed in a cell lysate. We believe these assays to be applicable to a wide range of TFs as a general approach towards rapid TF detection and quantitation.

While these assays provide rapid TF detection outside the cell, of events that have happened, we wanted to ask whether it was possible, using our probes, to monitor real-time active TF concentrations inside the cell as it happens. TF activity *in vivo* may be affected by a number of factors such as crowding, DNA topology, ionic concentrations, pH, or unknown cofactors or inhibitors [29, 30]. Thus it is important to quantify *in vivo* activity to complement *in vitro* data. In Chapter 4, we described the use of chemical modifications to DNA constructs to reduce their degradation susceptibility to nucleases for use in an *in vivo* context. It was found that phosphorothioate (PTO) linkers contained at either end of our TF bending sensor were a simple modification that dramatically increased nuclease resistance. We then investigated methods to introduce such sensors into cells.

To introduce our sensors into *E.coli* cells, we developed a successful method of fluorescent molecule internalisation using electroporation (Chapter 5). We demonstrate internalisation of labelled single- and double- stranded DNA and nucleoside tri-phosphates (NTPs). Using this method

we achieved an internalisation efficiency for a Cy3B-labelled ssDNA of ~10% whilst maintaining a cell viability of ~80%. Importantly in this chapter, we observe single-molecule fluorescence FRET inside cells. This capability would allow techniques such as ALEX to be used inside cells, dramatically increasing the information available from the internalised fluorescent molecules. The final step in this chapter was to internalised the PTO linked TF-sensor developed in Chapter 4. Initial results for TF detection *in vivo* proved promising, but work is needed to control active cellular TF concentrations to fully characterise the sensor *in vivo*. These methods are not only limited to TF detection; a wide variety of exciting experiments should now become possible. We discuss some of these ideas in section 7.2.

Moving away from TF detection, we introduce the possibility of controllable TF delivery in cells using nanodevices made from DNA. In Chapter 6, as a proof-of-principle, I describe the design of a novel non-covalent TF encapsulation and release system. For this goal, we produced two different DNA nanodevices based on the DNA tetrahedron [46]; one design was made to encapsulate the TF of interest, and the other to have the ability to reversibly open and close. Using PAGE, ensemble and single-molecule FRET we confirmed TF encapsulation in the first design, and show a dramatically slowed TF dissociation rate from the cage compared with linear DNA with the same binding sequence. The second design was initially validated using PAGE, and then reversible opening and closing up to three times was demonstrated using ensemble FRET. The work demonstrates the versatility with which DNA can be employed to make such rationally engineered structures. Some of the implications and future directions of this work are discussed in the next section.

7.2 Future Work

Having demonstrated the versatility of our TF sensing assays using bacterial TFs as a model system, it would be an attractive goal to extend our assays to detect eukaryotic TFs of clinical relevance. Using surface based approaches such as TIRF where hundreds or even thousands of molecules can be imaged at once, would allow rapid detection of multiple targets allowing sample TF profiling to be achieved. Targets need not necessarily be limited to TFs or even proteins. Other assays for the detection of RNA or DNA can and have been successfully implemented at

the single-molecule level [170, 20]. Target diseases can also extend beyond those due to inherent genetic problems but also to infection by foreign organisms e.g. bacteria; experiments in our lab are being carried out to develop novel assays to detect bacterial DNA. These assays in combination with a cheap, portable and easy-to-use monitoring system may allow for rapid on-site diagnostics of a library of targets. Towards this goal, efforts in our lab are currently being directed towards the development of a scaled down, affordable version of our TIRF microscope.

For more scientific questions, the ability to internalise fluorescently labelled probes and other molecules into *E.coli* will open many possibilities of exciting *in vivo* experiments. Beyond the advantages of real-time active TF concentration monitoring already discussed, the ability to internalise fluorescent NTPs may allow these molecules to be incorporated into newly replicated DNA or messenger RNA (mRNA). Incorporation into these relatively static molecules (in comparison to free diffusion) would allow their visualisation elucidating the spatial distribution of the bacterial chromosome, demonstrated for *Bacillus Subtilis* [125], or of mRNA localization, recently demonstrated in *Caulobacter crescentus* in fixed cells [171]. If the fluorescent labels could be controllably photoswitched then this approach in combination with super-resolution techniques such as stochastic optical reconstruction microscopy (STORM) [122] would allow sub-diffraction imaging of such localisations.

Internalisation of molecules could also be extended to short polypeptides or full proteins. As a means of protein tagging, short polypeptides such as GFP fragments [142] could be attached to quantum dots which would then attach to the larger GFP fragment fused to a protein of interest (POI). As quantum dots have several advantages over organic fluorophores in terms of photostability and quantum yield [172], there has been much interest in using these constructs to monitor biological processes *in vivo* [173, 174]. The ability to internalise site-specifically labelled proteins into cells would be of obvious interest. Assuming no effect on activity had occurred during internalisation, protein conformational dynamics [16] could be directly compared between *in vitro* and *in vivo* experiments, essential to our understanding of how proteins operate in their natural environment. This ability is only made possible through the use of single-molecule FRET (smFRET) in vivo. Our demonstration of single-molecule fluorescence and smFRET in live *E.coli* cells paves the way for such work. We believe the methods developed here and significantly contribute to the single-molecule toolbox.

Finally the development of the TF encapsulation and release system, in combination with the internalisation approaches, may well allow its use inside cells. Such a device could be triggered to open by a particular mRNA transcript given an external stimulus. In this way not only would we be able to monitor active TF levels *in vivo* using the methods described above, but we could also control TF levels to alter gene expression outcomes [38]. Another application of this technology, due to the tight encapsulation observed, may be the use of the DNA cage as reference point to measure protein dynamics, protein folding or determination of residue positions through triangulation [146].

In conclusion, I hope the methods and technologies I've developed in this thesis go some way in contributing to our understanding of the vastly complex process of gene regulation. The more we learn about this process the more accurately we will be able to accurately diagnose and treat disorders, such that we can prevent suffering before symptoms even arise.

Appendix A

Supplementary Material

A.1 DNA Used

A.1.1 Transcription Factor Detection: Stoichiometry Approach

Oligodeoxyribonucleotides were prepared by automated synthesis (IBA GmbH, Germany).

DNA fragments were labelled using amino-modifying groups attached to the sites of interest and N-hydroxy-succinimidyl esters of ATTO647N (ATTO-TEC GmbH, Siegen, Germany); or Cy3B (GE Healthcare, Uppsala, Sweden) using manufacturer's instructions, and HPLC or PAGE purified.

DNA sequences:
For all DNAs in all sections, X: amino-modifier C6 linker; Y: amino-modifier C2-dT DNA naming convention: 5'-A—B-3' 3'-D—C-5'. Strands AD form $H1^G$; strands BC form $H2^R$
DNAs for LacR-specific biosensor:

LacR A: 5'-*Cy3B* TG GTG TGT GGA ATT GTG A-3'
LacR B: 5'-GCG YAT AAC AAT TTC ACA CAG G-3'
LacR C: 5'-*ATTO647N* CC TGT GTG AAA TTG TT-3'
LacR D: 5'-AYA CGC TCA CAA TTC CAC ACA CCA-3'

DNAs for CAP-specific biosensor:
CAP A: 5'-*Cy3B* AA CGC AAT AAA TGT GA-3'
CAP B: 5'-AGY AGA TCA CAT TTT AGG CAC CA-3'
CAP C: 5'-*ATTO647N* TG GTG CCT AAA ATG TGA-3'
CAP D: 5'-TCY ACT TCA CAT TTA TTG CGT T-3'

Bridge DNAs for anisotropy measurements / AND gate work

FL-CAP/LacR-T: 5'-AGT AGA TCA CAT TTT AGG CAC CAT TAG GCA CCC CAG GCT TTA CAT TTA TGC TTC CGG CTC GTA TGT TGT GTG GAA TTG TGA-3'
FL-CAP/LacR-B: 5'-TTA CGC TCA CAA TTC CAC ACA ACA TAC GAG CCG GAA GCA TAA ATG TAA AGC CGT GGG TGC CTA ATG GTG CCT AAA ATG TGA-3'

DNAs for TIRF studies:
Biotinylated LacR-D: 5'-ATA CGC TCA CAA TTC CAC ACA CCA *Biotin*-3'
Biotinylated CAP-D: 5'-TCT ACT TCA CAT TTA TTG CGT T *Biotin*-3'

A.1.2 Transcription Factor Detection: FRET Approach

CAP Bend Top: 5'-*Cy3B* CCC ACT GCC GAA TGT GAG TTA GCT CAC TCA CGA AAA ACC ACT GTC GAA AAA GCT CTA CGG GCT CTG GCG TCG G *Biotin*-3'

CAP Bend Bot: 5'-*ATTO647N* CCG ACG CCA GAG CCC GTA GAG CCG ACA GTG GCG TGA GTG AGC TAA CTC ACA TTC GAA AAA GCA GTG GG-3'

High FRET Standard (Figure 4.3.3) Top: 5'-AGG CTT GAC ACT TTA TGC TTC GCG TCG TAT AAT G *Cy3B* G TGG AAT TGT GAG AGC GGA TAA CAA TTT C-3'

High FRET Standard (Figure 4.3.3) Bottom: 5'-GAA ATT GTT ATC CGC TCT CAC AAT TCC *ATTO647N* CAC ATT ATA CGA GCC GAA GCA TAA AGT GTC AAG CCT-3'

Linear Construct Top: 5'-*Cy3B* TAA ATC TAA AGT AAC ATA AGG TAA CAT AAC GTA AGC TCA TTC GCG-3'

Linear Construct Bottom: 5'-CGC GAA TGA GCT TAC GTT ATG TTA CCT TAT GTT ACT T *ATTO647N* A GAT TTA-3'

Linear PTO Construct Top: 5'-*Cy3B* T*A*A*A*T CT AAA GTA ACA TAA GGT AAC ATA ACG TAA GCT CAT TCG CG-3'

Linear PTO Construct Bottom: 5'-C*G*C*G*A AT GAG CTT ACG TTA TGT TAC CTT ATG TTA CTT *ATTO647N* AG ATT TA-3'

Dumbbell Top: 5'-TGC TTC GGC TCG TAT AAT G *Cy3B* G TGG AAT TGT GAG AGC GGA TAA CAA TTT CGC GCT TTG CGC CGA AAT TGT TAT CCG C-3'

Dumbbell Bottom: 5'-TCT CAC AAT TCC *ATT647N* CA CAT TAT ACG AGC CGA AGC ATA AAG TGT CAA GCC TGC GCT TTG CGC AGG CTT GAC ACT TTA-3'

CAP Bend PTO Top: 5'-*Cy3B* C*C*C*A*C T GCC GAA TGT GAG TTA GCT CAC TCA CGA AAA ACC ACT GTC GAA AAA GCT CTA CGG GCT CTG GCG TCG G *Biotin*-3'

CAP Bend PTO Bot: 5'-*ATTO647N* C*C*G*A* CG CCA GAG CCC GTA GAG CCG ACA GTG GCG TGA GTG AGC TAA CTC ACA TTC GAA AAA GCA GTG GG-3'

(* - indicates phosphorothioate linker)

A.1.3 Towards Transcription Factor Detection In-vivo

FRET Standard top (same for all): 5'-*Cy3B* T*A*A*A*TCT AAA GTA ACA TAA GGT AAC ATA ACG TAA GCT CAT T*C*G*C*G

High FRET Standard bottom: 5'-C*G*C*G*AAT GAG CTT ACG TTA TGT TAC CTT ATG TTA CTT *ATTO647N* AG A*T*T*T*A

Low FRET Standard bottom: 5'-C*G*C*G*AAT GAG CT *ATTO647N* ACG TTA TGT TAC CTT ATG TTA CTT TAG A*T*T*T*A

High FRET Standard (Figure 5.3.5) Top: 5'-AGG CTT GAC ACT TTA TGC TTC GCG TCG TAT AAT G *Cy3B* G TGG AAT TGT GAG AGC GGA TAA CAA TTT C-3'

High FRET Standard (Figure 5.3.5) Bottom: 5'-GAA ATT GTT ATC CGC TCT CAC AAT TCC *ATTO647N* CAC ATT ATA CGA GCC GAA GCA TAA AGT GTC AAG CCT-3'

CAP Bend PTO Top: 5'-*Cy3B* C*C*C*A*C T GCC GAA TGT GAG TTA GCT CAC TCA CGA AAA ACC ACT GTC GAA AAA GCT CTA CGG GCT CTG GCG TCG G *Biotin*-3'

CAP Bend PTO Bot: 5'-*ATTO647N* C*C*G*A* CG CCA GAG CCC GTA GAG CCG ACA GTG GCG TGA GTG AGC TAA CTC ACA TTC GAA AAA GCA GTG GG-3'

(* - indicates phosphorothioate linker)

A.1.4 Transcription Factor Encapsulation And Release Within A DNA Cage

A.1.4.1 Protein-Encapsulating DNA Cage

Normal Tetrahedron Strands from [46]:

s1: 5'-AGG CAG TTG AGA CGA ACA TTC CTA AGT CTG AAA TTT ATC ACC CGC CAT AGT AGA CGT ATC ACC-3'

s2: 5'-CTT GCT ACA CGA TTC AGA CTT AGG AAT GTT CGA CAT GCG AGG GTC CAA TAC CGA CGA TTA CAG-3'

s3: 5'-GGT GAT AAA ACG TGT AGC AAG CTG TAA TCG ACG GGA AGA GCA TGC CCA TCC ACT ACT ATG GCG-3'

s4: 5'-CCT CGC ATG ACT CAA CTG CCT GGT GAT ACG AGG ATG GGC ATG CTC TTC CCG ACG GTA TTG GAC-3'

Cage strands used for Protein Encapsulation:

s1 iCy3: 5'-AGG CAG TTG AG iCy3 CGA ACA TTC CTA AGT CTG AAA TTT ATC ACC CGC CAT AGT AGA CGT ATC ACC-3'

s2 Btn: 5'-CTT GCT ACA CG BTN TTC AGA CTT AGG AAT GTT CGA CAT GCG AGG GTC CAA TAC CGA CGA TTA CAG-3'

s3 CAP Inward: 5'-GGT GAT AAA ACG TGT AGC AAG CTG TAA TCG ACA ATA AAT GTG ATC TAG ATC ACA TTT TAG GAC TAC TAT GGC G-3'

s3 CAP Outward: 5'-GGT GAT AAA ACG TGT AGC AAG CTG TAA TCG AAA TGT GAT CTA GAT CAC ATT TTA GGC AAT AAC TAC TAT GGC G-3'

s4 CAP Inward: 5'-CCT CGC ATG ACT CAA CTG CCT GGT GAT ACG ACC TAA AAT GTG ATC TAG ATC ACA TTT ATT GAC GGT ATT GGA C-3'

s4 CAP Outward: 5'-CCT CGC ATG ACT CAA CTG CCT GGT GAT ACG ATA TTG CCT AAA ATG TGA TCT AGA TCA CAT TAC GGT ATT GGA C-3'

CAP consensus site containing strands used in competition experiment:

FL CAP Top: 5'-Cy3B AA CGC AAT AAA TGT GAA GTA GAT CAC ATT TTA GGC ACC A-3'

FL CAP Bottom: 5'-ATTO647N TT GGT GCC TAA AAT GTG ATC TAC TTC ACA TTT ATT GCG TT-3'

A.1.4.2 Reversibly Opening Tetrahedron

Normal Original Strands from [46] listed above (section A.1.4.1).

Edge Stability Testing Stands:

s1(-4): 5'-AGT TGA GAC GAA CAT TCC TAA GTC TGA AAT TTA TCA CCC GCC ATA GTA GAC GTA TCA CCA GGC-3'

s2(-4): 5'-CTA CAC GAT TCA GAC TTA GGA ATG TTC GAC ATG CGA GGG TCC AAT ACC GAC GAT TAC AGC TTG-3'

s3(-4): 5'-TGA TAA AAC GTG TAG CAA GCT GTA ATC GAC GGG AAG AGC ATG CCC ATC CAC TAC TAT GGC GGG-3'

s4(-4): 5'-TCG CAT GAC TCA ACT GCC TGG TGA TAC GAG GAT GGG CAT GCT CTT CCC GAC GGT ATT GGA CCC-3'

s2(-4) Black Phos 5'-Phos: AAT ACC GAC GAT TAC AGC TTG CTA CAC GAT TCA GAC TTA GGA ATG TTC GAC ATG CGA GGG TCC-3'

s2(-4)5Black-yellow-blue(9): 5'-AAT ACC GAC GAT TAC AGC TTG CTA CAC GAT TCA GAC TT-3'

s2(-4)3Blue(11)-Black Phos: 5'-Phos AGG AAT GTT CGA CAT GCG AGG GTC C-3'

Opening Design:

s2(-4) Black as above

s2(-4) Bl 5' Cy5: 5'-*Cy5* AAT ACC GAC GAT TAC AGC TTG CTA CAC GAT TCA GAC TT-3'

s2(-4) Bl 3’ Cy3: 5’-AGG AAT GTT CGA CAT GCG AGG GTC C *Cy3*-3’

s2(-4) 5’ iCy5: 5’-AAT ACC G *Cy5* C GAT TAC AGC TTG CTA CAC GAT TCA GAC TT-3’

s2(-4) 3’ iCy3: 5’-AGG AAT GTT CG *Cy3* CAT GCG AGG GTC C-3’

s4(-4) OVERHANG: 5’-TCG CAT GAC TCA ACT GCC TGG TGA TAC GAG GAT GGG CAT GCT CTT CCC GAC GGT ATT GGA CCC ACG GAC AAC GCT CT-3’

s4(-4)OH Purple: 5’-CTT CCC GAC GGT ATT GGA CCC TCG CAT GAC TCA ACT GCC TGG TGA TAC GAG GAT GGG CAT GCT ACG GAC AAC GCT CT-3’

Stripping Strands:

Strip 20: 5’-AGA GCG TTG TCC GTG GGT CC-3’

Strip Wrong: 5’-AGA GCG TTG TCC GTC CCA GA-3’

Strip Wrong 2: 5’-AGA GCG TTG TCC GT-3’

Strip Fuel: 5’-TGC CAA ACC AAG AGC GTT GTC CGT GGG TCC-3’

Strip Anti-fuel: 5’-GGA CCC ACG GAC AAC GCT CTT GGT TTG GCA-3’

A.2 Experimental Protocols & Supplementary Data

A.2.1 Transcription Factor Detection: Stoichiometry Approach

Half-sites and dsDNA fragments were formed by hybridization of labeled DNA oligos with their complementary unlabelled strand at a final concentration of 1 μM in 40 μl of annealing buffer (20 mM Tris-Cl pH 8.0, 1 mM EDTA and 500 mM NaCl) after heating for 4 min at 90°C and cooling to 25°C at a rate of 1°C/30 s.

Binding reactions (20 μl) between half-sites and TF (LacR as a tetramer or CAP dimer) were performed using 0.5-100 nM of each half-site and 0.5-400 nM TF at 14°C for 10-60 min in KG7 buffer (20 mM HEPES-NaOH pH 7.0, 100 mM potassium-L-glutamate, 10 mM MgCl2, 1 mM DTT, 100 μg/ml BSA, 1 mM MEA, 5% glycerol). The reaction mixture was then diluted in KG7 (supplemented with 0.2 mM cAMP in cases of CAP detection) and single-molecule measurements performed within 5 min (see also [175]). For detection of 200 pM LacR, complexes were formed in low salt buffer, KG7-LS (20 mM HEPES-NaOH pH 7.0, 10 mM potassium-L-glutamate, 10 mM MgCl2, 1 mM DTT, 100 μg/ml BSA, 1 mM MEA, 5% glycerol).

For IPTG detection, 200 nM LacR was incubated with 80 nM H1G, 20 nM H2R and 0.2-100 µM IPTG at 14°C for 40 min in KG7-LS. The sample was diluted 200-fold in KG7-LS plus an IPTG concentration identical to the initial incubation, and was examined using ALEX spectroscopy.

For experiments on surfaces, 50 pM of biotinylated $H1^G$ half-site was immobilised on a glass coverslip coated with PEG-Biotin and PEG (1.25:100 ratio) after it had been coated with 0.5 mg/ml neutravidin [13]. The protocols for slide cleaning and immobilization of biotinylated DNA have previously been described [176, 177]. Subsequently, 0-5.8 nM LacR (or CAP) and 1 nM of H2R half-site were incubated for 15-30 min in KG7 at room temperature; KG7 was then replaced by imaging buffer (1% Glucose, 165 U/ml glucose oxidase, 2170 U/ml catalase, and 2 mM Trolox diluted in Phosphate Buffered Saline [PBS]) and sealed with a glass coverslip before taking measurements. Each measurement consisted of 1 s movies with 20 Hz frame rate. For each sample, a total of 30 movies were collected.

For experiments in HeLa cell extracts, 10 nM of $H1^G$ and $H2^R$ were added to a final volume of 10 μl KG7 buffer containing 1 μg HelaScribe Nuclear Extract (Promega) and 10 nM CAP. The mixture was incubated for 10 min at 14˚C, diluted to ~20 pM DNA and examined using ALEX spectroscopy.

For experiments in bacterial lysates, *E.coli* BL21(DE3) pLysS carrying plasmid pAKCRP was grown in 500 ml fresh LB at 37°C to an OD_{600} of ~0.4. The culture was then induced with 1 mM IPTG. Samples of 50 ml cells were taken at 0 min (before induction), 5, 10, 30, 60, 120, and 180 min after induction. Cells were collected by centrifugation, the supernatant was discarded and the pellets resuspended in PBS. Cells were diluted to an OD of 0.5 in 2 ml KGE (20 mM HEPES-NaOH pH 7.0, 100 mM potassium-L-glutamate, 1 mM DTT, 100 μg/ml BSA, 1 mM MEA, 10 mM EDTA, 5% glycerol and supplemented with Roche Complete Protease Inhibitor). Cells were sonicated twice (30 s each) and cell debris was removed by centrifugation at 30,000 g for 20

min, and 10% glycerol was added to the supernatant. Protein in the supernatant was quantified using 280-nm absorption using a Nanodrop ND-1000 spectrophotometer. A dilution factor was determined to find the protein concentration required using the 3 hr sample to achieve ≈90% saturation of the DNA bound fraction as measured by ALEX spectroscopy. The incubations were performed using 100 nM DNA half-sites. Each time point was then diluted to the same protein concentration before measurement.

Ensemble Fluorescence Measurements

Ensemble fluorescence intensity measurements were performed on a fluorescence spectrophotometer (PTI, Birmingham, NJ). For fluorescence-anisotropy measurements [**?**, 178], the spectrophotometer was fitted with L-format Glan-Thomson polarizers. To obtain equilibrium constant K_{D1}, 100 nM of H1G was titrated with 0-5 μM of bridge DNA (K. Lymperopoulos), measured fluorescence anisotropy for each concentration, and fitted the concentration dependence of the relative anisotropy change to a rectangular hyperbola (titration was performed in triplicate; data not shown). It was found K_{D1} = 300±100 nM.

Stoichiometry Assay Kinetic Model

TF binding was modeled using a set of coupled reversible reactions representing the equilibrium between free and transiently-annealed half sites (Eq. A1, $K_{D1} = k_{-1}/k_1$):

$$H1^G + H2^R \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} H1^G - H2^R$$

and between the annealed half-sites and CAP (Eq. A2, $K_{D2} = k_{-2}/k_2$):

$$H1^G - H2^R + CAP \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} H1^G - CAP - H2^R$$

As half-sites in the quantitation experiments are supplied at equimolar ratios, we assume Eq. A3. Finally, Eq. A4 and Eq. A5 account for conservation of the G-only half-site and the CAP concentration, respectively:

$$\left[H1^G\right]_{eq}\left[H2^R\right]_{eq}k_1 = \left[H1^G - H2^R\right]_{eq}k_{-1} \tag{A1}$$

$$\left[H1^G\right]_{eq}\left[H2^R\right]_{eq}k_1 + \left[H1^G - CAP - H2^R\right]_{eq}k_{-2} = \left[H1^G - H2^R\right]_{eq}\left(k_{-1} + [CAP]_{eq}k_2\right) \tag{A2}$$

$$\left[H1^G\right]_{eq} = \left[H2^R\right]_{eq} \tag{A3}$$

$$\left[H1^G\right] = \left[H1^G\right]_{eq} + \left[H1^G - H2^R\right]_{eq} + \left[H1^G - CAP - H2^R\right]_{eq} \tag{A4}$$

$$[CAP] = [CAP]_{eq} + \left[H1^G - CAP - H2^R\right]_{eq} \tag{A5}$$

To relate ALEX data to the model, we determined the bound fraction, F_B, for a given experiment:

$$F_B = \frac{\left[H1^G - CAP - H2^R\right]_{eq}}{[H1^G]_{eq} + [H1^G - H2^R]_{eq} + [H1^G - CAP - H2^R]_{eq}} = \frac{\left[H1^G - CAP - H2^R\right]_{eq}}{[H1^G]}$$

Together, Eq.A1-A6 allows us to calculate the total TF concentration as a function of F_B (Eq. A6):

$$[CAP] = \left[H1^G\right] F_B + \frac{F_B K_{D2}\left(2\left[H1^G\right](1-F_B) + K_{D1} + \sqrt{K_{D1}}\sqrt{4\left[H1^G\right](1-F_B) + K_{D1}}\right)}{2\left[H1^G\right](1-F_B)^2} \quad \text{(A6)}$$

Measurement of F_B. Because reactions must be diluted to $\approx$ 50 pM prior to the ALEX assay, we assume that half-sites with $K_{D1} > 1$ nM dissociate completely (for our CAP experiments, $K_{D1} \approx$ 300 nM). The observed R-only bursts, N_A, are therefore proportional to $\left[H1^G\right]_{eq} + \left[H1^G - H2^R\right]_{eq}$, while the coincident bursts, N_C, are proportional to $\left[H1^G - CAP - H2^R\right]_{eq}$. This also requires that $H1^G - CAP - H2^R$ does not appreciably dissociate during the experiment, which is reasonable under our conditions: given the dissociation kinetics for the CAP-DNA complex ($k_{off} \approx 10^{-4}\ s^{-1}$; see [27]), F_B will decrease by < 7% due to dissociation of the $H1^G - CAP - H2^R$ complex during a 10-minute ALEX experiment.

Although the bound fraction in an ideal experiment would be $F_B = N_C/(N_C + N_A)$, in practice, the equation for F_B is:

$$F_B = N_C/(N_C + N_A - c) \quad \text{(A7)}$$

Parameter c is a constant describing the average number of "spurious" acceptor-only bursts per experiment. These occur as a result of G-fluorophore photobleaching of $H1^G - CAP - H2^R$ complexes, from improperly annealed half-sites, and other species incapable of participating in TF binding. Parameter c can be obtained easily, as it equals N_A when the total TF is in excess of half-site (where all $H1^G$ should, therefore, be bound in $H1^G - CAP - H2^R$ complexes).

"Normalized DNA coincidence" as described in the text refers to a double Gaussian fit to the one-dimensional S data, normalized to the saturation point (i.e., accounting for the spurious R-only bursts). Fig. A1 shows the Normalized DNA Coincidence (F_B) predicted by the model for various K_{D1}, K_{D2} and $H1^G$ values given increasing CAP concentration.

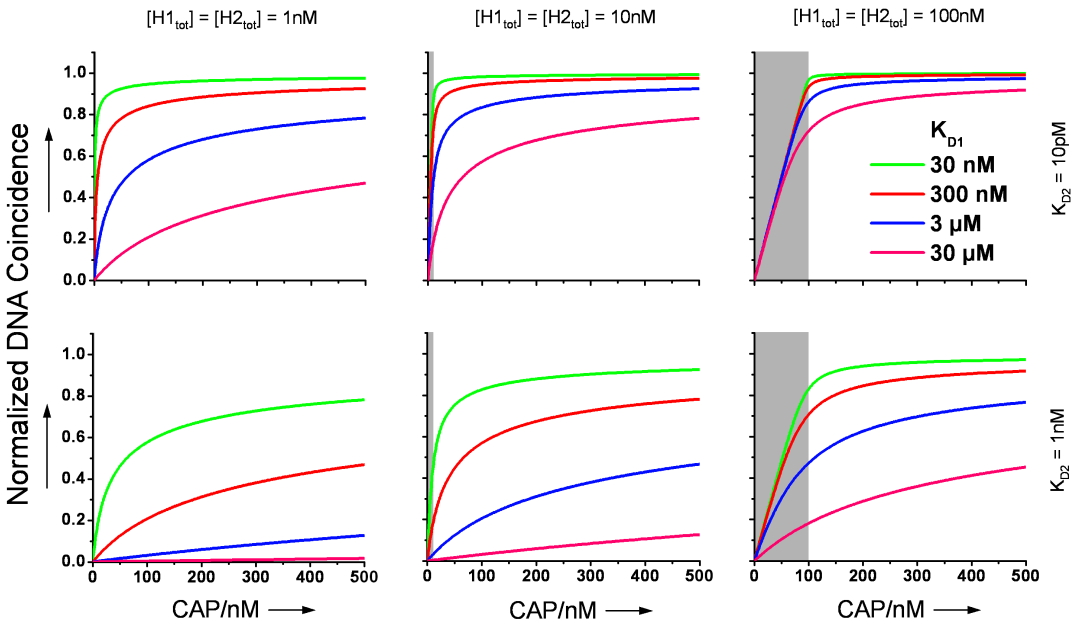


Figure A1: Model predictions for the dependence of the CAP-bound fraction (F_B) of DNA half-sites on CAP concentration, as a function of K_{D1}, K_{D2} and $H1^G$. Generally, lower K_{D1} values, lower K_{D2} values and higher $H1^G$ concentrations increase the bound fraction achieved for a given concentration of CAP. In the special case of low K_{D2} and high $H1^G$ (top right) the bound fraction increases in a nearly linear manner between 0 nM and 100nM (= $[H1^G]$) CAP. Above 100 nM, however, detection saturates. Increasing K_{D2} (bottom right) or decreasing $H1^G$ (top middle, left) generates more hyperbolic curves, which saturate for significantly higher CAP concentrations.

Transcription Factor Multiplexing

Multiplexed detection of more than one transcription factor was explored using differences in donor probe brightness, and therefore probe stoichiometry, S.

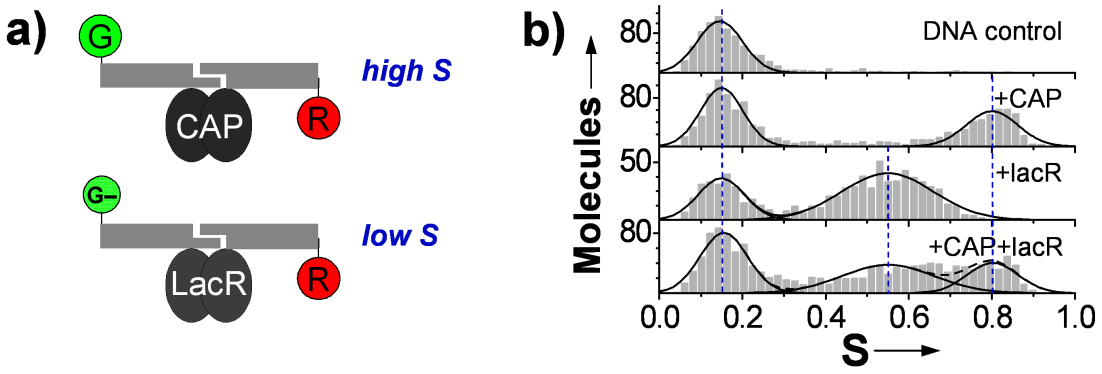


Figure A2: Stoichiometry-based Multiplexing. (a) Schematic describing multiplexing using half-sites labelled with fluorophores of different brightness. Due to the definition of S (see section 2.5), a G-R fluorophore pair with a "bright" G fluorophore will yield a higher S value than a G-R pair with a "dim" G fluorophore. Accordingly, we prepared CAP-specific half-sites using a bright G fluorophore (Cy3B) and LacR-specific half-sites using a dim G fluorophore (G-; tetramethyl-rhodamine, TMR). (b) All panels refer to assays in solutions containing both CAP-specific and LacR-specific DNA half-sites. Top: In the absence of CAP and LacR, only free half-sites are observed (the S≈0.15 species is due to free R-only half-sites). In the presence of 20 nM CAP (2nd panel), CAP-specific half-sites (10nM) associate to yield a population with S≈0.8 (due to the use of Cy3B as the bright G fluorophore). In the presence of 20 nM LacR (3rd panel), LacR-DNA half-sites (10nM) associate to yield a population with S≈0.55 (due to the use of TMR as the dim G fluorophore). In the presence of both TFs (bottom panel), both CAP- and LacR-specific half-sites associate, resulting in a histogram that corresponds to the sum of the individual distributions and enabling simultaneous detection and quantification of both CAP and LacR. Dotted vertical lines: mean S values for the 3 main species observed.

Detection of CAP on solid surfaces (via TIRF) was also achieved:

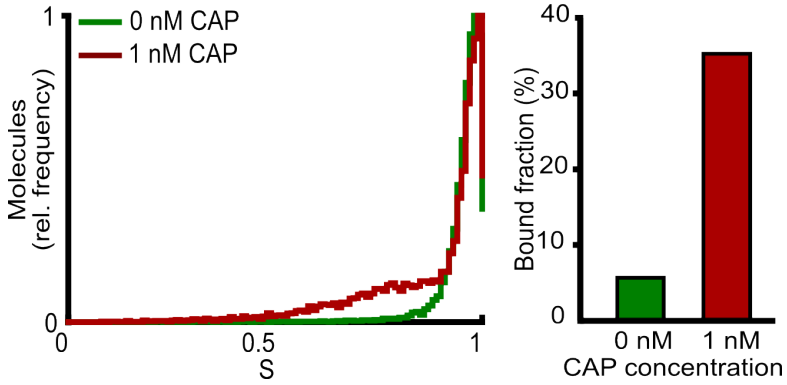


Figure A3: Detection of CAP on surfaces using TIRF. Incubation, imaging and analysis were as in Fig. 3.3.4(c), main text. Left: Relative molecular frequency vs. stoichiometry in the presence and absence of 1 nM CAP. Right: Bound fraction (the fraction of G fluorophores that co-localize with R fluorophores), calculated by summing the molecular frequency in the S histogram, for $S < 0.87$ (co-localized G and R) and $S > 0.87$ (G-only).

A.2.2 Transcription Factor Detection: FRET Approach

Dissociation Constant Determination

Measurements were performed on a fluorescence spectrophotometer (PTI, Birmingham, NJ). For fluorescence-anisotropy measurements [**?**, 178], the spectrophotometer was fitted with L-format Glan-Thomson polarizers. Here we titrated 0-3 μM of CAP in the presence of 0.2 mM cAMP with 10 nM of the bending sensor in binding buffer (KG7-LS; section A.2.1). A fluorescence scan between 555-700 nm given excitation at 525 nm, and anisotropy value were recorded for each point. Using FRET, the K_D of the interaction was ~790 nM and with anisotropy, the K_D was ~600 nM.

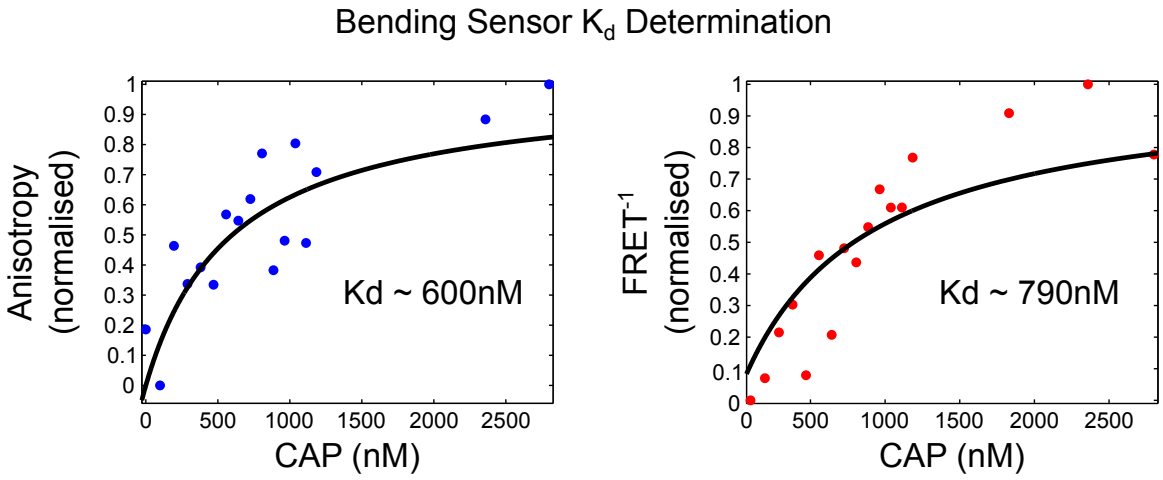


Figure A4: K_D Determination of Bending sensor using anisotropy (left, ~ 600nM) and FRET (right, ~790 nM). 0-3 μM of CAP in the presence of 0.2 mM cAMP with 10 nM of the bending sensor in binding buffer (KG7-LS; section A.2.1). A fluorescence scan between 555-700 nm given excitation at 525 nm, and anisotropy value were recorded for each point.

TIRF Measurements

For experiments on surfaces, 50 pM of bending sensor was immobilised on a glass coverslip coated with PEG-Biotin and PEG (1.25:100 ratio) after it had been coated with 0.5 mg/ml neutravidin [13]. The protocols for slide cleaning and immobilization of biotinylated DNA have previously been described [176, 177]. Subsequently, purified CAP or CAP lysate at concentrations indicated in the text were added in KG7 buffer (section A.2.1) to the immobilised sensor. Each measurement consisted of 1 s movies with 20 Hz frame rate. For each sample, a total of ~30 movies were collected.

Lysate Preparation

To generate CAP expressing bacterial lysates, Escherichia coli BL21(DE3) pLysS carrying plasmid pAKCRP was grown in 500 ml fresh LB at 37°C to an OD_{600} of ~0.4. The culture was then induced with 1 mM IPTG and grown for a further 3 hrs. Cells were collected by centrifugation, the supernatant discarded and pellets resuspended in KGE buffer (20 mM HEPES-NaOH pH 7.0, 100 mM potassium-L-glutamate, 1 mM DTT, 100 μg/ml BSA, 1 mM MEA, 10 mM EDTA, 5% glycerol and supplemented with Roche Complete Protease Inhibitor). Cells were sonicated 6x (30 s on, 30 s ice) and cell debris removed by centrifugation at 30,000 g for 20 min. 10% glycerol was added to the supernatant. Protein in the supernatant was quantified using 280-nm absorption using a Nanodrop ND-1000 spectrophotometer.

To generate cell lysates from the three genetically modified strains (DH5α, MG1655/pKD46 and Sure2), the same protocol as above was used only with cell growth stopped at OD_{600} of ~0.4 and with no induction with IPTG. Cell pellets were also resuspended in KG7 (no EDTA; section A.2.1) not KGE.

Ensemble FRET measurements

Ensemble fluorescence measurements were performed as described above for K_D determination. Before each experiment a spectra of buffer only was taken for later background subtraction. Spectra were corrected for wavelength dependent detector efficiencies using the manufacturer's software.

A.2.3 Towards Transcription Factor Detection In vivo

A.2.3.1 Preparation Of Electrocompetent *E.coli*

Items required (keep everything cold!):

- Overnight culture of *E.coli* (can include antibiotic here if needed)
- Sterile autoclaved 400mL LB in 2L xflask
- 500mL ice-cold mqH_20
- 100μL RT DMSO
- 2x 250mL pre-chilled flasks
- 3x pre-chilled 50mL falcon tubes
- ~30 sterile pre-chilled labelled 1.7mL eppis
- Ice water bath

- Pre-chill sigma floor and eppendorf table-top centrifuges to 4˚C

Procedure:

- Dilute approx 1:100 the overnight culture into the 400mL LB (to achieve OD_{600} ~0.01) and grow in the shaking incubator at 37˚, 250rpm until OD_{600} 0.4-0.5 (*no* antibiotics here)
- Transfer 200mL of culture to each of the 250mL pre-chilled flasks, and place them in an ice-water bath for 15 min with periodic swirling to evenly cool cells.
- Add 50mL culture to each of the 3 pre-chilled 50mL falcons and spin down at 1000g for 15 min. Discard supernatant. Repeat until all of the culture has been centrifuged into a pellet.
- Quickly add, then decant, 25mL ice-cold mqH2O, to rinse each pellet.
- Resuspend each pellet in 25mL ice-cold mqH2O by swirling/inversion. Spin, 15 min, 4°C, 1000g. Repeat 3x.
- Resuspend each pellet in 280μL ice-cold mqH2O, and transfer to chilled eppi.
- Spin the eppis in a pre-chilled table top centrifuge, 15 min, 1000g, 4°C.
- Carefully remove supernatant, resuspend each pellet in another 280uL ice-cold mqH2O.
- Add 11uL DMSO to each eppi, gently mix, and leave on ice for 5 minutes. Repeat, adding another 11uL DMSO, and incubating for 5 more minutes.
- Aliquot 40μL of the cells to each sterile labelled pre-chilled 1.7mL eppi and put in the -80°C freezer for later use.

A.2.3.2 Electroporation Of *E.coli* With Fluorescent Molecules

Items required:

- Long-stem (gel loading) pipette tips.
- A 1mm electroporation cuvette, kept on ice.
- 40uL aliquots of electrocompetent *E.coli* (kept at -80°C).
- At least 2μL of a 5μM sample of fluorescent molecules to be internalised, in a low-salt buffer (in general, should keep NaCl < 50mM).
- 900μL aliquots of sterile SOC (Super Optimal broth with Catabolite repression) media, pre-warmed to 37°C.
- ~200μL 1% Polyethyleneimine (PEI).
- Coverslip (thickness no.1).

- Silicon gasket.
- 0.22μm filtered sterile PBS or M9.

Procedure:

- Add silicon gasket (4 or more wells) to coverslip. In each well add ~30μL 1% PEI. Incubate RT 5 min.
- Wash each well 2x PBS
- Remove aliquot of electrocompetent cells from the -80°C freezer and place on ice to slowly melt.
- Once melted, add 2μL of the fluorescent sample (at least 5μM, in a low salt medium) to the cells and mix gently with pipette tip. Incubate on ice for 10 minutes.
- Add cell/fluorophore mix to the bottom of the pre-chilled cuvette with a long-stem pipette tip. Avoid bubbles, get even spread at bottom, can tap cuvette.
- Quickly dry metal surface of electroporation cuvette with a KimWipe, and place in Bio-Rad Electroporator. Electroporate on setting 'Ec1' (18kV/cm). Watch out for arcing.
- Rapidly add 900μL pre-warmed SOC media to cuvette and pipette up and down to flush out cells.
- Remove the SOC media and cells from the cuvette and place in 1.7mL eppi. Recover at 37°C for at least 20 min.
- Pellet cells by centrifugation at 6,000rpm for 1 min.
- Remove supernatant, resuspend in 500μL PBS or M9, spin again 6,000rpm for 1 min. Repeat at least 3x to remove free (non-internalised) fluorescent molecules.
- Finally, resuspend cells in 500μL PBS or M9 and add ~30uL to one well of the PEI coated coverslip. Incubate 5 min RT.
- Wash 2x in PBS or M9. Cells are ready for imaging.

Alternative attachment to slides can be through poly-lysine in place of PEI, or using low fluorescence M9 agarose (sandwiched in a gasket between two coverslips).

A.2.3.3 Widefield Imaging Of *E.coli* With Internalised Fluorescent Molecules

- Set green power to 200μW, turn on brighfield and mercury lamp (if needed). Start LabView/Data Acquisition and the Andor Ixon software.
- Add a drop of immersion oil to the objective. Block the laser.
- Place coverslip with attached cells onto stage

- Using the brightfield lamp (no gain!) find surface-adhered cells by focusing in the z-dimension. Take brightfield image (short 0.04s exposure).
- Turn off the brightfield lamp and unblock the laser to mage the cells using widefield fluorescence. Turn on 300x gain.
- Acquired 2s exposure (50x 0.04 frames). Save all files in .fits format.

A.2.3.4 Data Analysis Of Widefield Fluorescent Cell Images

To measure fluorescence from widefield fluorescence images from cells, a short function in MATLAB was written. The program would display the setup-acquired image of the flurescent cells. The user would select a region around the cell of interest, and then select a ring around this initial region. The average pixel intensity of the outside ring (with inner ring subtracted) was used to measure background. The inner region was then background subtracted by taking its total intensity minus the average background pixel intensity x number of pixel inside the inner region. Thus background subtracted intensities of both the green and red channels could be aquired, and FRET calculated according to equation 2.3.6. Leakage and direct excitation values were also calculated using this approach.

A.2.3.5 Single-molecule Confocal Spectroscopy Of Live *E.coli* With Internalised Fluorescent Molecules

- Having acquired widefield images of cells using section A.2.3.3, turn on the APDs, the translation stage and open the translation stage software (APT user).
- Switch to confocal mode by flipping down the widefield lens. Identify the bright spot generated by the laser and click this location in the Ixon software to mark the spot with crosshairs. This is where the confocal spot will end up on your sample, and where cells should be positioned for confocal imaging (in the x-y position).
- Flip the widefield lens back up and identify a large, bright *E.coli*. Use xy translation stage to position the crosshairs over the center of the cell.
- Switch back to confocal mode. An intense white spot where the laser intersects the *E.coli* should be clear. Block the laser.
- Stop recording with the camera, and move the APD/camera emission path mirror using the translation stage software. Take note of the current z-position of your sample.
- Start LabView data acquisition with the laser still blocked, then unblock the laser. A large amount of fluorescence should be visible in the data acquisition time trace. Move your sample slightly in either direction in the z-dimension to maximize the fluorescence. Note that as the fluorophores are actively photobleaching, the fluorescence will decrease over time regardless of what direction you move it in.

- Once you find a maximum, block the laser. Take note of the z-position of your stage. This should give you an idea of how many micrometers up or down you will have to move the stage when you want to switch from widefield to confocal imaging.
- Record a single-molecule time trace using the LabView software.

A.2.4 Transcription Factor Encapsulation and Release

A.2.4.1 Formation Of Cages and Complexes

To form the DNA cages, equal amounts of the four component strands (s1, s2, s3CAP and s4CAP) were mixed in annealing buffer (section A.2.1) to a final concentration of 1 μM, heated to 90˚C for 3 min before cooling to 4˚C over ~ 30 s. Samples concentrations for PAGE were 200 nM unless specified. For imaging unlabelled strands in PAGE gels, Sybr gold stain was used and gels were scanned in a BioRad PharosFX gel scanner.

CAP binding to the cage was performed with a specified excess of CAP at 14˚C for 10 min in KG7 buffer (section A.2.1).

A.2.4.2 Encapsulation Of CAP Within A DNA Cage

Figure A5 confirms correct binding of the Cy3 labelled protein cage by CAP labelled with Alexa 647. Lane 1 contains unlabelled cage, bound by unlabelled CAP in lane 2 and labelled CAP in lane 3. Lane 4 contains Cy3 labelled cage, bound by unlabelled CAP in lane 5 and Alexa 647 labelled CAP in lane 6.

	Lane					
	1	2	3	4	5	6
TH30	+	+	+			
Cy3 TH30				+	+	+
CAP		+			+	
A647 CAP			+			+
cAMP		+	+		+	+

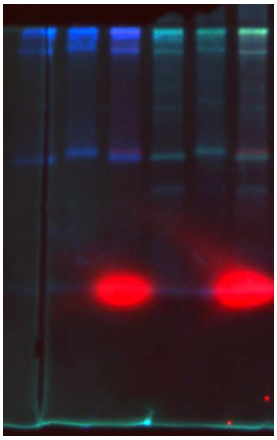


Figure A5: Confirmation of labelled CAP binding to labelled protein cage using PAGE. Contents shown above each lane. TH30 refers to protein cage. Gel conditions: 7.5% Tris-HCl (pH 8.8) separating, 4% Tris-HCl (pH 6.8) stacking, run in 1x Tris-Glycine at 4˚C.

A.2.4.3 Removal Of cAMP From CAP-Cage Complexes

Having formed the CAP-Cage complex as described above, cAMP was removed using a 'Micro Bio-Spin Size Exclusion Column' (Bio-Rad) according to the manufacturer's instructions. Samples were eluted into KG7 buffer (section A.2.1) containing no cAMP.

A.2.4.4 CAP Off-rate Competition Assay

CAP-Cage complexes were formed as described above. In this case ~0.8 pmoles cage was mixed with ~2.4 pmoles of Alexa647-labelled CAP. To this a 50x excess of unlabelled consensus site containing dsDNA (FL CAP DNA) was added. Donor unquenching over time was recorded.

As a comparison the same procedure, using the same concentrations using labelled FL CAP DNA was performed.

A.2.4.5 Opening Cage Ligation And Exonuclease Procedure

Ligase: 'T4 DNA Ligase' from NEB)

- 10μL - 2uM DNA (5' Phos'd) (1uM final)
- 2μL - 10x T4 DNA Ligase Reaction Buffer
- 1μL - 10x T4 DNA Ligase (400 units)
- 7μL - ddH20
- Place in thermal cycler at 16°C for 4 hours.

Exonuclease: 'Exo III' from NEB

- 10μL - 1uM DNA
- 5μL - 10x NEBuffer 1
- 1μL – ExoIII (100 units)
- 34μL – ddH20
- Place in thermal cycler at 37°C for 2 hours.

A.2.4.6 Opening Cage Bulk Fluorescence Measurements

Bulk FRET measurements were performed in a FluoroMax-3 fluorimeter (Horiba Jobin Yvon). Experiment were carried out at RT, using a magnetic stirrer bar in teh cuvette. Concentrations were as mentioned in the text. Strand addition was carried out by pausing the measurement, adding the strand, mixing, and resuming data acquisition.

A.3 Scientific Contributions Arising From This Thesis

Papers:

- Lymperopoulos K*, Crawford R*, Torella JP, Heilemann M, Hwang LC, Holden SJ, and Kapanidis AN, Single-molecule DNA biosensors for protein and ligand detection. Angew. Chem. 2010; 49 (7), pp 1316-1320. *equal contribution.

Talks:

- **2010 (March)** Winner of the 2010 Idea Idol competition run by Oxford Entrepreneurs (University of Oxford). Proposed a start-up company: 'Oxford Nanosense' to detect biomarkers of disease with high sensitivity, speed and at very low cost.
- **2010 (Feb)** Biophysical Society 54th Annual Meeting, San Francisco, California. Delivered talk entitled 'Single-Molecule DNA Biosensors for Quantitative Transcription Factor Detection'
- **2009 (Sept)** European Molecular Biology Organization (EMBO) Practical Course: 'Quantitative FRET, FRAP and FCS', BioQuant, University of Heidelberg, Germany. Designed and ran a four day practical demonstrating surface based single-molecule spectroscopy.
- **2009 (Mar)** 21st Annual RNA Polymerase Workshop, University of Bristol Delivered talk titled 'Ultra-sensitive detection of Transcription Factors and Small Molecules'. Presented poster on same topic.

Posters:

- **2010 (Sept)** Picoquant - 16th International Workshop on "Single-Molecule Spectroscopy and Ultrasensitive Analysis in the Life Sciences", Berlin, Germany. Presented: 'Protein sensing in vitro and in vivo using single-molecule FRET'.
- **2009 (Feb)** 'Advances in Biosensors for Point of Care Applications' by the Biosensing Special Interest Group, National Physics Laboratory, London. Presented: 'Single-Molecule Biosensors for Transcription Factor Detection'.

Bibliography

[1] Jeremy M Berg, John L Tymoczko, and Lubert Stryer. Biochemistry. *5th Edition, W.H. Freeman and Company, New York*, 2002.

[2] C Sheridan. Gene therapy finds its niche. *Nature Biotechnology*, 29(2):121–128, Jan 2011.

[3] C Branden and J Tooze. Garland publishing, new york. *Introduction to Protein Structure*, 1999.

[4] Amanda M Goh, Cynthia R Coffill, and David P Lane. The role of mutant p53 in human cancer. *Journal of Pathology*, 223(2):116–126, Oct 2010.

[5] Hongyan Zhou, Shili Wu, Jin Young Joo, Saiyong Zhu, Dong Wook Han, Tongxiang Lin, Sunia Trauger, Geoffery Bien, Susan Yao, Yong Zhu, Gary Siuzdak, Hans R Scholer, Lingxun Duan, and Sheng Ding. Generation of induced pluripotent stem cells using recombinant proteins. *Stem Cell*, 4(5):381–384, Aug 2009.

[6] Dohoon Kim, Chun-Hyung Kim, Jung-Il Moon, Young-Gie Chung, Mi-Yoon Chang, Baek-Soo Han, Sanghyeok Ko, Eungi Yang, Kwang Yul Cha, Robert Lanza, and Kwang-Soo Kim. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Stem Cell*, 4(6):472–476, May 2009.

[7] G Gahrton and B Björkstrand. Progress in haematopoietic stem cell transplantation for multiple myeloma. *Journal of internal medicine*, 248:185–201, Jan 2000.

[8] Olle Lindvall and Zaal Kokaia. Stem cells for the treatment of neurological disorders. *Nature*, 441:1094–1096, Jan 2006.

[9] S Goldman and M Windrem. Cell replacement therapy in neurological disease. *Philosophical Transactions of the Royal Society*, 361:1463–1475, Jan 2006.

[10] B Vogelstein, D Lane, and Arnold J Levine. Surfing the p53 network. *Nature*, Jan 2000.

[11] W E Moerner and L Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical Review Letters*, 62(21):2535–2538, Jan 1989.

[12] C Joo, H Balci, Y Ishitsuka, Chittanon Buranachai, and Taekjip Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annual review of Biochemistry*, 77:51–76, Jan 2008.

[13] C Joo and T Ha. Cold spring harbor laboratory press, new york. *Single-molecule Techniques: a laboratory manual*, 2008.

[14] X Sunney Xie and H Peter Lu. Single-molecule enzymology. *Journal of Biological Chemistry*, 274(23):15967–15970, Jan 1999.

[15] Brian P English, Wei Min, Antoine M Van Oijen, Kang Taek Lee, Guobin Luo, Hongye Sun, Binny J Cherayil, S C Kou, and X Sunney Xie. Ever-fluctuating single enzyme molecules: Michaelis-menten equation revisited. *Nat Chem Biol*, 2(2):87–94, Feb 2006.

[16] Yusdi Santoso, Catherine M Joyce, Olga Potapova, Ludovic Le Reste, Johannes Hohlbein, Joseph P Torella, Nigel D F Grindley, and Achillefs N Kapanidis. Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *PNAS*, 107(2):715–720, Dec 2010.

[17] Yusdi Santoso, Joseph P Torella, and Achillefs N Kapanidis. Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem*, Jun 2010.

[18] Lori A Neely, Sonal Patel, Joanne Garver, Michael Gallo, Maria Hackett, Stephen McLaughlin, Mark Nadel, John Harris, Steve Gullans, and Jenny Rooke. A single-molecule method for the quantitation of microRNA gene expression. *Nat Methods*, 3(1):41–6, Jan 2006.

[19] Konstantinos Lymperopoulos, R Crawford, J Torella, M Heilemann, Ling Chin Hwang, Seamus J Holden, and Achillefs N Kapanidis. Single-molecule DNA biosensors for protein and ligand detection. *Angewandte Chemie International Edition*, 49:1316–1320, Jan 2010.

[20] Z Peng, S Soper, M Pingle, Francis Barany, and Lloyd M Davis. Ligase detection reaction generation of reverse molecular beacons for near real-time analysis of bacterial pathogens using single-pair fluorescence resonance energy transfer and a cyclic olefin copolymer microfluidic chip. *Analytical Chemistry*, 82(23):9727–9735, Jan 2010.

[21] J William Efcavitch and John F Thompson. Single-molecule DNA analysis. *Annual Review Analytical Chemistry*, 3:109–128, Sep 2010.

[22] S Maurer, Jurgen Fritz, and Georgi Muskhelishvili. A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of dna organization. *Journal of molecular biology*, 387:1261–1276, Jan 2009.

[23] H Dietz and M Rief. Exploring the energy landscape of gfp by single-molecule mechanical experiments. *Proceedings of the National Academy of . . .*, Jan 2004.

[24] F Vanzi, C Broggio, Leonardo Sacconi, and Francesco Saverio Pavone. Lac repressor hinge flexibility and dna looping: single molecule kinetics by tethered particle motion. *Nucleic Acids Research*, 34(12):3409–3420, Jan 2006.

[25] R Dame, Maarten C Noom, and Gijs J L Wuite. Bacterial chromatin organization by h-ns protein unravelled using dual dna manipulation. *Nature*, 444(387-390), Jan 2006.

[26] Dunja Skoko, Ben Wong, Reid C Johnson, and John F Marko. Micromechanical analysis of the binding of DNA-bending proteins HMGB1, NHP6A, and HU reveals their ability to form highly stable DNA-protein complexes. *Biochemistry*, 43:13867–13874, Jan 2004.

[27] Achillefs N Kapanidis, Nam Ki Lee, Ted A Laurence, Sören Doose, Emmanuel Margeat, and Shimon Weiss. Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *PNAS*, 101(24):8936–8941, Jun 2004.

[28] A Kapanidis, T Laurence, Nam Ki Lee, Emmanuel Margeat, Xiangxu Kong, and Shimon Weiss. Alternating-laser excitation of single molecules. *Accounts of Chemical Research*, 38(7):523–533, Jan 2005.

[29] Ido Golding and Edward C Cox. Physical nature of bacterial cytoplasm. *Physical Review Letters*, 96, Jan 2006.

[30] H Zhou, German Rivas, and Allen P Minton. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual review of biophysics*, 37:375–397, Jan 2008.

[31] Nathan C Shaner, Paul A Steinbach, and Roger Y Tsien. A guide to choosing fluorescent proteins. *Nature Methods*, 2(12):905–909, Nov 2005.

[32] J Elf, G.-W Li, and X. S Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, May 2007.

[33] R Reyes-Lamothe, D. J Sherratt, and M. C Leake. Stoichiometry and architecture of active DNA replication machinery in escherichia coli. *Science*, 328(5977):498–501, Apr 2010.

[34] Y Taniguchi, P. J Choi, G.-W Li, H Chen, M Babu, J Hearn, A Emili, and X. S Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, Jul 2010.

[35] A Keppler, S Gendreizig, Thomas Gronemeyer, Horst Pick, Horst Vogel, and Kai Johnsson. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol*, 21:86–89, Jan 2002.

[36] John J Sakon and Keith R Weninger. Detecting the conformation of individual proteins in live cells. *Nature Methods*, 7(3):203–208, Jan 2010.

[37] Avigdor Eldar and Michael B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, Sep 2010.

[38] P. J Choi, L Cai, K Frieda, and X. S Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900):442–446, Oct 2008.

[39] P Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, Jan 2006.

[40] H Dietz, S Douglas, and W Shih. Folding DNA into twisted and curved nanoscale shapes. *Science*, 325(725-730), Jan 2009.

[41] J Shin and N Pierce. A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc*, Jan 2004.

[42] P Yin, H Yan, X Daniell, Andrew J Turberfield, and John H Reif. A unidirectional DNA walker that moves autonomously along a track. *Angewandte Chemie International Edition*, Jan 2004.

[43] H Baytekin and E Akkaya. A molecular NAND gate based on watson-crick base pairing. *Organic Letters*, Jan 2000.

[44] Milan N Stojanovic, Tiffany Elizabeth Mitchell, and Darko Stefanovic. Deoxyribozyme-based logic gates. *J. Am. Chem. Soc.*, 124(14):3555–3561, 2002.

[45] G Seelig, D Soloveichik, D. Y Zhang, and E Winfree. Enzyme-free nucleic acid logic circuits. *Science*, 314(5805):1585–1588, Dec 2006.

[46] R. P Goodman, I A T Schaap, C F Tardin, C M Erben, R M Berry, C F Schmidt, and A J Turberfield. Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science*, 310:1661–1665, Dec 2005.

[47] T Förster. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, Jan 1948.

[48] Lubert Stryer and Richard P Haugland. Energy transfer: A spectroscopic ruler. *PNAS*, 58:719–726, Nov 1967.

[49] Joseph R Lakowicz. Principles of fluorescence spectroscopy, third edition. *Springer*, 2006.

[50] George Gabriel Stokes. On the change of the refrangibility of light. *Philosophical Transactions of the Royal Society*, 142:463–562, 1852.

[51] Robert M Clegg. Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymology*, 211:353–388, Feb 1992.

[52] R Clegg. Fluorescence resonance energy transfer. *Current opinion in biotechnology*, 6:103–110, Jan 1995.

[53] E Tomko, Christopher J Fischer, and Timothy M Lohman. Ensemble methods for monitoring enzyme translocation along single stranded nucleic acids. *Methods*, 20:269–276, Jan 2010.

[54] Seamus J Holden, Stephan Uphoff, Johannes Hohlbein, David Yadin, Ludovic Le Reste, Oliver J Britton, and Achillefs N Kapanidis. Defining the limits of single-molecule fret resolution in TIRF microscopy. *Biophysical Journal*, 99(9):3102–3111, Mar 2010.

[55] M Orrit and J Bernard. Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal. *Physical Review Letters*, 65(21):2716–2719, Jan 1990.

[56] E Brooks Shera, Newton K Seitzinger, Lloyd M Davis, Richard A Keller, and Steven A Soper. Detection of single fluorescent molecules. *Chemical Physics Letters*, 174(6):553–557, Feb 1990.

[57] Shuming Nie, DanielT Chiu, and RichardN Zar. Probing individual molecules with confocal fluorescence microscopy. *Science*, 266(5187):1018–1021, Apr 1994.

[58] Nancy L Thompson, T Burghardt, and D Axelrod. Measuring surface dynamics of biomolecules by total internal reflection fluorescence with photobleaching recovery or correlation spectroscopy. *Biophysical Journal*, 33:435–454, Jan 1981.

[59] H Schneckenburger. Total internal reflection fluorescence microscopy: technical innovations and novel applications. *Current opinion in biotechnology*, 16:13–18, Jan 2005.

[60] R Lamichhane, A Solem, Will Black, and David Rueda. Single molecule FRET of protein-nucleic acid and protein-protein complexes: Surface passivation and immobilization. *Methods*, 52:192–200, Jan 2010.

[61] J Fries, L Brand, C Eggeling, Malte Köllner, and Claus A M Seidel. Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *The Journal of Physical Chemistry B*, 102:6601–6613, Jan 1998.

[62] N Lee, Achillefs N Kapanidis, You Wang, Xavier Michalet, Jayanta Mukhopadhyay, Richard H Ebright, and Shimon Weiss. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophysical Journal*, 88(4):2939–2953, Apr 2005.

[63] S Fredriksson, M Gullberg, J Jarvius, and C Olsson. Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*, 20:473–477, Jan 2002.

[64] Tomasz Heyduk and Ewa Heyduk. Molecular beacons for detecting DNA binding proteins. *Nature Biotechnology*, 20:171–176, Jan 2002.

[65] Ewa Heyduk, Eric Knoll, and Tomasz Heyduk. Molecular beacons for detecting DNA binding proteins: mechanism of action. *Analytical Biochemistry*, 316:1–10, Apr 2003.

[66] Y Santoso, L Hwang, L Le Reste, and Achillefs N Kapanidis. Red light, green light: probing single molecules using alternating-laser excitation. *Biochemical Society Transactions*, 36(4):738–744, Jan 2008.

[67] S Busby and R Ebright. Transcription activation by catabolite activator protein (cap). *Journal of Molecular Biology*, Jan 1999.

[68] K Gaston, A Bell, A Kolb, Henri Buc, and Stephen Busby. Stringent spacing requirements for transcription activation by crp. *Cell*, 62:733–743, Jan 1990.

[69] Shaun Scott, Stephen Busby, and Ifor Beacham. Transcriptional co-activation at the ansb promoters: involvement of the activating regions of crp and fnr when bound in tandem. 18(3):521–531, May 1995.

[70] Ira Pastan and Sankar Adhya. Cyclic adenosine 5'-monophosphate in escherichia coli. *Bacteriological Reviews*, 40(3):527–551, Nov 1976.

[71] Richard H Ebright, Yon W Ebright, and Angelo Gunasekera. 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*, 17:10295–10305, Jun 1989.

[72] M Lewis. The lac repressor. *C R Biol*, 328(6):521–548, Jun 2005.

[73] Catherine M Falcon and Kathleen S Matthews. Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. *Biochemistry*, 39:11074–11083, Feb 2000.

[74] Steve C Schultz, George C Shields, and Thomas A Steitz. Crystal structure of a CAP-DNA complex: The DNA is bent by 90. *Science*, 253(5023):1001–1007, Mar 1991.

[75] Michael J Lane, Teodoro Paner, Irina Kashin, Brian D Faldasz, Bo Li, Frank J Gallo, and Albert S Benight. The thermodynamic advantage of DNA oligonucleotide 'stacking hybridization' reactions: energetics of a DNA nick. *Nucleic Acids Research*, 25(3):611–616, Jan 1997.

[76] Mary D Barkley, Arthur D Riggs, Alan Jobe, and Suzanne Bourgeois. Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry*, 14(8):1700–1712, Nov 1975.

[77] A Riggs, R Newby, and S Bourgeois. lac repressor–operator interaction. ii. effect of galactosides and other ligands. *Journal of Molecular Biology*, Jan 1970.

[78] Keller Rinaudo, Leonidas Bleris, Rohan Maddamsetti, Sairam Subramanian, Ron Weiss, and Yaakov Benenson. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol*, 25(7):795–801, Jul 2007.

[79] Nicolas E Buchler, Ulrich Gerland, and Terence Hwa. On schemes of combinatorial transcription logic. *PNAS*, 100(9):5136–5141, Apr 2003.

[80] Edwin Southern, Kalim Mir, and Mikhail Shchepinov. Molecular interactions on microarrays. *Nature Genetics*, 21(1):5–9, Dec 1999.

[81] S Zimmerman and S Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of escherichia coli. *Journal of Molecular Biology*, Jan 1991.

[82] N Nichols and K Matthews. Protein-DNA binding correlates with structural thermostability for the full-length human p53 protein. *Biochemistry*, Jan 2001.

[83] Marie de Lumley, Darren J Hart, Matthew A Cooper, Stefan Symeonides, and Jonathan M Blackburn. A biophysical characterisation of factors controlling dimerisation and selectivity in the NF-kappab and NFAT families. *Journal of Molecular Biology*, 339(5):1059–75, Jun 2004.

[84] Mary Szatkowski Ozers, John J Hill, K Ervin, J Wood, A Nardulli, C Royer, and Jack Gorski. Equilibrium binding of estrogen receptor with DNA using fluorescence anisotropy. *Journal of Biological Chemisty*, 272(48):30405–30411, Jan 1997.

[85] Lan Ma, John Wagner, John Jeremy Rice, Wenwei Hu, Arnold J Levine, and Gustavo A Stolovitzky. A plausible model for the digital response of p53 to DNA damage. *Proceedings of the National Academy of Sciences*, 102(40):14266–14271, Jan 2005.

[86] Peter C van der Vliet and C.Peter Verrijzer. Bending of DNA by transcription factors. *BioEssays*, 15(1):25–32, Dec 1993.

[87] Achillefs N Kapanidis, Yon W Ebright, Richard D Ludescher, Shirley Chan, and Richard H Ebright. Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution. *Journal of Molecular Biology*, 312:453–468, May 2001.

[88] Serguei V Kuznetsov, Sawako Sugimura, Paula Vivas, Donald M Crothers, and Anjum Ansari. Direct observation of DNA bending/unbending kinetics in complex with dna-bending protein ihf. *PNAS*, 103(49):18515–18520, Jan 2006.

[89] P Rice, S Yang, K Mizuuchi, and H Nash. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, Jan 1996.

[90] Frank Stühmeier, Alexander Hillisch, Robert M Clegg, and Stephan Stephan Diekmann. Fluorescence energy transfer analysis of DNA structures containing several bulges and their interaction with CAP. *Journal of Molecular Biology*, 302:1081–1100, Jun 2000.

[91] Utz Dornberger, Alexander Hillisch, Friedrich A Gollmick, Hartmut Fritzsche, and Stephan Diekmann. Solution structure of a five-adenine bulge loop within a DNA duplex. *Biochemistry*, 38(39):12860–12868, Sep 1999.

[92] Alexander Hillisch, Mike Lorenz, and Stephan Diekmann. Recent advances in fret: distance determination in protein-dna complexes. *Current Opinion in Structural Biology*, 11:201–207, Mar 2001.

[93] J Passner and T Steitz. The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. *Proceedings of the National Academy of Sciences*, 94:2843–2847, Jan 1997.

[94] H Ishizuka, Akemi Hanamura, Tadashi Kunimura, and Hiroji Aiba. A lowered concentration of cAMP receptor protein caused by glucose is an important determinant for catabolite repression in escherichia coli. *Molecular Microbiology*, 10(2):341–350, Jan 1993.

[95] Robert Crawford, Douglas J Kelly, and Achillefs N Kapanidis. A protein biosensor that relies on bending of single dna molecules. *ChemPhysChem*, 13:918–922, Feb 2012.

[96] A Gunasekera, Yon W Ebright, and Richard H Ebright. Dna sequence determinants for binding of the escherichia coli catabolite gene activator protein. *Journal of Biological Chemistry*, 267(21):14713–14720, Jan 1992.

[97] Adi Stern and Rotem Sorek. The phage-host arms race: Shaping the evolution of microbes. *BioEssays*, 33:43–51, Jan 2011.

[98] Mohan Viswanathan and Susan T Lovett. Single-strand DNA-specific exonucleases in escherichia coli: roles in repair and mutation avoidance. *Genetics*, 149:7–16, Jan 1998.

[99] Michael McArthur and Mervyn J Bibb. Manipulating and understanding antibiotic production in streptomyces coelicolor A3 (2) with decoy oligonucleotides. *Proceedings of the National Academy of Sciences*, 105(3), Jan 2008.

[100] Fritz Eckstein and Gerald Gish. Phosphorothioates in molecular biology. *Trends in Biochemical Sciences*, pages 97–100, Jan 1989.

[101] Dwaine A Braasch and David R Corey. Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochemistry*, 41(14):4503–4510, Jan 2002.

[102] J Kurreck, E Wyszko, Clemens Gillen, and Volker A Erdmann. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Research*, 30(9):1911–1918, Jan 2002.

[103] B Richey, D Cayley, M Mossing, Corrine Kolka, Charles F Anderson, Thomas C Farrar, and M Thomas Record. Variability of the intracellular ionic environment of escherichia coli. differences between in vitro and in vivo effects of ion concentrations on protein-DNA interactions and gene expression. *Journal of Biological Chemistry*, 262(15):7157–7164, Jan 1987.

[104] L Stagg, S Zhang, Margaret S Cheung, and Pernilla Wittung-Stafshede. Molecular crowding enhances native structure and stability of a/b protein flavodoxin. *Proceedings of the National Academy of Sciences*, 104(48):18976–18981, Jan 2007.

[105] Ingmar Schoen, Hubert Krammer, and Dieter Braun. Hybridization kinetics is different inside cells. *PNAS*, early edition:1–6, Dec 2009.

[106] D Johnson, A Mortazavi, Richard M Myers, and Barbara Wold. Genome-wide mapping of in vivo protein-dna interactions. *Science*, 316:1497–1501, Jan 2007.

[107] Osamu Shimomura, Frank H Johnson, and Yo Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. *Journal of Cellular and Comparative Physiology*, 3:223–239, Nov 1962.

[108] Dmitriy M Chudakov, Sergey Lukyanov, and Konstantin A Lukyanov. Fluorescent proteins as a toolkit for in vivo imaging. *Trends in Biotechnology*, 23(12):605–613, Nov 2005.

[109] Ji Yu, Jie Xiao, Xiaojia Ren, Kaiqin Lao, and X Sunney Xie. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311:1600–1603, Jun 2006.

[110] Bas Ponsioen, Jun Zhao, Jurgen Riedl, Fried Zwartkruis, Gerard van der Krogt, Manuela Zaccolo, Wouter H Moolenaar, Johannes L Bos, and Kees Jalink. Detecting cAMP-induced epac activation by fluorescence resonance energy transfer: Epac as a novel cAMP indicator. *EMBO Reports*, 5(12):1176–1180, Nov 2004.

[111] Nataliya Panchuk-Voloshina, Rosaria P Haugland, Janell Bishop-Stewart, Mahesh K Bhalgat, Paul J Millard, Fei Mao, Wai-Yee Leung, and Richard P Haugland. Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *The Journal of Histochemistry & Cytochemistry*, 47(9):1179–1188, Jul 1999.

[112] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, Aug 2002.

[113] H Murakoshi, R Iino, Takeshi Kobayashi, Takahiro Fujiwara, Chika Ohshima, Akihiko Yoshimura, and Akihiro Kusumi. Single-molecule imaging analysis of ras activation in living cells. *Proceedings of the National Academy of Sciences*, Jan 2004.

[114] Wolfgang Schumann. Dynamics of the bacterial chromosome. structure and function. *Wiley-VCH*, 2006.

[115] MUNREET CHEHAL, JENNIFER COOK, AZRA REKIC, and ADELE WANG. Natural competency and electrotransformation efficiencies in escherichia coli DH5a spheroplasts. *Journal of Experimental Microbiology and Immunology*, 11:81–85, Sep 2007.

[116] Adam P Silverman and Eric T Kool. Quenched autoligation probes allow discrimination of live bacterial species by single nucleotide differences in rRNA. *Nucleic Acids Research*, 33(15):4978–4986, Jan 2005.

[117] J Wyber, Juile Andrews, and Antony D'Emanuele. The use of sonication for the efficient delivery of plasmid DNA into cells. *Pharmaceutical research*, 14(6):750–756, Jan 1997.

[118] Yizhi Song, Thomas Hahn, Ian P Thompson, Timothy J Mason, Gail M Preston, Guanghe Li, Larysa Paniwnyk, and Wei E Huang. Ultrasound-mediated DNA transfer for bacteria. *Nucleic Acids Research*, 35(19):e129, Jan 2007.

[119] William J Dower, Jeff F Miller, and Charles W Ragsdale. High efficiency transformation of e.coli by high voltage electroporation. *Nucleic Acids Research*, 16(13):6127–6145, Dec 1988.

[120] James C Weaver and Yu.A Chizmadzhev. Theory of electroporation: a review. *Bioelectrochemistry and Bioenergetics*, 41:135–160, Jan 1996.

[121] Joseph Peter Torella. Confocal single-molecule fluorescence as a tool for investigating biomolecular dynamics in vitro and in vivo. *Doctoral Thesis*, 2010.

[122] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, pages 1–3, Aug 2006.

[123] Yitzhak Tor and Peter B Dervan. Site-specific enzymic incorporation of an unnatural base, n6-(6-aminohexyl) isoguanosine, into rna. *Journal of the American Chemical Society*, 115:4461–4467, Jan 1993.

[124] Y Ito, N Teramoto, Naoki Kawazoe, and Kojiro Inada. Modified nucleic acid for systematic evolution of rna ligands by exponential enrichment. *Journal of Bioactive and Compatible Polymers*, 13:114–123, Jan 1998.

[125] I Berlatzky, Alex Rouvinski, and Sigal Ben-Yehuda. Spatial organization of a replicating bacterial chromosome. *Proceedings of the National Academy of Sciences*, 105(37):14136–14140, Jan 2008.

[126] H Kubitschek. Cell volume increase in escherichia coli after shifts to richer media. *Journal Of Bacteriology*, 172(1):94–101, Jan 1990.

[127] M Levene, J Korlach, S Turner, M Foquet, H G Craighead, and W W Webb. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 299:682–686, Jan 2003.

[128] J Eid, A Fehr, J Gray, K Luong, J Lyle, and G Otto.... Real-time DNA sequencing from single polymerase molecules. *Science*, 323:133–138, Jan 2009.

[129] K Samiee, J M Moran-Mirabal, Y K Cheung, and H G Craighead. Zero mode waveguides for single-molecule spectroscopy on lipid membranes. *Biophysical Journal*, 90(9):3288–3299, May 2006.

[130] M Elowitz, M Surette, Pierre-Etienne Wolf, Jeffry B Stock, and Stanislas Leibler. Protein mobility in the cytoplasm of escherichia coli. *Journal Of Bacteriology*, 181(1):197–203, Jan 1999.

[131] C Mullineaux, A Nenninger, Nicola Ray, and Colin Robinson. Diffusion of green fluorescent protein in three cell environments in escherichia coli. *Journal Of Bacteriology*, 188(10):3442–3448, Jan 2006.

[132] Robert Crawford, Joseph P Torella, Louise Aigrain, Anne Plochowietz, Kristofer Gryte, Stephan Uphoff, and Achillefs N Kapanidis. Long-lived intracellular single-molecule fluorescence using electroporated molecules. *Biophysical Journal*, 105:2439–2450, Dec 2013.

[133] John J Iandolo and William M Shafer. Regulation of staphylococcal enterotoxin b. *Infection and Immunity*, 16(2):610–616, Jan 1977.

[134] N Popovych, S Tzeng, Marco Tonelli, Richard H Ebright, and Charalampos G Kalodimos. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences*, 106(17):6927–6932, Jan 2009.

[135] H Sharma, S Yu, J Kong, J Wang, and T Steitz. Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. *Proceedings of the National Academy of Sciences*, Jan 2009.

[136] Tomasz Heyduk and James C Lee. Escherichia coli cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. *Biochemistry*, 28:6914–6924, Jan 1989.

[137] Tomasz Heyduk and James C Lee. Application of fluorescence energy transfer and polarization to monitor escherichia coli cAMP receptor protein and lac promoter interaction. *Proceedings of the National Academy of Sciences*, 87:1744–1748, Jan 1990.

[138] Erica A Pyles and J Ching Lee. Mode of selectivity in cyclic AMP receptor protein-dependent promoters in escherichia coli. *Biochemistry*, 35:1162–1172, Jan 1996.

[139] S. W Hell. Far-field optical nanoscopy. *Science*, 316(5828):1153–1158, May 2007.

[140] Achillefs N Kapanidis, Emmanuel Margeat, Sam On Ho, Ekaterine Kortkhonjia, Shimon Weiss, and Richard H Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314:1144–1147, Nov 2006.

[141] T Cordes, Y Santoso, A Tomescu, and K Gryte.... Sensing DNA opening in transcription using quenchable FRET. *Biochemistry*, 49:9171–9180, Jan 2010.

[142] Stéphanie Cabantous, Thomas C Terwilliger, and Geoffrey S Waldo. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnology*, 23(1):102–107, Jan 2005.

[143] Hagen Hofmann, Frank Hillger, Shawn H Pfeil, Armin Hoffmann, Daniel Streich, Dominik Haenni, Daniel Nettels, Everett A Lipman, and Benjamin Schuler. Single-molecule spectroscopy of protein folding in a chaperonin cage. *PNAS*, 107(26):11793–11798, Jun 2010.

[144] Stephan Uphoff, Seamus J Holden, Ludovic Le Reste, Javier Periz, Sebastian van de Linde, Mike Heilemann, and Achillefs N Kapanidis. Monitoring multiple distances within a single molecule using switchable FRET. *Nature Methods*, pages 1–9, May 2010.

[145] Jinwoo Lee, Sanghwa Lee, Kaushik Ragunathan, Chirlmin Joo, Taekjip Ha, and Sungchul Hohng. Single-molecule four-color FRET. *Angewandte Chemie International Edition*, 49:1–5, Nov 2010.

[146] H Dietz and M Rief. Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences*, 103(5):1244–1247, Jan 2006.

[147] N Seeman. Nucleic acid junctions and lattices. *Journal of Theoretical Biology*, Jan 1982.

[148] E Andersen, M Dong, M Nielsen, and K Jahn. Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, Jan 2009.

[149] J Heath, S O'brien, Q Zhang, and Y Liu.... Lanthanum complexes of spheroidal carbon shells. *Journal of the American Chemical Society*, Jan 1985.

[150] F Weiss, J Elkind, S C OBrien, R F Curl, and R E Smalley. Photophysics of metal complexes of spheroidal carbon shells. *Journal of the American Chemical Society*, Jan 1988.

[151] R Kumar, A Maitra, and P Patanjali.... Hollow gold nanoparticles encapsulating horseradish peroxidase. *Biomaterials*, Jan 2005.

[152] V Torchilin. Recent advances with liposomes as pharmaceutical carriers. *Nature Reviews Drug Discovery*, Jan 2005.

[153] B Okumus, S Arslan, M Fengler, Sua Myong, and Taekjip Ha. Single molecule nanocontainers made porous using a bacterial toxin. *Journal of the American Chemical Society*, Jan 2009.

[154] Yuji Ishitsuka, Burak Okumus, Sinan Arslan, Kok Hao Chen, and Taekjip Ha. Temperature-independent porous nanocontainers for single-molecule fluorescence studies. *Analytical Chemistry*, Fix this ref, not out yet:1–8, Nov 2010.

[155] Christoph M Erben, Russell P Goodman, and Andrew J Turberfield. Single-molecule protein encapsulation in a rigid DNA cage. *Angew. Chem. Int. Ed.*, 45(44):7414–7417, Nov 2006.

[156] R Goodman, M Heilemann, and S Doose. Reconfigurable, braced, three-dimensional DNA nanostructures. *Nature Nanotechnology*, 3, Jan 2008.

[157] T Kato, R Goodman, C Erben, Andrew J Turberfield, and Keiichi Namba. High-resolution structural analysis of a DNA nanostructure by cryoEM. *Nano Letters*, 9(7):2747–2750, Jan 2009.

[158] A Napoli, C Lawson, Richard H Ebright, and Helen M Berman. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: recognition of pyrimidine-purine and purine-purine steps. *Journal of Molecular Biology*, 357:173–183, Jan 2006.

[159] Robert Crawford, Christoph M Erben, Javier Periz, Lucy M Hall, Tom Brown, Andrew J Turberfield, and Achillefs N Kapanidis. Non-covalent single transcription factor encapsulation inside a dna cage. *Angewandte Chemie International Edition*, 52:2284–2288, Jan 2013.

[160] Harold Scott Macdonald Coxeter. Regular polytopes. *Methuen and Co.*, 1948.

[161] Christoph M Erben. Self-assembled DNA cages. *DPhil Thesis*, 2007.

[162] S Chen, J Vojtechovsky, Gary N Parkinson, Richard H Ebright, and Helen M Berman. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. *Journal of Molecular Biology*, 314:63–74, Jan 2001.

[163] Pramod Kumar, Dhananjay C Joshi, Mohd Akif, Yusuf Akhter, Seyed E Hasnain, and Shekhar C Mande. Mapping conformational transitions in cyclic AMP receptor protein: Crystal structure and normal-mode analysis of mycobacterium tuberculosis apo-cAMP receptor protein. *Biophysical Journal*, 98(2):305–314, Feb 2010.

[164] S Marras, Fred Russell Kramer, and Sanjay Tyagi. Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Research*, 30(21), Jan 2002.

[165] C Seidel, Andreas Schulz, and Markus H M Sauer. Nucleobase-specific quenching of fluorescent dyes. 1. nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *The Journal of Physical ...*, 100:5541–5553, Jan 1996.

[166] Johannes Hohlbein, Kristofer Gryte, Mike Heilemann, and Achillefs N Kapanidis. Surfing on a new wave of single-molecule fluorescence methods. *Physical Biology*, 7, Aug 2010.

[167] Carlos Guijarro and Jesús Egido. Transcription factor-kappa b (nf-kappa b) and renal disease. *Kidney international*, 59:415–424, Jan 2001.

[168] Isaac S Kohane, Alvin T Kho, and Atul J Butte. Microarrays for an integrative genomics. *The MIT Press*, 2003.

[169] Miguel Angel Hernández-Presa, Carmen Gómez-Guerrero, and Jesús Egido. In situ nonradioactive detection of nuclear factors in paraffin sections by southwestern histochemistry. *Kidney . . .*, 55:209–214, Jan 1999.

[170] G Yao, X Fang, Hiroaki Yokota, Toshio Yanagida, and Weihong Tan. Monitoring molecular beacon dna probe hybridization at the single-molecule level. *Chemistry-A European Journal*, 9:5686–5692, Jan 2003.

[171] Paula Montero Llopis, Audrey F Jackson, Oleksii Sliusarenko, Ivan Surovtsev, Jennifer Heinritz, Thierry Emonet, and Christine Jacobs-Wagner. Spatial organization of the flow of genetic information in bacteria. *Nature*, 466(7302):77–81, Jul 2010.

[172] U Resch-Genger, M Grabolle, Sara Cavaliere-Jaricot, Roland Nitschke, and Thomas Nann. Quantum dots versus organic dyes as fluorescent labels. *Nature Methods*, 5(9):763–775, Jan 2008.

[173] X Michalet, F Pinaud, L Bentolila, J M Tsay, S Doose, J J Li, G Sundaresan, A M Wu, S S Gambhir, and S Weiss. Quantum dots for live cells, in vivo imaging, and diagnostics. *Science*, 307:538–544, Jan 2005.

[174] Yun Xing and Jianghong Rao. Quantum dot bioconjugates for in vitro diagnostics & in vivo imaging. *Cancer Biomarkers*, 4:307–319, Dec 2008.

[175] Kapanidis, E Margeat, T Laurence, S Doose, and S Ho. Retention of transcription initiation factor [sigma] 70 in transcription elongation: single-molecule analysis. *Molecular Cell*, 20:347–356, Jan 2005.

[176] E Margeat, A Kapanidis, P Tinnefeld, and Y Wang. Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophysical Journal*, 90:1419–1431, Jan 2006.

[177] T Ha, I Rasnik, W Cheng, H Babcock, and G Gauss. Initiation and re-initiation of DNA unwinding by the escherichia coli rep helicase. *Nature*, 419, Jan 2002.

[178] Tomasz Heyduk, Yuexing Ma, Hong Tang, and Richard H Ebright. Fluorescence anisotropy: Rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods in Enzymology*, 274:492–503, Sep 1996.