



Federico Garza de León

Oriel College

Supervisor

Prof. Achillefs N. Kapanidis

SINGLE-MOLECULE FLUORESCENCE
ANALYSIS OF TRANSCRIPTION
FACTORS IN LIVING BACTERIA

**A thesis submitted for the degree of
Doctor of Philosophy**

Trinity 2015

A mis abuelos, que aunque ya no esten con nosotros,
seguramente lo estarán, orgullosos, el día de mi graduación.

A thesis submitted for the degree of Doctor of Philosophy.

Single-Molecule Fluorescence Analysis of Transcription Factors in Living

Bacteria

By Federico Garza de León, Oriel College.

Submitted Trinity 2015.

Extended Abstract

Super-resolution microscopy has become an important tool to image cells at ever increasing resolution. Specifically, localization microscopy, in which single emitters are localized with high accuracy, has allowed us to follow and track single-molecules in live cells. The technique has revealed aspects in the proteins' real-time function that were impossible to study previously. Despite their extensive use in recent years, there are still possibilities to improve and validate the technique and the resulting data. This thesis presents the first tracking photo-activation localization microscopy (tracking PALM) on three transcription factors (TFs): the lac repressor (LacI), araC protein and the cAMP receptor protein (CRP). With this work I expand the range of applicability of localization methods and increase in detail an important part of cellular function.

TFs control the expression of genes so that cells can adapt to external

conditions. In *Escherichia coli* (*E. coli*), previous studies have shown that more than 60 % of TFs have less than 100 monomers per genome copy. For example LacI, involved in the lactose utilization control, exists in *E. coli* in around 40 monomers in contrast to other proteins that exist in the tens of thousands. To study such low-copy numbers we found it necessary increase the throughput of tracking PALM and this thesis summarises our efforts to increase throughput.

After constructing a setup capable of tracking PALM with a 3X increased throughput, I studied the diffusion and localization of LacI and LacI relative to its operators. By studying LacI, I found that there were a number of challenges related to low copy numbers, specifically background noise that leads to localisations that can affect the results. I characterized the background tracks and tested the ability of clustering to find a number of tandem operators. This knowledge allowed us to study other TFs: araC, involved in the control of arabinose utilization which also exists in low-copy numbers but functions in multiple genes; and the CRP, which both regulates multiple sites and exists in large copy numbers.

In addition I worked towards the understanding of TF binding by simulating the combination of non-specific binding and free diffusion that emulates the search process of TFs. I obtained insight into the analysis of the data generated from tracking PALM. Finally I focused on creating a strategy to use our existing data to extract blinking rates from the fluorescent protein.

Our insights into TFs diffusion and binding have shown that tracking PALM can yield important insights at this regime of protein concentrations. Our work can be applied to other TFs and proteins that function in low copy numbers.

Statement of Originality

I declare that the work in this thesis is mine alone, except where indicated in the text and as described below:

The constructed setup was planned, ordered and built in collaboration with Stephan Uphoff and Mathew Stracy.

The image registration software was programmed by Bo Jing.

Elements of fig. 3.1 were adapted from the ASI microscope manual, specifically the RAMM side-view configuration.

The experiments on RNAP used to analyse blinking were prepared and performed by Mathew Stracy and I analysed the data.

Strains for LacI, araC and CRP fused to PAmCherry were developed by Laura Sellars, University of Birmingham.

Plasmid with LacI-PAmCherry and CRP-PAmCherry strain were developed by Riddhi Shah, , University of Birmingham.

Plasmid for LacIMut-PAmCherry was created by me with supervision of Pawel Zawadzki, , University of Oxford.

The 2D-Gaussian fitting algorithms were developed by Seamus Holden and Oliver Britton. Analysis software in collaboration with Stephan Uphoff and cell segmentation published work microbe-Tracker with analysis software developed by Mathew Stracy and modified by me.

The one diffusion species simulations were developed by Stephan Uphoff for Uphoff et al. (2013) and modified to simulate two-diffusion coefficients by me.

Figures 2.5 and 3.9 were adapted from a figure I created and were published in the review paper Stracy et al. (2015).

Federico Garza de León

October 2015

Acknowledgements

I would like to thank everyone who has put me in this path that I could never have imagined the first day I stepped into my first physics class. To Hector Garcia-Molina for his advise that lead me first into research. To Paul Hansma, my first research lab supervisor whose advise I hold eternally valuable to this day and started my path into biophysics. To Mikael Kall who supported me throughout my Masters thesis and who first led me to molecular biophysics.

I would like to specially thank my supervisor Achillefs Kapanidis who allowed me to break into an amazing field of research. You've have had to patience to help me along the way and the kindness to cheer me on my successes. My time here will be forever shaped by you.

I would like to thank everyone that throughout the years have been in the Kapanidis group whilst I've been working here. I've learnt a lot by spending time with you and have made valuable friendships. I would like to thank specially all the people in the downstairs office, who've made this adventure an enjoyable experience. In the middle

of writing many of these chapters I couldn't but remember all the silliness in our office and crack the occasional smile.

I would like to acknowledge Seamus Holden, Sarah Matthews, Robert Crawford, Stephan Uphoff, Bo Jing, Mathew Stracy, Pawel Zawadzki, Riddhi Shah and Laura Sellars for providing samples, software, and help with experiments. I would to thank Anouk Van Laer, Stephan Uphoff, Mathew Stracy and Robert Crawford for help in writing this thesis.

Anouk Van Laer, die de meeste instrumentale persoon is geweest in mijn tijd als een DPhil student en voor het schrijven van deze thesis. Dank u voor uw hulp en geduld. Ik zal nooit vergeten dat we ruzie hadden de dag dat ik heb gesolliciteerd voor Oxford, omdat ik het geheim had gehouden. Maar die dag heeft geleid tot waar we vandaag staan, zowel samen als op vlak van het nastreven van een doctoraat in het Verenigd Koninkrijk. Ik hoop dat ik binnenkort je kan bedanken voor al jouw hulp en die ann jou kan geven.

Y lo más importante lo dejo para el final. Quiero dar las gracias a las personas más importantes de mi vida, que me han ayudado no solo durante mi doctorado sino por muchos años antes. A mis padres que siempre nos han empujado a mi hermana y a mi a percibir la mejor educación posible. Tal vez hubo ocasiones que no les llamaba semanalmente y sé que no les describo mucho de lo

que me pasa en Oxford. Pero el saber que ahí están apoyandome desde lejos y solo están a una llamada esperando a ayudarme, no existen palabras para describir ese sentimiento. Gracias por empujarnos a cumplir objetivos que mi hermana y yo nunca hubiéramos imaginado y que lo hayan hecho sin importar que para cumplirlo tengamos que vivir a muchos km de distancia.

Gracias a todos.

I was supported by Consejo Nacional de Ciencia y Tecnología/I2T2 and Engineering and Physical Sciences Research Council through the Life Sciences Interface programme Doctoral Training Centre.

This thesis template was created by Brett Tully (2010) '*Allostasis of Cerebral Water: Modelling the Transport of Cerebrospinal Fluid*', PhD Thesis, University of Oxford.

Dissemination

Peer-Reviewed Journal Papers

Stracy M, Uphoff S, Lesterlin CO, **Garza de Leon F**, Uphoff S, Zaszki P & Kapanidis AN (2015) **Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid.** *Proceedings of the National Academy of Sciences*

Stracy M, Uphoff S, **Garza de Leon F**, & Kapanidis AN (2014) **In vivo single-molecule imaging of bacterial DNA replication, transcription, and repair.** *FEBS Letters*

Uphoff S, Reyes-Lamothe R, **Garza de Leon F**, Sherratt DJ, & Kapanidis AN. (2013) **Single-molecule DNA repair in live bacteria.** *Proceedings of the National Academy of Sciences*

Conference Presentations and Posters

Poster: **Garza de Leon F**, Stracy M, Sellars L, Busby SJW, & Kapanidis AN. (2014) **Super-resolution imaging of transcription**

factors in living bacteria. *26th & 27th UK RNA Polymerase Workshop.* Newcastle, UK & Birmingham, UK.

Talk: **Determining the blinking behaviour of photoactivatable fluorescent proteins in live bacteria** (Sep 2013) *19th International Workshop on "Single Molecule Spectroscopy and Ultra Sensitive Analysis in the Life Sciences".* Berlin, Germany.

Invited software workshop: **DAOSTORM** (June 2013) *Superresolution Microscopy Workshop: Seeing Biology at the Nanoscale.* Singapore.

Invited talk: **Super-resolution Imaging of Polymerases in Living Bacteria** (June 2013) *Superresolution Microscopy Workshop: Seeing Biology at the Nanoscale.* Singapore.

Talk: **Super-resolution imaging of transcription factors in live E. coli cells** (Mar 2013) *25th UK RNA Polymerase Workshop.* Milton Keynes, UK.

Journal Papers In Preparation

Garza de Leon F, Sellars L, Busby SJW, & Kapanidis AN. (In Preparation) **Tracking slow and low copy number proteins in live cells using the lac repressor**

CONTENTS

Contents	xvii
List of Figures	xxii
List of Tables	xxv
Acronyms	xxvi
1 Motivation	1
2 Imaging single-molecules in live bacteria	3
2.1 Introduction	3
2.1.1 Transcription factors	3
2.2 Fluorescence	6
2.2.1 Single-molecule detection	8
2.2.2 Microscopy	9
2.2.3 Single-molecule detection by light microscopy	10
2.3 Single-molecule imaging of TFs <i>in vivo</i>	13
2.3.1 Localising single-molecules <i>in vivo</i>	13

2.3.2	Enhanced localisation of LacI	14
2.3.3	Super-resolution microscopy	19
2.3.4	Photo-activatable proteins	20
3	tracking PALM	23
3.1	Introduction	23
3.2	Microscope Design and Construction	24
3.2.1	Single-molecule localisation microscopes	24
3.2.2	Materials and Methods	24
3.2.2.1	Design and construction	25
3.2.3	User interface	28
3.3	Single-molecule experiments and PSF measurement	29
3.3.1	Estimating chromatic aberrations	30
3.3.2	3D calibration and software	34
3.4	Other TIRF microscope used in this thesis	37
3.4.1	Imaging	37
3.5	tracking PALM	38
3.5.1	Counting molecules	39
3.5.2	Measuring the diffusion coefficient	39
3.5.3	Data visualisation and Analysis	42
3.5.3.1	Blinking	44
3.5.3.2	Simulations	44

4 Characterising fluorescent protein blinking in live cells	45
4.1 Introduction	45
4.2 Blinking Characterisation	47
4.3 Method and Materials	50
4.3.1 Identifying the presence of temporal clusters	50
4.3.2 Rate extraction	50
4.4 Blinking Model and rates	52
4.5 Memory parameter estimation	56
4.6 Discussion	60
4.7 Conclusions	64
 5 Tracking slow and low copy number proteins in live cells using the lac repressor	 65
5.1 Introduction	65
5.2 Results	68
5.2.1 Diffusion analysis of single LacI molecules using tracking PALM	68
5.2.2 The sub-cellular distribution of LacI in living <i>E. coli</i>	74
5.2.3 Auto-fluorescent species in <i>E. coli</i>	77
5.2.4 The <i>in vivo</i> diffusion profile of LacI expressed at native copy numbers	80
5.2.5 Bound LacI associated with the expected chromosomal locus	84

5.3	Discussion	88
5.3.1	LacI Diffusion	88
5.3.2	Spatial Distribution	89
5.3.3	Challenges with low to moderate copy numbers	90
5.3.4	Small FROS	91
5.4	Methods	91
5.4.1	Cell culture	91
5.4.2	Analysis	92
5.5	tracking PALM of other TFs	93
5.5.1	The AraC protein	93
5.5.2	Method	93
5.5.3	Diffusion profile AraC	94
5.5.4	CRP	97
5.5.5	Diffusion profile of CRP	97
5.5.6	Discussion	97
5.6	Conclusion	98
6	Simulating Diffusion of Transcription Factors	99
6.1	Introduction	99
6.1.1	Fluorescence	100
6.2	Materials and Methods	102
6.3	Results	105
6.3.1	Simulating one mode of diffusion	105

6.3.2	Simulating Two Modes of Diffusion	107
6.3.3	Simulating the <i>lac</i> repressor	110
6.4	Discussion	113
6.4.1	Conclusions	115
7	Conclusions	117
	References	119
A	Strain Information	127
A.1	Other Methods	127
A.1.1	ssDNA Cy3b experiments	127
A.2	Strain construction	128
A.2.0.1	6xlacO operator sequence	128
A.2.0.2	6xlacO operator sequence + 20x Mall operator sequence	129
A.3	Plasmid construction	130
A.3.0.3	LacIMut strain construction	130

LIST OF FIGURES

2.1	The <i>lac</i> operon and the structure of LacI.	4
2.2	Jablonski diagram and spectrum of the enhanced green fluorescent protein	6
2.3	Excitation beam position in fluorescence microscopy.	11
2.4	Resolution and the distance between the images of PSFs.	12
2.5	Localising single-molecules in live cells.	15
2.6	Structure and spectra of PAmCherry.	20
2.7	Resolving structures at sub-diffraction resolutions.	21
3.1	Drawing of the constructed TIRF microscope.	26
3.2	Picture of the TIRF microscope.	29
3.3	Microscope user interface.	30
3.4	Imaging single-molecules and their experimental point-spread function size.	31
3.5	Overlaid field of view of fluorescent beads.	31
3.6	Two-colour registration.	32

3.7 GUI for 3-D calibration.	34
3.8 Calibration curve for astigmatic 3-D localisation.	35
3.9 Single-molecule tracking of LacI-PAmCherry.	39
3.10 Analysing tracking PALM data.	41
3.11 Mobility of Pol1 in live and fixed cells.	43
4.1 Identifying blinking species in <i>E. coli</i>	49
4.2 Photo-activation density in one cell.	52
4.3 PAmCherry's behaviour in live cells.	53
4.4 Simulating blinking with HMM.	58
4.5 Optimal memory parameter and effect with varying photo-activation density.	59
4.6 The effects of corrections on bound molecules.	62
5.1 Single-molecule tracking of over expressed LacI-PAmCherry and LacIMut-PAmCherry.	71
5.2 Single fits for Plasmid LacI with ONPG and IPTG.	72
5.3 Spatial distribution of the localisation along the short cell axis.	75
5.4 WT cells with no fluorescent proteins.	78
5.5 Copy numbers per cell length.	79
5.6 Spot intensities from the localisations for each experiment.	80
5.7 Native copy numbers with tandem operators 6xlacO and all opera- tors removed NolacO.	81

5.8	6xlacO and NolacO strains description.	82
5.9	6xlacO strain with 20 Mal operators description.	84
5.10	Clustering of LacI-PAmCherry near the FROS marker.	86
5.11	Number of LacI monomers relative to FROS position.	87
5.12	Sub-diffraction positions of LacI.	90
5.13	D* and spatial distribution of AraC-PAmCherry <i>in vivo</i>	95
5.14	Copy numbers per cell length for the AraC-PAmCherry strain.	96
5.15	D* and spatial distribution of CRP-PAmCherry <i>in vivo</i>	98
6.1	Simulating the diffusion of single-proteins in live cells.	103
6.2	Simulating tracks with one diffusion coefficient	106
6.3	The relationship between unbiased diffusion coefficient and the ap- parent diffusion coefficient.	107
6.4	Changing the ratio of interaction between mobile and binding.	109
6.5	Simulating LacI D* distribution	112

LIST OF TABLES

4.1	Measured rates for PAmCherry in live <i>E. coli</i>	55
-----	---	----

ACRONYMS

cAMP	cyclic adenosine monophosphate
CRISP	continuous reflection interface sampling and position
CRP	cAMP receptor protein
D	diffusion coefficient
D*	apparent diffusion coefficient
<i>E. coli</i>	<i>Escherichia coli</i>
EGFP	enhanced green fluorescent protein
EMCCD	electron multiplying charge coupled device
FCS	fluorescence correlation spectroscopy
FP	fluorescent protein
FRAP	fluorescent recovery after photobleaching
FROS	fluorescent-operator repressor system
FWHM	full width half maximum
GFP	green fluorescent protein
HMM	hidden markov model
IPTG	isopropyl β -D-1-thiogalactopyranoside
<i>/lac operon</i>	lactose operon
LacI	Lac Repressor
MSD	mean squared displacement
NA	numerical aperture

ONPG	2-Nitrophenyl β -D-galactopyranoside
PAFP	photo-activatable fluorescent protein
PALM	photo-activated localisation microscopy
Pol1	DNA Polymerase I
PDF	probability distribution function
PSF	point spread function
RAMM	rapid automated modular microscope
RNAP	RNA Polymerase
TF	transcription factor
TIR	total internal reflection
TIRF	total internal reflection fluorescence

MOTIVATION

Imagine a man wakes up in a large room with what seems to be an infinitely large and rolled out scroll with innumerable amount of letters written on it. He somehow remembers that he was set the task to find one word in this scroll with almost 5 million letters. How should he do this? It will take some time, but it might not be so difficult.

Then he realises that he's not alone in the room, it is crowded, very crowded with people doing all types of things. There's people organising the scroll so it occupies less space, there's people copying the whole scroll unto another scroll and there's also some large people, the transcribers, who are reading the scroll. "They're the ones I'm supposed to stop from reading the code that comes after my word", he remembers. "But where is that word?!" All these people they're constantly bumping him around, he has no control of where to go in the room. But somehow he is able to find the scroll every now and again. He is able to read 40 letters at a time before being bounced away from the scroll once again, until finally he finds the word and he can at last do what he

was meant to do.

This is the *lac* repressor's and other transcription factors daily task. This thesis is about the technologies that have allowed us to observe how these fundamental proteins act in real-time. Solving and understanding some of the limitations of the techniques.

This thesis is about placing a searchlight on the man's head to observe how fast, where and in which way he is searching for that specific word.

IMAGING SINGLE-MOLECULES IN LIVE BACTERIA

2.1 INTRODUCTION

In this chapter I introduce the proteins studied in this work and the challenges involved in imaging them at the single-molecule level. I summarise previous single-molecule methods and super-resolution microscopy which are used to obtain and analyse transcription factors *in vivo* in bacteria.

2.1.1 Transcription factors

The bacterial DNA contains the information to make thousands of molecules. Each set of genes in the genome can express several molecules related to a cell's function. For example a single gene can express molecules in a pathway for carbon utilisation, the breaking down of sugar for cellular energy. But bacteria do not express all the genes contained in their DNA at all times, some are controlled and only activated in response to environmental changes.

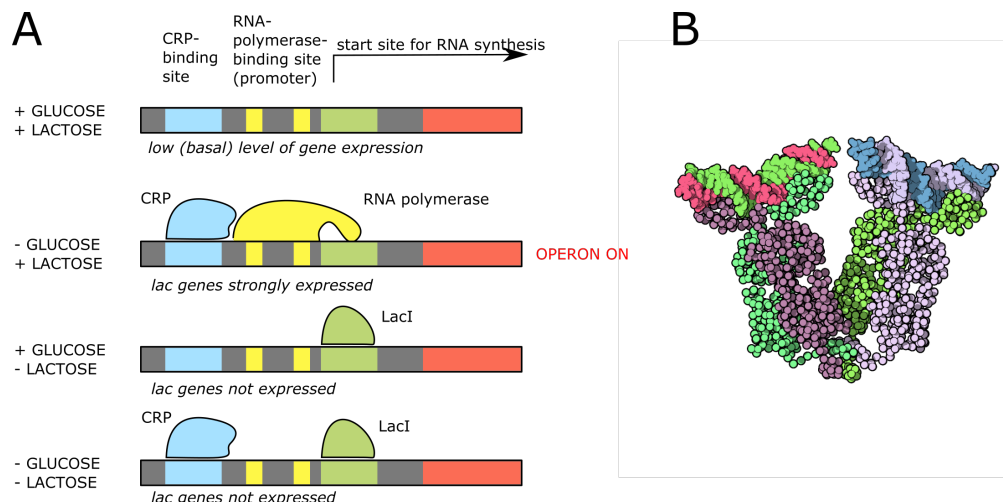


Figure 2.1: **The *lac* operon and the structure of LacI.** In A) the activation and repression of the *lac* operon is shown. In B) the LacI tetramer is shown binding to two operator regions. Each LacI monomer is shown in a different colour. A) based on (Alberts et al., 2008) and (Wikimedia-Commons, 2010)

One way of controlling gene expression is by interacting directly on the gene itself by the use of transcription factors (TFs), which are proteins that bind a specific sequence in a gene to repress or promote its transcription.

The first transcriptional repressor ever discovered was the *lac* repressor (LacI) (Gilbert and Muller-Hill, 1966; Lewis, 2005). It represses the *lac* operon, a set of back-to-back genes that encode proteins involved in the transport and digestion of lactose. In the operon, there is a regulatory sequence called the operator that LacI recognises and binds to repress the gene by preventing the RNA Polymerase (RNAP) from transcribing it.

The *lac* operon (fig. 2.1) combines positive and negative controls; the LacI stops repressing the gene when there is presence of lactose in the environment. Allolactose, whose concentration increases with the presence of lactose, binds to LacI thereby preventing the complex from binding DNA specifically. Basically, LacI negatively controls the operon, until it's signalled to stop. The

absence of a bound repressor to the operator is not sufficient for transcription as the operator also responds to a second signal, the presence or absence of glucose.

Even in the presence of lactose, *E. coli* prefers glucose as its carbon source (Alberts et al., 2008). In this case transcribing the *lac* operon would be a waste. To activate the operon a second transcription factor comes into play: the cyclic adenosine monophosphate (cAMP) receptor protein (CRP). CRP is a global regulator, it is involved not only in activating the *lac* operon but other operons involved in carbon utilisation.

When glucose levels are low, increasing concentrations of the signalling molecule cAMP binds to CRP enabling it to bind to a specific sequence in the DNA. With regards to the *lac* operon, CRP binds and activates the gene by directly interacting with RNAP. CRP positively controls the operon when lactose is present and glucose is absent. This is one example of many of the TF controlled expression that occur in *E. coli* (Alberts et al., 2008).

The search processes by which TFs find their specific targets have come into great scrutiny in recent years (Mueller et al., 2013). A great amount of *in vivo* single-molecule work on LacI has shown that it combines sliding on DNA and 3D diffusion to find its target (Redding and Greene, 2013). Some of this work is reviewed in section 2.3.

In this thesis I show and discuss novel strategies to study TFs *in vivo* and the limitations of using super-resolution microscopy techniques, which are capable of imaging single-molecules at unprecedented resolutions. In the following sections I review the main concepts involved in achieving single-molecule detection.

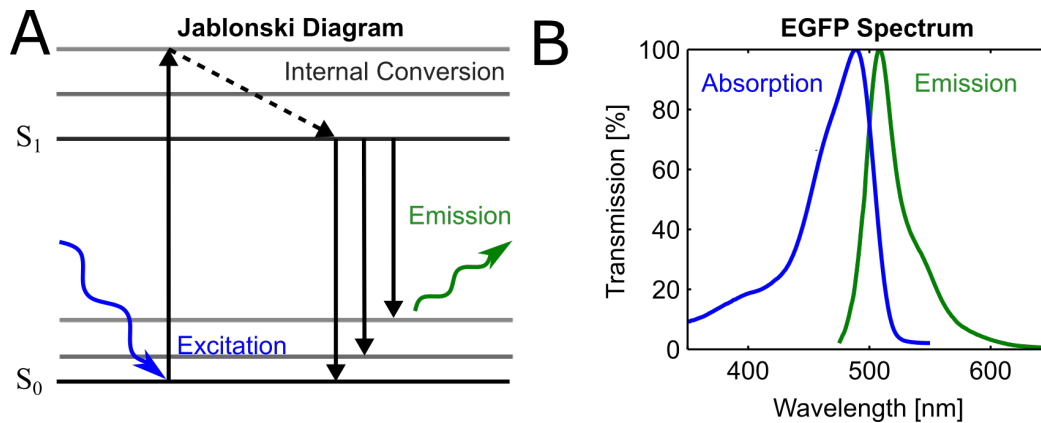


Figure 2.2: Jablonski diagram and spectrum of the enhanced green fluorescent protein. A) Energy levels and transitions schematic of fluorescence. The incoming light excites an electron from the S_0 singlet ground state to a state higher than the first excited state S_1 . Some energy is lost through internal conversions as shown by the dashed line, which causes the excitation energy to be higher than the emission. Most often, the excited electron relaxes to the lowest excited state S_1 , from the on fluorescence can occur from the S_1 state and S_0 and its excited vibrational states. In B) the absorption and emission spectrum of a typically used fluorescent protein is shown. Absorbed photons are later emitted shifted to longer wavelengths, but the emission spectrum has close to mirror symmetry from the absorption spectrum. The symmetry is due to the fact that the vibrational transitions (internal conversions) that occurred during absorbance are the same than from the excited state to the ground state (Jennison and a.F. Morgan, 1950).

2.2 FLUORESCENCE

The method by which we are able to directly observe proteins in cells is fluorescence. In fluorescence, incoming photons excite electrons to a higher energy level. The electron may relax to lower levels by internal conversions but they may emit a photon which is spectrally different (higher wavelength) than the excitation photons (fig. 2.2A).

Ever since the discovery of the green fluorescent protein (GFP) (fig. 2.2B, EGFP Spectrum), many applications have been discovered for FPs. The FPs can be used as expression reporters by inserting a gene encoding for an FP in

an operon. By measuring the bulk fluorescence you obtain a direct measurement of the operon's transcription, albeit delayed because of FP maturation (protein folding time). Another use for FPs is to insert them at the start or end of a protein of interest. In most cases the protein function will not be impeded and will reveal its function and location in live cells (Alberts et al., 2008).

Using FPs can also probe the interactions of two proteins. By labelling the proteins of interest with two different fluorophores whose emission of one overlaps with the absorption spectrum of the other. Monitoring two proteins in this way is called fluorescence resonance energy transfer (FRET). When the two proteins come into close proximity of each other (around 2 nm) the absorbed excitation of one FP can be transferred to the other which will subsequently emit a photon within the range of its emission spectrum. This phenomenon reports on the proximity of two FPs that may be interpreted as a direct interaction between the proteins (Alberts et al., 2008).

Another useful application involves tagging sections of the genome with FPs, which is done by using TFs as an imaging tool. Because TFs bind to a sequence specifically it's possible to insert a number of tandem operators in any region of the genome as long as it doesn't affect the cell growth and function and that they are still viable. By expressing a TF fused with an FP from a plasmid you are able to see the specifically bound molecules fluorescing. This technique is called fluorescent repressor-operator system (FROS) (Reyes-Lamothe et al., 2008).

For *in vitro* work synthetic molecules have been used to tag proteins and study them at the single-molecule level due to their longer photo-bleaching and brightness. Using synthetic dyes has been particularly helpful in single-molecule FRET (Holden et al., 2010), this time not to measure the interaction

between two proteins but to measure the distance between two amino acids on a protein. New efforts have been made to investigate inserting proteins labelled with synthetic dyes into live cells. Micro-injecting eukaryotic cells (Speil et al., 2011) or electroporating bacteria with a solution with labelled proteins (Crawford et al., 2013; Plochowietz et al., 2014).

Both approaches for introducing labelled proteins with synthetic dyes are technically challenging (Mueller et al., 2013), tagging molecules with FPs remains the best way of imaging the spatial distribution and function of proteins in live cells. They have become even more important with the emergence of photo-activatable variants of FPs for use with techniques that depend on stochastic activation such as single-molecule super-resolution microscopy.

2.2.1 Single-molecule detection

Single-molecule studies have gained relevance for their ability to measure the properties of individual molecules in contrast to a population as whole. Furthermore, (Walter et al., 2008) as technology has advanced camera sensitivity has allowed us to image dimmer fluorophores and at faster exposures. Single-molecule techniques have the advantage of observing heterogeneities that would be otherwise averaged in an ensemble measurement. It is also possible to observe real-time dynamics within a single-molecule which would otherwise be impossible in bulk averaging.

It is now possible to image single-molecules in live cells in optical microscopes, by reducing background from the sample, minimising stray light ending up on the camera and having a sensitive, low-noise camera. These can be attained mainly by the illumination of the sample, high-quality optical components and EMCCD cameras.

In *in vitro* work, it is possible to observe single-molecules diffusing in solution or by immobilising them on a surface. By minimising the background and optimising the experiment to spatially separate the molecules it is possible to observe single-molecules.

It is not the same when considering the observation in a bacterial cell. The cells are so small that distinguishing more than one fluorophore emitting at a time per cell might be impossible. The issue is resolution, the smallest measurable distance between two single-fluorescent probes; at best microscopes can have a resolution of around ~ 200 nm. In the following sections I explain the techniques that have allowed us to resolve single-molecules in live *E. coli* via fluorescence microscopy.

2.2.2 Microscopy

When using a fluorescence microscope there are methods to reduce the background of the sample by how you illuminate it. Where you position the incident excitation beam will have a large effect on the background collected from the sample.

In epifluorescence microscopy the excitation beam (fig. 2.3, red beam) is focused at the back focal plane (fig. 2.3, dashed line) of the objective and the subsequent fluorescence is collected from the same objective. This creates a collimated beam that deeply penetrates the sample. This penetration generates a lot of background from out of focus regions in the sample, which causes non-negligible amount of autofluorescence to be collected.

A different technique uses high numerical aperture objectives to create a beam that will have a high incident angle on the slide surface. This is achieved

by positioning the incident beam near the side of the objective. This causes total internal reflection (TIR) at the slide surface, creating an evanescent wave capable of exciting molecules up-to around 200 nm into the sample (fig. 2.3, blue beam). This dramatically decreases the background since very little of the out of focus sample is illuminated. Unfortunately, for most cells this is not deep enough to image as much as the cell volume as desirable but ideal for membrane studies (Manley et al., 2008).

Finally a new technique finds a middle ground between these two modes. By placing the incident beam near the objective edge but not far enough to create TIR, you can reduce background and penetrate through samples farther than TIR fluorescence (TIRF) microscopy. This technique is termed near-TIRF but is also known as highly-inclined and laminated optical sheet (HILO) microscopy (Tokunaga et al., 2008) when the illumination area is also minimised (fig. 2.3, green beam). Placing the excitation beam near the edge leads to a highly inclined beam at the sample. The beam penetrates as far as the epifluorescence mode beam, but less out-of-focus background is collected by the objective. In our work we use near-TIRF to image single-molecules in live cells, because it gives the best balance between sample penetration and low background to increase the signal-to-noise ratio of our experiments.

2.2.3 Single-molecule detection by light microscopy

By going through a microscope's optics, single-point emitter of fluorescence will be spread by diffraction into what is called the point-spread function (PSF, fig. 2.4) that is seen on the camera. As an illustrative example, two emitters are simulated without any camera noise and are sufficiently spaced apart. Each emitter is centred in the approximate 2-D Gaussian shown in fig. 2.4. When

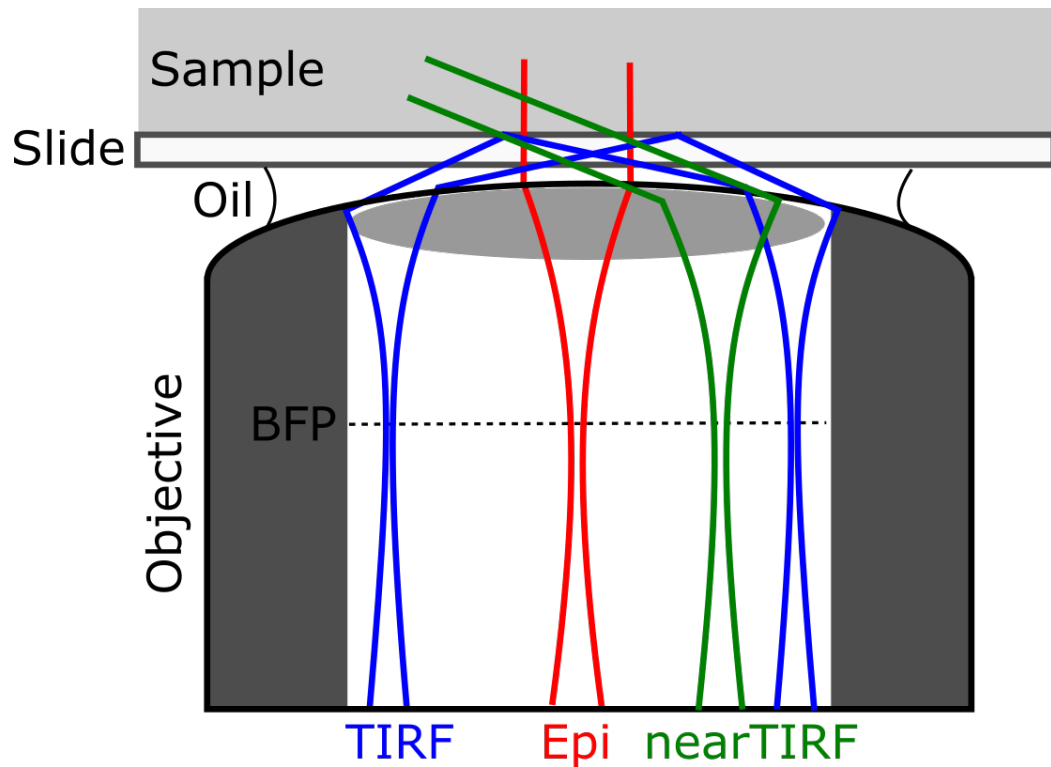


Figure 2.3: **Excitation beam position in fluorescence microscopy.** For a fluorescence microscope with a high NA, multiple illumination modes are possible depending on the experiment. Epifluorescence microscopy (red beam) consists of placing the excitation beam through the centre and focusing it at the back focal plane of the objective. This leads to a excitation beam that penetrates deeply into the sample but causes out-of-focus background. TIRF microscopy (blue beam) consists of placing the excitation beam on the edge of the objective. In this case the beam will have a high incidence angle on the slide causing the sample to be illuminated with an evanescent wave penetrating upto around 200 nm into the sample. Finally, near-TIRF microscopy (green beam) consists of placing the excitation beam on the edge of the objective but not far enough to enter TIRF. This causes a beam with a high incident angle that deeply penetrates the sample but causes less out-of-focus background.

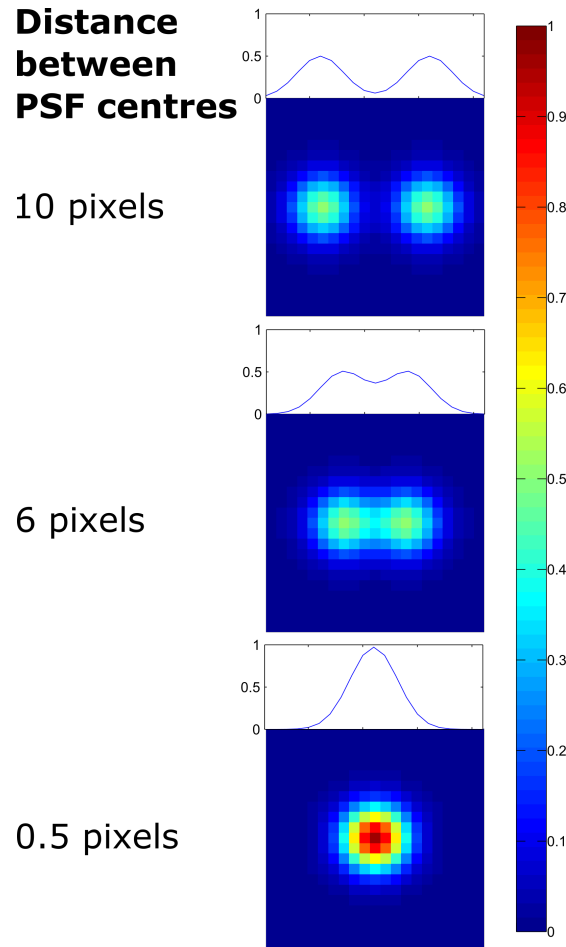


Figure 2.4: **Resolution and the distance between PSFs.** Two PSFs can be identified individually if they are separated by enough distance. In this figure two simulated PSF images with no noise and a full width half maximum of 1 pixel are shown with 10, 6 and 0.5 pixels apart. Although the two single-emitters cannot be distinguished from each other the brightness of the resulting image of the PSF is double that of the single-image. This fact is used in algorithms like DAOSTORM (Holden et al., 2011), where fitting two PSF models results in a better fit than a single model.

the spread is large we can easily estimate their centres (or localise) and count them by eye (fig. 2.4 10px separation; the image of the PSF can be distinguished).

Once the separation between the different emitters is shortened the task is more complicated because the image of the PSFs overlap, but both can still be

distinguished from each other, that is we can still estimate that there are two emitters whose centre separated by 6 pixels (fig. 2.4 6 px separation). But once the separation becomes too small, the task is impossible, we are unable to distinguish two clear centres, and it is difficult to measure the distance between them. We may still be able to count two emitters from the intensity of the large PSF. The intensity can sometimes be used as a tool to distinguish a single PSF from two, and this is used for dense localisation algorithms like DAOSTORM (Holden et al., 2011).

The simplified versions of the resolution problem of fluorescence microscopy is shown in fig. 2.4. In the example, we only needed to localise two emitters, but in live *E. coli*, copies of proteins go from tens to thousands in very small areas where two molecules may be undistinguishable from each other.

2.3 SINGLE-MOLECULE IMAGING OF TFs *in vivo*

2.3.1 Localising single-molecules *in vivo*

With such high concentrations of proteins in live cells how would we be able to identify single-molecules? To localise a single-molecule tagged with an FP we need to be able to resolve the PSF image from the noise (fig. 2.5A). To localise the molecule we fit the PSF with a 2-D Gaussian (fig. 2.5B) which yields sub-diffraction level positional information (Stracy et al., 2014; Betzig et al., 2006; Gahlmann and Moerner, 2013; Hensel and Xiao, 2013; Li and Elf, 2009). Isolated and bright enough PSFs can be localised over successive frames (fig. 2.5C, top). By joining together the successive localisation we can track the molecule (fig. 2.5C, bottom, arrows).

But as we mentioned before, most proteins exist in high concentrations,

with PSFs being resolvable if they are more than 200 nm apart and the *E. coli* cell having around 1 μm in width by 2 μm in length there is not much space to work with (fig. 2.5C, bottom, cell).

One technique to image DNA-binding proteins uses their properties to image them (Elf et al., 2007; Hammar et al., 2012). First you ensure low enough concentration of the protein tagged with an FP (copy number < 10). At large enough exposure times proteins that diffuse fast and throughout the cell are blurred and appear on the camera as added background in the cell (fig. 2.5D, top). When the protein only binds to DNA for enough time, you can image them without the added background of the diffusing species (fig. 2.5D, middle). But even in a combination of both, you can still localise single-proteins *in vivo* (fig. 2.5D, bottom). This technique has been applied to TFs as I review in the next section.

2.3.2 Enhanced localisation of LacI

As I introduced before, TFs control the expression of genes by binding and unbinding sites in DNA. In prokaryotes they play a key role in adaptation to environmental changes or internal triggers. How TFs find sites to bind to along with their residence times has been a matter of discussion (Mueller et al., 2013). Their binding is related to their function; they may be bound, specifically repressing or activating a gene, or non-specifically searching for a site. Measuring their activity *in vivo* is crucial to find the kinetics of the TF.

Fluorescent proteins allow direct tagging of the TFs; fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) are commonly used to measure the kinetics of TFs in live cells. Because of its inherent averaging the techniques measure the mean of the diffusion coef-

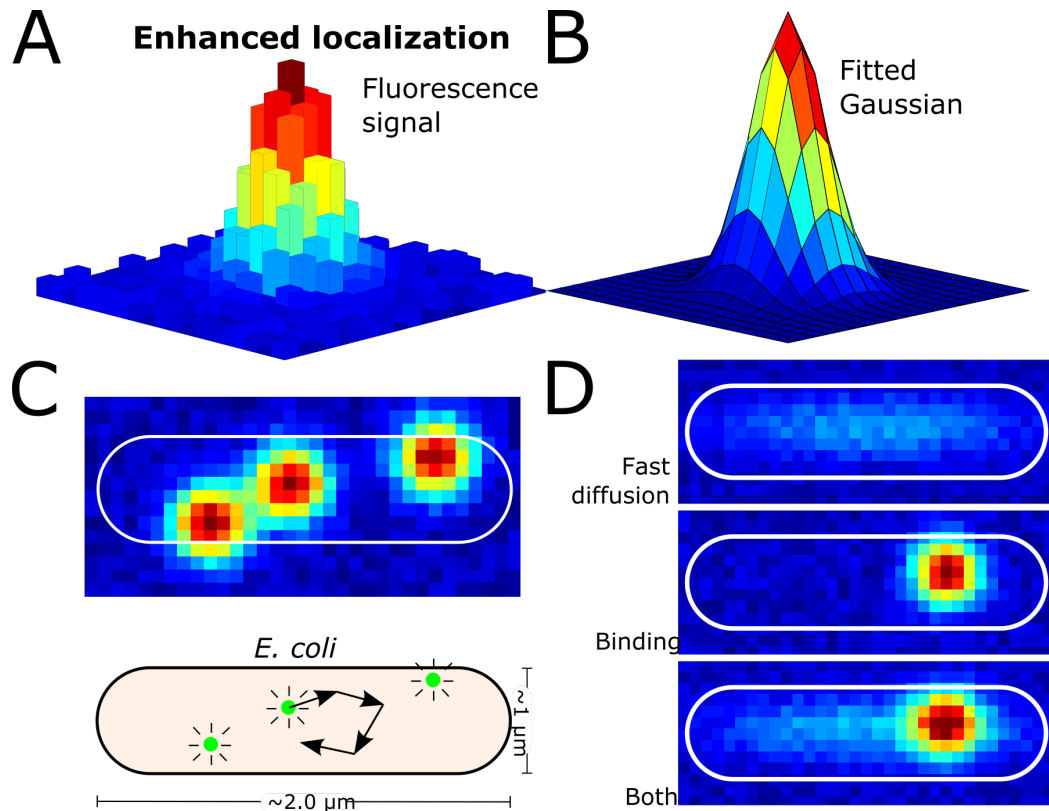


Figure 2.5: **Localising single-molecules in live cells.** The pixelated fluorescence image in A) is fit with a 2-D Gaussian to extract the localisation at sub-diffraction resolution, typically 40 nm for FPs. In C) (top) the images of PSFs that are isolated and bright enough are localised over successive frames and tracked by joining together the successive localisations (bottom). In D) the diffusive properties of DNA-binding proteins can be used to localise them at long exposure times. When a protein diffuses fast, in a long exposure, the signal will appear blurred (top). If the protein binds DNA for enough time, it's possible to resolve the PSF image (middle). It may also be the case that a fraction of molecules may be diffusing and another fraction bound (bottom).

ficient and the association and dissociation of the TFs. The limitations of these techniques are significant: FRAP and FCS are bulk methods of measuring the diffusion and binding of the TFs. Because of this they may be measuring a combination of specific and non-specific binding; if non-specific interactions with DNA dominate in the pool of probed proteins the details of the specific binding will be lost.

Single-molecule experiments in live cells, combined with single-particle tracking, can fill in the holes that FRAP and FCS leave behind. The techniques are able to extract information from a single-TF measuring the distribution of binding times and diffusion. But to be able to localise the proteins in a live cell, a low concentration of the protein is needed. This has been achieved lowering the copy number of the protein of interest that is natively expressed in the cell. In addition to low copy numbers, having enough signal to localise a molecule is important, thus *in vivo* experiments concentrate on bound proteins as shown in fig. 2.5D.

In (Elf et al., 2007) they created strains that generate lower copy numbers to be able to distinguish single-lac repressor (LacI) protein specifically bound to DNA in the cell (as in fig. 2.5D). The LacI is a model protein to study other more complex transcription factors. It controls the expression of genes that digest lactose. Under the absence of the sugar the LacI dimer binds the *lacO1* operator in the gene together with a second dimer binding the weaker *lacO2* and *lacO3* operators. This may cause looping by the tetramerisation of the dimers, but Elf *et al* fusion with a FP prevents this from happening.

The *lacO3* gene overlaps part of the gene that expresses *LacI*; the LacI's transcription is auto-repressed by the auxiliary operator. To achieve lower copy numbers per cell, Elf *et al* changed the *lacO3* operator by the stronger *lacO1*,

increasing LacI's repression of itself. By using 1 second exposure per frame they were able to discern a single bound LacI, whilst the PSF images of the non-specific and faster molecules were blurred. Further they treated the cells with isopropyl β -D-1-thiogalactopyranoside (IPTG), which binds LacI and prevents it from binding specifically to any of its operators. This led to the amount of bound molecules to lower; in high concentrations of IPTG no spots appeared in the cell and just like in strains where all native operators in the genome are deleted (i.e. no operators to bind to in a cell). This method allowed them to quantify the amount of bound LacI at any certain time.

In addition by reducing the exposure time and imaging with IPTG induction, they were able to measure the amount of time LacI spends non-specifically bound and the effective diffusion from the non-specific binding. When you measure a PSF, if its width is close to the diffraction limit it is bound or the exposure is fast enough that it looks bound. But if the exposure time is higher than the residence time the LacI will leave an image of the PSF with higher widths. They discovered the non-specific binding time by finding the exposure at which the PSF width changes between an IPTG treated cell and an untreated one.

The diffusion was done in a similar manner; they calculated at different exposures the distance travelled by the *lac* repressor, from that they constructed their mean squared displacement at different time lags and calculated an effective diffusion coefficient of $0.4 \mu\text{m}^2 \text{s}^{-1}$. By taking FCS measurements of the LacI without its DNA binding domain, they found that the diffusion coefficient was $3 \mu\text{m}^2 \text{s}^{-1}$, they concluded that the LacI spends 90 % of the time sliding on DNA and the rest diffusing in 3D. This was the first *in vivo* results that proved that the *lac* repressor has facilitated diffusion, and the first technique that can

discriminate between specific and non-specific binding.

These results were complemented with other work (Hammar et al., 2012), in which they showed that the repressor slides for an average of 45 base pairs. This was achieved through an *in vivo* assay in which they insert mutations in which two identical lac operator sequences are separated by varied distances. By placing the operators closer to each other, they can find the point where the 2 operators have the same rate of binding as one operator, which gives you an approximation of the sliding length. Further, they measured the association rate of the repressors by removing IPTG.

A likely scenario in cells is that there are multiple proteins binding DNA close to an operator of a TF. An assay where they add TetR, another TF, so that it binds closely to a single operator showed reduced rate of association. They found implications that there is a probability related to the binding of an operator and that it is not solely limited by diffusion. There is probability associated to the binding to an operator before the LacI slides to the adjacent base pair. This leads to further complications of TFs search mechanisms.

Single-molecule tracking techniques can distinguish non-specific from specific binding because they can measure the properties of a single-TF instead of the mean. The technique is limited to fast binding times, if a TF has a residence time of the order of minutes, observing TFs is limited by photo-bleaching. The issue may be circumvented with time-lapse experiments to extend the observation time but the use of synthetic dyes may be able to aid in this respect, at this moment it is a field in development (Crawford et al., 2013). In addition, using photo-activatable fluorescent proteins to temporally control the observable number can be equally applied to this type of studies without the need to limit copy numbers.

There is much to be studied in the function of transcription factors. How residence times and search modes affect the function of repressors or activators is still an open question (Mueller et al., 2013). The kinetic influence on how a TF finds an operator, and how the specific binding residence times influence the repression, are part of the future of field.

2.3.3 Super-resolution microscopy

If we were to tag proteins endogenously in an *E. coli* and excite them, all of the FPs would emit at the same time and it would be impossible to recognise a labelled cellular structure (fig. 2.7A) from the fluorescence image (fig. 2.7B) or any diffusion information. Previously this problem was solved by studying low copy number proteins (section 2.3). To be able to reconstruct the structure a new set of techniques were invented to circumvent the resolution limit of light microscopes. These techniques are named super-resolution microscopy.

Super-resolution is preceded by great efforts and advancement in the quality of cameras and processing power. The imaging of single-fluorescent dyes was first by Funatsu et al. (1995), where they could identify myosin molecules bound to actin filaments and image them at video rate speeds. Further to this work Schmidt et al. (1996) realised the potential to extract diffusion of a protein in a phospholipid membrane. Since then a number of techniques and analysis methods were devised to analyse localisations of single-molecules (Saxton and Jacobson, 1997). But all this work is in membranes or *in vitro*, live cells work, specifically bacterial work remains difficult. Super-resolution techniques are able to solve this problem.

Super-resolution microscopy encompasses many different techniques that image fluorophores at sub-diffraction resolutions (Huang et al., 2010). One

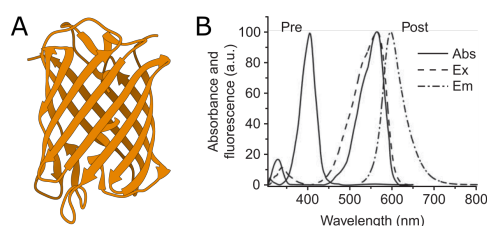


Figure 2.6: **Structure and spectra of PAmCherry.** In A) the structure of PAmCherry is shown with the barrel structural typical of FPs. In B) the spectra is shown for its pre- and post-activation states.

subset of these techniques, termed single-molecule localisation microscopy, activate or blink fluorescent probes to obtain a sparse field of view to be able to resolve them individually. Photo-activated localisation microscopy (PALM) aims to circumvent the resolution of a microscope by using photo-activatable fluorescent proteins (PAFP). It is possible to turn on proteins stochastically (Betzig et al., 2006). A labelled structure (fig. 2.7A) cannot be resolved if all fluorophores are activated (fig. 2.7B). If we can have a sparse activation of the fluorescent proteins then they can be individually localised (fig. 2.7C)

2.3.4 Photo-activatable proteins

To perform PALM experiments on LacI, the repressor carried a C-terminal fusion with PAmCherry fig. 2.6 (Subach et al., 2009), a photo-activatable fluorescent protein that emits in the red spectrum. Its spectrum allows us to do two-colour experiments with GFP, although in our case not simultaneously.

In the next chapter I present a microscope constructed to achieve super-resolution. I go through all the important parts that help us achieve the signal to noise to localise single-molecules in live *E. coli*. I also explain how we are able to track the proteins, analyse the diffusion and interpret the results.

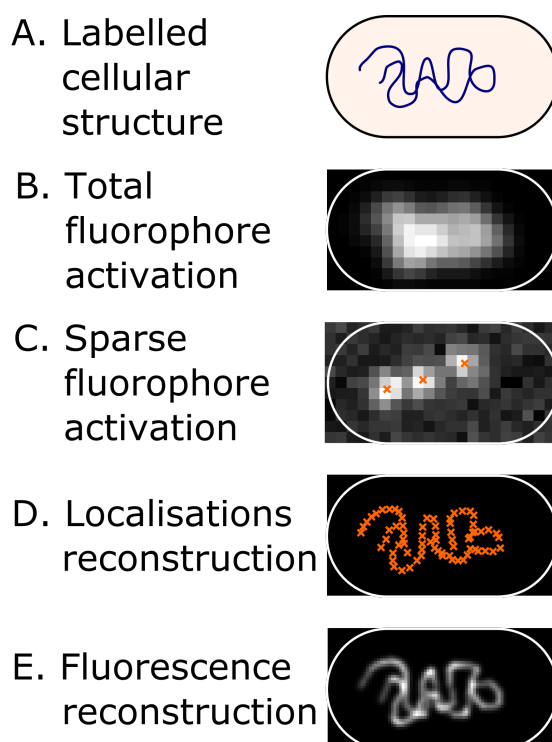


Figure 2.7: **Resolving structures at sub-diffraction resolutions.** Cartoon and simulated example of the detail that is possible to extract from a fluorescence image. In A), any cellular structure is labelled with photo-activatable protein or other synthetic dyes. The cartoon shows the shape of the structure. In B) all fluorophores emit simultaneously, the detail in the structure is lost. In C) the general method of localisation microscopy is shown. The fluorophores are photo-activated sparsely and localised through fitting algorithms. In D) the fitted localisations are collected and displayed revealing the underlying structure. In E) each localisation can then be reconstructed with a model PSF with a shorter width that represents the uncertainty in the localisation error.

TRACKING PALM: MICROSCOPY AND TECHNIQUE

3.1 INTRODUCTION

In this chapter I present our efforts to design and construct a localisation super-resolution microscope that is higher throughput, easy to use and customisable for a variety of experiments. Our work yielded a microscope currently housed in the Department of Biochemistry at the University of Oxford.

In particular a lot of time is taken to explain the main method in this work, tracking Photo-activated Localisation Microscopy (tracking PALM).

3.2 MICROSCOPE DESIGN AND CONSTRUCTION

3.2.1 Single-molecule localisation microscopes

In the medical and biological sciences, fluorescence microscopy has established itself as an essential technique, especially in the study of the molecular biology of the cell. Limited in the past by the resolution limit of light microscopy, over the past 10 years there has been great progress in circumventing this limit through clever techniques. A great amount of research has made it possible to observe a single-fluorescent molecule and to be able to apply this to biological sciences. Using super-resolution localisation microscopy we are able observe single-proteins and study their dynamics in prokaryotes (Stracy et al., 2014) and other living organisms (Liu et al., 2015) in real-time and at the nanometre scale.

Different methods and techniques can be used to construct a microscope capable of detecting single-molecules, but the ability to selectively illuminate a small section of a sample to reduce the signal-to-noise ratio makes TIRF microscopes a common choice in the field (as described in section 2.2.2) (Reck-Peterson et al., 2010). For *E. coli* on a slide these microscopes are ideal to minimise the background from the solution or agarose pad on which its attached.

3.2.2 Materials and Methods

I designed, constructed, aligned, and characterised, in collaboration with Stephan Uphoff and Mathew Stracy, a super-resolution capable TIRF microscope. Because we wanted to increase the ease of use and the ability to use different combinations of laser lines, we opted for the Rapid Automated Modu-

lar Microscope (RAMM) System from Applied Scientific Instrumentation (ASI). The RAMM system functions as a backbone for modular design and provides compatibility with other ASI products like stages and automated focus correction. Mirrors and filters can be exchanged easily to cater for different emission lines.

3.2.2.1 Design and construction

To minimise the loss of excitation power, good alignment of the excitation path to couple the multiple lasers in to the fibre is essential. In a bespoke microscope, misalignment of the emission path before the fibre can occasionally cause trouble. Fortunately, the multi-laser engine greatly simplifies the excitation path (ICHROME MLE-LFA; fig. 3.1). The wavelengths of the excitation beams are dictated by the dyes and fluorescent proteins to be used. The multi-laser engine provides four laser lines at wavelengths 405 nm-488 nm-561 nm-640 nm at ≥ 80 mW. It provides software control of the lasers to vary the powers and shutter control.

The illumination from the fibre needs to be focused at the back-focal plane of the objective. To achieve objective-type TIRF (section 2.2.2), the excitation needs to be collimated by the objective. Failure to do this will cause stray excitation rays at other angles that will increase the background observed on the camera (Reck-Peterson et al., 2010). In our microscope, the output of the laser engine is a single-mode optical fibre which is mounted to a manual stage that also mounts the collimation lens (L_1) and the TIRF lens (L_{TIRF}) focuses the excitation at the back focal plane of the objective. The manual stage can traverse perpendicular to the light path, this way we can switch the microscope from epifluorescence to near-TIRF and finally to TIRF. Finally, if laser powers

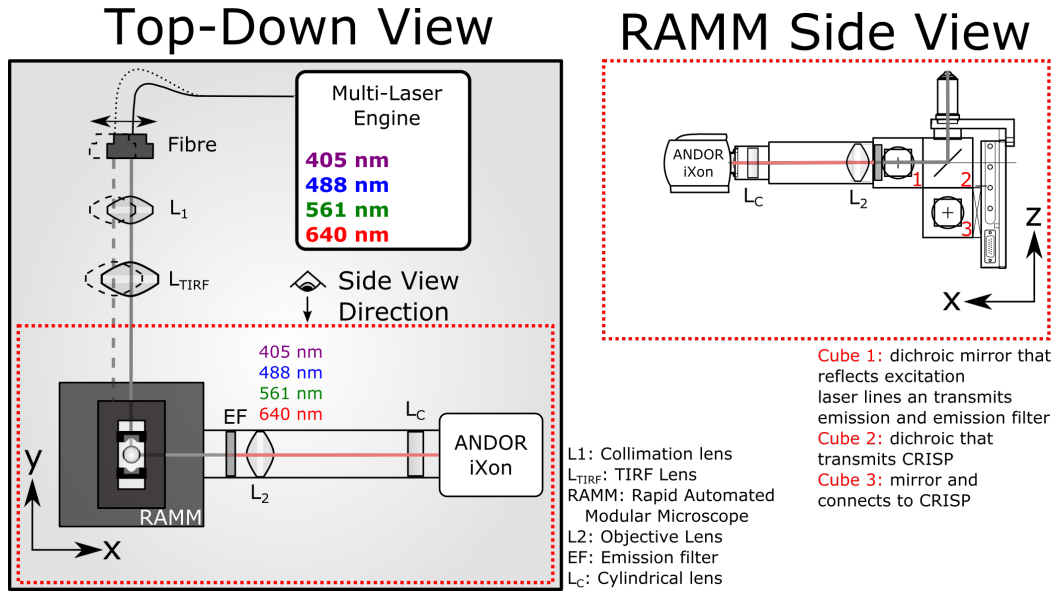


Figure 3.1: **Drawing of the constructed TIRF microscope.** The top-down view shows the setup as if it was observed over the table. The side-view shows the setup from the side as shown in the top-down view. The RAMM side view was adapted from the ASI manual and I added the Andor camera, lenses and light path.

are too high we add a neutral density filter (Thorlabs) in front of the fibre if needed.

Once the excitation line is past L_{TIRF} , it enters cube 1 as shown in the schematic fig. 3.1. Cube 1 contains the interchangeable dichroic that reflects the excitation laser lines (ZT405/488/561rpc, Chroma) and transmits the emission of the fluorophore of choice, it also contains the emission filter (ZET405/488/561NF, Chroma). The design of the RAMM allows for a different combination of laser lines to be used. Usage of the 640 nm filter and dichroic cuts off emission of the mCherry family of FPs and thus in this work no imaging was done with 640 nm laser line.

Excellent dichroic mirrors and filters are needed to successfully reduce background of the experiment. The dichroic mirror must reflect as much as

possible of the excitation light whilst transmitting as much of the emission. Because TIRF reflects the excitation light on the slide, a lot of the light comes back to the dichroic that separates excitation from emission. Unfortunately, the dichroic mirror does not reflect all light and there is a fraction that gets transmitted. Thus the emission filter functions as a final clean-up of any residual excitation light; in TIRF because the excitation follows a different path than the emission a light stop or block may be used to remove it.

Past Cube 1, the excitation light enters Cube 2 that contains a dichroic mirror that reflects the light up to the objective and transmits (T800lpxrxt, Chroma) that transmits the Continuous Reflection Interface Sampling and Position (CRISP, ASI) 890 nm LED. For localisation microscopy experiments, stability of the microscope is essential, the CRISP transmits the infra-red light into the microscope and detects the reflected beam from the slide at a position-sensitive detector. The CRISP then changes the focal-position of the microscope with piezo-stage plate (PZ-2150, ASI) mounted on the RAMM. Cube 3 contains a broadband mirror to reflect the 890 nm LED into the objective and connects to the CRISP module. Cube 2's dichroic mirror allows the infra-red light to pass through whilst reflecting all other light with a lower wavelength.

Both infra-red and excitation light go through the high NA objective (100x, NA 1.4 Olympus), through the immersion oil, slide and the sample. The emission from the sample is collected by the objective, reflected by the dichroic in cube 2, transmitted through the dichroic and emission filter in cube 3. Cube 1 is attached to a tube connected to the EMCCD camera (Andor iXon Ultra, 512x512, 16 μm pixels). In the tube there is the L_2 tube lens (300mm, final magnification 167X) that focuses the image on the EMCCD's chip (final pixel size 96 nm). Optionally a cylindrical lens (L_c) can be placed in front of the

camera for 3D localisation microscopy(see section 3.3.2 for more information).

For tracking PALM we need fast imaging ($\leq 0.5\text{Hz}$) and sensitivity to detect single molecules. Readout noise impedes single-molecule imaging at these levels, the faster the readout the higher the noise's effect. To be able to image dim fluorophores, EMCCDs use electron-multiplication gain by amplifying the signal on chip.

The camera, tube and cubes in the RAMM Side View (fig. 3.1) are attached to a static section of the motorised stage LR50 through cube 3; LR50 is in turn attached to the RAMM. The objective is attached to the mobile section of the LR50 and functions as the coarse z -focusing. In addition to the motorised stage, attached to the RAMM there's a stage to move the sample in the $x-y$ plane (S31121010, ASI) and piezo stage for fine z -focusing (PZ-2150, ASI).

For brightfield imaging we have a condenser that is not on the shown schematic 3.1 but appears on 3.2. The brightfield module is based upon Olympus IX2-LWUCD condenser which is attached to a green LED lamp (MIM-LEDLAMP, ASI).

For PAmCherry we controlled photo-activation through the 405 nm laser line and excited the proteins with 561 nm. Proteins fused to GFP were excited with the 488 nm laser line.

3.2.3 User interface

To control the the smTIRF microscope we use the native software packages that come with the EMCCD camera (Andor SOLIS for Imaging; fig. 3.3, left) and the multi-laser engine (TOPAS iChrome MLE; fig. 3.3, top right). For 3-D calibrations, Micro-Manager and Matlab were used instead of the Andor soft-

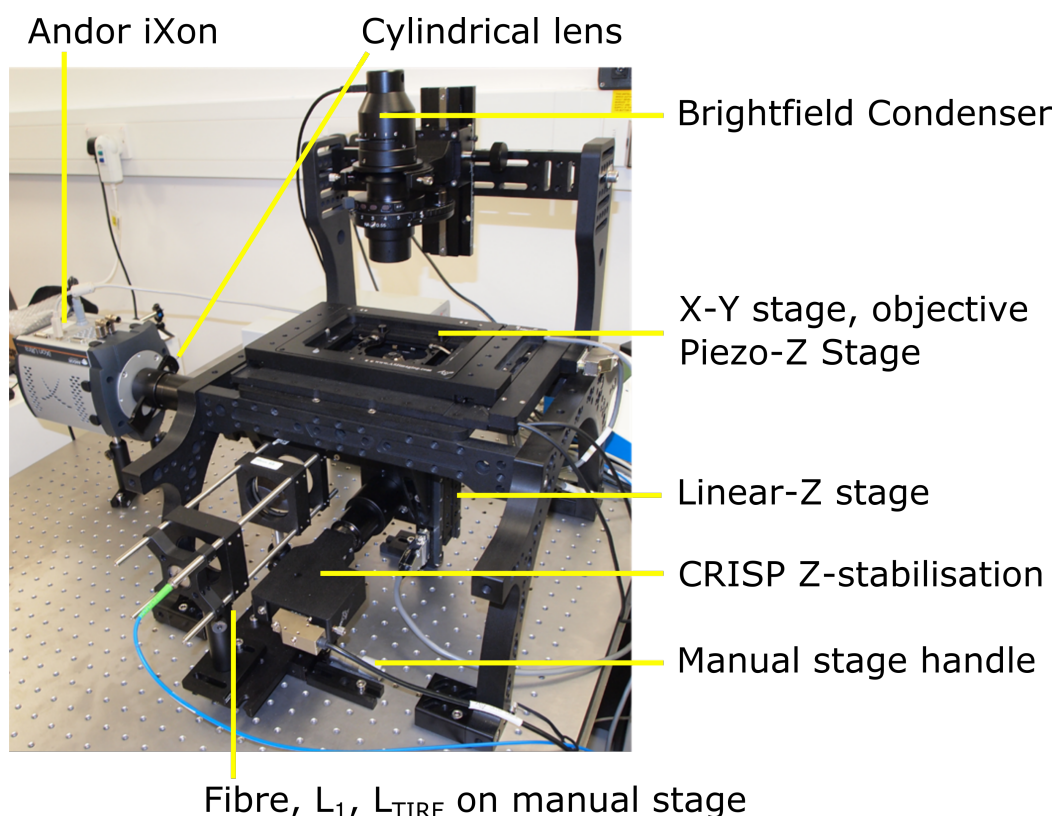


Figure 3.2: **Picture of the TIRF microscope.**

ware (section 3.3.2). Unfortunately there was no method to switch or change the intensity of the LED brightfield lamp. I wrote a GUI (Matlab2011b, Mathworks) that interacts with the stage control to facilitate usage (fig. 3.3, bottom right).

3.3 SINGLE-MOLECULE EXPERIMENTS AND PSF MEASUREMENT

To show that we could image single-molecules, we imaged Cy3b using biotin-BSA+neutravidin slide with a short biotinylated single stranded DNA-Cy3b (more information on technique appendix A.1.1). The molecule was imaged with 3.9 mW (183 mW cm^{-2}) of 561 nm laser line in TIRF mode to reduce back-

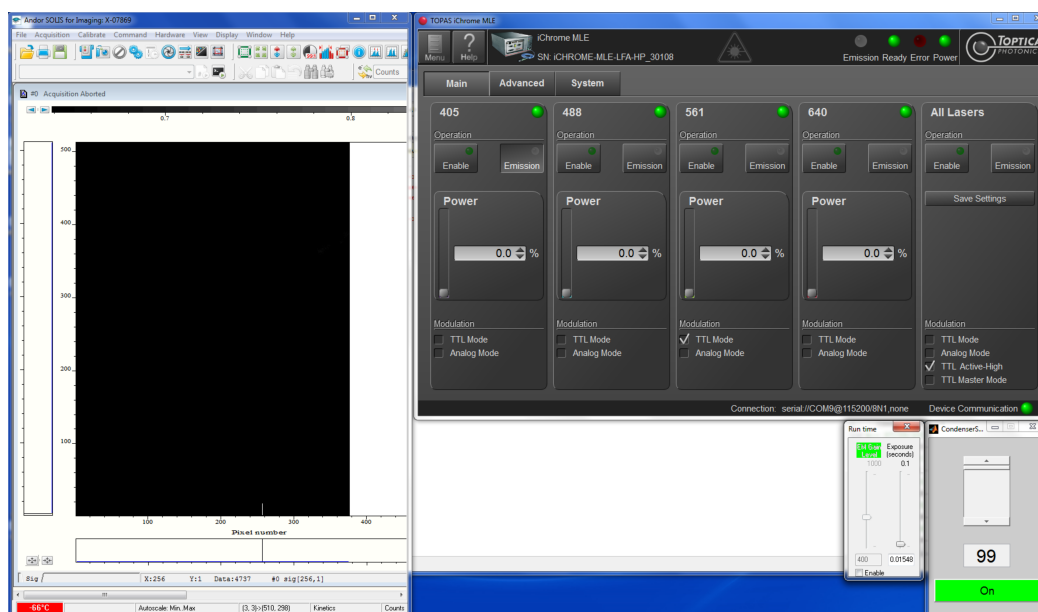


Figure 3.3: **Microscope user interface.** On the left the Andor SOLIS for capturing the images is shown. On the top right, the TOPAS iChrome software for controlling the lasers is shown. Lastly, at the bottom right the gain parameter from the Andor Solis programme and our own MATLAB coded brightfield controller.

ground from the solution.

Each PSF on the surface is fit with a 2-D Gaussian and the full width half maximum (FWHM) of the fitted function, which is a good approximation of the resolution. In this case the resolution is around 270 nm which is around what is expected for emission of Cy3b and an NA of 1.4. The fitted PSF images with lower FWHM than the expected for Cy3b probably represent spots that were fitted erroneously or fitting pixels that arise from noise that are not thresholded and removed before fitting.

3.3.1 Estimating chromatic aberrations

We work with different fluorescent probes which will emit at different colours. In some cases we will image samples for two-colour measurements. In exper-

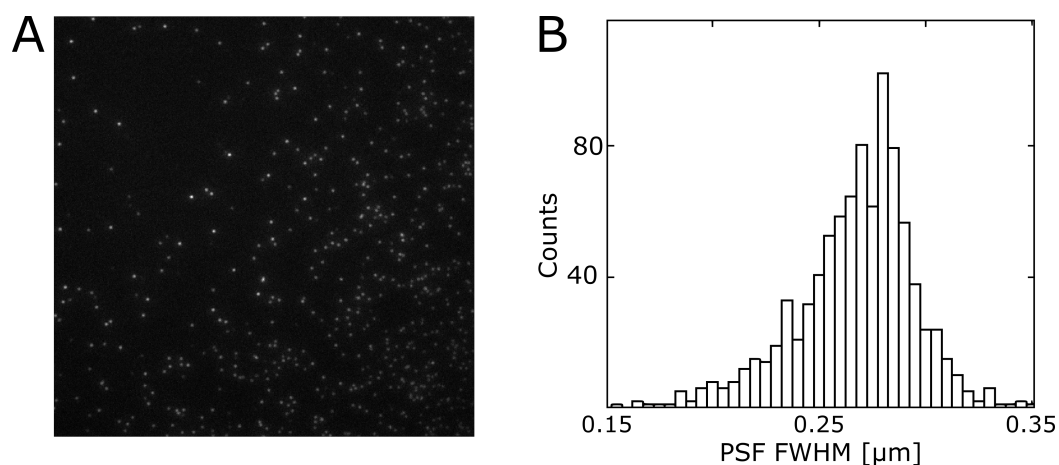


Figure 3.4: **Imaging single-molecules and their experimental point-spread function size.** A) shows a field of view of single-stranded DNA immobilised on a surface. B) the size of each PSF of the ssDNA-Cy3b molecules was measured by Gaussian fitting.

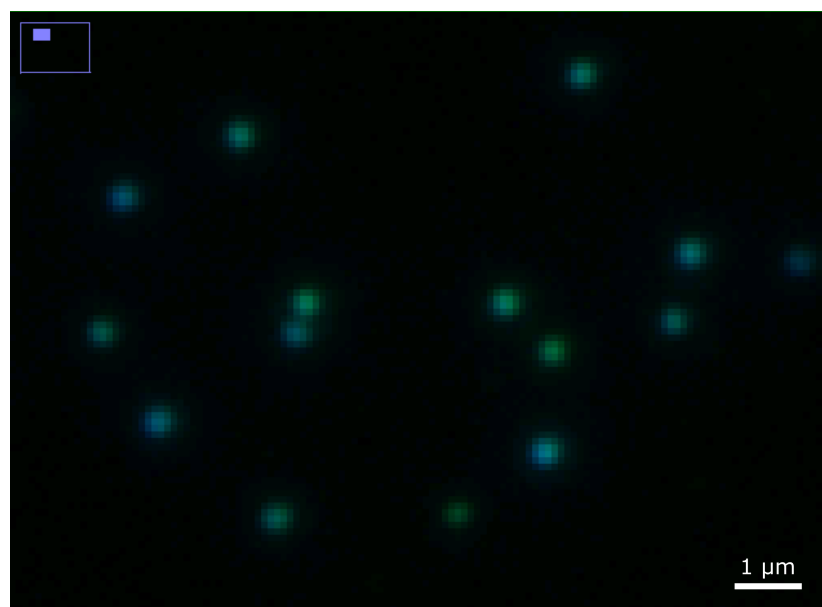


Figure 3.5: **Overlaid field of view of fluorescent beads.** The same field of view of multi-fluorescent beads is imaged at 488 nm (blue PSFs) and 561 nm (green PSFs).

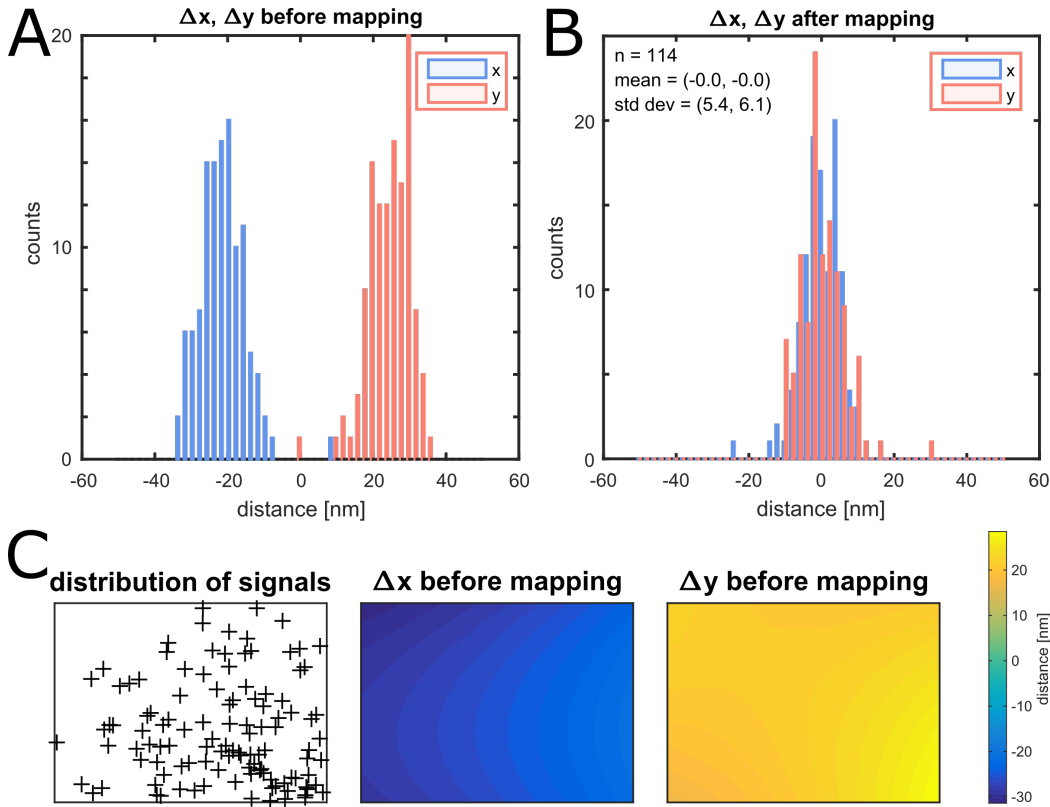


Figure 3.6: Two-colour registration. In A) we localised each PSF and calculated the closest distance between each field of view. The average aberrations are around 20 nm (fig. 3.6A) for both lateral dimensions x and y . In B) the same registered fields of view have an average deviation of 0 nm. In C) the distribution of the analysed signals for all fields of view and their positions on the camera are shown (left). How the Δx (centre) and Δy (right) deviations depend on the camera chip are shown ().

iments with split channels multi-fluorescent beads are (Barysch et al., 2010; Metcalf et al., 2013) used to register the two channels. In our case we designed our microscope to only have one channel and multi-colour images are acquired in succession instead of simultaneously.

To assess the chromatic aberrations of two channels (488 and 561 nm excitation) I prepared 100 nm TetraSpeck beads fig. 3.5 by placing spreading 1 μL on a slide and allowing it to dry. I imaged the beads in 10 fields of views totalling 114 localised bead pairs. Each field of view was first imaged for 488 nm

excitation, quickly followed by imaging with 561 nm. We registered the images with software written by Bo Jing our own Kapanidis Lab in MATLAB.

By overlaying both images it is possible to distinguish some deviations between the two colours (fig. 3.5). I measured both channels by localising each PSF and calculating the closest distance between them fig. 3.6. On average the aberrations amount to around 20 nm (fig. 3.6A) for both lateral dimensions x and y .

To calculate the aberration we find the pairs of beads that correspond to each other automatically. The program finds the closest bead in the opposite channel, in our case since deviations are small there are no false positives. To register an image we generate spatial mapping so that:

$$x_G = \mathbf{T}x_B \quad (3.1)$$

Where $\mathbf{T}$ is a projective transformation matrix where x_B are the points using excitation 488 nm excitation and x_G are the points with 561 nm.

By registering the images, the deviations were reduced to an average of close to 0 nm with most pairs showing a maximum deviations of 10 nm (fig. 3.6B). The distribution of analysed signals (fig. 3.6C) shows where in the field of view the deviations are the largest. For x the greatest deviations are on the left side of the camera chip, whilst for y it's the opposite, with the lower right being the lowest.

The chromatic aberrations from our microscope are non-negligible. In our work we don't make use of the registration for different colours. Localising FPs in live cells usually yields a localisation error of around 40 nm, the deviations

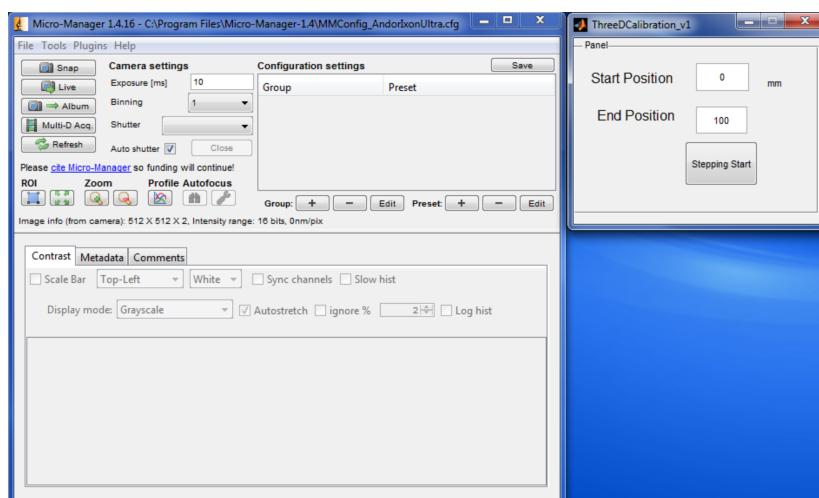


Figure 3.7: **GUI for 3-D calibration.** On the left is the Micro-manager software that interacts with the MATLAB code.

from chromatic aberrations could add a further 50 % when co-localising two probes. Calibrating and registering two-colour experiments is crucial for co-localisation in super-resolution microscopy.

3.3.2 3D calibration and software

I have explained the main concepts behind sub-diffraction localisation of FPs. Fitting the PSF with a Gaussian yields the lateral position of the protein, which is parallel to the slide surface. Strategies to extract the axial position are based on modifying the PSF shape or its focus. Without modifying the PSF, any molecules that appear over the focus or under the focus look like two very similar PSFs, in this case if the focus is $z = 0$ we are unable to distinguish $+z$ and $-z$ (fig. 3.8). This is because the blurring of the PSF from defocusing in any direction is equal or very similar.

To solve this issue the emission path needs to be modified to break the asymmetry. In the time when the first single-dyes were being imaged and tracked (Saxton and Jacobson, 1997), Kao and Verkman (1994) placed a weak

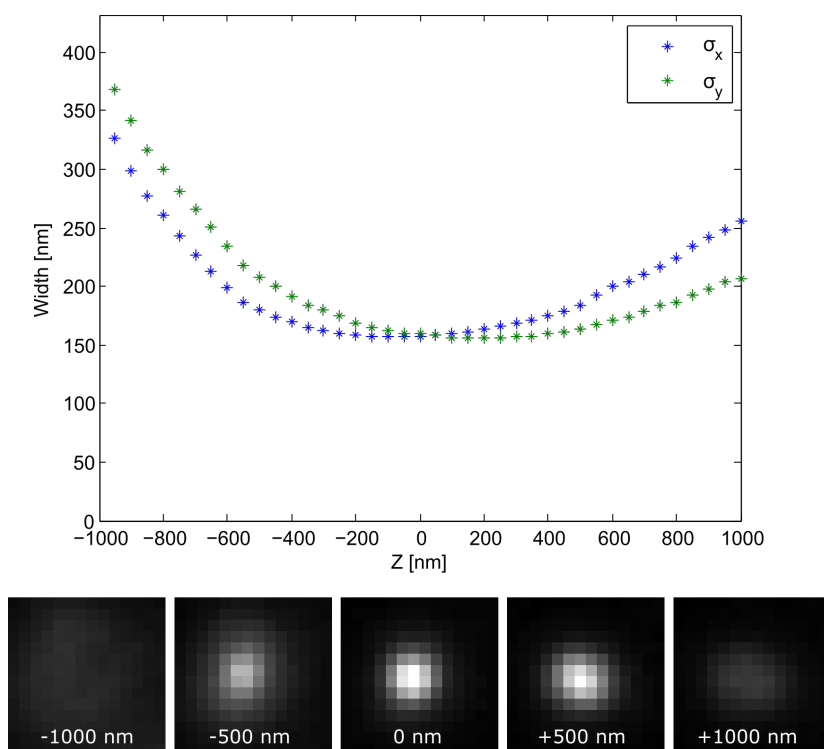


Figure 3.8: **Calibration curve for astigmatic 3-D localisation.** A slide of 100 nm Tetraspeck beads were imaged at different focus positions, stepping 50 nm at a time. The PSFs were fit with elliptical 2-D Gaussians and their standard deviations σ_x and σ_y calculated. The PSF of one bead is shown at, above and below the focus at 1000 nm, 500 nm and 0 nm.

cylindrical lens in the emission path. This technique introduces astigmatism into the microscope's optics (Huang et al., 2008). By inserting a cylindrical lens into the emission path, the PSF image is deformed asymmetrically. This way, by calibrating the system, it is possible to position the molecule in the axial dimension.

One technique, named bi-plane detection, splits the emission path using a mirror that transmits half of the light and reflects the other half (Prabhat et al., 2004; Mlodzianoski et al., 2009). One half of the light is imaged on one side (left) of the camera chip and the other half on the other side (right; two cameras may also be used). Light in these two paths will travel different distances and

thus will be focused at different points. When one image of the PSF is focused on the left side of the camera the other will be out of focus, whilst when the image of the PSF on the right side is focused the one on the left. This breaks the symmetry of the PSF image allowing us to position the FP.

Astigmatism is very limited since the deformation of the image leads to worsening lateral resolution. Bi-plane detection can localise molecules at around the same lateral resolution throughout a range twice the size of astigmatic localisation. Further because of the larger axial range, bi-plane can also image further into samples at around $2\mu\text{m}$ axial range (Mlodzianoski et al., 2009).

Even with these advantages, there are reasons led us to choose astigmatism: the cost of bi-plane, flexibility and unneeded advantage of axial imaging. Our purpose was to increase throughput of imaged bacteria, a using bi-plane detection we would have to split our image into two channels on the camera. There is the option of using two cameras, but each one represents about a third of the cost of the microscope parts. If we chose to have only one camera, the the full chip would not be used to image one field of view. If any user would not need to use 3D imaging, it would be very difficult to remove it to use the full field of view. Its limited flexibility is another reason not to use bi-plane, with astigmatism we only need to remove the cylindrical lens from the path, to have a full field of view 2D imaging. Finally for bacterial work, we do not need to image large axial ranges since bacteria have around $\sim 1\mu\text{m}$.

Even then, the design of the microscope does not allow us much flexibility to add a cylindrical lens ($f = 1000\text{ mm}$) to achieve astigmatic 3-D super-resolution (Huang et al., 2008). To test the possibility of the microscope to do 3D super-resolution. I prepared a sample of beads 100nm multifluorescent (Tetraspeck)

in the same way as was described before.

To create the calibration curve I wrote a controller for Matlab to calibrate for 3D Super-Resolution. The controller sends commands to create image slices at every 50 nm. At every slice the controller sends commands to micro-manager to take any amount of images. We imaged the beads at an exposure of 100 ms. The PSFs show asymmetry, which is plotted with the σ_x and σ_y for each fitted Gaussian. In this trial case, the astigmatism is not large enough for optimised 3-D localisation.

In this thesis 3-D localisation was not used or considered in any simulations so further efforts to work with other cylindrical lenses with lower focal lengths were not attempted.

3.4 OTHER TIRF MICROSCOPE USED IN THIS THESIS

3.4.1 Imaging

Some experiments were imaged in an older TIRF microscope, which was constructed originally for three-colour FRET. The setup had three separate channels imaged on the EMCCD thus the field of view was limited. Using the same objective as our constructed setup, its other main difference is a 250mm tube lens for a final 138x magnification (115 nm/pixel). With our newly constructed setup's higher magnification and no split channels, we achieve around 3 times the throughput than before because of larger fields of view and a camera capable of imaging a full frame at 4 Hz. Because some experiments in this thesis were done in the 138x magnification setup I quickly detail other technical aspects of the microscope.

The bespoke TIRF microscope was used to image PAmCherry fusions with 0-1 W cm⁻² 405 nm photoactivation laser (MLL-III-405, 100 mW, CNI), 475nm (Vortran 70mW Stradus) and 400 W cm⁻² 561 nm excitation laser (SLIM-561, 200 mW, Oxxius). The 405 nm laser was controlled between 0 and 10 μ W to limit the photoactivation rate; 475 nm at a constant 1 mW ; the 561 nm laser was a constant of 3.5 mW. A dichroic mirror (ZT405/473/561rpc, Chroma) and a notch filter (ZET405/473/561NF, Chroma) were used to filter the emission. The objective is a 100x oil-immersion microscope (Olympus UPLANSAPO 100x) and the camera EMCCD (Andor, iXon 897, 512x512 pixels) with 15.26 ms/frame for the tracking PALM experiment and 100 ms/frame for the FROS foci imaging.

3.5 TRACKING PALM

Using photo-activatable proteins, we are able to localise and track a single protein with up to $\sim$ 40 nm spatial precision Uphoff et al. (2013). After starting with a pool of inactive LacI-PAmCherry (fig. 3.9A), photoactivation using a 405 nm laser and readout using a 561-nm laser allowed us to observe and track a single LacI protein until it bleaches (fig. 3.9B), with every frame (15 ms exposure) yielding a localisation (fig. 3.9C). This series of steps was repeated stochastically until the entire pool of PAmCherry was tracked (fig. 3.9D). To implement tracking PALM, the density of observable LacI-PAmCherry in the cell must be low, preferably a single photoactivated protein per cell; this is achieved by manually regulating the power of the 405-nm photoactivation laser to control the photo-activation rate in order to meet the required density.

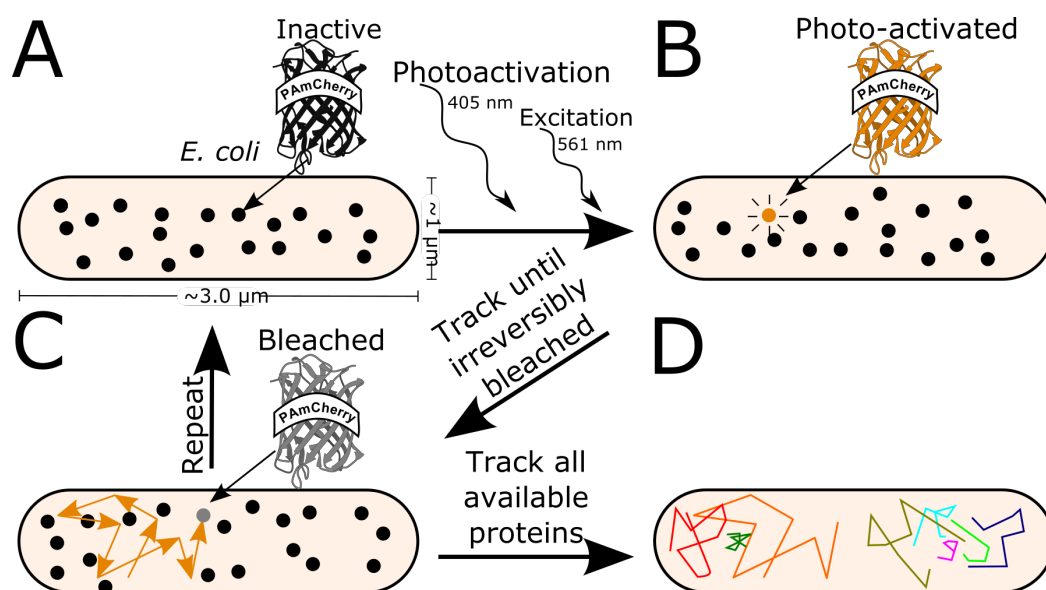


Figure 3.9: **Single-molecule tracking of LacI-PAmCherry.** We use tracking PALM to observe single LacIs by photo-activating, exciting and bleaching PAmCherry with 15 ms exposure. Shown is a schematic of the technique through different steps: A) your protein of interest (LacI) is tagged with PAmCherry and is initially photo-inactive. Going from A) to B), we control the rate of photo-activation by varying the power of the photo-activation laser line (405nm) and excite the photoactive PAmCherry with 561nm laser. From step B) to C), we track it until it's bleached and repeat this process until all available PAmCherry's are bleached step D).

3.5.1 Counting molecules

Because each molecule is represented by a track, this technique can be used to count the number of monomers that have been labelled the fluorescent protein.

3.5.2 Measuring the diffusion coefficient

In a thought experiment of a tracking PALM experiment, we have a track that arose from an *E. coli* cell in fig. 3.9. In that experiment a single-PAFP is photo-activated and localised five times (labelled 1-5, fig. 3.10A). Each one of these localisations yields displacements at different time lags. The displacement be-

tween localisation 1 and 2 happened in a time lag $1 * \Delta t$, the displacement between localisation 1 and 3 happened in a time lag $2 * \Delta t$ and so on.

Each of these displacements is used to calculate the mean-squared displacement (MSD) of the track at different time lags. As you can see each time lag contains the amount of data points generated from the short track shown in fig. 3.10B. If the molecule is not confined or bound it will have Brownian diffusion which in this case would appear as linear in the MSD curve. For a molecule that is confined or bound, this will not be the case (fig. 3.10C)

For our tracking PALM work we use the first time lag (fig. 3.10D) to calculate an apparent diffusion coefficient (D^*), which has the least error on the mean because it uses all the points of the track. The estimation of this apparent diffusion can be estimated by a linear fit to the MSD vs lag time or by using the first time point and using the point $MSD = 0$ at time = 0 to acquire a regression line. Then for each track we display the D^* in a histogram its contribution for the whole population of tracks and cells (fig. 3.10E-F).

To turn an apparent diffusion D^* to a real D a number of factors need to be considered: localisation error from the fitting algorithm, localisation error from the frame exposure and the non-Brownian behaviour from the confinement to the volume of the cell. Localisation error from the fitting and the frame time cause an offset from D^* . Localisation error from frame time is directly proportional to the magnitude of D . The relationship between MSD, D and its related errors is given by (Michalet, 2010):

$$MSD = 4D\Delta t + \frac{4}{3}D\Delta t - 4\sigma^2 \quad (3.2)$$

Where D is the diffusion coefficient, Δt is the camera exposure and σ the

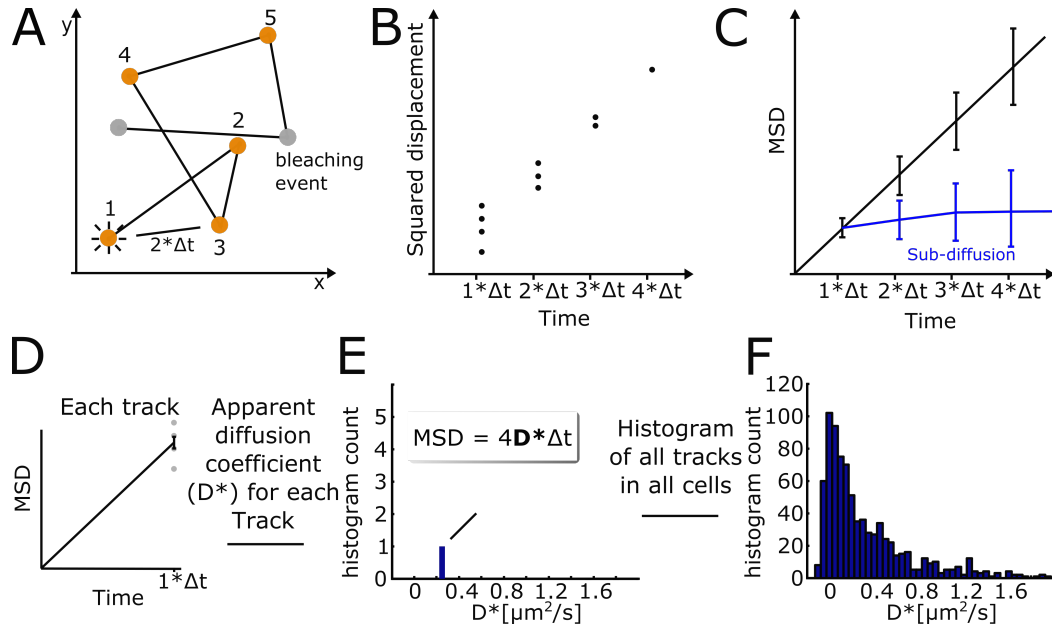


Figure 3.10: **Analysing tracking PALM data.** A single-PAFP is photo-activated and localised five times (labelled 1-5). In B) the mean-squared displacement (MSD) between time lags is plotted from the track shown in A). In C) the MSD is shown with bars representing the standard error on the mean. The blue line represents a molecule that has sub-diffusion such as a bound or confined molecule. In D-F) we show the general technique we use to analyse and visualise tracking PALM data. We calculate an apparent diffusion coefficient (D^*) for each track from the first time lag of the MSD. Then we calculate the probability function for all tracks and cells. Because the estimate of D^* is increased by localisation error eq. (3.2), in some cases localisation error contribution can be subtracted from the D^* histograms leading to negative D^* as in F).

localisation error. Each one of this factors may lead to $D^* > D$ as in the case where $D=0$, $D^* = 4\sigma^2$ localisation error; $D^* < D$ when D is large D^* will be smaller because the molecule is confined.

$$\text{If: } D^* = \frac{MSD}{4\Delta t} \quad (3.3)$$

$$\text{Then: } D = \frac{D^* 4\Delta t}{4\Delta t + \frac{4}{3} - 4\sigma^2} \quad (3.4)$$

This means that our calculated D^* is related to D by a scaling factor. It would be simple to calculate the D just by using this scaling factor, but there is no accounting for confinement. Currently there is no analytical equation to account for confinement, what we do know is that it is non-linearly dependent on the molecules' real D . To account for confinement we use simulations as described in the chapter 6.

Even if confinement is taken into account, this is still an incomplete picture because DNA-binding proteins are further confined to the nucleoid and can during the exposure of one frame, bind DNA multiple times. We try to address and understand the D^* distributions arising from these cases by simulations chapter 6.

3.5.3 Data visualisation and Analysis

Considering all the factors that I've presented paints a complicated picture of the real D . But working with the D^* has proved useful to deduce function from a single-protein. In a previous paper we show that using D^* we could successfully classify tracks as bound and mobile (Uphoff et al., 2013). In fig. 3.11 I briefly show one method to distinguish the bound and mobile tracks. Bound tracks will tend to have short displacements whilst mobile tracks will be larger on average (fig. 3.11A). If we prepare a slide of live cells and a slide of fixed cells we can see the differences in mobility (fig. 3.11B) by visualisation the classified tracks by colour overlayed with the brightfield image of the cells. Live cells have proteins that are more mobile and as we showed (Uphoff et al., 2013), bind DNA to repair it.

The diffusion distribution can be visualised as described before fig. 3.10. Each track generates a D^* value and we aggregate all the tracks informa-

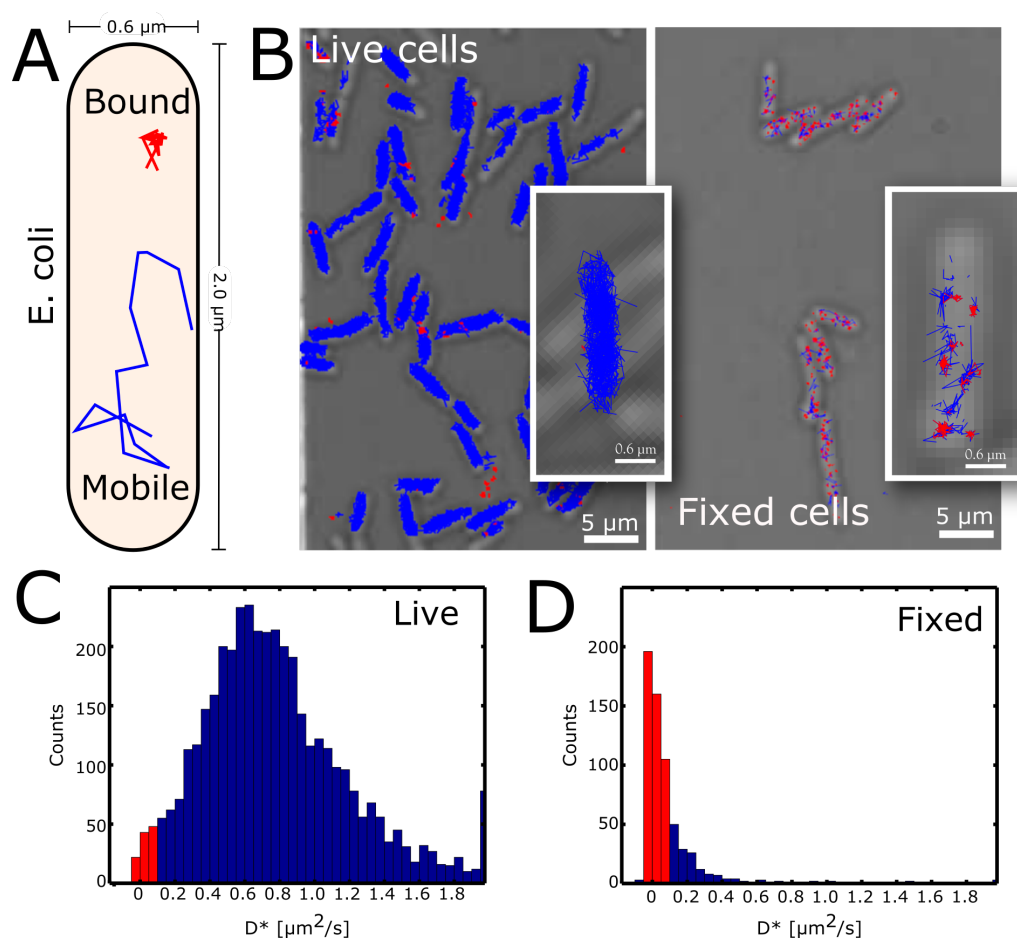


Figure 3.11: Mobility of Pol1 in live and fixed cells. Using tracking PALM we are also able to classify molecules with their diffusive properties (Uphoff et al., 2013). In A) two examples tracks are shown: one which is bound (red) and one that is mobile (mobile). In B), depending on the protein the ratio between bound and mobile will be different. Pol1 is mostly mobile and binds DNA to repair damaged DNA. If we fix cells with para-formaldehyde most tracks will be bound from cross-linking in the cytoplasm of the cell. This effect will show in the D^* histograms with live cells having a higher mobility C) than the fixed cells D).

tion and see their distribution. Live cells show high mobility (fig. 3.11C) with broad distribution and high mean of D^* , whilst understandably so, fixed cells are bound (fig. 3.11D).

3.5.3.1 Blinking

An inherent part of fluorophores is their photoblinking activity (Annibale et al., 2011). This can be used as an advantage because it reduces the amount of fluorescent probes emitting at any time. But when photo-activation controls the density of emitting PAFPs and we are using tracking PALM to extract diffusion information, it can be very damaging for the experiment analysis. In chapter 4 I address these issues and analyse our own experiment's blinking characteristics.

3.5.3.2 Simulations

Using simulations can be helpful to understand the distribution of diffusion coefficients. In the simulation we include exposure time effects and localisation error but additionally the size of the cell can be included to successfully simulate the distribution of D^* from a simulated D . Using the data of short tracks complicates the estimation of the D molecule.

CHARACTERISING FLUORESCENT PROTEIN BLINKING

4.1 INTRODUCTION

As we have discussed before single-molecule techniques allow the observation of heterogeneous activity within a protein population. Super-resolution techniques have allowed light microscopy to elucidate structure and function of proteins *in vivo* (Stracy et al., 2014; Liu et al., 2015). In addition, PALM can be used to study protein oligomerisation and stoichiometry in live bacteria (Stracy et al., 2014). Counting proteins can reveal protein complexes or larger stoichiometry that are otherwise unknown to occur in live cells or have only been determined *in vitro*. In the future PALM may become the golden standard for counting molecules (Coffman and Wu, 2014) but one of the main issues preventing this is fluorescent protein blinking.

Blinking is at the heart of super-resolution localization microscopy tech-

niques. It is as much a tool, like STORM where blinking reduces the temporal density of emitting probes, as an issue for quantitative interpretation of data. Disappearance of the probe due to blinking can split up a track and cause over-counting. But characterising and accounting for any blinking may help tracking PALM's ability to count molecules correctly.

Previously, to count molecules with fluorescence microscopes two methods (Coffman and Wu, 2014) have been implemented: counting photobleaching steps and comparing fluorescence intensity to a measured basal fluorescence for a single molecule. For example counting through photobleaching has elucidated the stoichiometry of components of the replication fork (Reyes-Lamothe et al., 2010) and of the bacterial flagellar motor (Leake et al., 2006). Through fluorescence intensity counting has led to estimates of the number of electroporated molecules tagged to a fluorophore (Crawford et al., 2013), both of these works were with live *E. coli*. Tracking PALM differs from these techniques in that it reveals not only stoichiometry but also diffusion and spatial distribution of proteins *in vivo* (English et al., 2011; Uphoff et al., 2013; Stracy et al., 2015).

In the past, groups have been able to study single-proteins in live bacteria, albeit at low copy numbers (Elf et al., 2007; Reyes-Lamothe et al., 2010). By photo-activating a sparse quantity of fluorescent proteins at a time, super-resolution techniques have enabled studying large copy number proteins both in live and fixed bacteria (Betzig et al., 2006). We can observe and track single-proteins using a combination of localization microscopy and single-particle tracking (tracking PALM) (Manley et al., 2008; Stracy et al., 2015; Uphoff et al., 2013; English et al., 2011).

By quantifying the fluorescent protein copy number we are able to make a direct measurement of the protein of interest in a single cell; but issues like

blinking and high photo-activation density, when too many PAFPs are activated at the same time, make it difficult to obtain reliable calculations without corrections. In tracking PALM we are able to distinguish a single-molecule by reducing the rate of photo-activation and new strategies to correct have been proposed (Lee et al., 2012).

In our work we use PAmCherry (Subach et al., 2009); other similar photo-activatable fluorescent proteins have been known to blink (Annibale et al., 2011; Lee et al., 2012). Experimental correction of blinking is complicated for a number of reasons. PAFPs are already specialised, engineered proteins, changing them at the protein level is a difficult task. Also, tracking PALM necessitates high temporal and spatial resolution leading to high intensities of excitation light that increases potential of photo-induced blinking. A post-processed correction is the best approach for the moment for tracking PALM work.

Based on previous *in vitro* work (Lee et al., 2012), we developed a strategy where we could measure the blinking of PAmCherry directly *in vivo*. With this strategy, we can study the photo physics of PAFPs *in vivo*, which has not been fully characterised and be applied to other PAFPs. By understanding the rates involved in the blinking and bleaching, we can characterise and minimise any errors related to over-counting and under-counting.

4.2 BLINKING CHARACTERISATION

Blinking has been recognised as an issue to extract protein diffusion and counts *in vivo*. Many strategies have been proposed to characterise and solve blinking. Most tracking work implements a correction to connect tracks from the same molecule that are separated by blinking. By using an allowable number of

frames where a molecule can disappear, we can still be considered connected to the track (fig. 4.1A). We call this range the memory parameter, defined as the number of frames over which tracks can be separated from each other and still be connected as one track (Stracy et al., 2015; Uphoff et al., 2013; Subach et al., 2009; Annibale et al., 2011).

Pair-correlation analysis PALM, is an alternative method which distinguishes between localisations arising from one molecule by how the density of localisations varies as a function of distance, has also been applied to count blinking molecules but can only work with slow or static molecules (Sengupta et al., 2013).

We attempted to find a solution that would work with tracking PALM. To test our method we used RNAP because it stably binds to DNA, which minimises the effects of proteins disappearing by diffusion. With absence of blinking, a single track in tracking PALM work will represent a single-molecule (fig. 4.1A, top cartoon trace). Without any method of connecting tracks, a single blink will divide the molecule into 2 tracks (fig. 4.1A, middle cartoon trace with 1 blink). For a single track, we identify multiple characteristics of blinking, on-time (length of emission between photo-activation, blinking or bleaching), off-time (length of absence of emission between emissions), number of blinks per molecule and bleaching time (from photo-activation to bleaching, fig. 4.1A bottom cartoon trace). This strategy builds on previous *in vitro* work characterising mEos2 and Dendra2 Lee et al. (2012).

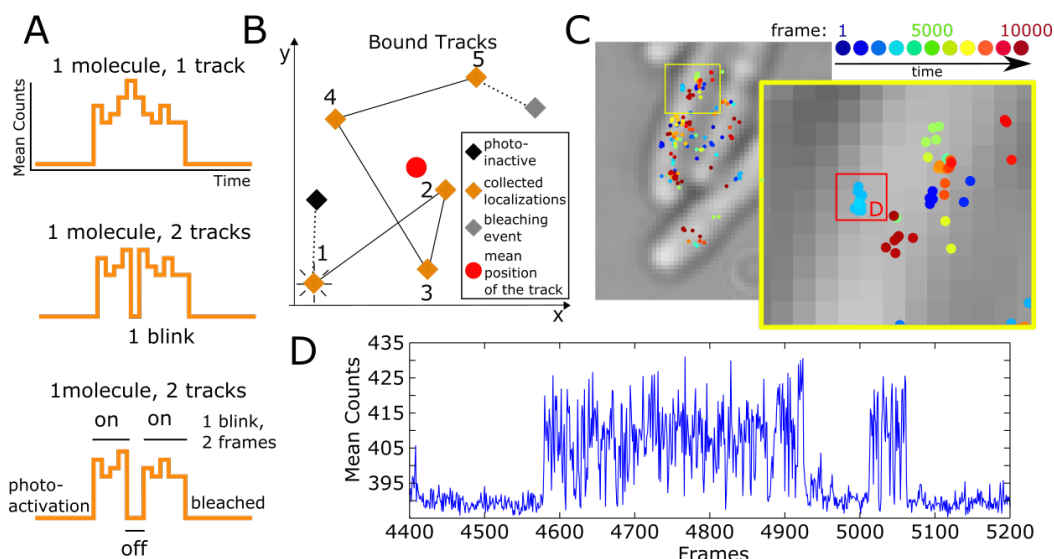


Figure 4.1: Identifying blinking species in *E. coli*. Blinking is a problem in *in vivo* single-molecule studies but essential for many super-resolution localization microscopy. A) Shows cartoon examples of a trace without a blink (top), a trace that blinks for a frame (middle) and one that may blink for more than a frame (bottom). In the last example the measured parameters are shown: photo-activation, on and off times, and bleaching. B) Shows a cartoon demonstration of the collected data: PAmCherry is photo-activated with 405 nm illumination. Then when it is excited with 561 nm, it emits light that is collected during a frame of the EMCCD. Each diamond figure represents the localisation result of a fitted PSF. The fluorescent protein is eventually bleached. To represent the position of immobile proteins the mean position is calculated from a track and represented by the red circle. C) Shows the mean positions of RNAP-PAmCherry in example cells. The zoomed in region shows how plotting these mean positions with temporal information, in this case the color, shows clustering that may be due to blinking. The red square denoted region in C) is a temporal cluster arising from the trace in D) calculated from aperture photometry.

4.3 METHOD AND MATERIALS

4.3.1 Identifying the presence of temporal clusters

To present further evidence of the problem, we classify our tracks as bound by using tracks with 5 localisations or more as described for our tracking PALM work (section 3.5). A single-track consists of an event of photo-activation, then subsequently the molecule is localised over a number of frames until it photo-bleaches (fig. 4.1B). To represent this track as one molecule, we use the average over all localisations (fig. 4.1B, red circle). If each mean position of the track is overlaid on the brightfield image showing the cell edges, we start to visualise the apparent clusters of molecules (fig. 4.1C). But if we represent each mean position by the colour of the frame where they first appeared in the movie (dark blue: early, green: middle and red: end), then we find that the localisations may not only be clustered in space but also in time (fig. 4.1C). If we analyse the fluorescence of a cluster by measuring the mean counts in each pixel on that section of 7x7 pixels (aperture photometry), we find clear transitions to dark states and eventually bleaching (fig. 4.1D). This could be due to low signal from the dimness of PAmCherry and blinking from the protein. In any case characterising the loss of signal and methods to correct for it is important to analyse tracking PALM experiments.

4.3.2 Rate extraction

To extract the photo-physics of PAmCherry, we apply tracking PALM to RNAP and make use of only the bound tracks; this emulates fluorescent proteins bound to a slide surface as in *in vitro* experiments (Lee et al., 2012). I chose RNAP (Stracy et al., 2015) to test the correction method because during tran-

scription the protein has a stable binding time of around ~ 2 s and has high copy numbers. When grown in rich medium up to 48 % of all RNAP is observed bound through tracking PALM (Stracy et al., 2015), with a total of around 3000 molecules of RNAP per cell.

For this work, *E. coli* cells were grown and imaged on a TIRF microscope (described section 3.4). For this experiment we used an endogenous fusion of PAmCherry with the β' subunit of RNAP (as in Stracy et al. (2015)). The cells were grown at 37 °C over-night in EZ Rich Defined Medium (Teknova) supplemented with 0.2 % glucose. Cells were then diluted and grown at 37 °C for around 2 hours until they reached an OD of ~ 0.2 . For the experiment the cells were immobilized for imaging on 1 % agarose pads.

In previous tracking PALM work (Stracy et al., 2015; Uphoff et al., 2013) we have used a memory parameter of 15 ms (one frame), to allow the molecule to disappear for 15 ms between tracks. In this work we are looking to extract dark states that are longer than one frame. To achieve this I track molecules with a memory parameter of 150 ms (10 frames) and a distance parameter of 5 pixels ($0.52 \mu\text{m}$). The number of frames was chosen empirically so that it would account for the majority of the τ_{on}, τ_{off} counts and while also minimising erroneously connecting separate molecules.

It is difficult to estimate how many molecules per second we are imaging at any time, specially if blinking will over-estimate the calculation. With no memory parameter we observe on average around 5 molecules/s (total localised molecules per cell/total imaging time) which represents one molecule per 200 ms. In practice the density is variable and we have bursts of photo-activation (fig. 4.2). It is possible that with a memory parameter of 150 ms (10 frames) we may be connecting separate molecules together. For further

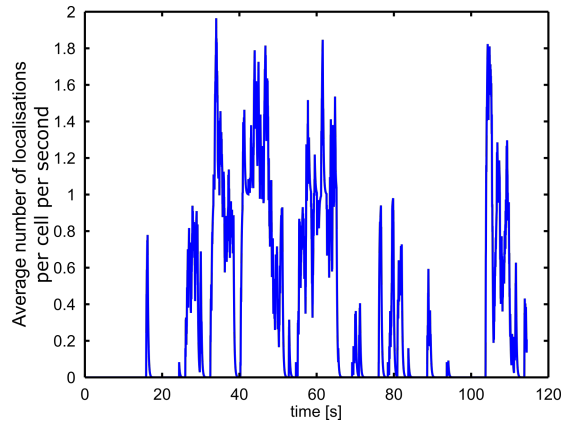


Figure 4.2: **Photo-activation density in one cell.** In each frame the number of localisations in a single-cell was measured and plotted in vertical axis. A moving average window of 180 ms (12 frames) was used to smooth the data.

discussion on the memory parameter choice, the effects on this work and the difference with *in vitro* characterisation (Lee et al., 2012) see section 4.6.

In the same way as described before we classify the tracks and only use the molecules that are bound after 4 displacements or more. From an experiment we extract the tracks of bound molecules with where and what times the molecules appeared, allowing us to assess the blinking. From this data set I extract the distributions of the number of blinks, the on, off and bleaching times (as described in section 4.4 and shown in (as shown in fig. 4.1A)) that will allow us to find the unknowns following previously published model (fig. 4.3A inset, (Lee et al., 2012)).

4.4 BLINKING MODEL AND RATES

The proposed model states that once a PAFP is photo-activated, it may transition to a non-bleached dark state. This is a simplified empirical model based on previous published work (Lee et al., 2012). It transitions from this active state to a dark state with a rate of k_d . The PAFP may recover from this state and will

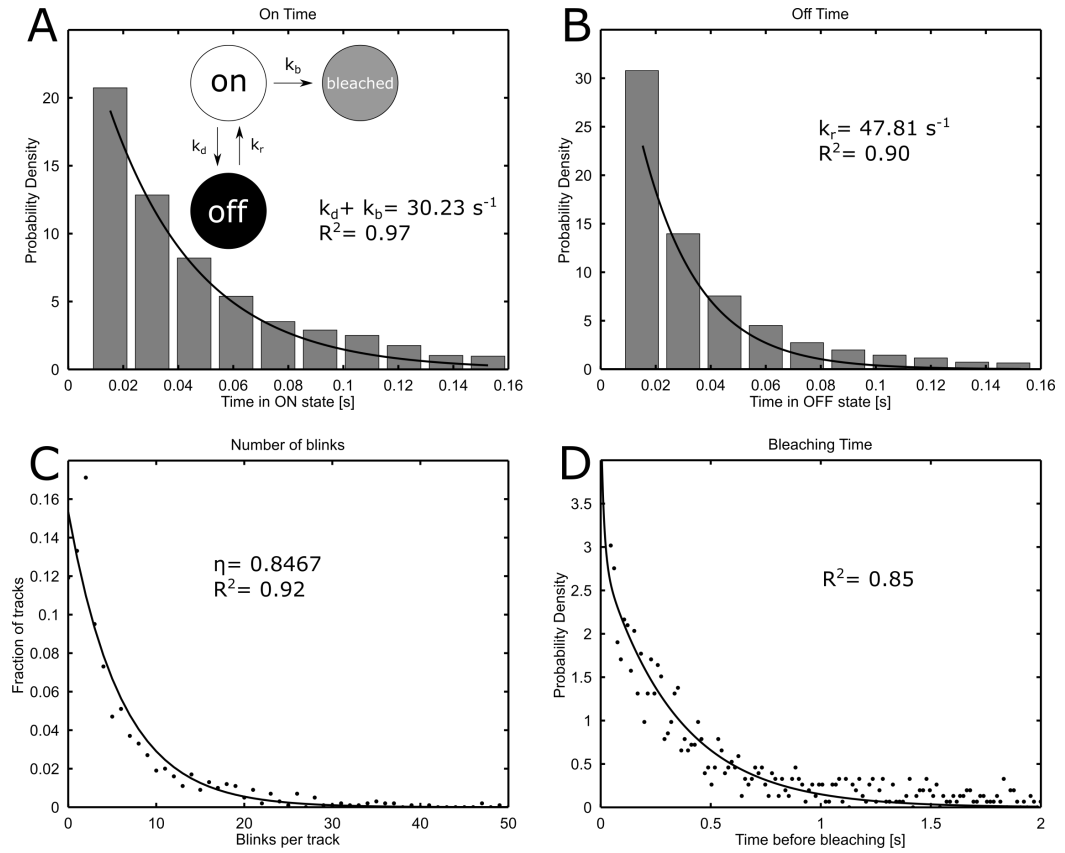


Figure 4.3: **PAmCherry's behaviour in live cells.** Data extracting from tracking PALM of RNAP-PAmCherry. A) shows the distribution of times spent in the active state for the RNAP-PAmCherry. A single-exponential is fit (black line) to predict the distribution and used to extract the combined rate of transitioning to a dark state and bleaching $k_d + k_b$. B) Shows the time spent in the dark state with a single-exponential fit to extract the rate of recovery k_r . C) Shows the number of blinks per track with a binomial distribution fit to the data to extract η probability of transitioning to a dark state and not bleaching. D) Shows the bleaching rate of PAmCherry and the predicted model from eq. (4.3) which describes the curve well.

do so at a rate of k_r . Finally at any time it is in the active state the PAFP can bleach and remain in that unrecoverable state at a rate of k_b (fig. 4.3A inset). The main photo-physical characteristics of PAmCherry that are needed to describe the previous model proposed (Lee et al., 2012) are: on time (time spent active), off time (time spent dark), number of blinks per track and bleaching time (figs. 4.1 and 4.3).

The extracted on times and off times distributions show a decaying probability as longer time-scales are accounted (fig. 4.3A-B, bar plots). In our case the on and off times can be described by a single exponential ($PDF = ke^{-kt}$, fig. 4.3A-B, black line) which was fitted by non-linear least squares method (MATLAB *lsqnonlin*). Both show a good fit (using $R^2 > 0.9$ metric) although both show an underestimation bias to the data points. A double exponential does not significantly improve the fits but we may be limited by how we extract the data (see section section 4.6).

Showing the single-exponential fit to be an good model of the data, we extract the rates at which PAmCherry recovers from a dark state (k_r) from the off time probability distribution function (PDF) and the total rate at which emission bleaches or transitions to a dark state ($k_b + k_d$, respectively) from the on time PDF. To finalise the calculation we extract the k_b from the number of blinks calculated. The distribution of the number of blinks is given by a geometric distribution of the probability to go from an active state to the dark state given by (Lee et al., 2012):

$$P(N_{blink} = n) = \eta^n (1 - \eta) \quad (4.1)$$

where:

$$\eta = \frac{k_d}{k_d + k_b} \quad (4.2)$$

In this model, η represents the probability of the transition from the active state to the dark state instead of bleaching. The geometric distribution represents the probability of n blinks with probability of η before the first bleaching

PAFP	k_d	k_r	k_b
<i>PAmCherry</i>	25.60 s^{-1}	47.81 s^{-1}	4.63 s^{-1}

Table 4.1: Measured rates for PAmCherry in live *E. coli*

occurs. By fitting a geometric distribution equation we can obtain the value for η and calculate the values for all three unknowns.

Using the model and the three calculated rates (k_d , k_r , k_b) would be able to predict the PDF of the photo-bleaching time. From *Lee et al.* The following equation should yield a good prediction of the photo-bleaching time without any additional values to fit (Lee et al., 2012):

$$\frac{P(t \leq T_{bleach} \leq t + dt)}{dt} = \frac{1}{1 + \alpha_r} \{k_1 e^{-k_1 t} + \alpha_r k_2 e^{-k_2 t}\} \quad (4.3)$$

This represents the probability distribution function that a molecule will bleach at a time T_{bleach} with any amount of blinks. Where in eq. (4.3):

$$k_1 = \frac{1}{2} \{(k_d + k_b + k_r) - \sqrt{(k_d + k_b + k_r) - 4k_b k_r}\} \quad (4.4a)$$

$$k_2 = \frac{1}{2} \{(k_d + k_b + k_r) + \sqrt{(k_d + k_b + k_r) - 4k_b k_r}\} \quad (4.4b)$$

$$\alpha_r = \frac{k_1(k_2 - k_1)}{k_2(k_r - k_1)} \quad (4.4c)$$

We input the known rates into eq. (4.3) and compare with our extracted data (fig. 4.3D, points: data, plot:black line). With a good coefficient of determination of the fit to the model, ($R^2 = 0.85$) there is also little under- or over-representation bias in the fit. The complete model is shown in table 4.1.

4.5 MEMORY PARAMETER ESTIMATION

With our calculated rates it is now possible to estimate the optimum memory parameter for counting (Lee et al., 2012). In our previous published tracking PALM work we implemented memory parameter of 1 frame (Uphoff et al., 2013; Stracy et al., 2015), a single-molecule may disappear for one frame and still be connected to the same track. The density of molecules at the start of the experiment is much higher than at the end when the pool of unactive PAmCherry is much smaller. Our method does not use constant photo-activation rates because we change our photo-activation laser power throughout the experiment (Uphoff et al., 2013) it is unsure if varying photo-activation intensity affects blinking (see section 4.6ref). Thus a different approach to find the optimal memory parameter Lee et al. (2012) work is required (Lee et al., 2012).

We use the extracted rates from section 4.4 to estimate what is the best balance between over-counting and under-counting. From the calculated rates we can obtain the probabilities of remaining in a state and transitioning to another state during the exposure time ΔT . I recreated our experimental tracks by simulating a Hidden Markov Model (HMM; MATLAB, hmmgenerate; based on model from fig. 4.3A). HMM is ideal to simulate blinking tracks because the transitions only depend on the current state, that is it is memoryless. To simulate using HMM, I used a transition probability matrix T_{ij} given by the following:

$$T = \begin{vmatrix} e^{-k_d \Delta T} + e^{-k_b \Delta T} - 1 & 1 - e^{-k_d \Delta T} & 1 - e^{-k_b \Delta T} \\ 1 - e^{-k_r \Delta T} & e^{-k_r \Delta T} & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

The transition matrix T_{ij} represents the probability of transition from a state

i to the state j . The emission probability matrix E_{kl} is given by:

$$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

E_{kl} is the probability that state k emits the symbol l . In our case there are 3 states (fig. 4.3A, inset): active (state 1); dark (state 2) and bleached state(3). In this emission matrix state 1 emits the property of a localised protein (logical 1; fig. 4.5A, trace); state 2 and 3 emit the same property: no localisation (logical 0; fig. 4.5A, trace).

I simulate a number of molecules (N_{mol} ; fig. 4.5A) each in a separate trace with 250 frames (every molecule will have bleached by then); I randomly overlay them on the same array with a larger number of frames ($N_f = 5000$, fig. 4.4, fig. 4.5A). The resulting array consists of a long binary trace of logical 1s and 0s with 5000 frames; 1 for a frame containing one or more localisations and 0 without (fig. 4.5A). Each array represents a cell and each overlaid HMM trace a photo-activated PAmCherry. With this array we can count molecules for different memory parameter values and different density of photo-activation; we asked if we could count molecules from the binary trace and if the number extracted is correct. Comparing the measured values with true amount allows a measure of the balance of the memory parameter for optimal counting.

For each simulated density of molecules (N_{mol}/N_f) per binary trace a memory parameter is used to connect it (fig. 4.5A); I also extract the calculated number of molecules (N_{count} ; fig. 4.5A). We can have a measure of the error by subtracting the number of simulated molecules with the number extracted after correction with a memory parameter ($(N_{count} - N_{mol})/N_{mol}$ in percent). If this

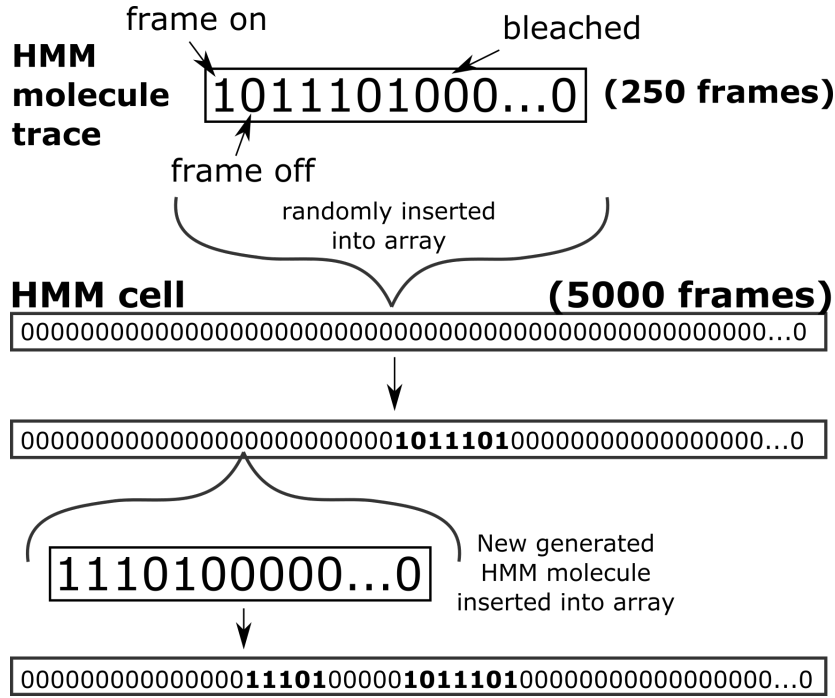


Figure 4.4: **Simulating blinking with HMM.** Each generated short binary trace of logical 1s and 0s represent the track of a molecule. Our HMM molecule contains 250 frames where it photo-activates, blinks, recovers and eventually bleaches. This simulated molecule is randomly inserted, representing random photo-activation, to a larger array of 5000 frames, which represents a single-cell. The process continues by simulating another HMM molecule and randomly inserted into the HMM cell until this is repeated N_{mol} , the number of simulated molecules. The HMM cell is then connected with a selected memory parameter and produces N_{count} , the counted number of molecules after connecting the binary trace.

error is positive then over-counting, from the inability to connect all the tracks successfully, is more prevalent; if it is negative then under-counting, from connecting two separate molecules together, is more prevalent. The density of the binary trace (how many molecules per frames) is what determines if a certain memory parameter will over-count or under-count the number of molecules.

For each density of molecules per binary trace we find the optimal memory parameter, i. e. the one that minimizes the error (fig. 4.5B). For the rates calculated from our scheme we find that the optimal choice for any density is

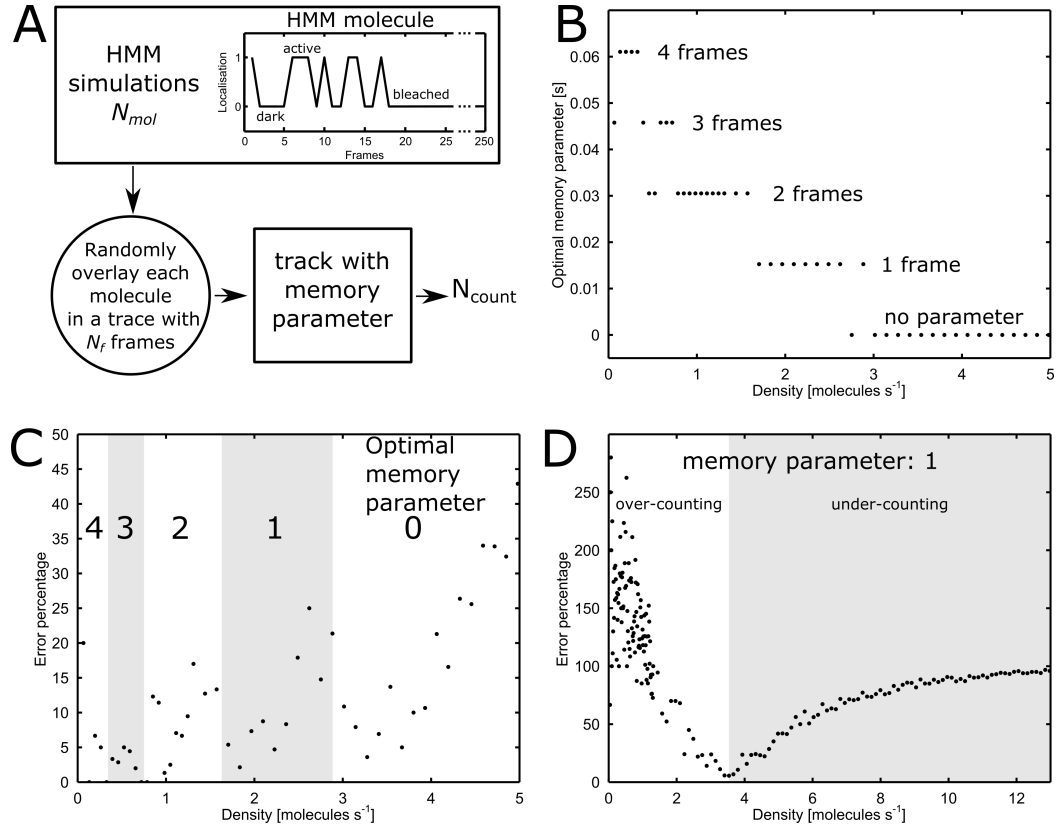


Figure 4.5: Optimal memory parameter and effect with varying photo-activation density of molecules. A) Each HMM trace containing N_{mol} molecules is tracked and the numbers of tracks counted N_{count} . The density of the trace is given by the number of molecules over the number of frames N_{mol}/N_f . In B) different measures of density were simulated with different memory parameters. For each density the optimal memory parameter was the one with the least deviation from the real number of molecules simulated. In C) the error percentage of each track in a given density with its optimal memory parameter are shown (inset text: best memory parameter is highlighted). In this case the error is the difference from the simulated number $abs(N_{mol} - N_{count})/N_{mol}$. In D) memory parameter of one frame (15.42 ms, used throughout the tracking PALM work in this thesis) was simulated for different densities. The Error was calculated and denotes the optimal density for the memory parameter.)

between 0 ms to 60 ms. These memory parameters achieve less than $\pm 20\%$ deviations from the correct number (fig. 4.5C).

In our previous work we use a memory parameter of 1 frame (15 ms, I tested what the error for this memory parameter would cause by varying densities (fig. 4.5D). At low densities (0–1 molecules/s) over-counting is very prevalent leading to errors of between 100–300 %. For a frame time of 15 ms in a movie with 5000 frames, these densities represent ≤ 80 molecules per movie for a single-cell.

We observe that the smallest error occurs at 3.5 molecules/s, it also serves as the transition between over-counting and under-counting. Higher densities than these lead to high under-counting eventually reaching the point where all molecules are connected into one track.

4.6 DISCUSSION

Our results show that, for our experiments, PAmCherry blinks on average 4 times more on average than mEos2 and 27 times more than Dendra2 when compared to the *in vitro* work (Lee et al., 2012). PAmCherry rates of transitioning to a dark state are 3 and 8 times higher than mEos2 and Dendra2, two other PAFPs used for counting or tracking PALM, but in contrast the rate of recovery from the dark state is 3 times faster in PAmCherry. PAmCherry shows the slowest rate of bleaching of the three, being 1.2x slower than mEos2 and 3.5x slower than Dendra2. It is unsure if these differences in photo-physics arise from our work being done in live cells instead of *in vitro*; Lee et al. (2012) apply their technique in an *in vivo* case as proof of principle but don't show differences in the rates between them.

We also consider that our experiment is optimised for tracking molecules and not counting. Because of cellular background we may also have worse signal-to-noise than the *in vitro* work; this may affect tracks where dimness of PAmCherry may appear as a "blink" in a single-frame. In any case we are characterising our tracking PALM experiment and we think *in vivo* rates will be more relevant for future experiments.

Recognizing that the differences for *in vitro* experiments of mEos2 and Dendra2, with our *in vivo* work with PAmCherry, if we compare both rates my work shows that PAmCherry is the best for tracking PALM because it bleaches the slowest of the three. Although it blinks a larger amount of times, the recovery rates are fast and implementing a correction for over- and under-counting a memory parameter of under 60 ms corrects a large amount of the error.

A major difference in our recovery rates is that we fit a single-exponential to extract one rate k_r , while Lee et al. (2012) extract a second slower rate. We are unable to quantify such recovery times because of the difference in experimental approach, whilst in the *in vitro* work all molecules are spatially separated, in our case we are unable to guarantee this. Thus to connect tracks that remain in the dark state would probably connect different molecules into 1 track. In the *in vitro* work the second exponential has a mean of 2 s whilst in our method we only calculate τ_{off} data up-to 0.15 s.

Another work studied PAmCherry *in vivo* using the human glycine receptor as a nano-template in *Xenopus* oocytes (Durisic and Laparra-Cuervo, 2014). The receptor is used as a nanotemplate because it stabilises in the oocyte membrane for several minutes and has a known stoichiometry. In that work they reported that 15 % of PAmCherry fused to a subunit of the receptor blinked and that on average it blinked once. This differs from our work where we ob-

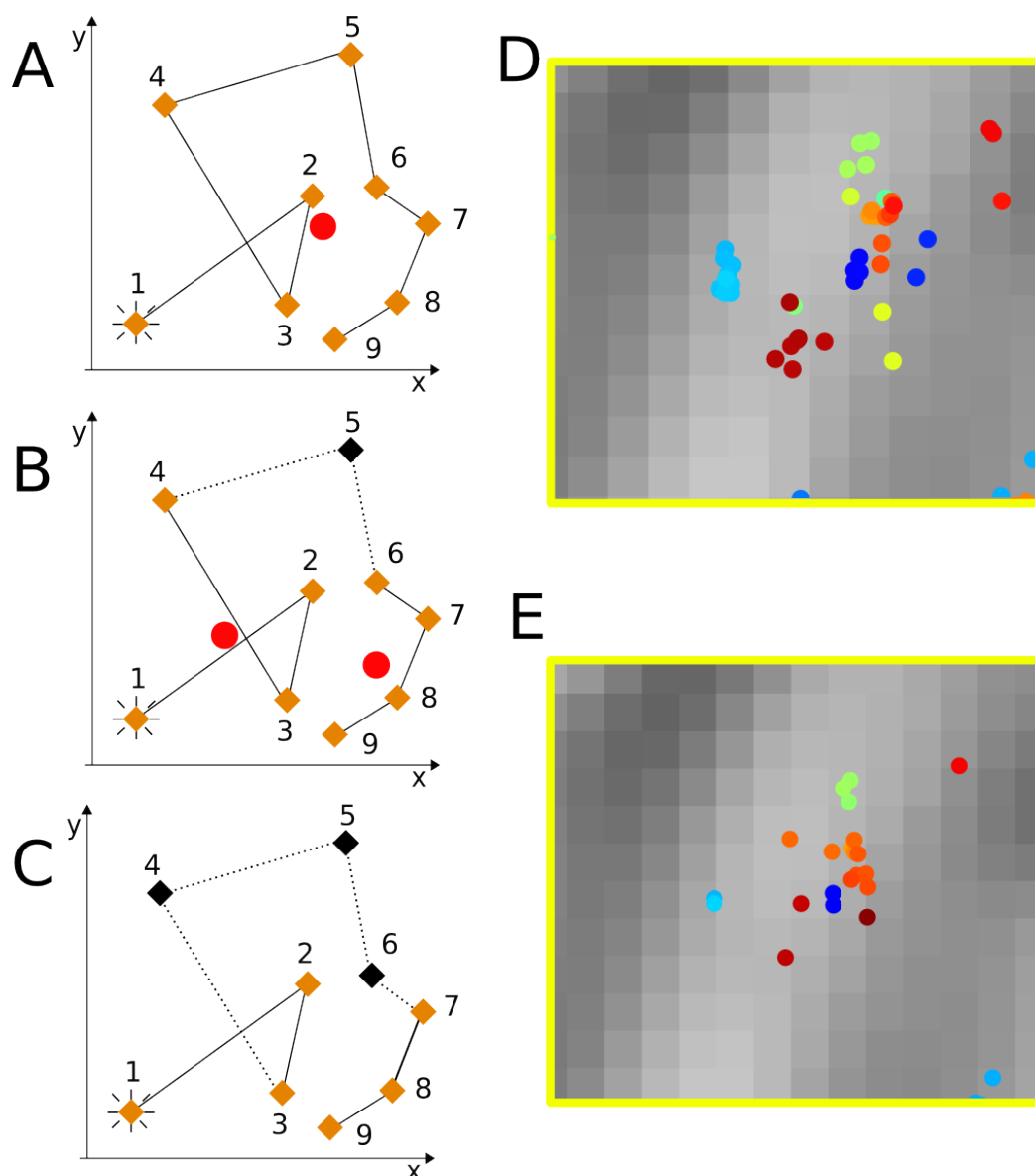


Figure 4.6: **The effects of corrections on bound molecules.** In A) there is one bound molecule corresponding to the track. In B) we see that for an identical track that has blinked for a frame, the number of bound molecules are doubled. If the blink is long enough and the on times are too small for the diffusive behaviour to be classified, there will be no bound positions. If these tracks are stitched together correctly they should all yield the same result within the localisation error. D) shows the results before connecting the tracks and E) shows after the corrections

serve PAmCherry blinking 5 times per track. This may be due to the differences in the experiment. Whilst the Durisic and Laparra-Cuervo (2014) work is optimised for counting PALM, imaging at 50 ms, we optimise our experiment for tracking PALM 15 ms. We may be seeing blinking that is averaged in a counting PALM experiment. Also the effect of photo-activation on the blinking rates needs to be further studied.

The technique described in this chapter is a limited approach for applications of PALM that track moving molecules. Balancing out under- and over-counting may work for counting proteins, but connecting two molecules diffusive properties will be detrimental for diffusion work. Spatial distribution of the mean positions of tracks will be affected too. If a track does not blink then the mean positions of bound molecules will be estimated as correctly as possible (fig. 4.6A). But if the track blinks then it may be split into two mean positions (fig. 4.6B), or failed to be counted completely for having too short tracks (fig. 4.6C). Even if the blinking cases would be connected with a memory parameter, the resulting mean position may deviate from the true position. In any case using a memory parameter to connect tracks has a dramatic effects on the number of mean positions (fig. 4.6E). Because of this it is recommended for spatial distribution to work with localisations and not mean positions.

In addition, this work may show some limitations. Recent work has shown that PAmCherry may dimerise and cause clustering artefacts (Wang et al., 2014), it is possible that this additional artefact may affect our approach. The best approach for this experiment would be to use the least dense photo-activation possible, this way higher memory parameters would not affect the rate equations. In addition a feedback loop to keep a constant rate of photo-activation would be necessary.

4.7 CONCLUSIONS

We show our efforts into using our current datasets to characterise the photo-physical behaviour of PAmCherry in a manner that can be replicated for other fluorescent proteins. I also discussed the effect on the mean tracks positions by correcting using a memory parameter.

The rates I extracted apply for our datasets and can help us understand the behaviour of PAFPs and how to use them in tracking PALM. It may not be the case that the rates we found can be generalised for other experiments. Each experiment design and PAFPs needs to be characterised and analysed, in this work we further build a framework to achieve this. We optimise our experiments depending on the required imaging speed and spatial resolution. For this reason the rates of photo-induced processes (activation, blinking, bleaching) will differ. Here we showed an example case and decided a general strategy that can be followed to optimise *in vivo* experiments.

The fact is that there is still not a clear strategy towards characterising photo-activatable proteins *in vivo* in *E. coli*, we believe this is an additional approach towards making tracking PALM the gold standard in protein counting and tracking in live cells.

CHAPTER 5

TRACKING SLOW AND LOW COPY NUMBER PROTEINS IN LIVE CELLS USING THE LAC REPRESSOR

In this chapter I image three TFs with tracking PALM, achieving sub-diffraction spatial and diffusion distribution. I also observe that there may be an inherent amount of background for a tracking PALM experiment. Because most TFs exist in less than 100 copy numbers per cell, the background may not be negligible. I characterise the background and extract positional and mobility information of the TFs.

5.1 INTRODUCTION: IMAGING THE *lac* REPRESSOR

TFs control the expression of genes by binding to specific sites in DNA and repressing or activating transcription in response to stimuli. The *lac* repressor (LacI) is a well characterised TF that regulates the ability of bacterial cells to uptake and metabolise lactose. In the absence of lactose, LacI specifically

binds a specific control region in DNA (the *lac* operator) and blocks transcription of the *lac* operon, thus repressing the genes for lactose utilisation. In the presence of lactose, LacI binds to allolactose (a metabolic intermediate of lactose) and shows markedly reduced affinity for its operator (Lewis, 2005).

Having a full understanding of gene regulation and its control elements can aid systems biology efforts to create models of larger systems. Furthermore, understanding control systems based on TFs will increase the knowledge on the available toolboxes for synthetic gene expression. TFs can also be a tool for imaging, as has been shown with the use of FROS to tag chromosomal locations (Lau et al., 2004; Robinett et al., 1996). Since LacI is one of the best characterised DNA-binding proteins and serves as an excellent model system, further research into its intracellular mobility, molecular interactions, and subcellular distribution of LacI should advance our understanding of many other DNA-binding proteins.

The mechanism by which LacI finds its target operator (21 bp) among the huge amount of non-specific chromosomal DNA (~4.5 Mbp for *E. coli*) has been studied extensively (Riggs et al., 1970; Wang et al., 1977; Redding and Greene, 2013; Elf et al., 2007; Wang et al., 2006, 2005). Initial biochemical work showed that LacI binds to its specific site at rates faster than what is expected from pure 3D diffusion (Riggs et al., 1970); such results led to the proposal of target location through facilitated diffusion, whereby DNA-binding proteins utilise a combination of 3D diffusion, non-specific DNA binding, and 1D sliding on DNA to locate their target swiftly. More recently, single-molecule studies, both *in vitro* and *in vivo*, have provided strong support for the facilitated diffusion model. Recently *in vitro* studies have shown that LacI can bind and slide on DNA with 1D diffusion constants mean of $0.021 \mu\text{m}^2 \text{s}^{-1}$ (Wang et al.,

2006, 2005). However, such techniques stretch out DNA from its native form, ignoring any effects of DNA deformation and bending on LacI sliding; further, sliding also depends on salt concentrations and buffer conditions used *in vitro*, which cannot replicate the environment encountered in living cells.

Single-molecule measurements in single living cells can address many of the limitations of *in vitro* work, since such measurements go beyond the ensemble averaging of cell populations, retain molecular resolution, and uncover any cell-to-cell heterogeneity (Mueller et al., 2013). Recent experiments have examined the behaviour of single fluorescently labeled LacI molecules inside live *E. coli* cells. Using strains with artificially low copy numbers (~ 7 monomers per *lacI* gene), it became possible to track single LacI molecules in live *E. coli* (Elf et al., 2007; Hammar et al., 2012, 2014), and determine the effective diffusion coefficient of non-specific binding $0.4 \mu\text{m}^2 \text{s}^{-1}$. These results, coupled with measurements of 1D diffusion coefficients *in vitro* $0.046 \mu\text{m}^2 \text{s}^{-1}$, led to the estimate that LacI spends $\sim 87\%$ of its time non-specifically bound to DNA (Elf et al., 2007), where it slides for an average of ~ 45 bp (Hammar et al., 2012). However, to date, no such work has been done on strains with the native LacI copy numbers (40 monomers, see Gilbert and Muller-Hill (1966)).

In addition to diffusion, the spatial distribution of LacI is also relevant to target location, since any LacI concentration gradients will influence the target search times, and the status of repression (Kuhlman and Cox, 2012; Pulkkinen and Metzler, 2013; Kuhlman and Cox, 2013). Previous work (Kuhlman and Cox, 2012) using fixed cells showed that the subcellular spatial distribution of LacI is non-uniform and it depends on growth rates and the position of the transcribing gene. In contrast, live cell work showed that LacI does not maintain high local concentrations (Hammar et al., 2014); as a result, it is currently

unclear whether LacI is uniformly distributed in the cell.

Here, we study the LacI intracellular mobility and spatial distribution of LacI in live bacteria using photoactivation localisation microscopy (PALM) combined with single-particle tracking (hereafter, tracking PALM) (Subach et al., 2009; Stracy et al., 2014). Since we track single LacI molecules in live cells by stochastically photo-activating and observing fluorescent proteins individually, we have no limitations in the copy number of the protein under study; as a result, we were able to study the behaviour of LacI in bacterial strains that contain the natural copy numbers for LacI, as well as much higher copy numbers due to *lacI* over expression. Our results allowed us to determine the relative abundance of specific, near-specific, and non-specific DNA binding modes of *lacI in vivo*, showing that all these modes are operational in cells. Further, we examined the spatial distribution of LacI in live cells, confirming that its specific binding is indeed to lac operator regions on the chromosome; we also showed that the mobile LacI is exploring the bacterial nucleoid, in a way similar to other DNA-binding proteins. Finally, our work provides procedures and recommendations for applying tracking PALM to studies of low copy number proteins in living bacteria, and shows that minimal FROS markers of down to 6 operators are detectable using tracking PALM.

5.2 RESULTS

5.2.1 Diffusion analysis of single LacI molecules using tracking PALM

. In *E. coli*, tracking PALM has been used to analyse several processes *in vivo* related to nucleic acid metabolism, such as stringent response (English

et al., 2011), DNA repair (Uphoff et al., 2013) and gene transcription (Stracy et al., 2015). All three processes were studied using tracking of proteins that are present in high copy numbers (100-10,000) in each cell. Here, we apply tracking PALM to the study of *lacI*, a transcriptional repressor that has low copy numbers in each cell (~40 monomers /cell). Since *lacI* has a well-studied function and intracellular mobility, it is ideal as a model protein for establishing methods for studying transcription factors with low-to-moderate copy numbers per cell (10-100 monomers); indeed, many TFs in bacteria fall in this range of copy numbers

To draw a parallel to the copy numbers in previous studies and to gain more statistical information per cell, we first studied cells with a high copy number for LacI. Specifically, we over expressed LacI from a *lac* gene placed on a plasmid and regulated by two *lac* operators; over expression was achieved by growing cells in the presence of 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG), a mimic of allolactose that binds LacI and prevents specific binding to the operator. To study the intracellular mobility of LacI under conditions where it is capable of specifically binding DNA we diluted the cells grown at 1 mM IPTG. They were diluted to a final concentration of 10 μ M IPTG and added 1 mM 2-Nitrophenyl β -D-galactopyranoside (ONPG), which out-competes IPTG, binds LacI, and restores specific DNA binding of LacI to operator regions, as well as to any near-operator sites present in the bacterial chromosome.

To perform tracking-PALM experiments on LacI, the repressor carried a C-terminal fusion with PAmCherry (Subach et al., 2009), a photoactivatable fluorescent protein that emits in the red spectrum. As with GFP-derivatives fusions with LacI (Elf et al., 2007), LacI-PAmCherry binds to a single operator as a dimer (and not a tetramer, as in the case of wild-type LacI). Using pho-

toactivatable proteins, we are able to localise and track a single protein with up to ~ 40 nm spatial precision (Uphoff et al., 2013; Stracy et al., 2015). After starting with a pool of inactive LacI-PAmCherry (fig. 3.9A), photoactivation using a 405-nm laser ($0-0.25 \text{ W cm}^{-2}$) and readout using a (184 W cm^{-2} 561-nm laser allowed us to observe and track a single LacI protein until it bleaches (fig. 3.9B), with every frame (15 ms exposure) yielding a localisation (fig. 5.1A, fig. 3.9C). This series of steps was repeated stochastically until the entire pool of PAmCherry was tracked (fig. 5.1B; fig. 3.9D). To implement tracking PALM, the density of observable LacI-PAmCherry in the cell must be low, preferably a single photoactivated protein per cell; this is achieved by regulating the power of the 405-nm photoactivation laser to control the photo-activation rate in order to meet the required density.

To measure the intracellular mobility of LacI mobility, we used the mean squared displacements of the tracked localisations to calculate the apparent diffusion coefficient (D^*) of each track that contains 4 steps or more (5 localisations) (Uphoff et al., 2013). The larger the amount of steps the less spread on the D^* distribution, that is the error on D^* depends on the number of steps analysed and its distribution is given by the gamma distribution. Because fluorescent proteins bleach quickly in tracking PALM experiments, we are only able to collect on average ~ 3.5 localisations per track; using 5 localisations allows us to classify as much tracks as possible as best as possible. If the track is longer than 5 localisations we only use the first 4 steps to analyse its diffusion.

The D^* distribution in the +ONPG sample is wide (with most values ranging from 0 to $1 \mu\text{m}^2 \text{ s}^{-1}$) and resembles a normal distribution with a long tail at higher D^* values (fig. 5.1C). Considering that the chosen tracks contains 4 steps (which are equivalent to four independent and exponentially distributed

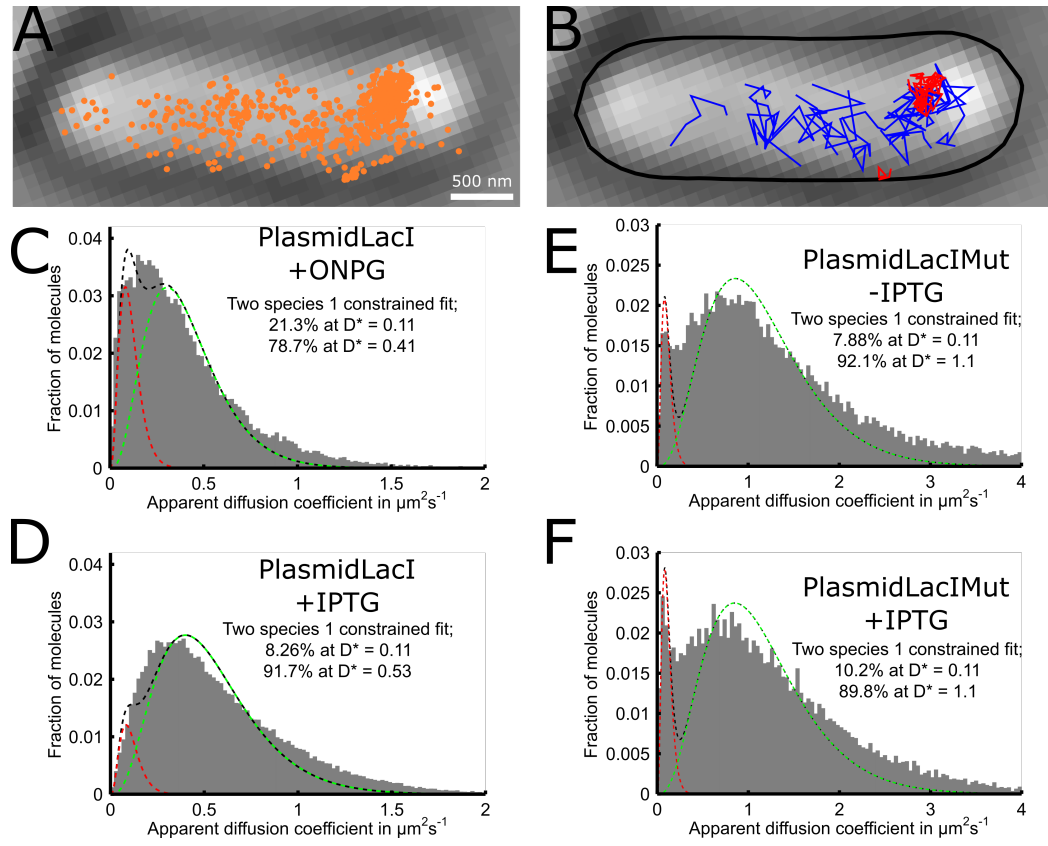


Figure 5.1: Single-molecule tracking of over expressed LacI-PAMCherry and LacIMut-PAMCherry. We use tracking PALM to observe single LacIs by photo-activating, exciting and bleaching PAMCherry with 15 ms exposure. A) shows an example cell and its localisation, B) the resulting tracks with 4 or more steps that can be classified as bound (red) and mobile (blue) by thresholding of their apparent diffusion (D^*). In C) we show the D^* histogram for the LacI without the DNA binding domain fit with two gamma distributions at $D^* = 0.11 \mu\text{m}^2\text{s}^{-1}$ and $D^* = 1.1 \mu\text{m}^2\text{s}^{-1}$. Addition of 1mM IPTG, shown in E), does not change the behaviour. D) Contains the full length LacI with ONPG to out compete IPTG used to over express the LacI. When IPTG is added F), the lac repressor has a larger D^* .

variables), we tried to fit the D^* distribution to a single gamma distribution for four such steps (fig. 5.2); despite the fact that the distribution is asymmetric and has a tailing trend, a single gamma distribution does not fit well to the D^* distribution.

However, the result above was not surprising considering the DNA-binding

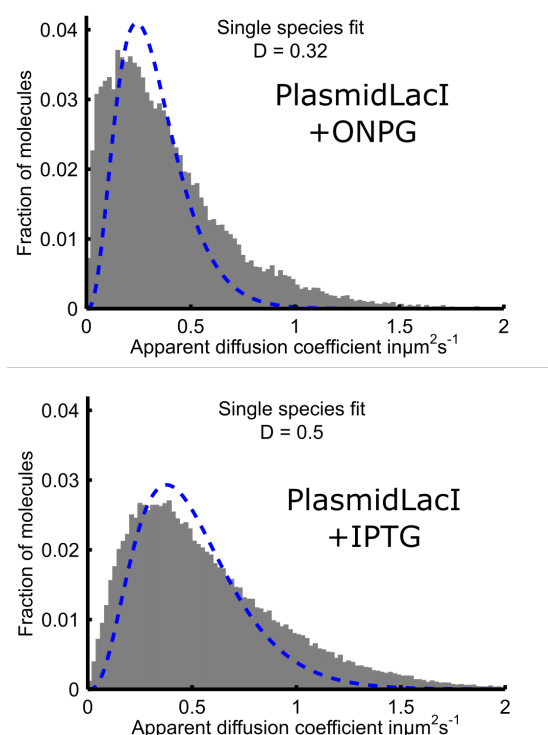


Figure 5.2: **Single fits for Plasmid LacI with ONPG and IPTG.** Single fits for the histograms shown in fig. 5.1.

properties of LacI, which dictate that a fraction of molecules is expected to bind to DNA within the average tracked window (75 ms). We then fit the D^* distribution to two gamma distributions (fig. 5.1C, red and green dashed line fits): a constrained bound fraction with a gamma distribution with a mean of $D^* = 0.11 \mu\text{m}^2 \text{s}^{-1}$ (red dashed line; representing PAmCherry-tagged proteins bound to DNA, and extracted using DNA Polymerase I and RNA Polymerase samples (Stracy et al., 2015)), and a freely fit gamma distribution which corresponds to the mobile fraction of LacI (green dashed line; note that the width of the bound and mobile fit differ because the width depends on the mean of the distribution, with lower D^* corresponding to a smaller width). This approach improves the quality of the fit substantially, although there are still deviations (possibly because there are more than 2 mobile species represented in the D^* distri-

bution); the bound population accounts for ~20% of the molecules, whereas ~80% has a mean D^* of $\sim 0.4 \mu\text{m}^2 \text{s}^{-1}$. The latter D^* is much lower than expected from free and unconstrained LacI diffusion in the cytoplasm (previously reported at $3.0 \mu\text{m}^2 \text{s}^{-1}$), consistent with expectations that the mobile species spends considerable time non-specifically and transiently bound to DNA (Elf et al., 2007).

To study LacI mobility in the presence of its inducer (IPTG), which is expected to abolish its specific and near-specific DNA binding modes of LacI, we examined cells grown and imaged in 1 mM IPTG using tracking PALM (fig. 5.1D). The D^* distribution is now wider (covering a range of $0 \mu\text{m}^2 \text{s}^{-1}$ - $1.5 \mu\text{m}^2 \text{s}^{-1}$), with the mean increasing significantly compared to the +ONPG sample (0.35 vs 0.2, respectively). Using the two-distribution fitting approach, we see that the mobile fraction (green dashed line) increased significantly, both in terms of fraction (~90% vs ~80% in the +ONPG sample), and in terms of diffusion coefficient ($\sim 0.53 \mu\text{m}^2 \text{s}^{-1}$ versus $\sim 0.4 \mu\text{m}^2 \text{s}^{-1}$ in the +ONPG sample).

To explore further the influence of non-specific DNA binding of LacI on its diffusion profile in live cells, we examined a LacI derivative where its DNA-binding domain (removing 41 amino acids from the N-terminal) has been removed; for this mutant LacI (hereafter, LacIMut), all specific, near-specific and non-specific binding modes should be non-operational. Indeed, the removal of the DNA-binding domain leads to a striking change in the D^* distribution under +ONPG conditions (fig. 5.1E), with the LacI mobility increasing substantially. Specifically, we observe a large increase in the mean of the mobile fraction, with the D^* reaching $1.1 \mu\text{m}^2 \text{s}^{-1}$ (fig. 5.1E, green dashed line; compare with the $\sim 0.4 \mu\text{m}^2 \text{s}^{-1}$ observed for the full-length LacI). This D^* value is in

good agreement to previously reported diffusion coefficient for LacI without its DNA-binding domain (converting D^* to D through simulations we find $D = 3.0 \mu\text{m}^2 \text{s}^{-1}$ as previously reported (Elf et al., 2007)). Interestingly, although the LacIMut should be entirely mobile, ~8% of tracks appear to remain bound.

Addition of IPTG (fig. 5.1F) to the cells expressing LacIMut are not expected to show any significant change, since IPTG does not bind to the mutant derivative. Indeed, IPTG addition does not change significantly either the bound/mobile distribution, or the mobility of the mobile fraction.

5.2.2 The sub-cellular distribution of LacI in living *E. coli*

Since our single-molecule localisations provide the location of LacI molecules in the cell with up to 40 nm precision, we are able to re-construct the spatial distribution of LacI from the localisations (fig. 5.1A). Further, to monitor the spatial distribution of the bound molecules separately from the mobile molecules, we sort our tracks into ones that belong to either bound or mobile molecules by using a threshold of $0.15 \mu\text{m}^2 \text{s}^{-1}$ (Uphoff et al., 2013) (Stracy et al., 2015). We reconstruct the probability density of LacI's spatial distribution by binning and then collapsing the LacI localisations along the short axis of cells, normalising for cell width (fig. 5.3; see also (Stracy et al., 2015)). If a protein covers the entire volume of the cell equally, the localisation probability along the short axis is expected to resemble a dome (fig. 5.3, black dashed line), since there is a higher probability for molecules of that protein to be in the mid cell region due to the approximate cylindrical geometry of *E. coli* cells.

For LacI with all of its DNA-binding activities intact (+ONPG conditions; fig. 5.3A), the mobile fraction differs from the uniform distribution. As with RNA polymerase and nucleoid-associated protein HU (Stracy et al., 2015), mobile

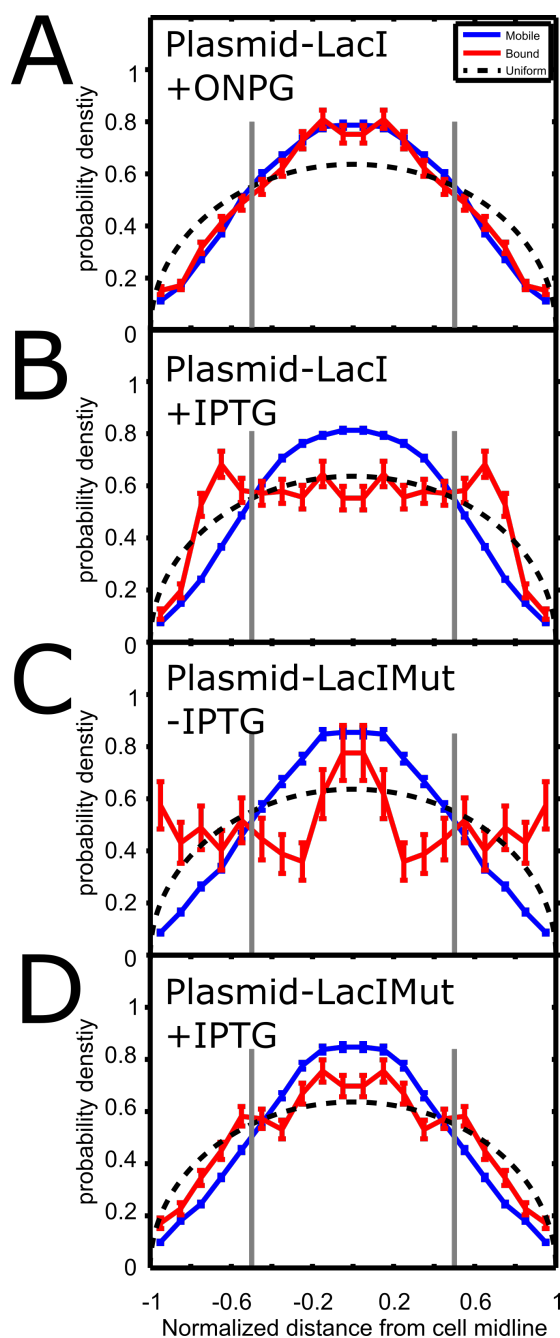


Figure 5.3: Spatial distribution of the localisation along the short cell axis. The short axis distribution of the molecules is measured with -1 and 1 representing the cell membrane in the normalised distance. Because of their single nucleoid, between $1.5\ \mu\text{m}$ and $2.5\ \mu\text{m}$ long were used to generate the plots. Error bars represent the error of a Poisson distribution ($\pm \text{normalised } \sqrt{N}$, N being the bin count number).

LacI tracks (~80% of tracks) do not distribute uniformly; instead, they follow the bacterial nucleoid (fig. 5.3A, blue line), where most of chromosomal DNA is concentrated, likely due to constant and transient non-specific DNA binding. The bound fraction (~20% of tracks; red line, Fig. 2A) also matches the spatial profile of the mobile population, and follows the distribution dictated by the nucleoid. Under conditions where the specific and near-specific DNA binding is abolished (+IPTG condition; fig. 5.3B), the mobile population (~90% of tracks) also follows the nucleoid; this result is consistent with the non-specific DNA binding mode dictating the localisation pattern of LacI on the nucleoid (as in the case of RNAP and HU; Stracy et al. (2015)). However, unlike the previous cases, the bound population (~10% of tracks) shows a probability density that differs from what is expected from either nucleoid binding or a uniform distribution; instead, bound molecules appear to be excluded from the nucleoid (red line, fig. 5.3B).

To dissect further the role of all DNA-binding on the spatial distribution of LacI, we examined the LacIMut spatial profile. Surprisingly, the mobile fraction under either -IPTG and +IPTG conditions (~90% of tracks) still showed a nucleoid-specific localisation (fig. 5.3C and fig. 5.3D, blue curves), even though LacI cannot bind DNA either specifically or non-specifically; these results imply that there are factors additional to DNA binding that result in an apparent co-localisation with the nucleoid (see Discussion section 5.3). The bound fraction (which is not as well sampled as the mobile fraction due to its low fraction) shows some exclusion from the nucleoid under +ONPG conditions, which is largely abolished upon IPTG addition

5.2.3 Auto-fluorescent species in *E. coli*

In control experiments with *E. coli* strains, both in this work, as well as in work related to electroporated molecules (Crawford et al., 2013), we did observe that cells lacking genes for photoactivatable FPs ("autofluorescence control cells") still carry autofluorescent particles. When we image such cells, we first expose them to the excitation laser (561 nm) to bleach any background generated before we turn on the photoactivation laser (405 nm); although most background bleaches using this protocol, we have noticed that photo-activation is associated with activation of auto-fluorescent species in the sample.

To account for autofluorescent species that may add noise to our D^* distribution and to the spatial distribution of localisations, we tracked autofluorescent molecules (or particles) in autofluorescence control cells. Tracking-PALM studies in such cells grown in media without IPTG generated several autofluorescent tracks (fig. 5.4) that appear immobile; they result in a narrow D^* distribution (fig. 5.4A), described by a single gamma distribution with a mean D^* of $0.067 \mu\text{m}^2 \text{s}^{-1}$. This value is lower than the $0.11 \mu\text{m}^2 \text{s}^{-1}$ obtained for immobile DNA-binding proteins (Stracy et al., 2015), and not due to differences in brightness of the autofluorescent localisations (fig. 5.6); this would effectively reduce or increase the localisation error of the fit and the apparent D^* for an immobile species. Although the presence of autofluorescent localisations will be more important for experiments with low copy number proteins (which is not the case for over expressed LacI), it is likely that these localisations also influenced the quality of the fit in (5.1), since the D^* distribution for the bound molecules has contributions both from bound LacI molecules and autofluorescent particles. We also examined the spatial distribution of the bound tracks, which appear to be distributing rather uniformly, and show a large amount of

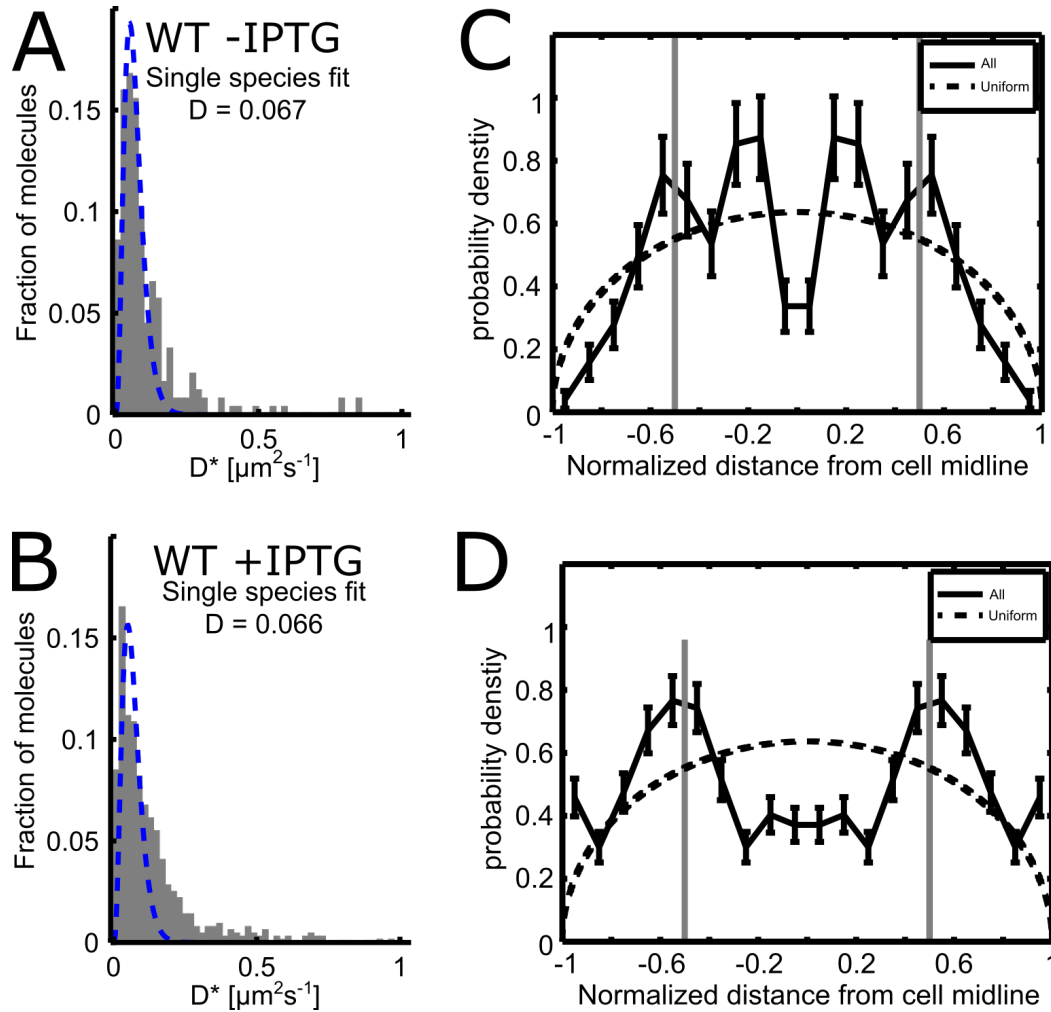


Figure 5.4: **WT cells with no fluorescent proteins.** A-B) shows the D^* of the background localisations found for WT cells without and with IPTG, in the inset of the figures the spatial distribution of the background is shown. C-D) Spatial distribution of the localisation along the short cell axis.

cell-to-cell variability (fig. 5.4C,D).

To examine whether IPTG addition has an effect on the population of autofluorescent particles, we also studied autofluorescence control cells in the presence of IPTG (fig. 5.4B). Addition of IPTG increased the number of autofluorescent tracks (fig. 5.6), which reached values that need to be considered carefully for low copy number proteins, such as LacI (with 40 monomers per cell in wild-type cells). Regarding the mobility of tracks, we observed a profile

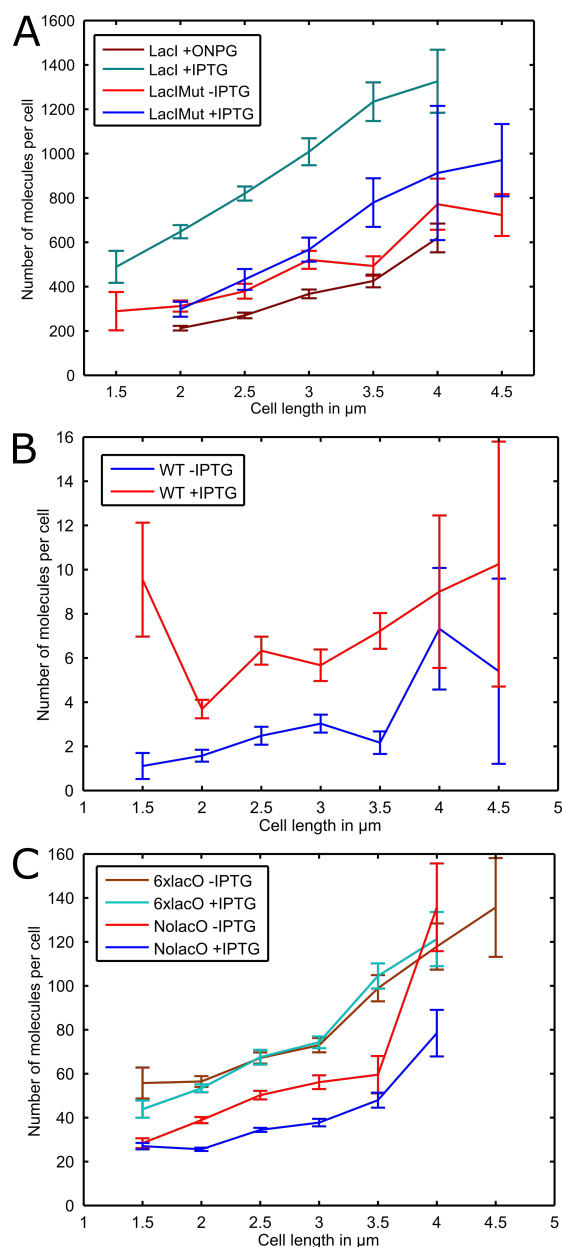


Figure 5.5: Copy numbers per cell length. In A) the copy numbers of the LacI expressed from a plasmid are shown. In B) the copy numbers for the wild-type cells are shown. In C) the copy numbers for the genomically modified strains are shown. Error bars are the standard error of the mean.

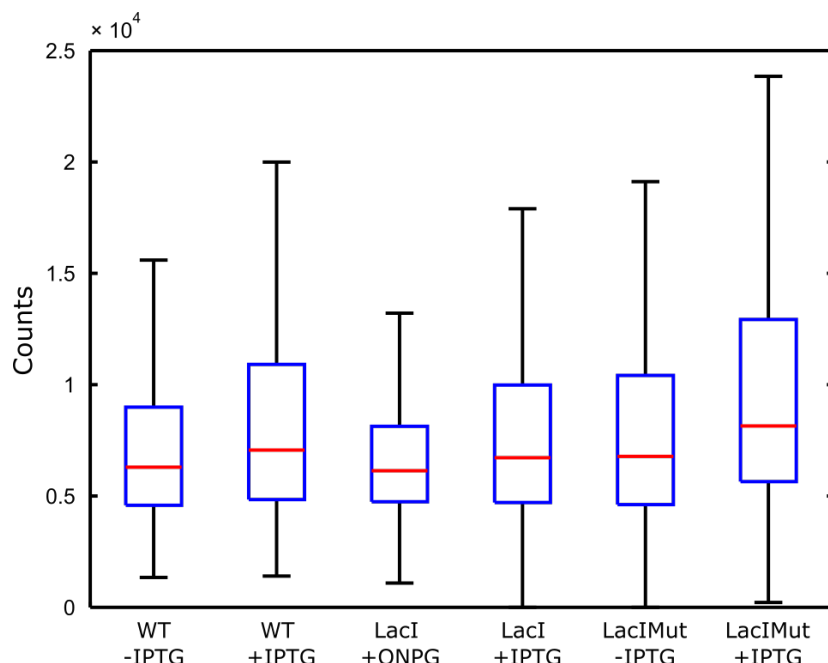


Figure 5.6: **Counts from the localisations for each experiment.** A 7 by 7 pixel window is used to fit a 2D Gaussian and the total number of counts under the area of the Gaussian calculated.

similar to the case without IPTG, albeit the D^* distribution was slightly wider, and the immobile particles tended to be excluded from the nucleoid (fig. 5.4B, inset).

5.2.4 The *in vivo* diffusion profile of LacI expressed at native copy numbers

We then measured the LacI mobility and spatial distribution in two *E. coli* strains carrying wild-type copy numbers for LacI, both strains have a chromosomal copy of full-length LacI fused to PAmCherry (and carry no plasmid for LacI over expression). In both strains, the wild-type operators (lacO1, O2, and O3) have been deleted (fig. 5.7, fig. 5.2). The first strain (named 6xlacO) has six proximal consensus LacI operators (a 21-bp DNA sequence) within a 194-bp DNA segment (fig. 5.8); the second strain (named NolacO) contains

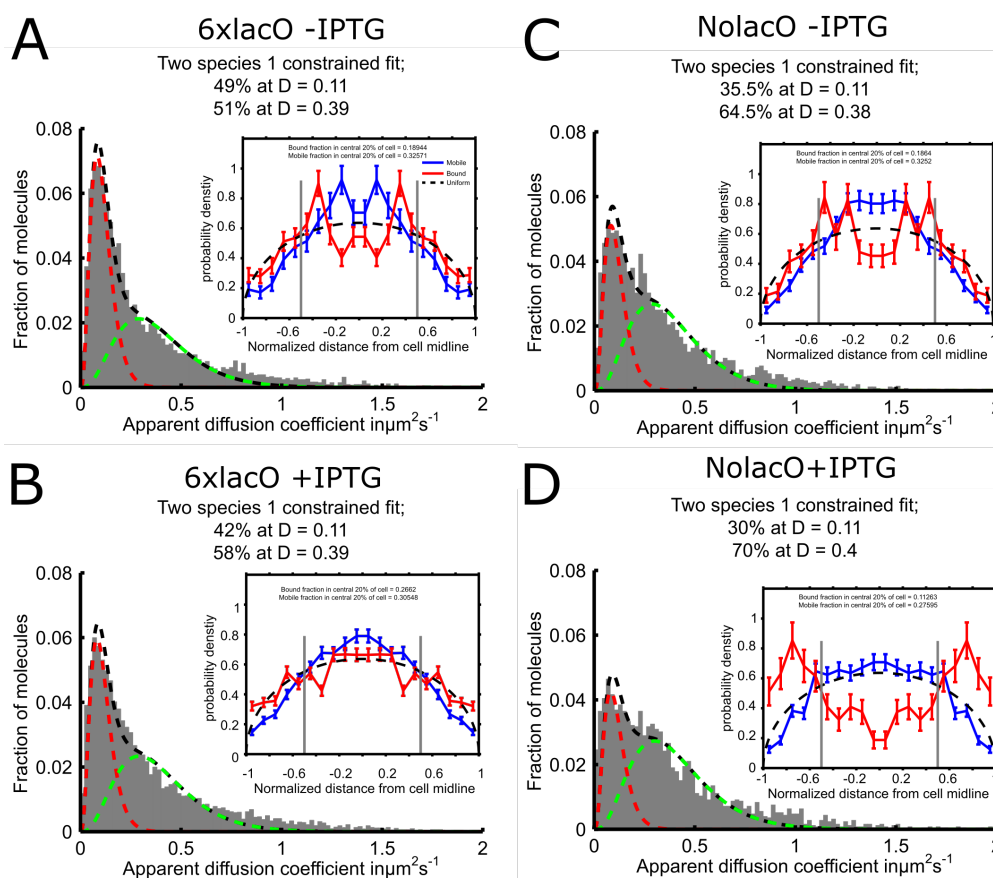


Figure 5.7: Native copy numbers with tandem operators 6xlacO and all operators removed NolacO. We tracked LacI-PAmCherry with and without a 1mM IPTG treatment using the 6xlacO and NolacO strains in live cells. A- D) diffusive behaviour from the calculated apparent diffusion of each track and displayed as the outline of the histogram of all the tracks with the spatial distribution of each as an inset. .

no lac operators. The strains allow us to observe an exaggerated amount of LacI binding to operators (up to 12 monomers, all of high affinity) compared to wild-type (6 monomers, with differing affinities).

We first tracked LacI in the 6xlacO strain in the absence of IPTG; in this case, all DNA binding modes are operational, a fact that should lead to reduced mobility for the many LacI molecules. Consistent with this expectation, we observed that ~50% of the tracks are bound (Fig. fig. 5.7A); further, the mobility of the mobile fraction (as reflected in its D^* distribution) is very similar to

5. TRACKING SLOW AND LOW COPY NUMBER PROTEINS IN LIVE CELLS USING THE LAC REPRESSOR

Federico Garza de León

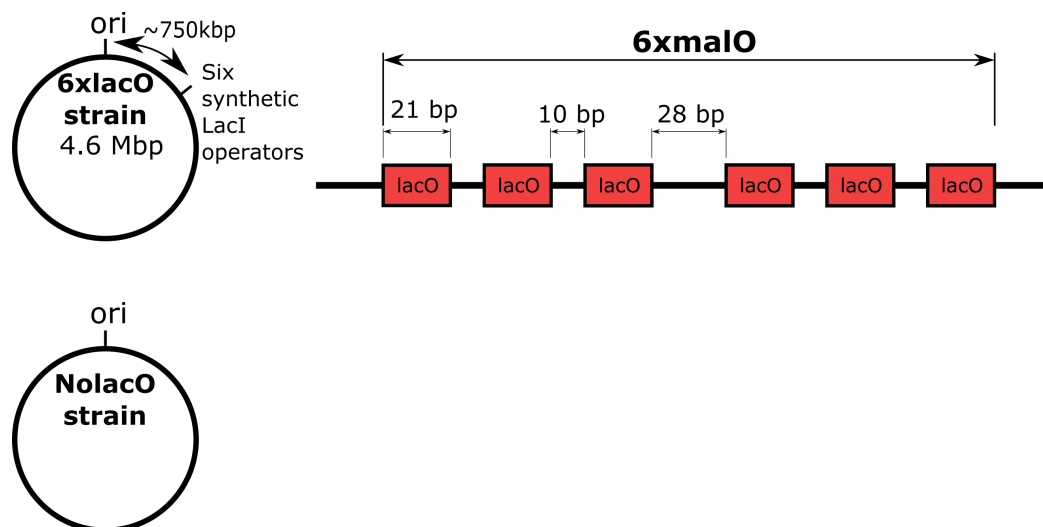


Figure 5.8: **6xlacO and NolacO strains description.** In 6xlacO, 6 tandem *lac* operators are inserted into the genome and the native operators removed. In NolacO there are no operators in the genome.

that from the overexpressed sample ($0.39 \mu\text{m}^2 \text{s}^{-1}$ vs. $0.41 \mu\text{m}^2 \text{s}^{-1}$, respectively). We also examined the spatial distribution of localisations; due to the low copy number of LacI in these cells (fig. 5.6), the distribution was more noisy than in the case of over expressed LacI. The mobile distribution (fig. 5.7A, inset) resembles that observed for the over-expressed LacI, i.e., preferential localisation within the bacterial nucleoid. By contrast, the bound localisations showed a preference for the nucleoid periphery, and are less abundant in the central region. We also examined the effect of IPTG addition to the 6xlacO strain, which is expected to lose the LacI molecules that bind to the 6 tandem operators, and therefore display decreased LacI binding. Consistent with expectations, the proportion of bound tracks decreased to from 49% to 42% which indicates that there is still non-specific binding (fig. 5.7B); the increased mobile fraction still showed mobility (with a mean D^* of 0.39) that matched that of the mobile fraction in 6xlacO in the absence of IPTG, as well as in the case of LacI over expressed from a plasmid in the presence of ONPG. The mobile

fraction once again matched the nucleoid spatial distribution, whilst the bound fraction seems to be uniformly distributed throughout the cell.

To further study the modes of LacI binding *in vivo*, we examined the NolaC0 strain, which contain no operators. When LacI has no operators and IPTG is absent, it can still bind non-operator DNA, such as near-specific sequences which may exist at various chromosomal loci, as well as non-specific DNA; however, specific binding such as the one anticipated for the 6xlaC0 strain will not be present. Consistent with this, there is a decrease in the proportion of bound tracks (from ~50% for the 6xlaC0 strain without IPTG, to ~35% for the NolaC0 strain without IPTG), although there is little change in spatial localisation; further, the mobile fraction is very similar in mobility and spatial localisation to that seen for the 6xlaC0 strain.

Finally, we examined how the LacI profile in NolaC0 is affected by IPTG, which is expected to remove all specific and near-specific binding, and reduce non-specific binding (Marklund et al., 2013). Indeed, IPTG addition (fig. 5.7D) reduces the bound fraction to ~30%, and shows localisations that imply some exclusion from the nucleoid; however, at these low copy numbers, the presence of autofluorescent background species may skew the data. The mobile fraction shows a mobility that is essentially identical to the three previous cases, reinforcing the conclusion that the mobility of the LacI species reflects a combination of 3D diffusion interrupted by non-specific binding to the bacterial nucleoid.

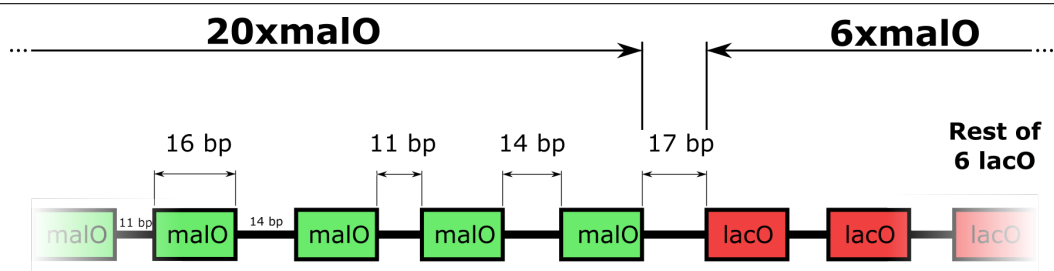


Figure 5.9: 6xlacO strain with 20 Mal operators description.

5.2.5 Bound LacI associated with the expected chromosomal locus

To ensure that the bound molecules associate with the desired chromosomal location, and to examine the presence of any LacI concentration gradient in living *E. coli*, we examined the location of the LacI molecules relative to the actual location of the six lacO operators within the cell. It has been possible in the past to image small FROS on this scale (Hensel et al., 2013; Wang et al., 2011). To label the chromosomal position of the 6xlacO tandem operators, we constructed a strain that carries a FROS (Lau et al., 2004; Robinett et al., 1996) marker based on Mall (fig. 5.9), which is another transcriptional repressor protein involved in the regulation of maltose. We placed 20 Mall operator sites near the six lac operators (with the boundaries of the two tandem operator systems separated by 17 bp; fig. 5.9) and expressed Mall-GFP from a plasmid; binding of the Mall molecules to their operators forms a single fluorescent spot per labelled genetic region, allowing us to check the number of copies and the location of these genes in individual *E. coli* cells. We imaged the GFP with a 473-nm laser line (115 W cm^{-2}). If LacI molecules bind to the 6xlacO operators, the LacI localisations will cluster near the FROS marker; further, the LacI clustering should disappear upon IPTG treatment.

After imaging the FROS foci to obtain the sub-cellular position of the 6xlacO

sequence (fig. 5.10A), and subsequent tracking PALM on LacI-PAMCherry on the same cells, we noticed that, as expected, many LacI localisations were found very close to the FROS marker in the absence of IPTG (fig. 5.10B). To quantify the LacI clustering relative to its operator sites, we applied the radial distribution function analysis $g(r)$ (Sengupta et al., 2011)(Veatch et al., 2012) to our PALM data. For each cell, the distance between the FROS marker and all LacI localisations is determined. To account for the geometry of *E. coli* cells, we generated simulated random localisations within the cell (using a uniform molecular density in the cell) and determined the distances of the localisations to the LacO sites. The radial distribution is calculated as the ratio of experimentally acquired distances to the simulated distances. Thus, randomly distributed data yield a $g(r)$ of 1, while clustered data yields $g(r)$ values > 1 .

Using radial distribution functions for the 6x_{lacO} strain (fig. 5.10C), we find that the LacI localisation exhibits non-random behavior in both untreated and IPTG-treated cells. In untreated cells, we find that the probability of finding LacI near its operators (blue line, Figure fig. 5.10C) is up to 16-fold higher than what would be expected for a simulated random distribution (black line, fig. 5.10C), clearly indicating LacI specifically binds to its tandem operators. In contrast, in IPTG-treated cells, this probability is only slightly higher (red line, fig. 5.10C). Treating the cells with IPTG makes it 4× more probable to find LacI molecules within 200 nm of the operator sites. The LacI in cells without IPTG were 11× more probable to be found on a radius of 200 nm of the FROS position (indicated in green in fig. 5.10B), whilst in the treated cells this was only 2.7× more probable. Using the colocalisation from the radial distribution function (200 nm), we counted the number of molecules that are within a certain distance from the FROS marker (fig. 5.5). This shows that there are, on av-

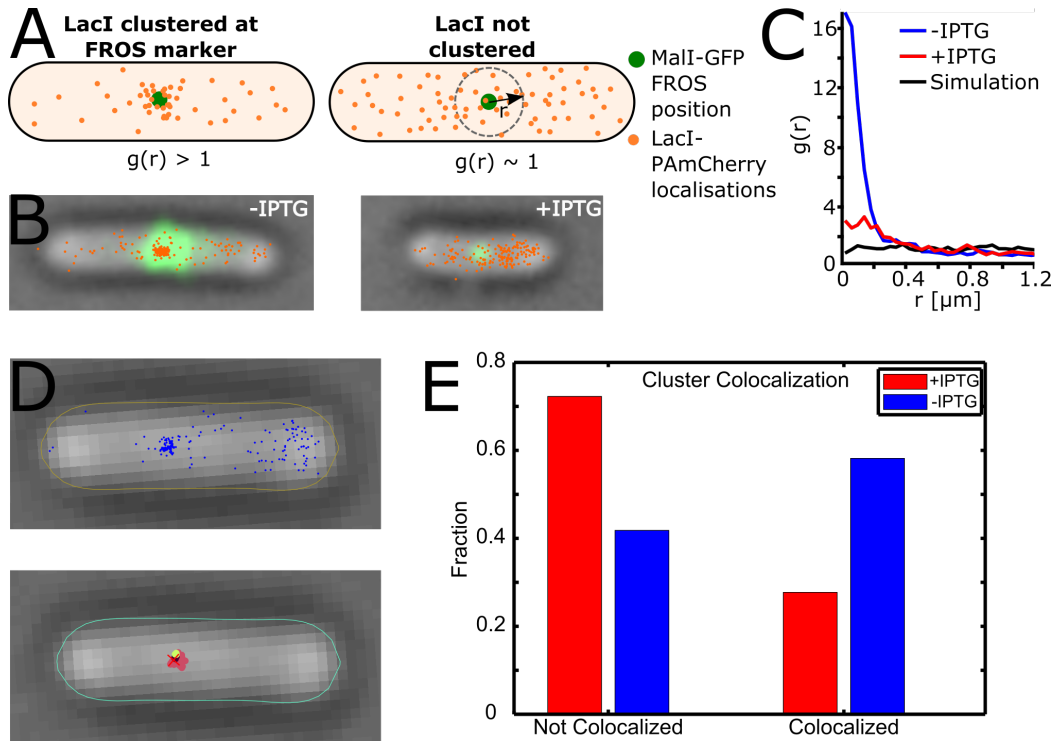


Figure 5.10: **Clustering of LacI-PAMCherry near the FROS marker.** We use pair correlation to determine the clustering of the LacI near its binding site by tagging the gene with a 20 tandem operators of Mall. To label the operators, Mall-GFP is expressed from a plasmid. In A) the clustered localisations will have a radial distribution $g(r)$ higher than 1. If it is unclustered and randomly distributed in the volume of the cell it will lead to a $g(r)$ that is close to 1. In B) Experimental data of the FROS marker is shown along with the identified localisations of the LacI-PAMCherry. In C) the radial distribution function are shown for the 6xlacO strain with and without treatment of IPTG. The black line plot that is close to $g(r) = 1$ are random simulated positions with the same number of localisation from the real data found in the cells. In D) the localisations for a different cell are shown on the top image and the results from the clustering algorithm (clusters of different colors, red cross is the localisation of the FROS marker). E) Shows the fractions of the clusters that are successfully colocalised to a FROS localisation.

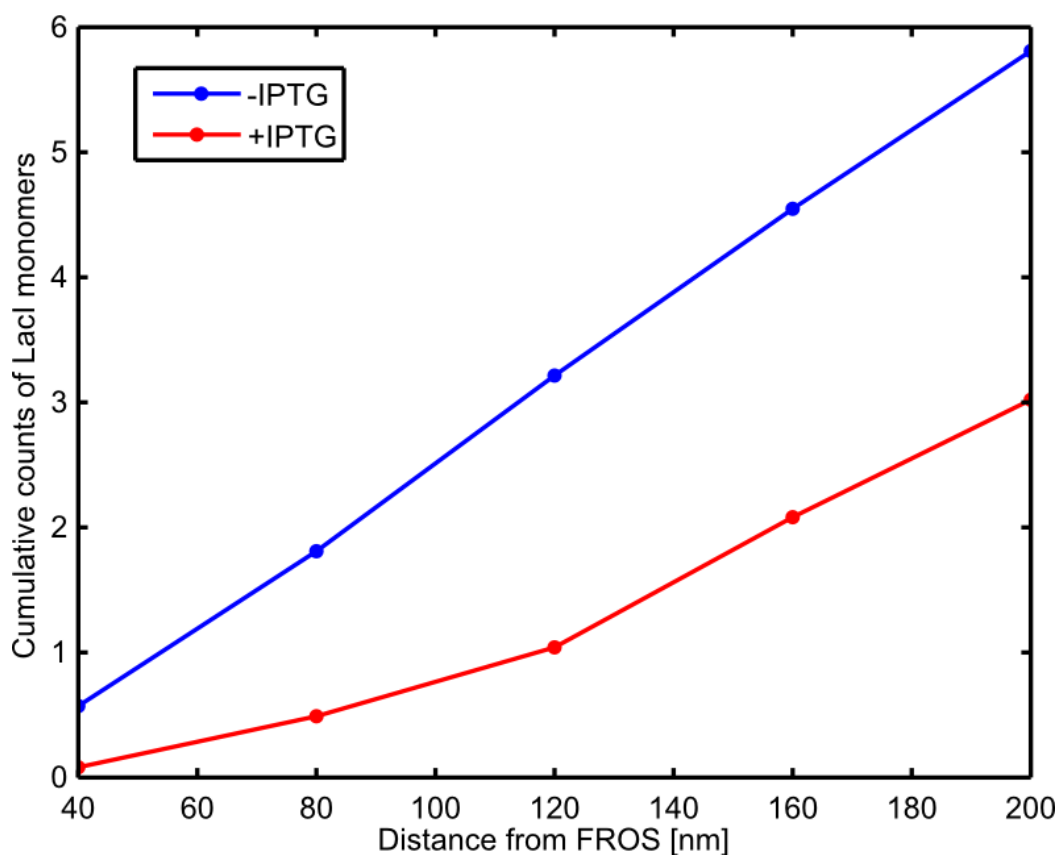


Figure 5.11: **Number of LacI monomers relative to FROS position.** On average, 5.8 monomers of LacI related to the FROS marker in absence of IPTG to a distance of 200 nm (blue line). In the presence of IPTG this number decreases to 3.0 for a distance of 200 nm (red line).

erage, 5.8 monomers of LacI related to the FROS marker in absence of IPTG (fig. 5.11, blue line). In the presence of IPTG this number decreases to 3.0 (fig. 5.11, red line).

We also tested the ability of using native copy numbers to identify operator regions. In a typical FROS experiment, free diffusing proteins are blurred and a cluster of bound proteins can be localised. In our case we can detect all molecules, free or bound; we wish to test the ability of identifying the tandem operators' position with clustering algorithms. The raw localisation as depicted in the upper cell in fig. 5.10D can be clustered into the red cluster depicted in

the lower cell. Once clustered, we colocalise the position of the cluster with the nearest FROS marker using the 200 nm threshold. We calculate what fractions of the clusters colocalise with a FROS with and without IPTG (fig. 5.10E). We find that without IPTG almost 60% of the clusters colocalise to a FROS position. Surprisingly around 30% of clusters colocalise to a FROS with addition of IPTG even though LacI loses the ability to bind the operators in this case.

5.3 DISCUSSION

Many transcription factors exist and work in the cell in low to medium copy numbers (1-100 monomers). PALM is one of the best methods to elucidate their activity and spatial distribution *in vivo*. In this work we applied tracking PALM to LacI to probe the technique on a low-copy number DNA-binding protein. We calculated the D^* profile and found that its distribution differs between treatment (+/- IPTG, fig. 5.1,4), LacI condition (LacIMut, over expressed LacI and LacI, fig. 5.1,4) and presence/absence of specific binding sites (WT, 6xlacO, NolacO, fig. 5.1,3-4). We also found that fluorescent background species that photo-activate can have a significant effect on low-copy number diffusion and spatial distribution (fig. 5.4). In addition we successfully showed colocalisation LacI to tandem operators by using FROS to image the chromosomal locus (fig. 5.10). With this we were able to probe the efficiency of identifying a tagged chromosomal locus purely from tracking PALM data.

5.3.1 LacI Diffusion

Our work on LacI is consistent with much of the previous live cell single-molecule work has shown in the past: LacI searches operators by facilitated diffusion. Our tracking PALM work shows that whilst LacI without its DNA-

binding domain has a fast diffusion ($\sim 3.0 \mu\text{m}^2 \text{s}^{-1}$) the full-length repressor's diffusion is reduced and the affinity towards DNA is shown with spatial bias towards the nucleoid. We also found that near-specific binding has a large effect when LacI is uninduced; this would affect current search times published which were estimated with LacI induced with IPTG. The decreased affinity of induced LacI towards non-specific DNA has been shown in the past (Marklund et al., 2013) which suggests that uninduced LacI may spend, on average, more than the ~ 5 ms bound to DNA. We estimate that LacI bound to IPTG spends 5% less time on DNA on average than the unbound LacI (+IPTG 82%, -IPTG 87%; $D_{\text{eff}} = (1 - f)D_{\text{free}}$).

5.3.2 Spatial Distribution

In addition to the diffusion work, we observe the spatial distribution of LacI and show the bias that LacI has towards the nucleoid, as has been previously observed (Kuhlman and Cox, 2012). We also found that LacIMut's short axis distribution seemed to follow the nucleoid shape but upon further examination of the long-axis distribution we found that LacIMut does not follow the double nucleoid shape as does LacI (fig. 5.12). We discard that LacI has any interaction with the nucleoid at this time-scale and that the bias may be due to our exposure time relative to the fast diffusion. At this time-scale, LacIMut may diffuse a large distance and the point-spread function will be biased toward the center of the cell because of the cell wall acting as a border. This will also bias the localisation of the point spread function towards the center of the cell. In addition we also found that bound molecules contain a large bias towards the edges of the cell which can partially be contributed by background fluorescence.

5. TRACKING SLOW AND LOW COPY NUMBER PROTEINS IN LIVE CELLS USING THE LAC REPRESSOR

Federico Garza de León

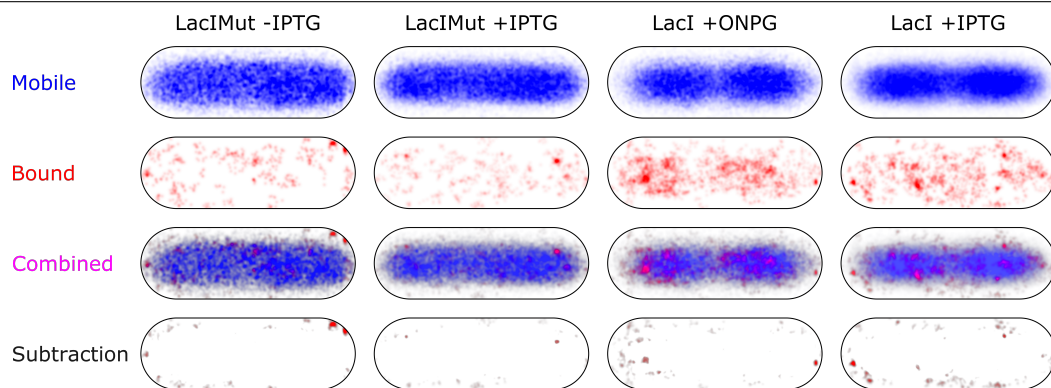


Figure 5.12: **Sub-diffraction positions of LacI.** All cells from a with lengths $2.5\mu\text{m}$ to $3.5\mu\text{m}$ are normalised and their positions aggregated. Mobile molecules are shown in blue and bound in red. I overlay both top cells for a combined image and the bottom cell is the subtraction of the mobile and the bound to show the number of bound molecules at the edges of the cell.

5.3.3 Challenges with low to moderate copy numbers

Working with low (1-5 monomers) to moderate (5-100 monomers) copy numbers can have many issues as the background will have a larger effect on the data and the number of classifiable tracks is very small per cell. Based on our work we found that the number of background molecules will affect the bound fraction of tracks. Because it is the bound fraction that is normally associated with function (LacI: repression, Pol I: DNA repair), it is important to assess the effect of background tracks. In addition to fractions of bound and mobile molecules, the background will also have an effect on the spatial distribution of the classified tracks. We found that background tracks differ in their distribution compared to DNA-binding proteins that follow the nucleoid or freely diffusing proteins, which diffuse throughout the cell. The background tracks seem to exclude themselves from the nucleoid volume. This can have an effect on the distribution of the bound tracks and caution must be taken when interpreting the results. At the same time a control experiment with WT cells helps elucidate errors and strategies for correction.

5.3.4 Small FROS

In our work we also study a small FROS of 6 lacO operators and investigate the ability of this experiment to identify the lacO positions *in vivo*. We found that with low copy numbers we are able to identify regions of operators, but may have a large amount of false positives and false negatives. We also found that we are able to colocalise around 60% of the clusters to a tagged chromosomal locus. This estimate gives us an approximation that is dependent on localising both the FROS marker and the cluster. It may be the case that we identify a cluster without identifying a FROS marker because of the efficiency of the Mall-GFP binding. The 30% colocalisation when treated with IPTG might be due to coincidence. Our experiment gave a measure of the reliability to find operator regions from clusters of molecules. We believe that the issues that we confronted in our work with LacI will be common in tracking low-to-moderate copy number proteins. Our approach and insights will be helpful when studying transcription factors or other proteins that exist in low concentrations in bacteria.

5.4 METHODS

5.4.1 Cell culture

A plate was streaked and placed overnight in 37 C and a single colony was placed into a culture tube in 5 ml of LB. The culture was grown for 4 hours at 37 C. The culture was diluted 10,000 times in M9 medium, supplemented with MEM amino acids and L-proline, MEM vitamins and 0.2% glucose. The diluted culture was grown overnight at 37C. The next day the cells were diluted to approximately 0.025 OD600 in the M9 medium stated above and grown at

37C for approximately 2 hours until they reached OD600 0.1. The cells were then concentrated by centrifugation and immobilised in 15 µL wells with 1% polyethylenimine or in 1% agarose pads. Induction of cells with IPTG was done by mixing 1mM of IPTG in the M9 media. To induce on the slide the cells were washed the M9 minimal media supplemented with 1 mM IPTG.

5.4.2 Analysis

MicrobeTacker (Sliusarenko et al., 2011) was used to segment cells and analyse tracks on a per cell basis. To track the single-molecules we used custom-written code in MATLAB. We localised the molecules with a precision of 20-nm using elliptical Gaussian fitting. Localisations that were consecutively within a radius of 0.57 µm were linked; we also allowed for the molecules to disappear (due to blinking or defocusing) for 1 frame. The mean squared displacement of each track was calculated to analyse the diffusion of the molecules. The apparent diffusion coefficient was calculated as follows:

$$D^* = \frac{MSD}{4\Delta t} - \frac{\sigma_{loc}^2}{\Delta t} \quad (5.1)$$

where D^* stand for the apparent diffusion coefficient (in $\mu\text{m s}^{-2}$), MSD is the mean square displacement, Δt stands for the time lag and σ_{loc}^2 represents the localisation error.

To define the apparent coefficient of an immobile molecule, cells were fixed in 2.5% paraformaldehyde as shown in (Uphoff et al., 2013). Thus a threshold was defined for the molecules that are bound: $D^*_{\text{thres}} = 0.15 \mu\text{m}^2 \text{s}^{-1}$. Any molecules below this D^* is deemed to be immobile. The radial distribution was constructed with radial binning $dr = 40 \text{ nm}$.

5.5 TRACKING PALM OF OTHER TFs

5.5.1 The AraC protein

The protein known as AraC was the first discovered positive regulator TF. Like LacI, AraC binds DNA as a dimer but unlike it, it is involved in the regulation of multiple genes. It is a positive regulator because it activates transcription of the genes it regulates when its bound by its inducer L-arabinose. The method by which AraC represses a gene in absence of arabinose is by tetramerising with an AraC dimer located further away from the gene and creating a repression loop. When arabinose is present it dimerises with a closer monomer that allows binding of RNAP and CRP. The genes are also regulated by CRP just as in the *lac* operon (Schleif, 2010; Stringer et al., 2014). Like LacI, AraC exists at low copy numbers; a typical cell has been reported to have 40 monomers of AraC (Kolodrubetz and Schleif, 1981).

Imaging AraC is a step forward from the well studied LacI. In this case it regulates multiple genes whilst being around the same size and concentration as LacI. To be able to image AraC, we label it with a C-terminal fusion with PAmCherry endogenously. In this case AraC-PAmCherry is also controlling its own expression. We perform tracking PALM in the same fashion as LacI experiments.

5.5.2 Method

We grow cells with the endogenous AraC-PAmCherry fusion in minimal media (M9) with either 0.2 % L-arabinose (sigma-Aldrich) or glucose. 30 min before spinning and placing on the slide, one batch of cells was treated with 50 $\mu\text{g } \mu\text{L}^{-1}$ Rifampicin final concentration. The cells were imaged as in LacI experiments.

5.5.3 Diffusion profile AraC

AraC binds its operators when both repressing and promoting transcription. In the presence of glucose AraC represses the *araBAD* and *araC* genes, which are also doubly repressed by low concentrations of cAMP since CRP is also involved in promoting transcription. In the absence of glucose and presence of arabinose AraC still binds its operators but CAP is now able to bind the gene and promote transcription.

When the cells are grown in 0.2 % glucose, most proteins are bound to DNA (fig. 5.13A). The width distribution of shows signs of AraC's continuous interaction just like with LacI (fig. 5.13B). Both bound and mobile molecules do not follow the uniform distribution and follow the shape of the nucleoid.

When the cells are grown in arabinose the number of bound molecules is reduced by 10 % (fig. 5.13C). This increase may be purely due to the increase in monomer concentration (fig. 5.14). Induction with arabinose causes activation of the *araBAD* gene that also expresses the *araC* gene, that is there simply is more freely AraC but not enough sites to bind. The cell width distribution's bound population follows the distribution for cells grown in glucose (fig. 5.13D), that is it follows the nucleoid shape. The bound molecules follow more uniform distribution with some periphery bias; because of AraC's low copy numbers it may still be affected from bound background molecules affecting the width distributions.

The similarities with LacI's diffusion and distribution yield us to conclude that AraC's search process may be very similar to the repressor's case. Mobile molecules seem to follow the nucleoid shape and even when mobile have a low mobile D^* of around $0.2 \mu\text{m}^2 \text{s}^{-1}$. We believe that AraC is continuously

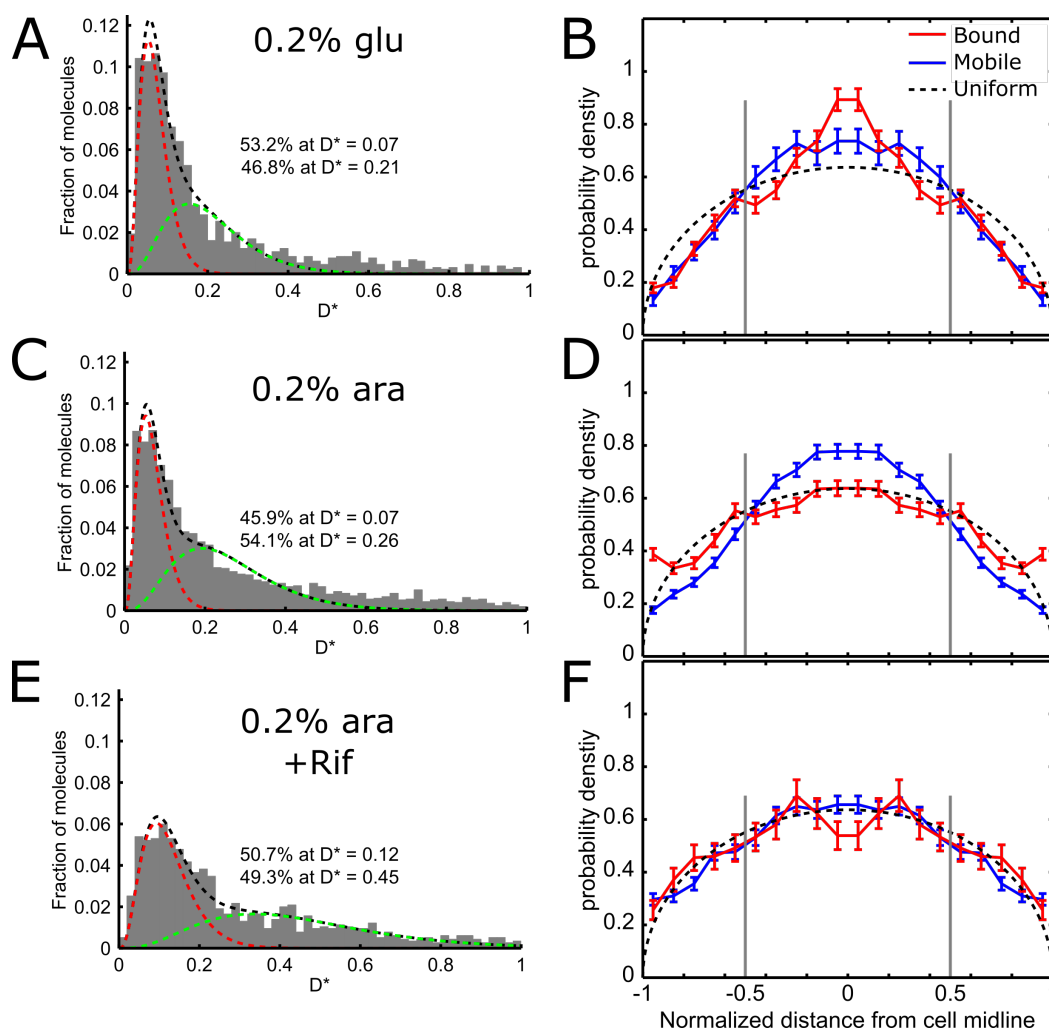


Figure 5.13: **D^* and spatial distribution of AraC-PAMCherry *in vivo*** In A) cells are grown in 0.2 % glucose, most proteins are bound to DNA. In B) the width distribution of shows signs of AraC's continuous interaction. Both bound and mobile molecules do not follow the uniform distribution and follow the shape of the nucleoid. In C) the cells are grown in arabinose the number of bound molecules is reduced by 10 %. In D) the cell width distribution's bound population follows the distribution for cells grown in glucose.

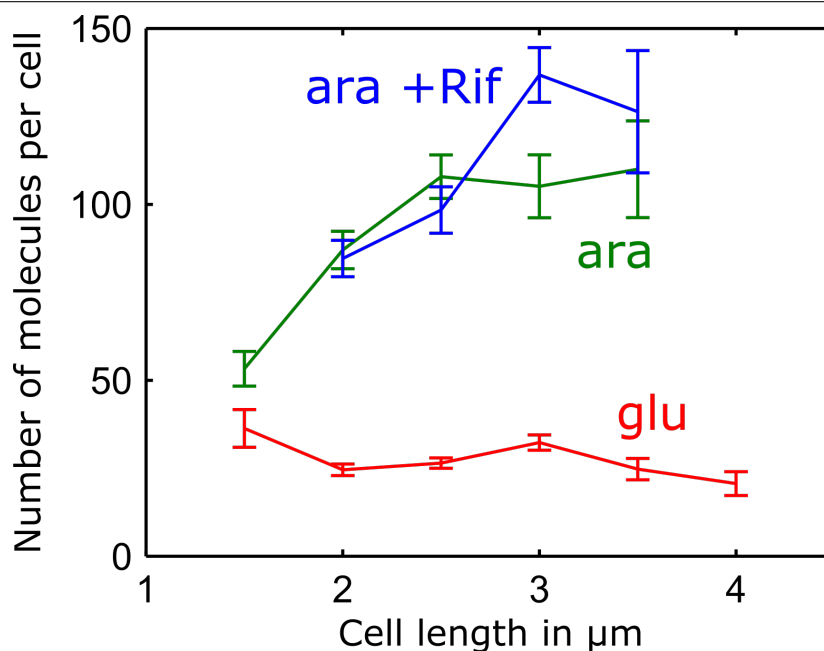


Figure 5.14: **Copy numbers per cell length for the AraC-PAMCherry strain.** The copy numbers for the AraC-PAMCherry strain are calculated for the cases grown 0.2 % glucose, 0.2 % arabinose and arabinose treated with $50 \mu\text{g } \mu\text{L}^{-1}$ Rifampicin.

binding DNA, but we don't report enough evidence to conclude if it slides or hops on DNA.

To show that the nucleoid shape of the width plots is not due solely to some other effect. We treated one set of cells with $50 \mu\text{g } \mu\text{L}^{-1}$ rifampicin, which is known to expand the nucleoid shape and make RNAP width distribution more uniform throughout the cell width (Stracy et al., 2015).

We find some other effects known for rifampicin treatment (Stracy et al., 2015), the slow and fast populations have increased in mobility (fig. 5.13E). The effect of the expansion of the nucleoid is also evident in the distribution of AraC in the cell width (fig. 5.13F).

5.5.4 CRP

As has been described before CRP is a positive regulator of many genes in *E. coli*. We attempted to image CRP with tracking PALM. We fused PAmCherry to the C-terminal of CRP endogenously.

5.5.5 Diffusion profile of CRP

In the absence of glucose cAMP binds to CRP and together bind DNA to promote transcription. In the presence of glucose, cAMP concentrations are lowered but there is still enough to bind CAP consensus sequence and for transcription to be promoted. We observe that 16 % of CRP seems to be bound to DNA and the rest have a high mobility (fig. 5.15A). The distribution of the CRP that is bound seems to have some bias to the outside of the nucleoid (fig. 5.15B).

When the cells are grown in fructose and induced with 5 mM cAMP the bound fractions increases 5 %. This is not the striking effect we expected. With such high concentrations of CAMP we were expecting to see a large change in mobility. There does not seem to be much different between the mobile and bound spatial distribution.

5.5.6 Discussion

We do not observe CRP transitioning to largely mobile state in our experiments or a large change in mobility. We tested CRPs ability to promote an operon on a plasmid by a β -galactosidase (provided by Riddhi Shah) and found that activity is lowered by 2/3 for the case of the fused CRP. This may not affect binding but we find that our fusion may be affecting the CRP.

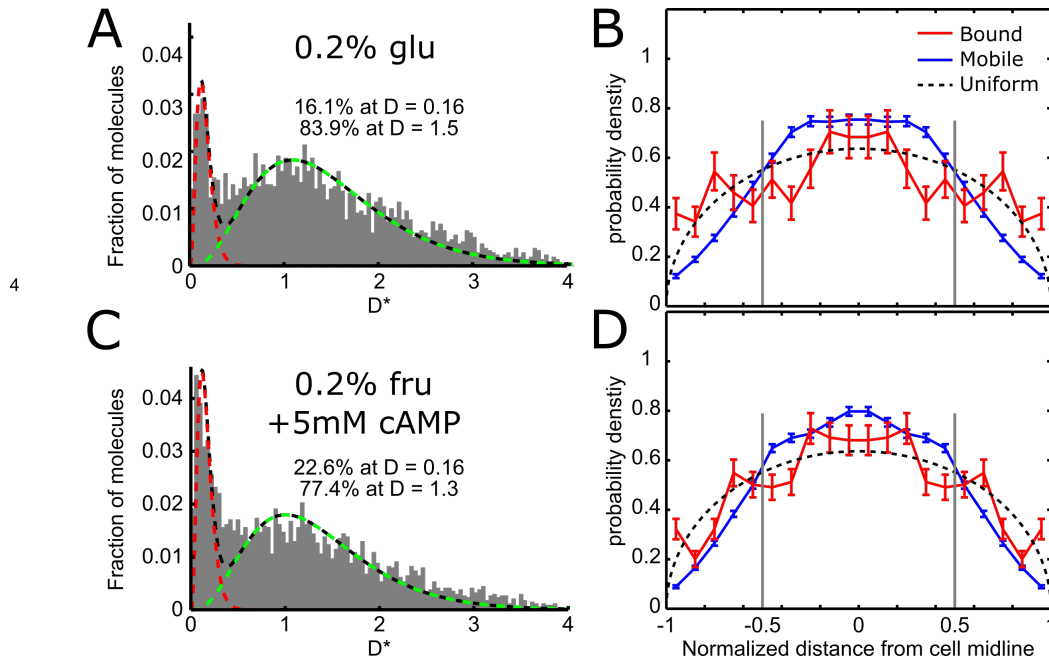


Figure 5.15: **D^* and spatial distribution of CRP-PAmCherry *in vivo*** In A) when grown 0.2 % glucose we observe that 16 % of CRP seems to be bound to DNA. In B) The distribution of the CRP that is bound seems to have some bias to the outside of the nucleoid (fig. 5.15B). In C) the cells are grown in 0.2 % fructose and induced with 5 mM cAMP. The observed bound fraction increases 5 % from the glucose grown cells. In D) there is little difference between the mobile and bound spatial distribution.

5.6 CONCLUSION

Studying TFs with tracking PALM has presented a challenge that wished to characterise. Because most TFs exist in copy numbers of less than 100 monomers per cell there may be affected by background from imaging. I showed that this background may affect interpretation of the spatial and diffusion distribution. But just as PSFs can be resolved with enough signal-to-noise, we can distinguish the data from our TFs from the "noisy" background tracks if we understand them better.

SIMULATING DIFFUSION OF TRANSCRIPTION FACTORS

6.1 INTRODUCTION

As I have shown in the previous chapters, single-molecule tracks of PAFPs are difficult to analyse; the tracks are short which increases the variance of the calculated apparent diffusion coefficient (D^*), are confined to the cell volume and are dependant on a variable localization error depending on the magnitude of their diffusion (Uphoff et al., 2013; Stracy et al., 2015). Until techniques that use long-lived fluorescent probes mature enough to be used with any protein (Crawford et al., 2013; Plochowietz et al., 2014), FPs remain the best method to tag and observe proteins *in vivo* at the single-molecule level.

To my knowledge there are no analytical equations to extract unbiased diffusion coefficients from tracking data of cytoplasm confined molecules; previous work has used simulations to fulfil this requirement (Uphoff et al., 2013; Stracy

et al., 2015). It is possible to account for the localization error and the exposure frames contribution to the apparent diffusion (Michalet, 2010), but there is little work to try and account for the variable confinement volume of the cell and the increase of localisation error for highly mobile molecules.

In this chapter I show that simulating the confinement in the cytoplasm of the cell is crucial to estimate the diffusion coefficient of the fluorescent probe. As an original contribution, we must increase the relevance of the simulation by emulating the diffusion of DNA-binding proteins like TFs, which diffuse in a combination of two states: mobile and bound to DNA (Mueller et al., 2013). Simulations can open the door to understand the complexities arising in experimental tracking PALM data of DNA-binding proteins (English et al., 2011; Uphoff et al., 2013; Stracy et al., 2015). I hope that it will also shed light into future studies of DNA-binding proteins in live cells.

6.1.1 Diffusion coefficient estimation

Without confinement or the presence of the nucleoid, DNA-binding proteins will diffuse in Brownian motion. Adding obstacles or confining molecules to the cytoplasm of the cell reduces the D^* mean of the tracks. Added to these issues are the scatter in the D^* distribution from the shortness of tracks (typically 4 displacements in our tracking PALM work) and any other noise from biological or instrumental sources (Qian et al., 1991; Saxton, 1997; Michalet, 2010). As has been shown the distribution of diffusion coefficients calculated from short tracks has a wide variance (Qian et al., 1991; Saxton, 1997), although even with broad distributions they have been used in previous work (Uphoff et al., 2013; Stracy et al., 2015). In addition, localisation error might be diffusion dependent, having proteins who have moved considerably in a frame time having

less well defined PSFs which are harder to fit and lead to a higher localisation error. Fitting the MSD and calculating the intercept in the y-axis, a method commonly used to calculate localisation error, can lead to imaginary localisations errors for short tracks. In most accounts the large variance is statistical and localisation error can be accounted for to study proteins (Uphoff et al., 2013; Stracy et al., 2015) but simulations can elucidate any other effects on D^* distribution. The only way to interpret the contribution of localisation error dependence and volume of the cell is to simulate the conditions of an experiment; first by simulating tracks in a confined volume and simulating the exposure of an EMCCD camera.

In previous work the best way of understanding tracking PALM results was by simulations. Initially Saxton (1997) looked at the effect of tracks and their length to the distribution of diffusion coefficients. Previous work performed simulations in the context of an *E. coli* cell and noted the spatial effects of confinement (English et al., 2011; Niu and Yu, 2008). I simulated transitions of different modes of diffusion in the cytoplasm of the cell. This a more adequate model since rarely will TFs have a single diffusion coefficient (Mueller et al., 2013). They may transition between specific binding, non-specific binding and free diffusion as has been shown before (Elf et al., 2007). Depending on the time-scale of the transitions it may be possible to analytically separate the different modes. However for LacI the transitions are on the time-scale shorter than our typical frame exposure (15 ms). To understand D^* distributions, it is necessary to simulate different modes and transition and analyse how the time-scales will affect the diffusion measurements. This will in turn aid future *in vivo* single-molecule work of diffusing TFs.

6.2 MATERIALS AND METHODS

To simulate the tracks I adapted previous simulation methods written for MATLAB and used to find the diffusion coefficient of Pol1 (Uphoff et al., 2013). Before this work a similar method had been implemented to analyse the stringent response machinery of *E. coli* (English et al., 2011) and the dynamics of the FtsZ cytoskeleton (Niu and Yu, 2008).

Our implementation mimics our experimental work: each ~ 15 ms step consists of any number of microsteps as necessary (Figure fig. 6.1A). Each microstep simulates the displacement of the molecule for a given diffusion coefficient. Afterwards the mean position for the displacements in number of microsteps probed from a Normal distribution centred at 0 with standard deviation $\sqrt{(2 * D * \Delta T / N_{steps})}$ (fig. 6.1B) is calculated and that represents the localisation for that frame. Where D is the current diffusion coefficient, ΔT the exposure of the frame and N_{steps} the number of microsteps. If the simulated molecule reaches an edge and the microstep takes the molecule out of the cytoplasm, that step is rejected. A new step is calculated until a step that doesn't cross the edge is found. No special compartments are simulated within the cell.

The novelty of our method is the addition of two-modes of diffusion to these microsteps. To simulate different modes of diffusion we probe from two diffusion coefficients for each microstep. By simulating the time being spent in each state (τ_{bound} , τ_{free}), we sample from an exponential distribution to calculate how many microsteps it will spend in that state. The fraction between the two time-scales is the time spent in a state.

$$f_{bound} = \frac{\tau_{bound}}{\tau_{bound} + \tau_{free}}; f_{free} = \frac{\tau_{free}}{\tau_{bound} + \tau_{free}} \quad (6.1)$$

For a 15 ms exposure of a simulated protein that spends on average 5 ms

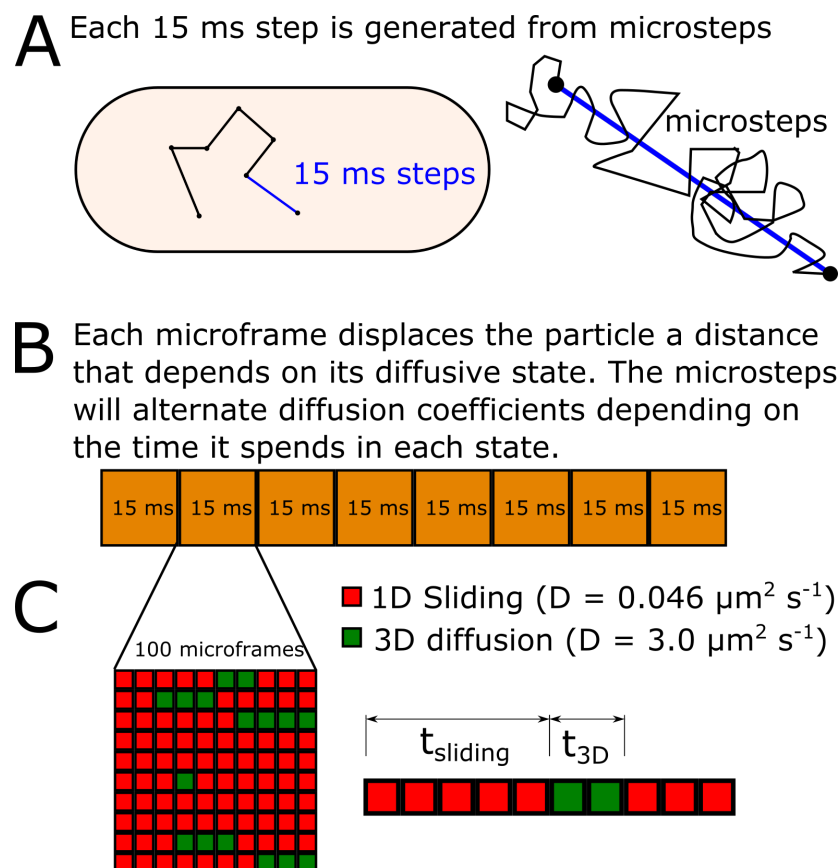


Figure 6.1: Simulating the diffusion of single-proteins in live cells. The figure shows a cartoon of the method implemented to simulate tracks with multiple modes of diffusion. A) Each step (exposure time) contains an amount of microsteps (number of steps per exposure frame) that can be changed depending on the simulated molecules. B) A typical tracking PALM molecule may be contained in a number of 15 ms frames. In our simulation, a frame consists of any number of microframe in which the simulation displaces the simulated molecule. C) Each one of the microframes can simulate a different diffusion coefficient to be probed from. In the example it can transition from 1D sliding and 3D diffusion continuously during a frame but can only have one diffusive state in each microframe. The average number of time it will remain bound or free state is determined by the inputs τ_{bound} and τ_{free} and is probed from an exponential distribution in each transition.

bound to DNA and 1 ms diffusing freely (cartoon example, probing from bound red square, probing from mobile green square; fig. 6.1C) we may find that one exposure may contain a molecule that transitions (bound to mobile; mobile to bound) between the two states. In our example (fig. 6.1) the molecule transitions 11 times and the time spent in each state depends on the given τ_{bound} and τ_{free} . In each microframe the simulation samples from the D related to its state and creates the microstep for that microframe.

To choose the initial diffusion state of the track a random number is generated. If the random number is smaller than f_{free} ($randnum < f_{free}$) then the starting state is mobile, otherwise it starts in the bound state. The initial position of a track is uniformly distributed. The bleaching time of each trajectory is taken from empirical data for our setup. To image similar conditions as the LacI, 40 trajectories were simulated per cell with 5000 frames. Afterwards all cell data was aggregated and its D^* distribution calculated as in our experimental tracking PALM work fig. 6.2. I simulated tracks with microframes of 75 μs (200 microframes per 15 ms) unless explicitly stated. After each position is generated, I added localisation noise by adding a sampled normally distributed with 44 nm (calibrated until D of $0 \mu m^2 s^{-1}$ produced D^* of $0.11 \mu m^2 s^{-1}$) . To account for the position noise we averaged the molecules position throughout an exposure cycle.

For our simulations different volumes can be probed. In our work the size of a single simulated cell is 2 μm for the long axis and 0.9 μm for the short axis (typical for the *E. coli* cells used in our experimental work (Uphoff et al., 2013; Stracy et al., 2015)). At the same time DNA-binding proteins mainly confine to the nucleoid volume, for this case we simulate a long-axis length of 1.35 μm and short-axis of 0.65 μm (Uphoff et al., 2013). For each simulation 200 cells

were simulated unless explicitly stated otherwise. Finally the noise is added the same way as for the tracks with one D.

6.3 RESULTS

6.3.1 Simulating one mode of diffusion

As a first test and to show the effect of confinement on the calculated D^* , I simulated a single diffusion coefficient. Using the first four steps of a simulated track with real D, we calculate the distribution of D^* just as in our tracking PALM work (fig. 6.2). We simulate a fast diffusing species of the size of LacI ($D = 3 \mu\text{m}^2 \text{s}^{-1}$) that leads to a broad and tailed distribution that follows the gamma distribution because it is a convolution of 4 exponentially, independently and identically distributed variables (fig. 6.2A, gamma distribution fit: blue dashed line). The fitted gamma distribution results in a D^* of $1.1 \mu\text{m}^2 \text{s}^{-1}$, the real D is underestimated by more than half its value. The high diffusion coefficient that was originally simulated, in this case is underestimated because of confinement.

We also simulate the case for any protein bound to DNA ($D = 0 \mu\text{m}^2 \text{s}^{-1}$, fig. 6.2B). For bound molecules, the distribution of D^* is narrow and tailed; a gamma distribution fit yields $D^* = 0.11 \mu\text{m}^2 \text{s}^{-1}$ which in this case overestimates the diffusion coefficient. Unlike fast molecules simulated beforehand, confinement has no effect on these tracks. Localization error contributes to the over-estimation of the diffusion coefficient. Simulations help us understand the effect of localization error to confinement and distribution width for tracking PALM distribution data.

With the previous two examples for a fast diffusing species and a bound

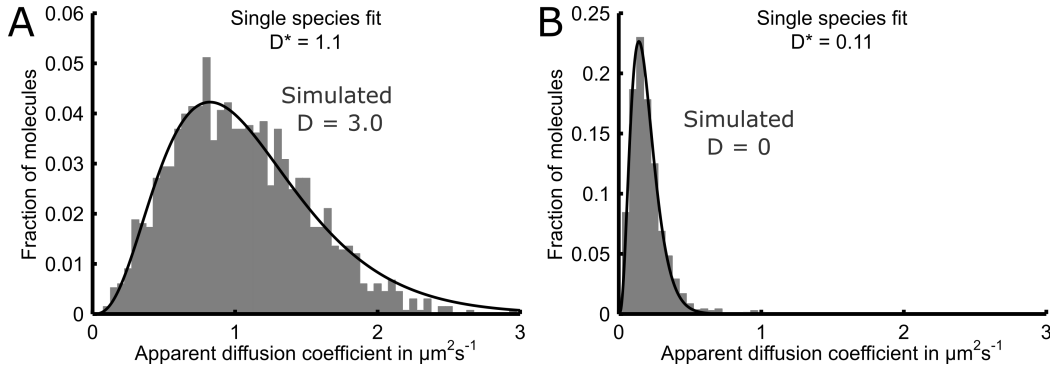


Figure 6.2: Simulating tracks with one diffusion coefficient. In A) I simulated tracks with a real diffusion coefficient $D = 3.0 \mu\text{m}^2 \text{s}^{-1}$ in the simulated volume of a cell. The resulting tracks follow the fitted gamma distribution for $D^* = 1.1 \mu\text{m}^2 \text{s}^{-1}$, which underestimated the diffusion coefficient due to cell confinement. In B) the simulated diffusion coefficient is $D = 0 \mu\text{m}^2 \text{s}^{-1}$. The resulting tracks follow the fitted gamma distribution for $D^* = 0.11 \mu\text{m}^2 \text{s}^{-1}$, which overestimates the diffusion coefficient due to localisation error.

one we show that we are able to simulate different modes of diffusion. We are able to scan through different values to find the relationship between the simulated D and the apparent diffusion that we calculate from our tracks. We simulate each D and we fit the D^* distribution; in the figure we report the D^* fit and the original simulated D fig. 6.3.

The relationship between D and D^* is not linear and starts plateauing for values higher than $2 \mu\text{m}^2 \text{s}^{-1}$. The simulations also show the threshold at which D^* goes from overestimating the simulated D to under estimating it (fig. 6.3, $D = 0.25 \mu\text{m}^2 \text{s}^{-1}$, dashed line intersection with data).

Our simulation also guides us on the limits of our experiment. Because of the non-linear relationship between D^* and D , the same changes in D have a better resolution for lower values of D . That is, for a change in D of $0.25 \mu\text{m}^2 \text{s}^{-1}$ in the whole cell simulations (from $0 \mu\text{m}^2 \text{s}^{-1}$ to $0.25 \mu\text{m}^2 \text{s}^{-1}$) leads to a change of $0.14 \mu\text{m}^2 \text{s}^{-1}$ in fitted D ; the same change from $2 \mu\text{m}^2 \text{s}^{-1}$ to $2.25 \mu\text{m}^2 \text{s}^{-1}$ in simulated D leads to a measured change of fitted D^* of

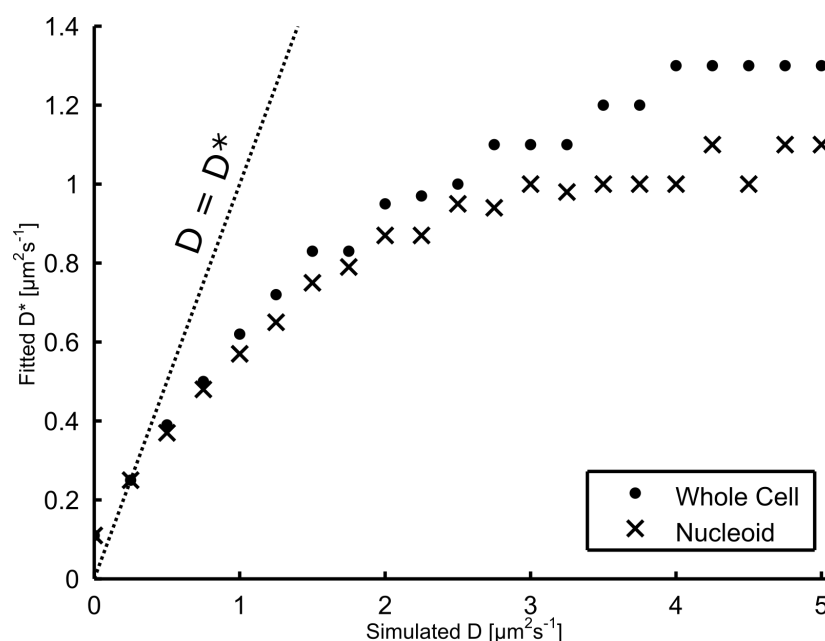


Figure 6.3: **The relationship between unbiased diffusion coefficient and the apparent diffusion coefficient.** In A) the fitted D^* is shown on the vertical axis versus the simulated D on the horizontal axis. The plot shows that the relationship between D and D^* is non-linear. Simulating the area of the whole cell versus the nucleoid shows the effect is related to the accessible volume for the protein. The case where the fitted D^* and the simulated D are equal is shown by the dashed line. It shows how the D^* over estimates D for values lower than around $D = 0.25 \mu\text{m}^2 \text{s}^{-1}$.

$0.02 \mu\text{m}^2 \text{s}^{-1}$. This case will change depending of the geometry of the cell volume, localisation error and exposure of the frame.

Simulating one mode of diffusion has helped us estimate the effects of localization error and confinement. After this we took the next step for this work, allowing a single simulated molecule to be able to diffuse with two diffusion coefficients, one for bound, one for free diffusion.

6.3.2 Simulating Two Modes of Diffusion

As their name suggests, DNA-binding proteins interact with DNA to function. It is still an open question if constant DNA-binding is necessary in the search

process or if they bind by only interacting with their sites of function and not to non-specific DNA (3D-capture) (Elf et al., 2007; Marklund et al., 2013; Mueller et al., 2013). Ideally we would have a consensus of the interaction of every protein, but up until now only a small amount of proteins have been tested and many of these in eukaryotic cells (Mueller et al., 2013). DNA Polymerase I diffuses fast, if it interacts constantly with DNA in its search process it does so at a much smaller time-scale than the exposure of 15 ms (Uphoff et al., 2013).

Most TFs, as a common characteristic, constantly interact with DNA for their search process (Mueller et al., 2013). It is important to include the constant DNA binding to the simulations if further tracking PALM data of transcription factors is to be interpreted. We thus simulated proteins with two modes of diffusion and at two different scales of interaction. We simulate molecules that are bound to DNA with a $D_{bound} = 0 \mu\text{m}^2 \text{s}^{-1}$ and freely diffusing in the cytoplasm $D_{free} = 3.0 \mu\text{m}^2 \text{s}^{-1}$; these are the two different modes we simulated previously when only one-diffusion coefficient was used (section 6.3.1). We simulated transitions at much longer time scales than the exposure times (no transitions on an average length track) and at much smaller scales (many transitions on an average length track).

We expect that in the case where the transitions are at longer time scales than the exposure times, they will remain in either the bound or free state throughout the frame ($\tau_{bound} + \tau_{mobile} \ll t_{exposure}$). In this case we will be able to observe two distinguishable populations in our D^* distribution when different fractions of mobile and bound are simulated. We confirm from simulations that this the case (6.4A); when our molecules are freely diffusing 20 % of the time (6.4A, light red line) there is a 20 % increase in the broader D^* distribution with a mean of $D^* = 1.1 \mu\text{m}^2 \text{s}^{-1}$ from the case were 0 % are mobile (fig. 6.4A, dark

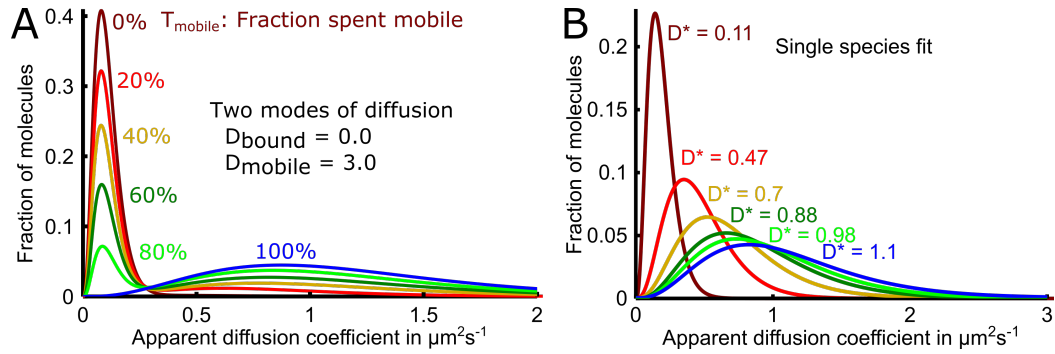


Figure 6.4: Changing the ratio of interaction between mobile and binding. How the changes in the time the molecule spends in free diffusion $\tau_{\text{free}}/(\tau_{\text{bound}} + \tau_{\text{mobile}})$ and keeping $\tau_{\text{bound}} + \tau_{\text{mobile}} = \text{Const}$ was simulated. In A) $\tau_{\text{bound}} + \tau_{\text{mobile}} \gg t_{\text{exposure}}$, so that if a molecule is found in a state it will probably stay in the same state for the whole observation period (average ~ 3 frames). The curves represent a two species fit to $D^* = 0.11 \mu\text{m}^2 \text{s}^{-1}$ and $D^* = 1.1 \mu\text{m}^2 \text{s}^{-1}$. In B) $\tau_{\text{bound}} + \tau_{\text{mobile}} \ll t_{\text{exposure}}$, that is the molecule will transition multiple times from the bound to the free state leading to an averaging of the D^* . In this case a single-species was able to capture the distribution and are shown for each curve.

red line). This happens at each 20 % increase until all molecules are in the free state (fig. 6.4A, blue line). We observe two distinguishable populations, similar to the case of Pol I were it spends a large amount of time freely diffusing as opposed to DNA-bound repair of damaged sites (Uphoff et al., 2013).

When the transitions are at scales lower than the exposure times ($\tau_{\text{bound}} + \tau_{\text{mobile}} \ll t_{\text{exposure}}$) the D^* distribution contrasts to that where the time is large. In one observation period the simulated molecule transitions from the bound to mobile state multiple times; in this case we expect a large amount of averaging in the D^* . Our simulations show how at these fast time-scales the D^* distribution changes (6.4B). To understand the D^* we have to find the effective diffusion found from averaging of the free diffusion and DNA-binding ($D_{\text{eff}} = f_{\text{free}}D_{\text{free}}$). After calculating the effective diffusion, we have to convert this value to the D^* following our previously shown relationship (fig. 6.3).

Following the same for D^* , that is that $D^* = f_{free}D_{free}^*$ yields incorrect values for the D^* because of the non-linear relationship of increasing diffusion (fig. 6.3). For example, in the case of $f_{free} = 20\%$ and $D^* = 1.1\mu\text{m}^2\text{s}^{-1}$ yields $D_{eff}^* = 20\%(1.1\mu\text{m}^2\text{s}^{-1}) = 0.22\mu\text{m}^2\text{s}^{-1}$. In reality $D^* = 0.47\mu\text{m}^2\text{s}^{-1}$ which is the converted value from $D_{eff} = 20\%(3.0\mu\text{m}^2\text{s}^{-1}) = 0.6\mu\text{m}^2\text{s}^{-1}$. This is important because any usage of D^* will yield incorrect results. All D^* need to be converted to real D to calculate any fractions of interaction from averaging of two diffusion coefficients. Unless you have a separate measurement of D_{free} , calculating f_{free} would not be possible. Recently new strategies to use tracking PALM have been developed by using an *E. coli* with minimal DNA and long DNA-free endcaps where it is possible to track any protein in the absence of DNA and extract D_{free} (Stracy et al., 2015).

6.3.3 Simulating the *lac* repressor

Our next objective is to simulate a real TF using our data from LacI. LacI has around 40 molecules per cell and exhibits facilitated diffusion, the combination of DNA-sliding and 3-D diffusion to find a target (Elf et al., 2007; Marklund et al., 2013). If it is possible to simulate two-modes of diffusion in the cellular volume we should be able to recreate the D^* distribution of LacI. The quantities needed to simulate a TF are the D_{free} , D_{bound} , how long it spends in each state (τ_{bound} , τ_{free}). Fortunately this information on how long LacI spends sliding on DNA and how long it spends freely diffusive have been published from single-molecule data and studied in previous tracking PALM work (Elf et al., 2007).

First I use the published work numbers to simulate LacI (Elf et al., 2007; Hammar et al., 2012; Marklund et al., 2013). From the single molecule work we find the diffusion coefficients for each state are $D_{free} = 3.0\mu\text{m}^2\text{s}^{-1}$, and in this

case when LacI is bound to DNA it slides with known $D_{sliding} = 0.045 \mu\text{m}^2 \text{s}^{-1}$. We also know that non-specific DNA binding is on the ~ 5 ms time-scale and that LacI spends 87 % of the time bound to non-specific DNA.

If we compare with the published values in our simulation we obtain a D^* distribution that differs from the LacI over-expressed from a plasmid (fig. 6.5A, simulation black line). The mean of the simulated molecule is closer to $D^* = 0.4 \mu\text{m}^2 \text{s}^{-1}$ and our molecules seem, on average more mobile than this. To obtain the best simulation we converted the $D^* = 0.5 \mu\text{m}^2 \text{s}^{-1}$ (mean of the fast population) to D value using fig. 6.3 for the distribution to a D_{eff} from our data fig. 6.3 and discussed in section 6.3.2. The D of $D^* = 0.5 \mu\text{m}^2 \text{s}^{-1}$ is around $0.75 \mu\text{m}^2 \text{s}^{-1}$; with this D_{eff} can be approximated as $f_{free} * D_{free}$ and the fraction of time spent mobile is 75 %, lower than the previously reported value. With this corrected fraction the simulation yields a closer distribution compared with the data (fig. 6.5, red line).

This yields the underlying interaction, according to our data and confirmed from our simulations, we observe that LacI interacts 75 % of the time with DNA. When we out compete 10 mM IPTG with 1 mM ONPG the over-expressed changes its diffusion distribution (fig. 6.5B). To be able to recreate LacI with ONPG we add a fraction of molecules that will remain bound throughout the simulated-exposed frame ($D = 0$). These molecules represent specifically bound LacI or near-specifically bound LacI (binding to near-operator sites), as these last molecules have been observed to remain on DNA for upto 10 s (Marklund et al., 2013).

We find that LacI over-expressed from a plasmid and treated with ONPG has 18 % of molecules that are strongly bound to DNA (fig. 6.5B). These include LacI bound to the natural operators in the nucleoid, the operators present

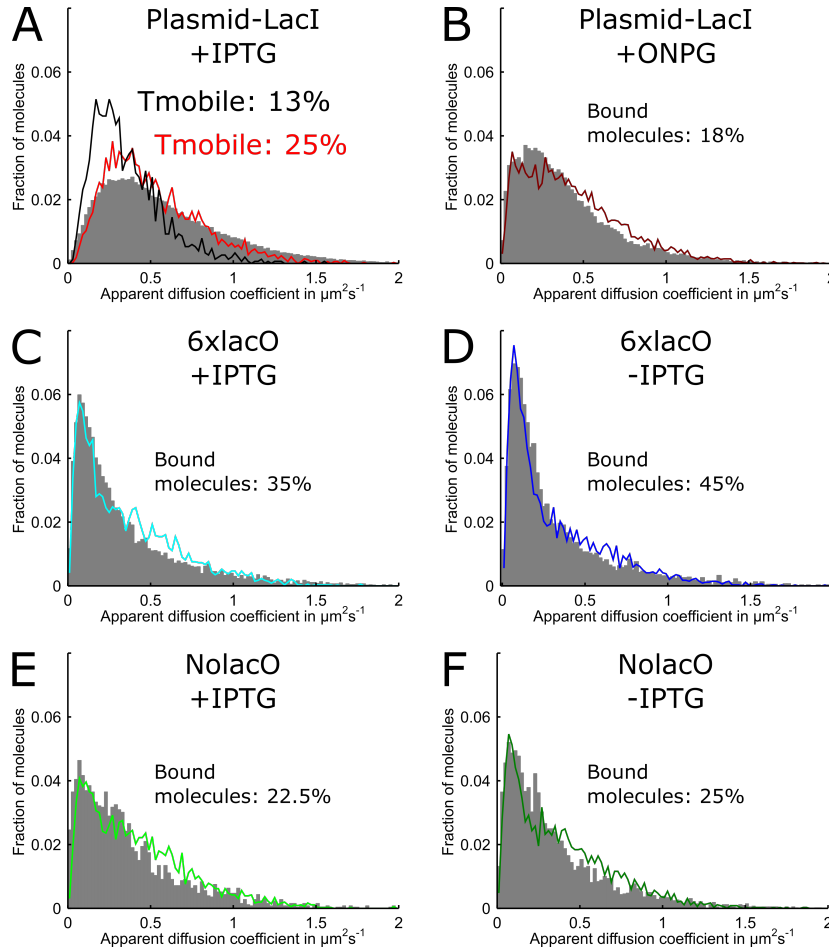


Figure 6.5: **Simulating LacI D* distribution** A) shows the LacI expressed from a plasmid. Using previous work data for DNA-binding times for LacI (5 ms) with IPTG and the fraction spent mobile (13%, Elf et al. (2007)), the simulated D* distribution deviates (black line) from the data. We measure a different fraction, by best fit, between bound and mobile ($f_{free} = 25\%$) and simulate the interaction to obtain a more accurate model (red line). We use this interaction to simulate all cases; in B) to account for the LacI that can bind the operators in the plasmid or partially-specific interactions (Marklund et al., 2013), we find 18% of molecules to be strongly bound to DNA. In C) the number of strongly bound molecules had to be increased to 35% for IPTG treated cells whilst without IPTG in D) it increases to 45%. In E) without any operators there are still 25% of molecules that look strongly bound whilst adding IPTG, which increases binding to near specific sites only yields a 2.5% increase

in the plasmid and near-specific binding.

For our strain with the genomic insertion of 6 tandem *lac* operators with only natively expressed LacI (6xlacO), we observe a higher number of molecules that strongly bound even in IPTG. For 6xlacO treated with 1 mM IPTG, we find 35 % of bound molecules to account for the D^* distribution. With IPTG, LacI cannot bind to any specific or near-specific sites on DNA, but background may cause for a fraction of tracks to be bound which we found to be more prevalent in IPTG treated cells. Without IPTG the number of bound molecules increases to 45 % due to a fraction of LacI binding to the tandem operators. For our strain with no operators and LacI expressed natively we observe around the same number of strongly bound molecules for either IPTG treated and untreated cells. They both differ by at least 10 % lower specifically-bound molecules from the IPTG treated 6xlacO sample.

6.4 DISCUSSION

Single-molecule simulations of diffusion in the cell cytoplasm are a necessary tool to further understand D^* distribution resulting from tracking PALM work. They are able to show if a molecule is purely bound, freely diffusing or a combination of both. We showed that we can measure how the diffusion coefficients will be either under-estimated due to confinement in the cytoplasm or over-estimated from localization error. By simulating one mode of diffusion, it has been possible to convert D^* measurements into reportable diffusion coefficients (Uphoff et al., 2013; Stracy et al., 2015). To add to this work we have shown simulations that include two possible modes of diffusion, which dramatically improve real diffusion coefficient estimation in TFs.

We also found that the D^* to D relationship is not linear and will depend on the PSFs exposure and localisation error. The relationship plateaus over $D = 2 \mu\text{m}^2 \text{s}^{-1}$, large changes in D do not yield changes in D^* for a typical *E. coli* size. If a sample contains two different fast states, at this exposure we are unable to distinguish them from each other. Although for DNA-binding proteins there isn't a reason to have two modes of fast diffusion, there is still a question whether proteins exist as monomers, dimers, tetramers in solution and at which fractions, our current tracking PALM protocol would be unable to distinguish them. It may be possible to lower the exposure time to better measure changes in fast diffusing molecules, but low D resolution with fast molecules may be due to cellular confinement. Also shorter exposure times will affect localisation errors from the number of photons collected because of limits in the emission of FPs (Michalet, 2010). In our case 15 ms is the fastest our EMCCD camera can image at full frame and has been helpful to identify function of other DNA-binding proteins (Uphoff et al., 2013; Stracy et al., 2015).

Each experiment with varying cell volumes, exposures and fluorescent probe will have a different D -vs- D^* curve so it is suggested to simulate an experiment to tailor for any specific molecule of interest.

Another interesting result resulting from the non-linear relationship between D^* and D is their effect on the calculation of fractions of interactions. To obtain accurate fractions of interaction for TFs, it is necessary to convert D^* numbers to real D estimations (section 6.3.2). This is crucial since these *in vivo* numbers are used to estimate DNA-binding times and target-search times. By converting D^* to D and applying it to our work in LacI, we obtained that when the repressor searches for specific site, it is interacting with DNA 75 % of the time.

From previous work (Elf et al., 2007; Marklund et al., 2013) and our LacI work we already had established parameters, $D_{sliding}$, D_{free} , τ_{bound} , τ_{free} , f_{free} , required for the simulation. We could build from previous work and narrow the parameters to our experimental data. It may not be possible to use simulations to find the optimal numbers for any TF but it is possible to show which are the plausible ranges given a set of tracking PALM data.

6.4.1 Conclusions

Simulations are an important part in the analysis of tracks of mobile FPs. Simulating two-modes of diffusion was crucial to understand the relationship of f_{free} , D and D^* .

Further work in the field would be to characterise how the nucleoid and the nucleoid organization proteins behave as an obstacle to diffusion, just as how we studied confinement in the cytoplasm. In addition a more realistic construction of the nucleoid would help understand its effect, at the moment it is only simulated as a compartment due to current limitations on knowledge of the structure. Joining blinking (chapter 4) and the diffusion would also lead to further understanding on the effect of blinking to diffusion estimates.

CONCLUSIONS

I have presented, in this thesis, an argument for studying TFs at the single-molecule level. I showed that tracking PALM is one of the best methods to observe the number, spatial and diffusive distributions of DNA-binding proteins.

We also present the limitations of tracking PALM. We attempted to characterise some of the issues that still arise in an experiment. Firstly we characterised our experiment's photo-physical behaviour of PAmCherry in a way that can be replicated for other fluorescent proteins. I extracted rates at which PAmCherry transitioned to dark states, recovered and bleach. This allowed us to simulate our experiment and obtain an estimate of the error in counting proteins. We hope that this additional approach helps to understand and characterise one of the main issues in tracking PALM.

After this I showed how tracking PALM is used to study TFs. Because of their low copy numbers, TFs may be prone to background molecules having a significant effect on the interpretation of the tracking data. I showed that

this background is significant at this level, and I characterised the effect when imaging LacI. We were able to extract diffusion information on LacI that may give more insight into its diffusion. We also showed that tracking PALM can be applied to other TFs by imaging AraC and CRP.

Other effects in tracking PALM work is that molecules are confined to the cytoplasm of the cell. There is no analytic expression to account for this so simulations is the best method to understand its effect. In addition I simulated TFs which have a search process with two modes of diffusion, sliding on DNA and free diffusion. Simulating two-modes of diffusion was crucial part in understanding the relationship between the two modes and the extracted data D^* and D .

I believe tracking PALM will remain relevant to image live-cell and real-time function. The discovery of better PAFP will keep the technique relevant and will make it a standard microscopy method for other cellular organisms. PAm-Cherry is limited and may cause artefacts that need to be considered, other PAFPs need to be tested and compared. At the same time, tracking proteins tagged with fluorescent dyes has a number of advantages because of their long-lived fluorescence. Techniques pushing for insertion of tagged proteins will improve and may remove some relevance to tracking PALM.

I hope that this work gives further insight into tracking PALM, imaging transcription factor and the analysis of tracks arising from single-molecules in live cells.

REFERENCES

- ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K., AND WALTER, P. *Molecular Biology of the Cell*. Garland Science, 5th edition, November 2008. ISBN 0815341059.
- ANNIBALE, P., VANNI, S., SCARSELLI, M., ROTH LISBERGER, U., AND RADENOVIC, A. Quantitative Photo Activated Localization Microscopy: Unraveling the effects of photoblinking. *PLoS ONE*, 6(7), 2011. ISSN 19326203.
- BARYSCH, S. V., JAHN, R., AND RIZZOLI, S. O. A fluorescence-based in vitro assay for investigating early endosome dynamics. *Nature Protocols*, 5(6):1127–1137, 2010. ISSN 1754-2189.
- BETZIG, E., PATTERSON, G. H., SOUGRAT, R., LINDWASSER, O. W., OLENYCH, S., BONIFACINO, J. S., DAVIDSON, M. W., LIPPINCOTT-SCHWARTZ, J., AND HESS, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science (New York, N.Y.)*, 313(5793):1642–5, September 2006. ISSN 1095-9203.
- COFFMAN, V. C. AND WU, J.-Q. Every laboratory with a fluorescence microscope should consider counting molecules. *Molecular biology of the cell*, 25(10):1545–8, May 2014. ISSN 1939-4586.
- CRAWFORD, R., TORELLA, J. P., AIGRAIN, L., PLOCHOWIETZ, A., GRYTE, K., UPHOFF, S., AND KAPANIDIS, A. N. Long-Lived Intracellular Single-Molecule Fluorescence Using Electroporated Molecules. *Biophysical Journal*, 105(11):2439–2450, 2013. ISSN 0006-3495.
- DURISIC, N. AND LAPARRA-CUERVO, L. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nature Methods*, 11(2), 2014.
- ELF, J., LI, G.-W., AND XIE, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science (New York, N.Y.)*, 316(5828):1191–4, May 2007. ISSN 1095-9203.

- ENGLISH, B. P., HAURYLIUK, V., SANAMRAD, A., TANKOV, S., DEKKER, N. H., AND ELF, J. Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108(31):E365–73, August 2011. ISSN 1091-6490.
- FUNATSU, T., HARADA, Y., TOKUNAGA, M., SAITO, K., AND YANAGIDA, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution., 1995. ISSN 0028-0836.
- GAHLMANN, A. AND MOERNER, W. E. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nature reviews. Microbiology*, 12(1):9–22, December 2013. ISSN 1740-1534.
- GILBERT, W. AND MULLER-HILL, B. Isolation of the Lac Repressor. *Proceedings of the National Academy of Sciences*, 56(6):1891–1898, 1966. ISSN 0027-8424.
- HAMMAR, P., LEROY, P., AND MAHMUTOVIC, A. The lac Repressor Displays Facilitated Diffusion in Living Cells. *Science*, (June):1595–1598, 2012.
- HAMMAR, P., WALLDÉN, M., FANGE, D., PERSSON, F., BALTEKIN, O., ULLMAN, G., LEROY, P., AND ELF, J. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nature genetics*, 46(4):405–8, 2014. ISSN 1546-1718.
- HENSEL, Z., WENG, X., LAGDA, A. C., AND XIAO, J. Transcription-factor-mediated DNA looping probed by high-resolution, single-molecule imaging in live *E. coli* cells. *PLoS biology*, 11(6):e1001591, June 2013. ISSN 1545-7885.
- HENSEL, Z. AND XIAO, J. Single-molecule methods for studying gene regulation in vivo. *Pflügers Archiv : European journal of physiology*, 465(3):383–95, March 2013. ISSN 1432-2013.
- HOLDEN, S. J., UPHOFF, S., HOHLBEIN, J., YADIN, D., LE RESTE, L., BRITTON, O. J., AND KAPANIDIS, A. N. Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophysical Journal*, 99(9):3102–11, 2010. ISSN 1542-0086.
- HOLDEN, S. J., UPHOFF, S., AND KAPANIDIS, A. N. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature methods*, 8(4):279–280, 2011. ISSN 1548-7091.

- HUANG, B., BABCOCK, H., AND ZHUANG, X. Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *CELL*, 143(7):1047–1058, December 2010. ISSN 0092-8674.
- HUANG, B., WANG, W., BATES, M., AND ZHUANG, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (New York, N.Y.)*, 319(5864):810–3, February 2008. ISSN 1095-9203.
- JENNISON, M. H. AND A.F. MORGAN. Fluorescence microscopy. *Tubercle*, 31(4):84–87, 1950. ISSN 00413879.
- KAO, H. P. AND VERKMAN, A. S. Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position. *Biophysical journal*, 67(3):1291–1300, 1994. ISSN 00063495.
- KOLODRUBETZ, D. AND SCHLEIF, R. Identification of araC protein and two-dimensional gels, its in vivo instability and normal level. *Journal of Molecular Biology*, 149(1):133–139, 1981. ISSN 0022-2836.
- KUHLMAN, T. E. AND COX, E. C. Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Molecular systems biology*, 8(610):610, September 2012. ISSN 1744-4292.
- KUHLMAN, T. E. AND COX, E. C. DNA-binding-protein inhomogeneity in E. coli modeled as biphasic facilitated diffusion. *Physical Review E*, 88(2):022701, August 2013. ISSN 1539-3755.
- LAU, I. F., FILIPE, S. R., SØ BALLE, B., Ø KSTAD, O.-A., BARRE, F.-X., AND SHERRATT, D. J. Spatial and temporal organization of replicating Escherichia coli chromosomes. *Molecular Microbiology*, 49(3):731–743, January 2004. ISSN 0950382X.
- LEAKE, M. C., CHANDLER, J. H., WADHAMS, G. H., BAI, F., BERRY, R. M., AND ARMITAGE, J. P. Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature*, 443(7109):355–358, 2006. ISSN 0028-0836.
- LEE, D. J., BINGLE, L. E. H., HEURLIER, K., PALLAN, M. J., PENN, C. W., BUSBY, S. J. W., AND HOBMAN, J. L. Gene doctoring: a method for recombineering in laboratory and pathogenic Escherichia coli strains. *BMC microbiology*, 9:252, January 2009. ISSN 1471-2180.
- LEE, S.-H. S.-H., SHIN, J. Y., LEE, A., AND BUSTAMANTE, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proceedings of the National Academy of Sciences of the United States of America*, 109(43):2–7, October 2012. ISSN 0027-8424.

- LEWIS, M. The lac repressor. *Comptes rendus biologiques*, 328(6):521–48, June 2005. ISSN 1631-0691.
- LI, G.-W. AND ELF, J. Single molecule approaches to transcription factor kinetics in living cells. *FEBS letters*, 583(24):3979–83, December 2009. ISSN 1873-3468.
- LIU, Z., LAVIS, L. D., AND BETZIG, E. Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. *Molecular cell*, 58(4):644–659, 2015. ISSN 1097-4164.
- MANLEY, S., GILLETTE, J. M., PATTERSON, G. H., SHROFF, H., HESS, H. F., BETZIG, E., AND LIPPINCOTT-SCHWARTZ, J. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature methods*, 5(2):155–157, 2008. ISSN 1548-7091.
- MARKLUND, E. G., MAHMUTOVIC, A., BERG, O. G., HAMMAR, P., VAN DER SPOEL, D., FANGE, D., AND ELF, J. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49):19796–801, 2013. ISSN 1091-6490.
- METCALF, D. J., EDWARDS, R., KUMARSWAMI, N., AND KNIGHT, A. E. Test Samples for Optimizing STORM Super-Resolution Microscopy. *Journal of Visualized Experiments*, (79):1–17, 2013. ISSN 1940-087X.
- MICHALET, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E*, 82(4):041914, October 2010. ISSN 1539-3755.
- MLODZIANOSKI, M. J., JUETTE, M. F., BEANE, G. L., AND BEWERSDORF, J. Experimental characterization of 3D localization techniques for particle-tracking and super-resolution microscopy. *Optics Express*, 17(10):8264–8277, May 2009. ISSN 1094-4087.
- MUELLER, F., STASEVICH, T. J., MAZZA, D., AND McNALLY, J. G. Quantifying transcription factor kinetics: at work or at play? *Critical reviews in biochemistry and molecular biology*, 48(5):492–514, 2013. ISSN 1549-7798.
- NIU, L. AND YU, J. Investigating Intracellular Dynamics of FtsZ Cytoskeleton with Photoactivation Single-Molecule Tracking. *Biophysical Journal*, 95(4):2009–2016, 2008. ISSN 00063495.

- PLOCHOWIETZ, A., CRAWFORD, R., AND KAPANIDIS, A. N. Characterization of organic fluorophores for in vivo FRET studies based on electroporated molecules. *Physical chemistry chemical physics : PCCP*, pages 1–3, 2014. ISSN 1463-9084.
- PRABHAT, P., RAM, S., SALLY WARD, E., AND OBER, R. J. Simultaneous imaging of different focal planes in fluorescence microscopy for the study of cellular dynamics in three dimensions. *IEEE Transactions on Nanobioscience*, 3(4):237–242, 2004. ISSN 15361241.
- PULKKINEN, O. AND METZLER, R. Distance Matters: The Impact of Gene Proximity in Bacterial Gene Regulation. *Physical Review Letters*, 110(19):198101, May 2013. ISSN 0031-9007.
- QIAN, H., SHEETZ, M. P., AND ELSON, E. L. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical journal*, 60(4):910–21, 1991. ISSN 0006-3495.
- RECK-PETERSON, S. L., DERR, N. D., AND STUURMAN, N. Imaging single molecules using total internal reflection fluorescence microscopy (TIRFM). *Cold Spring Harbor protocols*, 2010(3):pdb.top73, March 2010. ISSN 1559-6095.
- REDDING, S. AND GREENE, E. C. How do proteins locate specific targets in DNA? *Chemical Physics Letters*, 570:1–11, May 2013. ISSN 00092614.
- REYES-LAMOTHE, R., SHERRATT, D. J., AND LEAKE, M. C. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science (New York, N.Y.)*, 328(5977):498–501, April 2010. ISSN 1095-9203.
- REYES-LAMOTHE, R., WANG, X., AND SHERRATT, D. Escherichia coli and its chromosome. *Trends in microbiology*, 16(5):238–45, May 2008. ISSN 0966-842X.
- RIGGS, A., BOURGEOIS, S., AND COHN, M. The lac repressor-operator interaction: III. Kinetic studies. *Journal of molecular biology*, pages 401–417, 1970.
- ROBINETT, C., STRAIGHT, A., AND LI, G. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *The Journal of cell ...*, 135(6), 1996.
- SAXTON, M. J. Single-particle tracking: the distribution of diffusion coefficients. *Biophysical journal*, 72(4):1744–53, April 1997. ISSN 0006-3495.

- SAXTON, M. J. AND JACOBSON, K. Single-particle tracking: applications to membrane dynamics. *Annual review of biophysics and biomolecular structure*, 26:373–399, 1997. ISSN 1056-8700.
- SCHLEIF, R. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS microbiology reviews*, 34(5):779–96, September 2010. ISSN 1574-6976.
- SCHMIDT, T., SCHÜTZ, G. J., BAUMGARTNER, W., GRUBER, H. J., AND SCHINDLER, H. Imaging of single molecule diffusion. *Proceedings of the National Academy of Sciences*, 93(7):2926–2929, 1996. ISSN 0027-8424.
- SENGUPTA, P., JOVANOVIĆ-TALISMAN, T., AND LIPPINCOTT-SCHWARTZ, J. Quantifying spatial organization in point-localization superresolution images using pair correlation analysis. *Nature protocols*, 8(2):345–54, February 2013. ISSN 1750-2799.
- SENGUPTA, P., JOVANOVIĆ-TALISMAN, T., SKOKO, D., RENZ, M., VEATCH, S. L., AND LIPPINCOTT-SCHWARTZ, J. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nature methods*, 8(11), 2011.
- SLIUSARENKO, O., HEINRITZ, J., EMONET, T., AND JACOBS-WAGNER, C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *MOLECULAR MICROBIOLOGY*, 80(3):612–627, May 2011. ISSN 0950-382X.
- SPEIL, J., BAUMGART, E., SIEBRASSE, J. P., VEITH, R., VINKEMEIER, U., AND KUBITSCHKE, U. Activated STAT1 transcription factors conduct distinct saltatory movements in the cell nucleus. *Biophysical Journal*, 101(11):2592–2600, 2011. ISSN 00063495.
- STRACY, M., LESTERLIN, C., GARZA DE LEON, F., UPHOFF, S., ZAWADZKI, P., AND KAPANIDIS, A. N. Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences of the United States of America*, page 201507592, 2015. ISSN 0027-8424.
- STRACY, M., UPHOFF, S., GARZA DE LEON, F., AND KAPANIDIS, A. N. In vivo single-molecule imaging of bacterial DNA replication, transcription, and repair. *FEBS letters*, 588(19):3585–94, 2014. ISSN 1873-3468.

- STRINGER, A. M., CURRENTI, S., BONOCORA, R. P., BARANOWSKI, C., PETRONE, B. L., PALUMBO, M. J., REILLY, A. A., ZHANG, Z., ERILL, I., AND WADE, J. T. Genome-Scale Analyses of *Escherichia coli* and *Salmonella enterica* AraC Reveal Noncanonical Targets and an Expanded Core Regulon. *Journal of Bacteriology*, 196(3):660–671, 2014. ISSN 0021-9193.
- SUBACH, F. V., PATTERSON, G. H., MANLEY, S., GILLETTE, J. M., LIPPINCOTT-SCHWARTZ, J., AND VERKHUSHA, V. V. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nature Methods*, 6(2):153–160, 2009.
- TOKUNAGA, M., IMAMOTO, N., AND SAKATA-SOGAWA, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature methods*, 5(2):159–161, 2008. ISSN 1548-7091.
- UPHOFF, S., HOLDEN, S. J., LE RESTE, L., PERIZ, J., VAN DE LINDE, S., HEILEMANN, M., AND KAPANIDIS, A. N. Monitoring multiple distances within a single molecule using switchable FRET. *Nature methods*, 7(10):831–836, 2010. ISSN 1548-7091.
- UPHOFF, S., REYES-LAMOTHE, R., GARZA DE LEON, F., SHERRATT, D. J., AND KAPANIDIS, A. N. Single-molecule DNA repair in live bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20):8063–8, May 2013. ISSN 1091-6490.
- VEATCH, S. L., MACHTA, B. B., SHELBY, S. A., CHIANG, E. N., HOLOWKA, D. A., AND BAIRD, B. A. Correlation functions quantify super-resolution images and estimate apparent clustering due to overcounting. *PloS one*, 7(2):e31457, January 2012. ISSN 1932-6203.
- WALTER, N. G., HUANG, C.-Y., MANZO, A. J., AND SOBHY, M. A. Do-it-yourself guide: how to use the modern single-molecule toolkit. *NATURE METHODS*, 5(6):475–489, June 2008. ISSN 1548-7091.
- WANG, A. C., REVZIN, A., BUTLER, A. P., AND VON HIPPEL, P. H. Binding of *E.coli* lac repressor to non-operator DNA. *Nucleic acids research*, 4(5):1579–93, 1977. ISSN 0305-1048.
- WANG, S., MOFFITT, J. R., DEMPSEY, G. T., XIE, X. S., AND ZHUANG, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8452–7, June 2014. ISSN 1091-6490.

WANG, W., LI, G.-W. G.-W., CHEN, C., XIE, X. S., AND ZHUANG, X. Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. *Science*, 333(6048):1445–1449, September 2011. ISSN 0036-8075.

WANG, Y. M., AUSTIN, R. H., AND COX, E. C. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Physical Review Letters*, 97(4):048302, July 2006. ISSN 0031-9007.

WANG, Y. M., TEGENFELDT, J. O., REISNER, W., RIEHN, R., GUAN, X.-J., GUO, L., GOLDING, I., COX, E. C., STURM, J., AND AUSTIN, R. H. Single-molecule studies of repressor-DNA interactions show long-range interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(28):9796–9801, 2005. ISSN 0027-8424.

WIKIMEDIA-COMMONS. Lac operon, 2010. File: Lac operon-2010-21-01.png.

STRAIN INFORMATION

A.1 OTHER METHODS

A.1.1 ssDNA Cy3b experiments

Labelled DNA provided by Stephan Uphoff from previously published work (Uphoff et al., 2010).

To bind the DNA to the slide surface we used wells with a burnt slide and used 30 μL of 10 mg mL^{-1} Biotin-BSA for 5 mins. Washed with PBS twice and place 20 μL of 0.5 mg mL^{-1} Neutravidin for 2mins and washed with PBS three times. Diluted the Cy3b labelled ssDNA to 100 pM and pipetted 15 μL into the well and let it rest for 1 minute then washed and filled with PBS.

The ssDNA used is the following:

5'-CAG ACG GCT GGC ACG ACT TX T GCA CGA TCA GCC CTG GGC
GAG CGC ATT GTT GCG C-Biotin

modified at the 3'-end with Biotin and at position 20 (X) with Cy3b.

A.2 STRAIN CONSTRUCTION

Strains were constructed as published in (Lee et al., 2009) and provided by the Steve Busby group in Birmingham. 6 LacI operator DNA sites (5'-AATTGTGAGCGGATAACAATT-3') were inserted adjacent to the araBAD promoter in the strain.

A.2.0.1 6x_{lacO} operator sequence

6 X LacI Binding Site

```
GAATTCGGATCCCTCGAGCCAGCACGTAGCTAGCAGAGAGTAAGGAATTGT-  
GAGCGGATAACAATTAAAAGTCAGAAATTGTGAGCG-  
GATAACAATTAAAGACGGAGAAATTGTGAGCGGATAA-  
CAATTAATAAAGCAGCTCTAGCATCAACGCGCAAATTGTGAGCGGATAA-  
CAATTAGCAAGACACAATTGTGAGCGGATAACAATTCTGGATAGGCAATTGT-  
GAGCGGATAACAATTCACAGCGCCCCTCTAGAGTCGACAGATCTAAGCTT
```


A.2.0.2 6x_{lac}O operator sequence + 20x Mall operator sequence**10 X Mall Binding site 6 X *LacI* Binding site**

GTCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGTCCTGATAAACGTTTATCACGTGTCGAGCATGGATAAACGTT-
 TATCGCTAGAGGGGGCGCTGTGAATTGTTATCCGCTCA-
 CAATTGCCTATCCAGAATTGTTATCCGCTCACAAATTGTGTCTTGCTAATTGT-
 TATCCGCTCACAAATTGCGCGTTGATGCTAGAGCTGCTTTATTAAATTGT-
 TATCCGCTCACAAATTCTCCGTCTTTAAATTGTTATCCGCTCA-
 CAATTCTGACTTTTAAATTGTTATCCGCTCACAAATTCCTTACTCTCTGCTAGC

A.3 PLASMID CONSTRUCTION

A.3.0.3 *LacI*Mut strain construction

LacI full length

```
GTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTT
ATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGA
AAACGCGGGGAAAAAGTGAAGCGGCGATGGCGGAGCTGAATTACATTC
CCAACCGCGTGGCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTG
GCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCTG
CGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGT
CGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACA
ATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGA
TGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGC
GTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTC
CCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTTCGCATTGGG
TCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCG
CGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTC
AGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTTC
AACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCT
GGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGA
GTCCGGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGA
TACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAG
GATTTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTC
TCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTG
GTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGA
CTGGAAAGCGGGCAGTGA
```

ForLacI-41KpnI GTCAGGTACC AATGGCGGAGCTGAATTACATT

RevLacIBamHI GTGCGGATCCCTGCCCGCTTTCCAGTCGGG

To be merged with PAmcherry with a linker from pROD85 (or 82) ForpamcherryBamHI GTGCGGATCC TCG GCT GGC TCC GCT GCT GGT TC

RevPamcherrySall GTTGGTTCGAC TCAGAGGATCCCATATGAATATCCTCC

Plasmid to put it in is pBad24: First enzyme to be used is KpnI(GGTACC), enzyme to merge CDSs will be BamHI (GGATCC), the end enzyme will be Sall(GTCGAC)

ATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACCTGGCG
GGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTG
CACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAA
CTGGGTGCCAGCGTGGTGGTGTGATGGTAGAACGAAGCGGCGTCGAA
GCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGG
CTGATCATTAACCTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAG
CTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACA
CCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCG
TGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGG
GCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAA
ATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGAC
TGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGG
GCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGG
GCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATA
TCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCC
GCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGC
GTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAAT
CAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCC
AATACGCAAACCGCCTCTCCCCGCGCGTGGCCGATTCATTAATGCAGC

TGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGGGATCCTCGGCTG
GCTCCGCTGCTGGTTC

Oligo ordered for pCYC ForLacI-41NcoI GATCCCCGGGATGGCGGAGCT-
GAATTACATT pCYCHindIIIrev GACTAAGCTTCCCCGTGGAGGTAATAATTGA
Sequence of a plasmid after ligation into pBad24:

AATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGGGATTCTGC
AAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATTCTGTTACCAAT
TATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACG
GAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAA
TAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGC
ATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCT
CGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGA
CAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATC
AAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGT
ACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGC
GCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATA
GCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTGCTG
AATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCA
AATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGA
AAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCG
TAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGC
AAACAAATTCTCGTCCCTGATTTTTTACCACCCCCTGACCGCGAATGGT
GAGATTGAGAATATAACCTTTCATTCCCAGCGGTGGTTCGATAAAAAAAT
CGAGATAACCGTTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGG

CATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGGCGCTTCAGCCAT
ACTTTTCATACTCCCGCCATTTCAGAGAAGAAACCAATTGTCCATATTGCA
TCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAA
CCGGTAACCCCGCTTATTAAGCATTCTGTAACAAAGCGGGACCAAAG
CCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGT
CCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATT
TATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCT
ACTGTTTCTCCATACCCGTTTTTTGGGCTAGCAGGAGGAATTCACCATG
GTACccggGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAAC
AACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCT
GGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGC
CGATCAACTGGGTGCCAGCGTGGTGGTGTGATGGTAGAACGAAGCGG
CGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGT
CAGTGGGCTGATCATTAAGTATCCGCTGGATGACCAGGATGCCATTGCT
GTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTG
ACCAGACACCCATCAACAGTATTATTTCTCCCATGAAGACGGTACGCG
ACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCT
GTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGG
CTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGG
GAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGC
TGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGAT
GGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTG
GTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCAT
GTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGG
GCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGT
GAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACC
CTGGCGCCCAATACGCAAACCGCCTCTCCCGCGCGTGGCGCGATTCA

TTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGGGT
ACCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATTAAGGAG
TTCATGCGCTTCAAGGTGCACATGGAGGGGTCCGTGAACGGCCACGTG
TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCA
GACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCACCTG
GGACATCCTGTCCCCTCAATTCATGTACGGCTCCAATGCCTACGTGAAG
CACCCCGCCGACATCCCCGACTACTTTAAGCTGTCCTTCCCCGAGGGC
TTCAAGTGGGAGCGCGTGATGAAATTCGAGGACGGCGGGCGTGGTGACC
GTGACCCAGGACTCCTCCCTGCAGGACGGTGAGTTCATCTACAAGGTG
AAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAG
AAGACCATGGGCTGGGAGGCCCTCTCCGAGCGGATGTACCCCGAGGA
CGGCGCCCTGAAGGGCGAGGTCAAGCCGAGAGTGAAGCTGAAGGACG
GCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGC
CCGTGCAGCTGCCCCGGCGCCTACAACGTCAACCGCAAGTTGGACATCA
CCTCACACAACGAGGACTACACCATCGTGGAACAGTACGAACGTGCCG
AGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAaccggt
CAATTGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAA
CTTCGGAATAGGAACTTCAAGATCCGGATCATATCGTCAATTATTACCTC
CACGGGGAagctTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGA
TACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCT
GGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAA
GTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGA
GTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGAC
TGGGCCTTTCGTTTTATCTGTTGTTTGTGGGTGAACGCTCTCCTGAGTAG
GACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGG
AGGGTGGCGGGCAGGACGCCCCGCCATAAACTGCCAGGCATCAAATTAA
GCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAACTCTT

TTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTC
AACATTTCCGTGTGCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCT
GTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTA
AGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCAC
TTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGG
CAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTG
AGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAG
AGAATTATGCAGTGCTGCCATAACCATGAGTGATAAACTGCGGCCAAC
TACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGC
ACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCT
GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGC
AATGGCAACAACGTTGCGCAAACCTATTAACCTGGCGAACTACTTACTCTAG
CTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGG
ACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAA
TCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGG
CCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTC
AGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTC
ACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTA
GATTGATTTACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCG
CTCCTTTGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCGGCTTTCCC
CGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTT
TACGGCACCTCGACCCCAAAAACTTGATTTGGGTGATGGTTCACGTAG
TGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCC
ACGTTCTTTAATAGTGGACTCTTGTTCCAACTTGAACAACACTCAACCC

TATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCCT
ATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACA
AAATATTAACGTTTACAATTTAAAAGGATCTAGGTGAAGATCCTTTTTTGAT
AATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGC
GCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGT
TTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCT
TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTT
AGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG
CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTA
CCGGGTTGGA CTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCG
GCTGAACGGGGGGTTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCT
ACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTC
GGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA
TCTTTATAGTCCTGTGCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTT
TGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG
CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTC
TTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
CAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTA
CGCATCTGTGCGGTATTTACACCGCATAGGGTCATGGCTGCGCCCCG
ACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCC
GGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTG
TCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCAAGGAGA
TGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCACG
CCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCA

TCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCG
GTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCTGCTCATGTTTG
ACAGCTTATC