

Single-molecule studies of DNA-binding proteins in live bacteria

Mathew Stracy

A thesis submitted for the degree of
Doctor of Philosophy

Wolfson College

University of Oxford

Trinity Term 2015

A thesis submitted for the degree of Doctor of Philosophy.

Single-molecule studies of DNA-binding proteins in live bacteria

By Mathew Stracy, Wolfson College.

Submitted Trinity Term 2015.

Abstract

Protein-DNA interactions are critical to many important biological functions, from transcription to DNA replication. To better understand these processes we need to look at molecular details, such as the stoichiometries and binding kinetics of these proteins. However, focusing on the molecular level can miss the bigger picture; we also need to understand how protein-DNA interactions shape the organisation of chromosomes and cause phenotypical changes over the whole cell.

In this thesis I describe the construction of a super-resolution fluorescence microscope to image single molecules in live bacteria, and show how analysis with tools like single-particle tracking allow individual proteins specifically bound to DNA to be distinguished from mobile molecules, offering a new perspective on protein-DNA interactions, from the molecular level to the length scale of whole bacterial cells. I detail how I have applied these techniques to answer key questions in transcription, chromosome organisation, and DNA segregation in *Escherichia coli*.

Firstly, I looked at RNA polymerase (RNAP) to study how transcription affects the organisation of the nucleoid. Discriminating specifically bound RNAPs showed that low levels of transcription can occur throughout the nucleoid, but clustering analysis and 3D Structured Illumination Microscopy (SIM) showed that dense clusters of transcribing RNAPs form at the nucleoid periphery, indicating a movement of gene loci out of the bulk of DNA as levels of transcription increase. Furthermore, I developed an assay to characterise the search process and non-specific DNA interactions of RNAP, which I also apply to a diverse selection of other DNA-binding proteins.

I also characterized the *in vivo* behaviour of the type II topoisomerase, TopoIV. Imaging both subunits of TopoIV, combined with over-expression of unlabelled subunits, allowed the fraction of functional enzymes to be determined. Measuring the duration of catalytic events indicated that the majority of active TopoIV molecules catalyse decatenation. Finally, I studied MukBEF, an SMC (Structural Maintenance of Chromosomes) complex that acts in chromosome segregation, to show that TopoIV and MukB interact directly *in vivo* and determine the dissociation constant and turnover of this TopoIV-MukB complex.

Statement of Originality

I hereby declare that the work in this thesis is my own. Any contribution made to this research by others, with whom I have collaborated at the University of Oxford, is explicitly acknowledged in the text and detailed below. The pronoun 'we' is used instead of 'I' throughout this thesis for stylistic reasons, although the latter would be more accurate.

- Chapters 1, 2, and 3 contain text and figures adapted from my review publication (Stracy et al., 2014).
- The single-molecule TIRF microscope was designed and built in collaboration with Stephan Uphoff and Federico Garza de Leon.
- Chapter 4 is an extended version of my publication (Stracy et al., 2015).
- Structured illumination experiments described in Chapter 4 were performed in collaboration with Christian Lesterlin (Oxford Biochemistry).
- Control proteins HU-PAmCherry and HNS-PAmCherry strains were created by Pawel Zawadzki (Oxford Biochemistry).
- Experiments with degraded chromosomes described in Chapter 5 were performed in collaboration with Christian Lesterlin and Stephan Uphoff. Christian Lesterlin constructed and prepared strains for microscopy. Stephan Uphoff helped perform preliminary experiments.
- The work presented in Chapters 6 and 7 was done in collaboration with Pawel Zawadzki, resulting in the joint-first author publication (Zawadzki et al., 2015). Figures and text in these chapters are adapted from this publication, which Pawel Zawadzki, Dave Sherratt and I wrote. Strain construction and the majority of experimental acquisition in these chapters were performed by Pawel Zawadski.
- The data in Figure 7.10 was collected and analysed by Kasia Ginda (Oxford Biochemistry) and is reproduced for completeness.

Mathew Stracy.
October 2015.

Acknowledgements

I am very grateful to my supervisor Achillefs Kapanidis for both his project ideas and his enthusiastic support. I would like to thank everyone in the Kapanidis group, in particular Stephan who started the exciting *in vivo* single-molecule imaging in the Kapanidis lab, and provided a great deal of help getting me started at the beginning of my DPhil and Fede, my fellow *in vivo* lab member.

The work in this thesis could not have been achieved without the collaborations with the microbiologists in the Sherratt Lab. I want to thank David Sherratt and Amy Upton who helped me learn to make strains. My thanks go in particular to both Christian for his help with SIM and degrading DNA and to Pawel for the fantastic collaboration on TopoIV and many useful discussions.

Lastly, I want to thank Enrichetta for putting up with me while I wrote this thesis.

Dissemination

Publications

- Stracy, M., Uphoff, S., Garza de Leon, F., Kapanidis, A.N. (2014) In vivo single-molecule imaging of bacterial DNA replication, transcription, and repair. *FEBS Letters* 588(19):3585-3594. May 2014.
- Stracy, M., Lesterlin, C., Garza de Leon, F., Uphoff, S., Zawadzki, P., Kapanidis A.N. Live cell super-resolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences*, 112(32):201507592. July 2015.
- Zawadzki, P*., Stracy, M*., Ginda, K., Zawadzka, K., Lesterlin, C., Kapanidis, A.N., Sherratt, D.J. The localization and action of topoisomerase IV in Escherichia coli chromosome segregation is coordinated by the SMC complex, MukBEF. *Cell Reports* (under review). 2015. *Equal contribution.
- Stracy, M., Lesterlin, C., Uphoff, S., Sherratt, D.J., Kapanidis, A.N. Single-molecule behaviour of DNA-binding proteins in live bacteria. (in preparation).

Conference Presentations and Posters

- RNA polymerase workshop (talk), Milton Keynes, 2013.
- Biophysics seminar series (talk), 2013, Oxford, 2013.
- PicoQuant Single Molecule Spectroscopy and Super-resolution Microscopy workshop (talk), Berlin, 2013.
- Biophysical Society meeting (poster), San Francisco, 2014.
- RNA polymerase workshop (talk), Newcastle, 2014.
- Quantitative Methods in Gene Regulation (talk), Cambridge, 2014.
- RNA polymerase workshop (talk), Birmingham, 2015.

Contents

Contents vi

List of Figures viii

Acronyms and Abbreviations x

1 Introduction 1

2 Super-resolution microscopy and single-molecule fluorescence 4
- 2.1 Fluorescence microscopy . . . 6
- 2.2 Beating the diffraction limit . . . 8
- 2.3 TIRF microscope design and construction . . . 11

3 Single-molecule transcription in live *E. coli* 19
- 3.1 Transcription . . . 19
- 3.2 Diffusion . . . 22
- 3.3 Imaging DNA binding . . . 23
- 3.4 Tracking and categorisation of RNA polymerase . . . 27
- 3.5 Counting protein copy numbers . . . 34
- 3.6 Custom analysis software . . . 37
- 3.7 Materials and methods . . . 37

4 RNA polymerase and its role in nucleoid organisation 40
- 4.1 Introduction . . . 40
- 4.2 Bound and mobile RNAPs show different spatial distributions . . . 43
- 4.3 Transcribing RNAPs tend to cluster at the edge of the nucleoid . . . 46
- 4.4 Transcription in rich media . . . 56
- 4.5 Discussion . . . 61
- 4.6 Materials and methods . . . 67

5 The target search process **70**
5.1 Introduction . . . 70
5.2 RNAP promoter-search . . . 72
5.3 Monte Carlo simulations of RNAP diffusion . . . 74
5.4 Changing environmental conditions . . . 79
5.5 A versatile assay for determining the DNA-free diffusion . . . 80
5.6 Search models of RNAP, Pol I and MukB . . . 84
5.7 Discussion . . . 87
5.8 Materials and methods . . . 90

6 Single-molecule characterisation of topoisomerase IV **92**
6.1 Introduction . . . 92
6.2 Single-molecule tracking of ParC and ParE . . . 95
6.3 TopoIV catalysis in non-replicating cells . . . 100
6.4 TopoIV catalytic binding time . . . 101
6.5 Discussion . . . 105
6.6 Materials and methods . . . 106

7 The action of TopoIV is directed by MukB **108**
7.1 Introduction . . . 108
7.2 Half of immobile TopoIV molecules result from Interaction with MukB 110
7.3 MukB directs immobile TopoIV molecules to centre of the nucleoid . 114
7.4 Two Populations of immobile TopoIV Molecules . . . 116
7.5 Catalytically active TopoIV . . . 120
7.6 The interaction between ParC and MukB facilitates *ori* decatenation . 123
7.7 Discussion . . . 125
7.8 Materials and methods . . . 127

8 Conclusion and outlook **129**
8.1 Organisation of enzyme activity in the nucleoid . . . 129
8.2 Dissemination of techniques . . . 131
8.3 Future outlook . . . 132

References **134**

A Characterisation of bacterial strains **142**

B Example Matlab code **145**

List of Figures

2.1 Jablonski diagram 7
2.2 Single-molecule localisation 10
2.3 Schematic of the custom-built single-molecule TIRF microscope 14
2.4 Pictures of the custom-built single-molecule TIRF microscope. 15
2.5 Acquisition software 17
2.6 Example images acquired using the custom-built microscope 18

3.1 The transcription cycle 21
3.2 Detection by immobilisation 25
3.3 Photoactivated single-molecule tracking and analysis 28
3.4 Immobile and mobile RNA polymerase 32
3.5 Sorting individual RNAP molecules. 33
3.6 Blocking transcription initiation 35
3.7 RNA polymerase copy number 36
3.8 LoColi analysis suite 38

4.1 Comparison of categorised RNAPs with the distribution of DNA 44
4.2 Radial distribution of mobile and bound RNAP. 45
4.3 RNAP in the nucleoid 47
4.4 Average spatial distribution of transcribing RNAP 49
4.5 3D structured illumination microscopy images of RNAP and DNA 50
4.6 Transcribing RNAPs cluster at the nucleoid periphery 52
4.7 Blocking active transcription 54
4.8 SIM image of an example cell treated with rifampicin 55
4.9 Blocking active translation 57
4.10 Mobility of RNAP clusters 59
4.11 Organisation of transcription in rich media. 60
4.12 RNAP-DNA segregation in rich media. 62
4.13 Mechanisms of gene spatial organisation. 65

5.1 Schematic of different protein-DNA interactions 71
5.2 DNA-free diffusion of RNAP 75
5.3 Monte Carlo simulations of RNAP diffusion 77
5.4 Partitioning RNAP at different growth rates 81
5.5 Generating degraded-DNA cells 82
5.6 DNA-free diffusion of RNAP, Pol I and MukB 84
5.7 Monte Carlo simulations of diffusion in degraded-DNA cells 86
5.8 Search models for RNAP, Pol I and MukB 88

6.1 Topoisomerase IV strand passage mechanism 94
6.2 Intracellular mobility of ParC 96
6.3 Intracellular mobility of ParE 97
6.4 TopoIV diffusion is cell cycle independent 100
6.5 Mobility of ParC in non-replicating cells 102
6.6 TopoIV binding time 104

7.1 ParC interacts with MukB 111
7.2 Radial distribution of ParC 112
7.3 Clustering of ParC near MukB foci. 114
7.4 MukB clusters. 115
7.5 Intracellular organisation of MukB and TopoIV 117
7.6 ParC shows a Muk-dependent and Muk-independent binding behaviour 118
7.7 TopoIV simulated diffusion 120
7.8 Diffusion of catalytically active TopoIV 122
7.9 Spatial distribution of TopoIV catalysis 124
7.10 The MukB-ParC interaction stimulates chromosome segregation at *ori* . . 125
7.11 Cartoon showing MukB-dependent and independent TopoIV activity. . . 127

A.1 TopoIV fusion functionality 143
A.2 Characterisation of fusion and overexpression strains 144

Acronyms and Abbreviations

Amp	Ampicillin
bp	Base pair
Cam	Chloramphenicol
CDF	Cumulative Distribution Function
ChIP	Chromatin ImmunoPrecipitation
CTD	C-terminal domain of ParC
D	Diffusion coefficient
D*	Apparent diffusion coefficient
DAPI	4',6-diamidino-2-phenylindole
DSB	Double Stranded Break
E.coli	*Escherichia coli*
EMCCD	Electron Multiplying Charge Coupled Device camera
Fis	Factor for inversion stimulation protein
FP	Fluorescent Protein
FRAP	Fluorescence Recovery after Photobleaching
Gen	Gentamicin
GFP	Green Fluorescent Protein
H-NS	Histone-like Nucleoid Structuring protein
HU	Heat Unstable protein
Kan	Kanamycin
LacI	*lac* repressor
LB	Luria Broth
M9	Minimal growth medium
MMS	Methyl MethaneSulfonate
MSD	Mean Squared Displacement
mYPet	Monomeric yellow fluorescent protein variant
NA	Numerical Aperture
NAP	Nucleoid Associated Protein
Ori	Origin of replication

PALM	PhotoActivated Localization Microscopy
PAmCherry	PhotoActivatable monomeric Cherry fluorescent protein
Pol I	DNA polymerase I
PSF	Point Spread Function
RDM	Rich Defined Media
Rif	Rifampicin
RNAP	RNA Polymerase
rpoC	Gene encoding the beta' subunit of RNA polymerase
SEM	Standard error of the mean
SIM	Structured Illumination Microscopy
SMC	Structural Maintenance of Chromosomes
STORM	STochastic Optical Reconstruction Microscopy
Tet	Tetracycline
TIRF	Total Internal Reflection Fluorescence microscopy
TopoIV	Topoisomerase IV



Chapter 1

Introduction

Protein-DNA interactions are critical to a huge range of important biological functions, from gene regulation and transcription to chromosome organisation and segregation. To better understand these processes we need to look at molecular details, such as the intracellular concentrations and stoichiometries of DNA-binding proteins, the location of DNA binding sites, and the binding kinetics of these proteins. However, focusing on the molecular level can miss the bigger picture; we also need to understand how protein-DNA interactions shape the organisation of chromosomes and cause phenotypical changes over the whole cell.

At the molecular level, much of our understanding of how proteins interact with DNA has come from *in vitro* experiments performed using purified proteins and DNA oligonucleotides. The finest level of detail has been achieved through X-ray crystallography of proteins in complexes with short DNA fragments. This can achieve atomic resolution, allowing the precise interactions between the bases on the DNA and the amino acid residues on the protein to be determined. However, the

'snapshots' from crystallography are poorly suited to studying dynamic behaviour. To complement structural information from crystallography, *in vitro* single-molecule experiments are becoming increasingly popular to determine the kinetics of these interactions, by directly observing the behaviour of individual molecules (Hohlbein et al., 2010). For example, quantum dot labelled proteins have been imaged binding to linear DNA 'curtains' to determine how proteins locate their target site (Wang et al., 2013). While *in vitro* single-molecule techniques have been used to great effect in elucidating molecular behaviour, care must be taken when inferring the physiological relevance, as the experiments are performed on highly simplified systems, in isolation from the rest of the cellular components. At a larger scale, protein-DNA interactions can be put in the context of the complete chromosome with tools like chromatin immunoprecipitation (ChIP), which uses lysates from a population of cells to determine the specific binding sequences of proteins of interest (Grainger et al., 2005).

On the other hand, DNA and DNA-binding proteins can both be imaged inside living cells with fluorescence microscopy. However, while conventional fluorescence microscopy can report on large cellular features, details are lost below the diffraction limit of light (~200 nm). Technological advances now allow single molecules to be imaged inside living cells, bridging the gap between molecular-level and cellular-level experiments, and opening the doors to the study of questions that neither *in vitro* studies nor conventional fluorescence microscopy can answer. With *in vivo* single-molecule techniques, individual DNA binding events to be observed (Elf et al., 2007; Hammar et al., 2012; Uphoff et al., 2013; Gahlmann and Moerner, 2013; Persson et al., 2013), shedding new light on binding behaviour and kinetics, and resolving heterogeneity that ensemble measurements can miss (Xie et al., 2008). Moreover, these techniques can report on the spatial locations of these actions in cells, and how

they change as cells respond to stimuli (Uphoff et al., 2013).

The aim of this thesis is to give an overview of the techniques involved in imaging single molecules in live bacteria and show how these can be combined with tools like single-particle tracking to offer a new perspective on protein-DNA interactions from the molecular length scale to the level of whole bacterial cells. We will then highlight broadly applicable tools, methods, and analysis we have developed to study protein-DNA interactions in living cells. Finally, we will apply these techniques and tools to answer key questions in *Escherichia coli* transcription, chromosome organisation, and DNA segregation.



Chapter 2

Super-resolution microscopy and single-molecule fluorescence

This chapter provides an introduction to the *in vivo* single-molecule and super-resolution fluorescence techniques. We also detail the design and construction of the single-molecule microscope used throughout this thesis.

Light microscopy is one of the most important tools for biologists, however it is limited by the fundamental physics of light. When a single point source of light is focused by the microscope objective, it does not converge to an infinitely sharp point. Instead, diffraction causes the light to form a spot with a finite size. The radius of this spot, r, is governed by the Rayleigh diffraction limit, and depends on the wavelength of the light, λ, the refractive index of the medium, n, and the angle at which the light wave converges, θ:

$$r = \frac{0.61\lambda}{n\sin\theta} \tag{2.1}$$

The denominator, $n\sin\theta$, is called the numerical aperture (NA) of the lens. A

modern objective can have an NA up to ~1.65, giving a radius between 175 nm and 300 nm for visible light. The intensity profile of this spot is known as the point spread function (PSF) and has roughly the same width as that of the focal spot described above. This width is typically used to define the resolution limit, hence two identical emitters separated by a distance less than the width of the PSF will appear as a single object, making them unresolvable from each other (Huang et al., 2010).

Several alternative imaging techniques can achieve significantly higher resolution than conventional light microscopy. Using electrons to probe biological structures allows greatly increased resolution due to the much shorter wavelength, but typically samples need enhanced contrast via, for example, immunogold labelling which involves chemical fixation and permeabilization which can distort biological structures (Schnell et al., 2012). Alternatively, samples can be prepared by cryogenic freezing, however, so far live cell imaging is impossible with electron microscopy, making it unsuitable for studying kinetics (Huang et al., 2010). Atomic force microscopy also allows very high resolution images to be made by raster scanning a physical probe across a surface. However, the need to place the probe in extremely close proximity to the structure under study makes it unsuited to imaging inside living cells.

Despite its limited resolution, light microscopy remains the most widely used technique for imaging live samples. However, over the past decade a number of new techniques have been developed to beat the diffraction limit allowing light microscopy to achieve much higher resolution than ever thought possible. The impact of these new techniques was recognised in the 2014 Nobel Prize for Chemistry. Different methods have accomplished this by approaching the problem from different angles, however they are all built on the foundations of one central technique: fluorescence microscopy.

2.1 Fluorescence microscopy

During the past 20 years there has been a remarkable growth in the use of fluorescence microscopy (Lakowicz, 2006). One of the key reasons for its popularity compared to conventional light microscopy is the ability to observe labelled molecules of interest inside cells with greatly reduced unwanted signal from other cellular components. This can be achieved due to the physics of fluorescence: As illustrated in Figure 2.1, a molecule can be excited from the ground electronic state (S_0) to one of several higher electronic states (S_1, S_2 etc) by absorbing a photon. For each of these electronic states there are higher vibrational energy levels. A molecule excited to a higher vibrational levels will rapidly undergo non-radiative decay to the lowest vibrational level of S_1. From here it can return to the ground state, releasing a photon with less energy than the absorbed photon (Lakowicz, 2006). The result of this effect, known as the Stokes shift, is that the emitted photons from fluorescent molecules are of a longer wavelength than the excitation photons. In fluorescence microscopy this can be exploited by filtering out the shorter wavelength excitation light leaving only the emitted signal.

Molecules excited in the S_1 state can also undergo a spin conversion into the triplet state T_1, a process called inter-system crossing. Emission from T_1 is called phosphorescence and occurs more slowly than fluorescence because transition between T_1 and S_0 is forbidden. For fluorescence imaging, population of the triplet state is generally undesirable as it is a relatively long lived state in comparison to the singlet state and hence increases lifetime of the molecule in an excited state. Excited molecules are much more likely to undergo irreversible chemical reactions with components in the environment causing the fluorophore to permanently lose the ability to fluoresce (photobleaching).

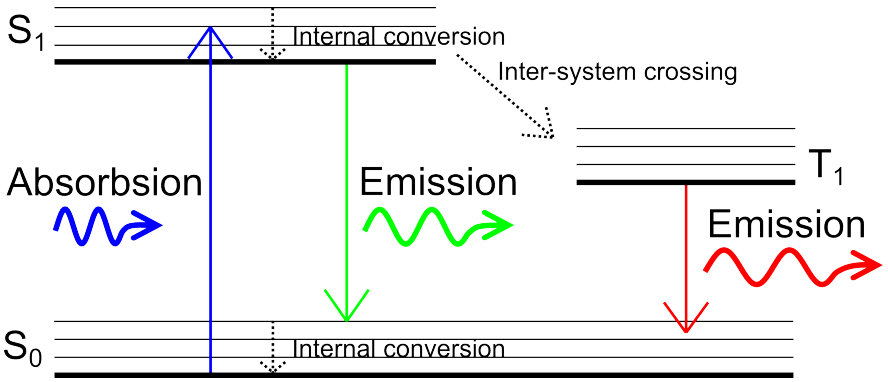


Figure 2.1: **Jablonski diagram illustrating the energy levels and transitions of a typical fluorophore.** The electronic states S_0, S_1, and T_1 represent the singlet ground state, the first excited singlet state and the excited triplet state respectively. The vibrational ground states of each electronic state are indicated with thick lines, the higher vibrational states with thinner lines. Solid arrows show radiative transitions, and dashed arrows show non-radiative transitions.

Labelling molecules

In order to use fluorescence to study a molecule or structure of interest, it normally has to be labelled with a fluorescent dye. The fluorophore used to label the protein of interest is of crucial importance to the sensitivity of the experiment. *In vitro*, synthetic dyes, such as cyanine dyes are particularly useful due to their small size and good photostability, and the ability to conjugate them directly to a protein of interest. For *in vivo* imaging these dyes can be conjugated to an antibody to immunostain structures of interest, however, this requires cells to be chemically fixed and permeabilized. Unknown labelling stoichiometries of fluorophores per antibody and incomplete staining of a particular protein with antibodies in the cell can hinder quantitative analysis of the microscopy images (Gruß mayer et al., 2014).

For live cell imaging, some synthetic dyes can naturally permeate the cell membrane and bind preferentially to particular molecules without the need for antibodies. For example, intercalating DNA stains such as 4',6-diamidino-2-phenylindole (DAPI) can be used to label the nucleoid. Progress is also being made on delivering syn-

thetic dyes into cells and targeting them to specific locations (Fernández-Suárez and Ting, 2008; Hinner and Johnsson, 2010; Zhang et al., 2003; van de Linde et al., 2012; Crawford et al., 2013). Although these techniques remain difficult and not yet widely adopted. Currently the most common method for labelling specific proteins in live cells relies on genetically encoded fluorescent proteins (FPs).

After the green fluorescent protein (GFP) was first identified from the *Aequorea victoria* jellyfish, it started a new era in cell biology by allowing researchers to use molecular cloning techniques to genetically fuse a protein, typically via the C or N terminal, to an FP (Tsien, 1998). While care has to be taken to characterise protein-FP fusions, many exhibit essentially no change in functionality or abundance. Although FPs are much larger and less photostable than synthetic dyes, the ability for cells to produce the fluorophore themselves makes them ideal for imaging in live cells. Fusion proteins can be expressed either on a plasmid, or the by replacing the chromosomal copy of the protein. The latter approach, combined with the near 1:1 labelling stoichiometry, facilitates determining intracellular copy numbers.

2.2 Beating the diffraction limit

The specificity of fluorescence microscopy combined with the reduced background signal naturally lends itself to single-molecule and super-resolution techniques. Super-resolution techniques fall broadly into two categories: single-molecule localisation methods and ensemble methods (Huang et al., 2010). Both techniques are employed in this thesis and each carries its own advantages and disadvantages.

Ensemble super-resolution techniques

The 'ensemble' methods for breaking the diffraction limit rely on illuminating the sample with patterned excitation. Stimulated emission depletion microscopy (STED) (Hell and Wichmann, 1994) increases resolution by suppressing unwanted fluorescence surrounding the small central excited area, achieving a resolution of 50 nm to 70 nm for fluorescent proteins (Huang et al., 2010).

Structured illumination microscopy (SIM) increases resolution by using sinusoidal patterned excitation light (Heintzmann et al., 2002; Schermelleh et al., 2008). The interference pattern of the sample structure and the excitation pattern contains otherwise-unobservable information about the sample. Multiple images are taken of the same sample but with different angles and phases of the excitation pattern. As the excitation pattern is known in all of the snapshots the final image can be computationally reconstructed from the multiple snapshots, allowing a resolution of ~100 nm.

One of the key advantages of these patterned illumination techniques is that they require no special fluorophores or sample preparations, they are hence 'back compatible' with previously labelled samples and can be readily used in live cells. On the other hand, these techniques require either strong excitation intensities to achieve maximum resolution (STED), or multiple images per field of view (SIM) and hence can cause rapid photobleaching, making them unsuited to samples with low copy numbers of labelled molecules.

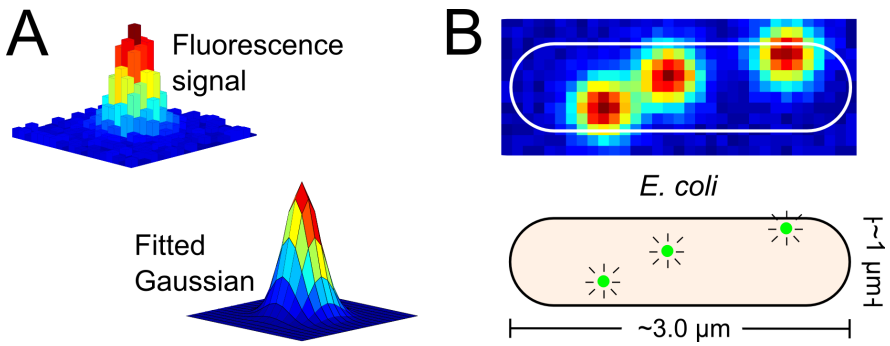


Figure 2.2: **Single-molecule localisation.** A) The position of a single fluorophore can be localised with much greater precision than the diffraction limit by fitting the intensity profile with a Gaussian mask which, in many cases, approximates the underlying point spread function well. B) In order to fit the fluorescence intensity profile of a molecule, it must be sufficiently separated from any other molecules to be clearly spatially resolved. In the case of bacteria, this means imaging only a few molecules per cell; a much lower density than most proteins in *E. coli*.

Localisation microscopy

In contrast to the 'physics' approach of ensemble super-resolution methods, localisation based methods have arisen in a large part due to breakthroughs in the chemistry of the fluorophores used to label the sample.

The basis of these approaches rely on the fact that the image produced by a single fluorophore can, in many cases, be well approximated by a Gaussian function (Thompson et al., 2002; Cheezum et al., 2001). The exact position of an isolated molecule can therefore be determined with much higher precision than the size of the original fluorescence spot by fitting to a Gaussian mask (Figure 2.2 A). The uncertainty of the fitted position depends mainly on the number of photons collected (Thompson et al., 2002; Smith et al., 2010).

However, in order to fit the fluorescence intensity profile of a molecule, it must be sufficiently separated from any other molecules to be clearly spatially resolved. In the case of bacteria, this means imaging only a few molecules per cell; a much lower density than most proteins in *E. coli* (Figure 2.2 B). The breakthrough in super-

resolution localisation microscopy came about with the ability to image only a small subset of fluorophores at any one time by exploiting photoswitching and photoactivation. Molecules are stochastically activated, imaged and localised over a movie with typically several thousand frames. The localisations from all frames can then be reconstructed into a super-resolved image (Betzig et al., 2006; Greenfield et al., 2009; Huang et al., 2010).

In the case of stochastic optical reconstruction microscopy (STORM; Rust et al., 2006), this was originally achieved with a pair of organic cyanine dyes, and later simplified by using a single fluorophore and imaging in an oxidizing and reducing buffer (dSTORM, Heilemann et al., 2008). The oxidising buffer induces the fluorophore to enter a dark triplet state. As the molecules return stochastically they can be excited to fluoresce so that single molecule positions can be recorded.

For photoactivated localisation microscopy (PALM, Betzig et al., 2006) the breakthrough came through the development of photoswitchable or photoactivatable fluorescent proteins, such as PA-GFP (Patterson and Lippincott-Schwartz, 2002), mEos2 (McKinney et al., 2009), or PAmCherry (Subach et al., 2009). These proteins can be photoactivated with (~400 nm) UV light, the intensity of which can be chosen to ensure that there are very few emitting molecules at any given time.

2.3 TIRF microscope design and construction

Here we describe the total internal reflection fluorescence (TIRF) microscope used in this thesis for super-resolution localisation imaging. This microscope was designed, constructed, and maintained in collaboration with Stephan Uphoff and Federico Garza de Leon (Kapanidis Group). A TIRF microscope was chosen to reduce the problem of out-of-focus fluorescence inevitable in epifluorescence systems. In TIRF

illumination, the excitation beam is reflected at the interface between the coverslip and the imaging medium so that the resulting evanescent excitation extends only ~150 nm into the sample (Paige et al., 2001).

Total internal reflection occurs when light passes from a medium of higher refractive index (n_2), to a lower refractive index (n_1), with an angle of incidence is greater than the critical angle, θ_C. The critical angel is given by:

$$\theta_C = \sin^{-1}\left(\frac{n_1}{n_2}\right) \tag{2.2}$$

The resulting evanescent field decays exponentially at increasing distance, z, above the coverslip:

$$I = I_0 \exp\left(-z/d\right) \tag{2.3}$$

The decay constant, d is dependent on the wavelength, λ, and angle, θ, of the incident light:

$$d = \frac{\lambda}{4\pi}\left(n_2^2 \sin\theta - n_1^2\right)^{-1/2} \tag{2.4}$$

TIRF is well suited to studies focused on the cell membrane, but less appropriate for imaging deeper into the cytoplasm. To image the bacterial nucleoid, TIRF systems can be used at sub-critical angles giving a thin sheet of excitation light at a shallow angle to coverslip (Tokunaga et al., 2008). This latter method is employed throughout this thesis.

The microscope was built to be run in the Micron advanced microscopy facility in the Biochemistry Department (Oxford). Central criteria in the design were therefore

reliability and ease of use by non-technical users. Standard 'objective-type' TIRF was chosen due to the reduced alignment required, the ease of working with conventional coverslips, and the ability to illuminate the sample from above for transmission microscopy of cells. A schematic of the microscope is shown in Figure 2.3, and photos of the completed set-up in Figure 2.4.

Laser excitation reduces unwanted background fluorescence due to the narrow frequency spectrum. To allow versatile imaging of a range of fluorophores spanning the visible spectrum, we used an multilaser engine (iChrome MLE, Toptica), providing 405 nm, 488 nm, 561 nm, and 640 nm lasers all with a power of 100 mW. The laser engine had a self-alignment feature to allow the laser coupling into the single mode optical fibre to be optimised with minimal input from the user. In order to maximise the signal collected from each fluorophore we used a high numerical aperture objective (100x oil immersion objective, NA 1.4, focal length 1.8 mm, Olympus) together with a back-illuminated electron-multiplying CCD camera (iXon Ultra 897, 512x512 pixels, Andor).

At the fibre output the excitation beam was collimated (50 mm achromatic lens, Thorlabs) and focused (250 mm achromatic lens, Thorlabs) in the back focal plane of the objective. For stability, the fibre output, collimation lens, and focusing lens were mounted in a cage system (Thorlabs). To allow adjustment of the incident angle of the beam at the coverslip from TIRF to epifluorescence, the whole unit was mounted on a translation stage controlling the position of the beam in the back focal plane of the objective. A custom stage inset with a thicker design and smaller aperture was made to increase stability, and was machined in the research workshop (Oxford Physics) from steel to allow magnetic slide holders to be used.

Fluorescence emission was collected by the same objective used to introduce the excitation light, and the excitation beam and emission signal were separated

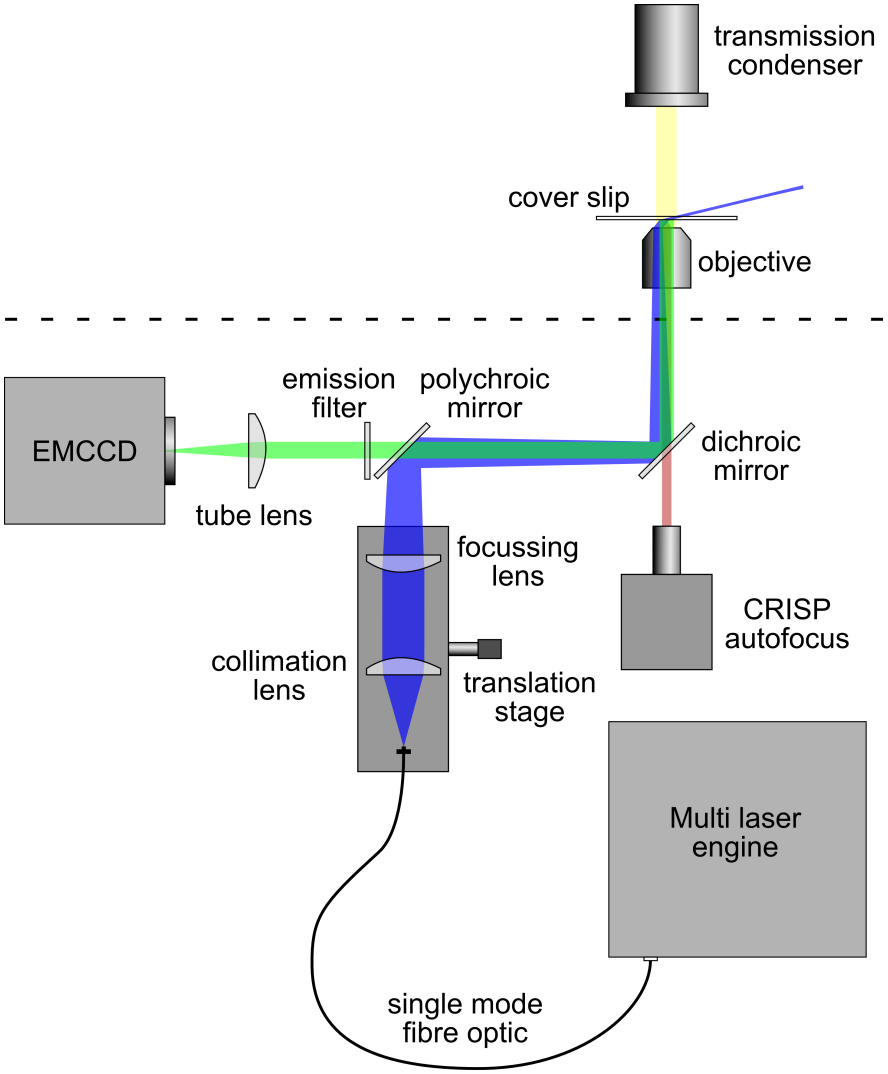


Figure 2.3: **Schematic of the custom-built single-molecule TIRF microscope.** Light from the fibre output of the multi laser engine is collimated and focused on the back focal plane of the objective. Translation of the fibre output, collimation and focusing lenses allows for control of the incident angle of the beam at the coverslip. The emission signal is filtered from the excitation light with a polychroic mirror and focused onto an EMCCD camera. Transmission light is provided by an LED above the sample, and autofocus is provided by an IR LED and a position sensitive photo detector.

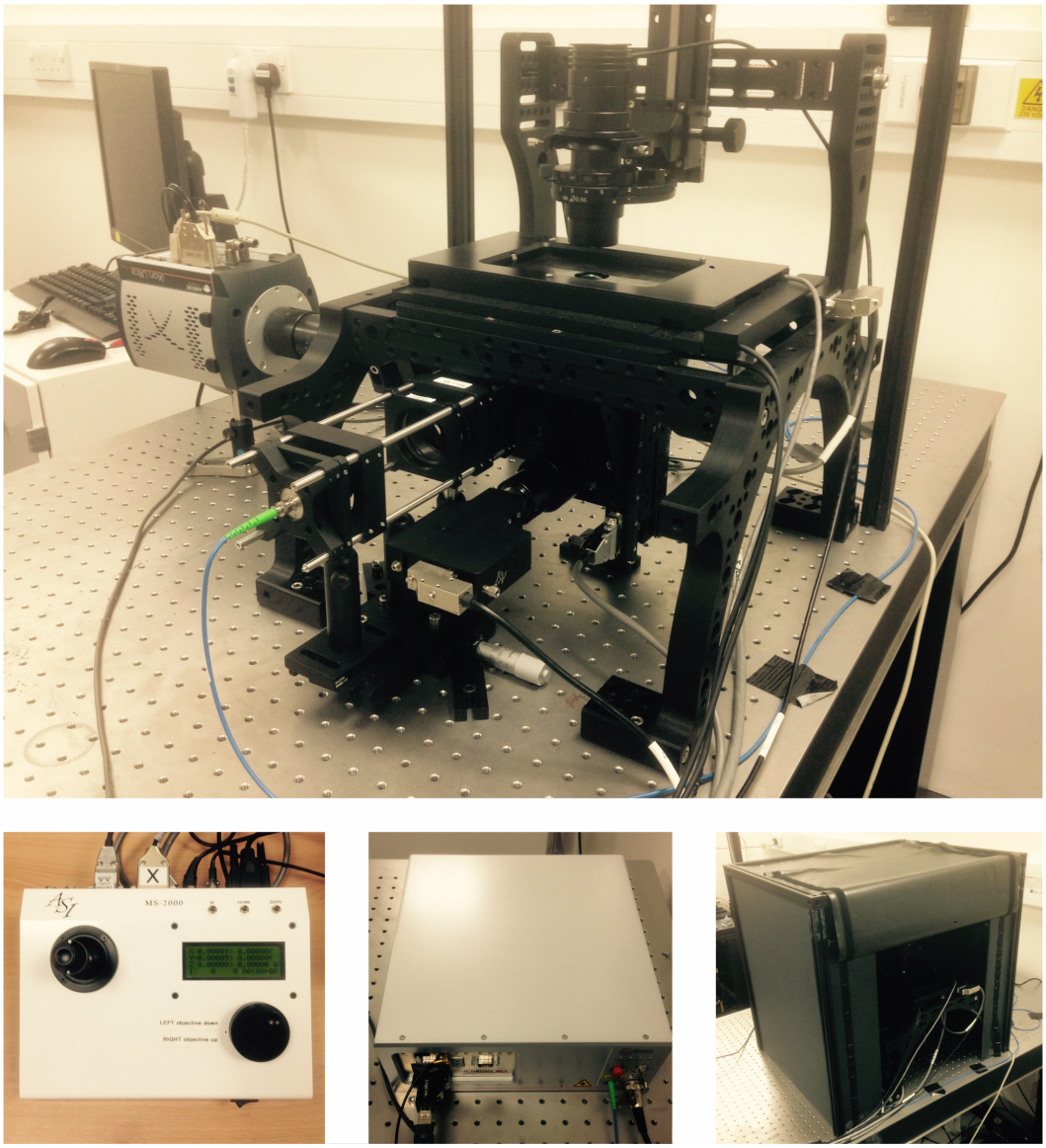


Figure 2.4: **Pictures of the custom-built single-molecule TIRF microscope.** (Top) The main microscope components including the stage, camera, excitation and emission optics, transmission condenser, and the autofocus system. (Bottom, from left to right) The stage and focus control unit, the multi laser engine, and the complete microscope with light shielding.

with a custom polychroic mirror and emission filter. Because the microscope was designed as facility being used for both PALM and STORM imaging, two polychroic mirrors were used, which could be interchanged via a magnetic cube mount. A four colour polychroic mirror and quad-band emission filter allowed the use of all four lasers (ZT405/488/561/640rpc and ZET405/488/561/640NF, Chroma). A three colour polychroic mirror (ZT405/488/561rpc and ZET405/488/561NF, Chroma) optimised for use with the 405 nm, 488 nm, 561 nm excitation maximised the signal collected from red fluorescent proteins (which have a long emission tail) and was used for the experiments in this thesis. A single tube lens (300 mm achromatic lens, Thorlabs) was used to focus all emission channels onto the camera. The 300 mm tube lens and 1.8 mm focal length objective gives 167 x magnification. Each 16 µm pixel of the camera therefore theoretically covers 96 nm on the image. The relative pixel size was verified by microruler measurement to be 95.7 nm.

To record transmitted light images of cells, we included an LED light source and condenser (IX2-LWUCD, Olympus). The transmission LED was controlled via a custom written Matlab GUI (Figure 2.5 A), and laser powers and alignment was controlled through the Toptica iChrome software (Figure 2.5 B). The camera acquisition was controlled with the AndorSolis software (Figure 2.5 C) which was programmable to allow triggering of the lasers by the camera shutter for time-lapse experiments.

The microscope was built around the Rapid Automated Modular Microscope (RAMM) System (ASI Imaging) mounted on an actively damped optical table. Coarse focussing was provided by mounting the objective on a z motor (LS50, ASI Imaging). Sample position and fine grain focusing was provided by a linearly encoded stage with z piezo (PZ-2000FT, ASI imaging). An autofocus module containing a 890 nm LED and clean-up filter and position sensitive photo detector (CRISP autofocus, ASI imaging), was added together with a 850 nm short pass dichroic mirror to allow focus



Figure 2.5: **Acquisition software.** A) Custom written Matlab GUI for control of transmission LED. B) Toptica iChrome software for control of laser intensity. C) AndorSolis software for control of EMCCD camera acquisition settings.

to be maintained over long time periods for time lapse imaging. The x-y and z control and autofocus lock were controlled with an ASI control unit (MS2000, ASI imaging).

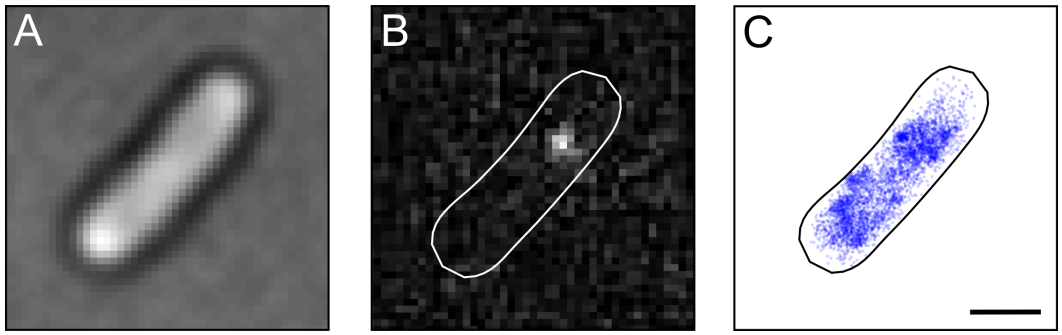


Figure 2.6: **Example images acquired using the custom-built microscope.** A) A transmission image of an example live *E. coli* cell. B) A single frame of a PALM movie showing the fluorescence image from a single labelled RNA polymerase molecule. C) Individual molecules are photoactivated using the 405 nm laser, and imaged for a number of frames using the 561 nm laser until they photobleach irreversibly. This process is repeated as other molecules are stochastically activated until all molecules of interest have been imaged. The Fluorescence images from each frame are fitted to a Gaussian to generate super-resolved localisations of all imaged RNA polymerase molecules in the cell. Scale bar 1 µm.

For localisation microscopy, fluorophore image fitting was performed using custom-written MATLAB software (MathWorks) adapted from Holden et al. (2011) and previously implemented by Uphoff et al. (2013). Fluorophore images were identified by band-pass filtering and applying an intensity threshold to each frame of the movie. Candidate positions were used as initial guesses in a two-dimensional elliptical Gaussian fit for high-precision localisation. An example transmission light image, fluorescent signal, and reconstructed PALM image taken with the microscope are shown in Figure 2.6.



Chapter 3

Single-molecule transcription in live *E. coli*

The diffusion of most proteins is much faster than the movement of DNA loci. Of central importance to this thesis is the idea that proteins bound to DNA can be resolved from the rest of the population based on their lower mobility. This chapter introduces some of the techniques used to image protein diffusion, such as single-particle tracking. Using RNA polymerase (RNAP) as an example, we demonstrate that the apparent diffusion of each tracked molecule can be used to discriminate mobile RNAP molecules from molecules specifically bound to DNA. This allows the spatial distribution of transcribing RNAPs to be seen in live *E. coli*.

3.1 Transcription

Genetic information in *E. coli* is stored on the 4.6 Mb chromosome, but this information can only be accessed via interactions between proteins and DNA. Genes are read by RNA polymerase (RNAP) which initiates transcription by binding to specific DNA sequences, known as promoters. A host of other proteins, called transcription factors regulate which genes are expressed and which are silenced, by binding in or near promoter regions and affecting the ability of RNAP to bind and initiate transcription.

Unlike eukaryotes, there is only a single type of RNAP in *E. coli*, which is responsible for transcription of mRNA and non-coding RNA. It is a multisubunit enzyme made up of a beta and beta' subunit, two alpha subunits and a smaller omega subunit. In order to bind promoters, the RNAP core associates with the transcription initiation factor sigma to form the 450 kDa RNA polymerase holoenzyme (Travers and Burgess, 1969). Sigma factors reduce the affinity of RNAP for non-specific DNA while increasing specificity for promoters. Once bound to a promoter the enzyme opens a bubble in the duplex DNA to form an 'open complex' (Figure 3.1). From here it can initiate transcription, however on some promoters the polymerase makes several attempts to start transcriptions generating short abortive RNAs (Kapanidis et al., 2006). Once past the 11th nucleotide the RNAP breaks its interactions with promoter DNA and enters into processive synthesis of RNA as an 'elongation complex'. In bacteria, transcription and translation are not segregated, and ribosomes can form on the nascent transcript as soon as the ribosome binding site has emerged from the exit channel of RNAP. At some point during transcription the sigma factor usually dissociates and is free to associate with another core enzyme (Kapanidis et al., 2005). Finally RNAP reaches the end of the gene and the RNA transcript and the core enzyme dissociate from DNA.

There are ~3000 promoters in *E. coli*, each containing a core sequence of ~35 base pairs (bp) in length, with two short sequences approximately -10 and -35 bp upstream of the transcription start site. Together these promoter sequences account for less than 2 % of the *E. coli* genome (Wang et al., 2013). For an RNAP molecule to locate a promoter it must therefore discriminate between vast amounts of non-specific DNA. Like all DNA-binding proteins, RNAP is expected to employ some form of diffusion to do this.

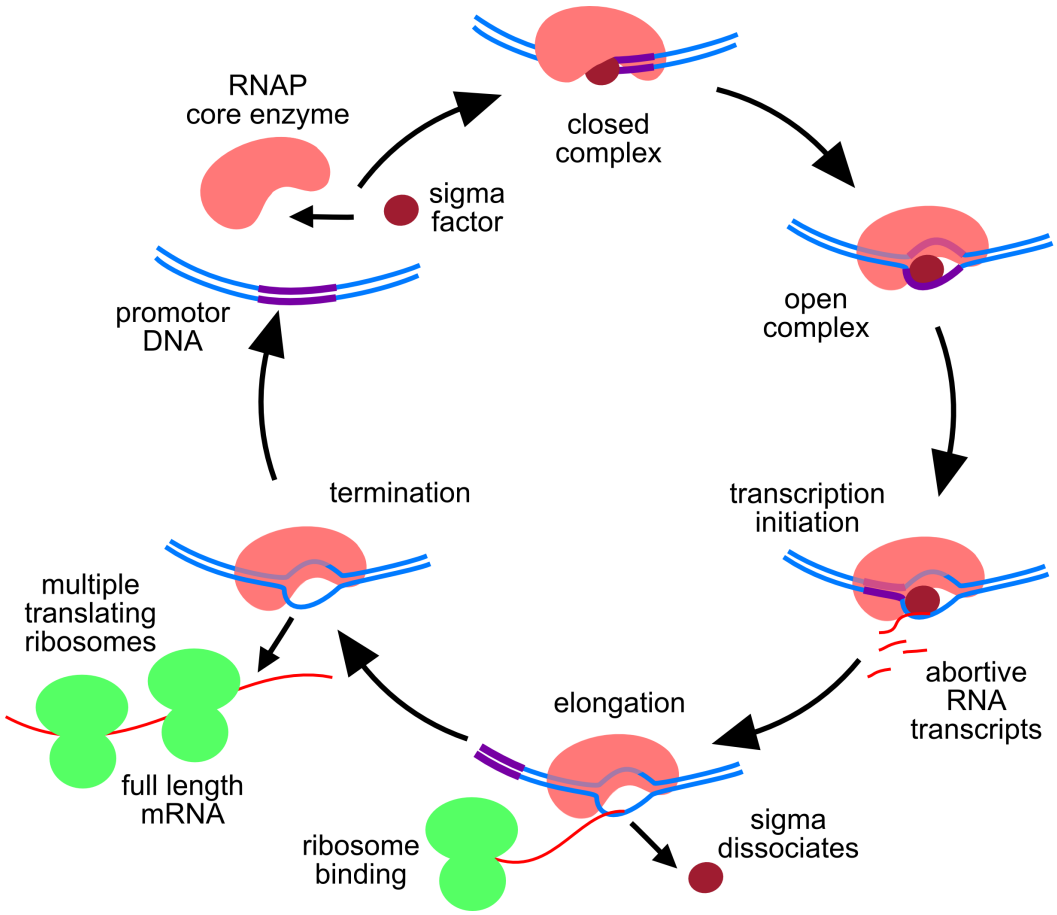


Figure 3.1: **The transcription cycle.** Multisubunit RNAPs carry out transcription by repeatedly cycling through initiation, elongation and termination phases. In order to bind to a promoter, the RNAP core enzyme first associates with a sigma factor. After promoter binding the enzyme opens a short section of the DNA duplex to form a more stable 'open complex'. It can then initiate transcription, however on some promoters the polymerase makes several attempts to start transcriptions generating short abortive RNAs. Once past the 11th nucleotide the RNAP is in the elongation complex and translocates along DNA generating the RNA transcript. Ribosomes can form on the nascent transcript as soon as the ribosome binding site has emerged from the exit channel of RNAP. At some point during transcription the sigma factor usually dissociates and is free to associate with another core enzyme. Finally RNAP reaches the end of the gene and the RNA transcript and the core enzyme dissociate from DNA.

3.2 Diffusion

Any molecule or small particle suspended in a fluid moves randomly due to collisions with the surrounding molecules, a process called Brownian motion. For a spherical particle with a hydrodynamic radius a, the diffusion coefficient can be calculated from the Einstein-Stokes relation:

$$D = \frac{k_B T}{6\pi\eta a} \tag{3.1}$$

Where T is the absolute temperature, η is the viscosity of the solution, and k_B is the Boltzmann constant. For a sphere diffusing in the same viscosity liquid, at the same temperature, the diffusion coefficient is therefore proportional to its radius. For a particle with diffusion coefficient D, the probability for it to be displaced from $x = 0$ to position x in time τ is given by the Gaussian distribution:

$$\rho(x,t) = \frac{1}{4\pi D\tau}\exp\left(-\frac{x^2}{4D\tau}\right) \tag{3.2}$$

The expectation value for the displacement, $E[x]$, at time τ is zero as the particle is equally likely to move left as it is to move right. However the expectation value for the squared displacement (i.e. the variance of the distribution as $E[x] = 0$) increases in proportion to the elapsed time:

$$E[x^2] + E^2[x] = E[x^2] = 2D\tau \tag{3.3}$$

This is important as the mean of squared displacement (*MSD*) of a particle can be experimentally measured at different time-points, Δt and used to determine the diffusion coefficient, D. For Brownian motion in two dimensions the equation becomes:

$$D = \frac{MSD}{4\Delta t} \tag{3.4}$$

For Brownian motion the MSD therefore increases linearly with increasing lag time. A confined molecule, however, will exhibit 'sub-diffusion', where the MSD scales with Δt^{α} where $0 < \alpha < 1$.

In *E. coli*, chromosomal loci have been shown to move sub-diffusively with $\alpha \approx 0.4$ (Weber et al., 2010; Javer et al., 2013), as might be expected for such a large molecule compacted between 10^3- to 10^4-fold compared to the equivalent linear DNA (Sherratt, 2003). For chromosomal loci measured with a lag time of one second the MSD $\approx 10^{-2}\,\mu m^2$ (Weber et al., 2010). In contrast, the mobility of a single protein not interacting with DNA can be several orders of magnitude higher (Elmore et al., 2005; Uphoff et al., 2013). For example, estimates of the diffusion coefficients of freely diffusing unconjugated fluorescent proteins in *E. coli* range between $7\,\mu m^2/s$ (Kumar et al., 2010), and $10\,\mu m^2/s$ (Sanamrad et al., 2014). The mobility of proteins bound to DNA therefore generally have a significantly lower mobility than those that are freely diffusing.

3.3 Imaging DNA binding

The difference in mobility between a protein specifically bound to DNA and one that is diffusing and binding only transiently can be exploited by several techniques to distinguish these two different populations.

Fluorescence recovery after photobleaching

Fluorescence recovery after photobleaching (FRAP), measures the time it takes for fluorescent molecules to diffuse into a previously photobleached area

(Reits and Neefjes, 2001). A small spot on a cell is bleached using a pulse from a high-powered focused laser beam. Subsequent diffusion of surrounding non-bleached fluorescent molecules into the bleached area leads to a recovery of fluorescence, which is recorded at low laser power and the rate or recovery can be used to establish the mobility of the protein being studied. The fraction of pre-bleached fluorescence intensity recovered can be used to establish the fraction of immobile molecules, as these molecules remain in the bleached area. This has been used to study fluorescently labelled RNAP in *E. coli* (Bratton et al., 2011). Averaged over many cells, 53 % ± 19 % of the RNAP molecules were mobile on the 3 s time-scale, and the remaining 47 % were immobile even on the 30 s time-scale.

While FRAP can be used to great effect, it is limited by the fact that it is an ensemble measurement. For samples with multiple molecular species with different diffusion behaviour, care must be taken in the model assumptions used to interpret the data (Lippincott-Schwartz et al., 2001). FRAP is also a low throughput technique which images a single spot in the cell at any given time. *In vivo* single-molecule techniques can overcome some of these limitations, by imaging many cells simultaneously (Uphoff et al., 2013). Most importantly, they provides information about each molecule and thus reveals information about heterogeneity in the sample, and rely less on pre-defined models for interpreting the data.

Detection by immobilisation

One of the first *in vivo* single-molecule techniques was detection by immobilisation (Xie et al., 2008).[1] This technique relies on the fact that, at long exposure times, the image produced by fast moving molecules is motion blurred into a broad spot covering a large fraction of the cell, whereas stationary molecules still appear as a

[1] This technique is also referred to as detection by localisation.

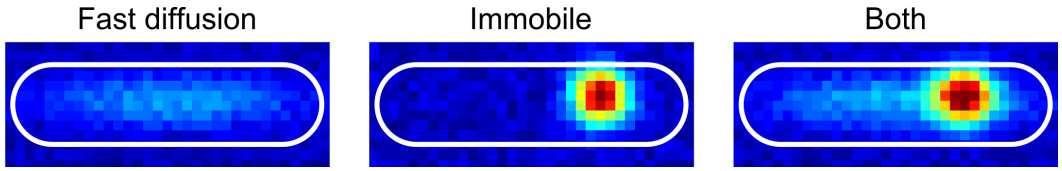


Figure 3.2: **Detection by immobilisation.** Immobile molecules can be distinguished from diffusing molecules by imaging at long exposure times, so that the fluorescent spot from fast diffusing molecules appears broader due to motion blurring, whereas bound molecules still appear as diffraction-limited point sources.

point source (Figure 3.2). For very long exposure times, the image from diffusing molecules is overwhelmed by the cellular autofluorescence, and only the signal from immobile molecules is detected (Yu et al., 2006). At somewhat shorter times both immobile and mobile molecules can be detected. When these fluorescent spots are fitted to a Gaussian, the width of the fit can be used to distinguish between proteins stably bound to DNA and diffusing proteins (Elf et al., 2007; Uphoff et al., 2013). This approach has been used to study the association and dissociation of transcription factors (Hammar et al., 2012, 2014).

Photoactivated single-molecule tracking

Tracking the movement of molecules in live cells is a powerful approach that in principle allows directly observing the kinetics and location of protein activities. Single-particle tracking (SPT) algorithms join together the positions of molecules over a series of images to form trajectories. The density of labelled molecules should be low so that their PSFs are resolved for localisation analysis and trajectories do not cross. This requirement limited early studies in bacterial cells to artificially low protein copy numbers (Elf et al., 2007).

Combining single-particle tracking with the strategy of photoactivation central to PALM allows many molecules to be tracked sequentially (Manley et al., 2008), making it possible to study a much broader range of biological problems (Figure 3.3 A). As in

typical PALM studies, single molecules are sparsely photoactivated and imaged for a number of frames. One of the key limitations of *in vivo* single-molecule techniques is photobleaching. Trajectories of single fluorescent proteins are typically limited to only a four or five frames before photobleaching, which makes observing processes with slow kinetics more difficult. Compared to ordinary PALM, lower excitation intensities are used to allow molecules to be tracked for a longer duration at the cost of decreased localisation precision.

Trajectories generated from single-molecule tracking experiments can be analysed in several ways. The ensemble mean squared displacement (*MSD*) of many trajectories at different lag times can be used to determine if the diffusion is Brownian or sub-diffusive, and the slope of the plot can be used to determine the mobility (Figure 3.3 B) . This can be used to determine if the population is generally slow or fast moving (English et al., 2011), however it becomes harder to interpret if multiple species with different diffusive behaviour are present. Alternatively, the distribution of displacements for molecules taken at one particular lag time can be plotted as an empirical cumulative distribution function (CDF) (Figure 3.3 C). The resulting curve can be fitted to an analytical expression for the CDF (obtained by integrating Equation 3.2) for single or multiple diffusing species to extract estimates for the diffusion coefficients of these molecules and determine the fractions of molecules in different states (Schütz et al., 1997).

Finally, an 'apparent' or 'nominal' diffusion coefficient (D^*) can be determined for each trajectory based on the single-step MSD (Uphoff et al., 2013). Generating a D^* value for each trajectory has two advantages; firstly, the mobility information remains linked to the spatial information for each molecule, secondly, different populations with a significant difference in mobility appear as resolvable peaks when the probability density of D^* values is plotted (Figure 3.3 D), allowing intuitive inter-

pretation of the data (Uphoff et al., 2013; Stracy et al., 2015). Using RNA polymerase as an example we will show that this approach can be generally applied to any DNA binding protein, and used to discriminate between mobile molecules and molecules specifically bound to DNA.

3.4 Tracking and categorisation of RNA polymerase

To track RNA polymerase molecules, we used an endogenous fusion of photoactivatable fluorescent protein PAmCherry (Subach et al., 2009) with the beta' subunit of RNAP (Endesfelder et al., 2013). As the beta' subunit is the last subunit in the assembly sequence of the core enzyme it is incorporated very quickly (1.5 min) into functional RNAP (Saitoh and Ishihama, 1977). The maturation time of PAmCherry has been measured at 23 min (Subach et al., 2009), hence almost all mature β'-PAmCherry fluorophores should be incorporated into core or holoenzymes.

PALM microscopy was performed on the custom-built TIRF microscope detailed in Chapter 2, in near-TIRF illumination. Live cells were imaged under continuous 5 mW 561 nm excitation at 15.48 ms per frame for 15,000 frames. PAmCherry photoactivation was controlled by adjusting the 405 nm excitation between 0 nW to 10 nW to limit the activated molecules to no more than one per cell.

To follow the movement of single molecules over multiple frames, we used an adapted version of the MATLAB implementation (Uphoff et al., 2013) of the single-particle tracking algorithm described in (Crocker and Grier, 1996). Positions were linked to a track if they appeared in consecutive frames within a window of 5 pixels (0.48 µm); this window size was chosen as > 1 % of steps are larger than this value, given the diffusion of RNAP. In rare cases when multiple localisations fell within the tracking radius, tracks were linked such that the sum of step distances was minimized.

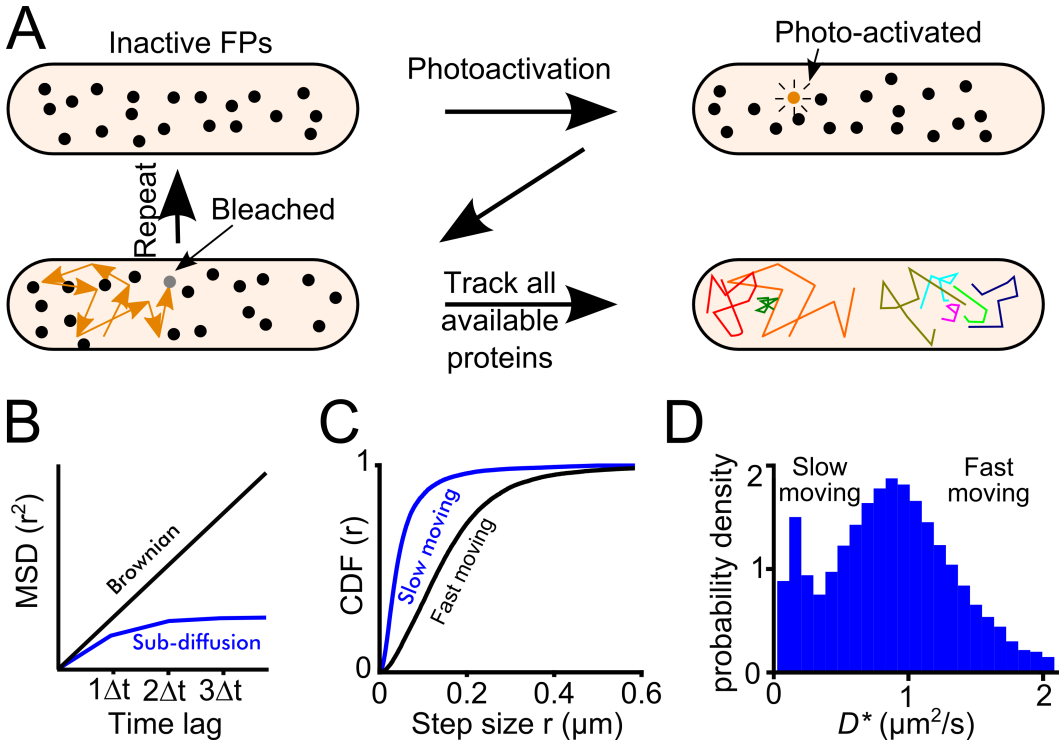


Figure 3.3: **Photoactivated single-molecule tracking and analysis.** A) Individual molecules are photoactivated using 405 nm light, then imaged for a number of frames until they photobleach. This process is repeated as other molecules are stochastically activated until all molecules of interest have been imaged. The fluorescence images from each frame are fitted to a Gaussian to generate super-resolved localisations all imaged molecules. The molecule localisations can be linked together to form a trajectory for each molecule. B) There are several ways to represent and analyse diffusion of tracked molecules. The mean squared displacement (MSD) of trajectories can be measured at different lag times (Δt) to infer diffusion characteristics. For molecules undergoing Brownian diffusion, MSDs increase linearly with lag time, whereas a plateau in the MSD curve is characteristic for sub-diffusion of molecules whose motion is confined. C) The distribution of displacements for a given lag time can be plotted as an empirical cumulative distribution function (CDF). Slowly moving molecules have, on average, smaller displacements than fast moving molecules. D) An apparent diffusion coefficient (D^*) can be calculated from the single-step MSD of each trajectory. The probability density of these D^* values can highlight slower (low D^*) and faster (high D^*) moving populations.

We used a memory parameter of 1 frame to allow for transient disappearance of the PSF within a track due to blinking or missed localisation.

An apparent diffusion coefficient, D^*, was calculated for each trajectory based on the single-step MSD using:

$$D^* = \frac{1}{4n\Delta t}\sum_{i=1}^{n}[x(i\Delta t) - x(i\Delta t + \Delta t)]^2 + \left[y(i\Delta t) - y(i\Delta t + \Delta t)\right]^2 \tag{3.5}$$

Where $x(t)$ and $y(t)$ are the coordinates of the molecule at time t, the frame time of the camera is Δt, and n is the number of steps in the trajectory. Tracks shorter than $n = 4$ have a larger statistical uncertainty and are discarded for this analysis. For a molecule with diffusion coefficient D, the probability distribution of obtaining a D^* value, ρ, is given by:

$$f(\rho; D, n) = \frac{(\frac{n}{D})^n \rho^{n-1} e^{\frac{-n\rho}{D}}}{(n-1)!} \tag{3.6}$$

Where n is the number of steps in the trajectory. The D value can therefore be obtained from the expectation value of the probability distribution of individual D^* values. To allow the experimental distribution to be fitted to this analytical expression tracks of less than 4 steps are not used, and tracks of more than 4 steps are truncated at the 4th step. The expression then becomes:

$$f(\rho; D) = \frac{(\frac{4}{D})^4 \rho^3 e^{\frac{-4\rho}{D}}}{6} \tag{3.7}$$

For D^* distributions with more than one species, more terms can be added:

$$f(\rho; D_1, A, D_1, D_2) = \frac{A(\frac{4}{D_1})^4 \rho^3 e^{\frac{-4\rho}{D_1}}}{6} + \frac{(1-A)(\frac{4}{D_2})^4 \rho^3 e^{\frac{-4\rho}{D_2}}}{6} \tag{3.8}$$

RNAP can be either specifically bound to DNA (while interacting with promoter regions or with transcribed genes during transcription elongation), or diffusing through the nucleoid searching for promoters to initiate transcription. The time to open a promoter and transcribe a gene (between 20 s and 100 s; Bremer and Dennis, 1996) is at least 500-fold longer than *in vitro* estimates of the non-specific DNA dwell times (30 ms; Wang et al., 2013). The population should therefore contain at least two species with different mobility, which was confirmed by the poor least squares fit of Equation 3.7 to the histogram of D^* values (Figure 3.4 A).

The localisation uncertainty in each measurement, σ_{loc}, manifests itself as a positive offset in the D^* value of ${\sigma_{loc}}^2/\Delta t$ (Michalet, 2010). The localisation uncertainty of ~40 nm for our measurements, estimated from the y intercept of the MSD plots, therefore result in a positive shift in the mean D^* value of immobile molecules to ~0.1 μm^2/s. We verified this by using a well characterised control protein DNA polymerase I (Pol I) which shows clearly distinct D^* populations for molecules specifically bound to DNA and those searching the chromosome for substrate (Figure 3.4 B). Inducing DNA methylation damage by incubating cells with MMS increases the fraction of specifically bound molecules (described previously by Uphoff et al., 2013), making it easier to resolve the distribution of bound molecules. We fitted this distribution to two species model (using Equation 3.8), which showed that immobile molecules specifically bound to DNA, had a D_{imm} of 0.11 μm^2/s.

Using this D_{imm} value for DNA-bound molecules to constrain one species and allowing a second unconstrained species fits well to the distribution of RNAP D^* values giving two populations with 48 % ± 1 % of RNAP molecules immobile and 52 % mobile diffusing at D = 0.36 μm^2/s (Figure 3.4 C). The fraction of immobile RNAP molecules agreed well with the ~47 % molecules immobile on the ~30 s timescale previously observed with FRAP, and agrees broadly with estimates of the fraction of

immobile molecules from single particle tracking data of 54 % ± 7 % (Bakshi et al., 2013).[2]

Distinguishing transcribing RNAPs

To sort individual molecules as immobile or mobile we determined a threshold of $D^* = 0.16\,\mu m^2/s$, which preserves the ratio established from fitting. We also introduced a second threshold using the D^* value calculated from the 2Δ t lag times. Molecules showing D^* values below both thresholds were categorised as bound, whereas molecules with D^* values above both thresholds we categorised as mobile; this sorting method gives an estimated categorisation precision of ~90 %. In order to minimize the uncertainty when determining the D^* for categorising individual molecules, only trajectories with at least 4 steps were used, and all longer steps were included.

To check whether most immobile RNAPs are indeed actively transcribing, we treated the cells with the antibiotic rifampicin, which binds to RNAP and blocks transcription beyond a three nucleotide RNA, but does not affect promoter binding, open complex formation, or transcription by RNAPs already in transcription elongation (Campbell et al., 2001). Incubation with rifampicin for 30 min is thus expected to cause most RNAPs to become mobile, with only promoter-bound RNAPs remaining bound to DNA, since any elongating RNAP will complete transcription and dissociate from the DNA. Consistent with these expectations, the rifampicin treatment led to a clear increase (from 52 % to 84 %) in the fraction of mobile RNAPs (Figure 3.6 A-D). In contrast, the rifampicin treatment did not change the fraction of immobile Pol I, confirming that the effect seen for RNAP is specific to blocking active transcription

[2] Estimates determined by Bakshi et al. (2013) were from cells grown in rich media compared to minimal media shown here. As detailed in Chapter 5, we see a larger fraction of transcribing RNAPs when cells are grown in rich media.

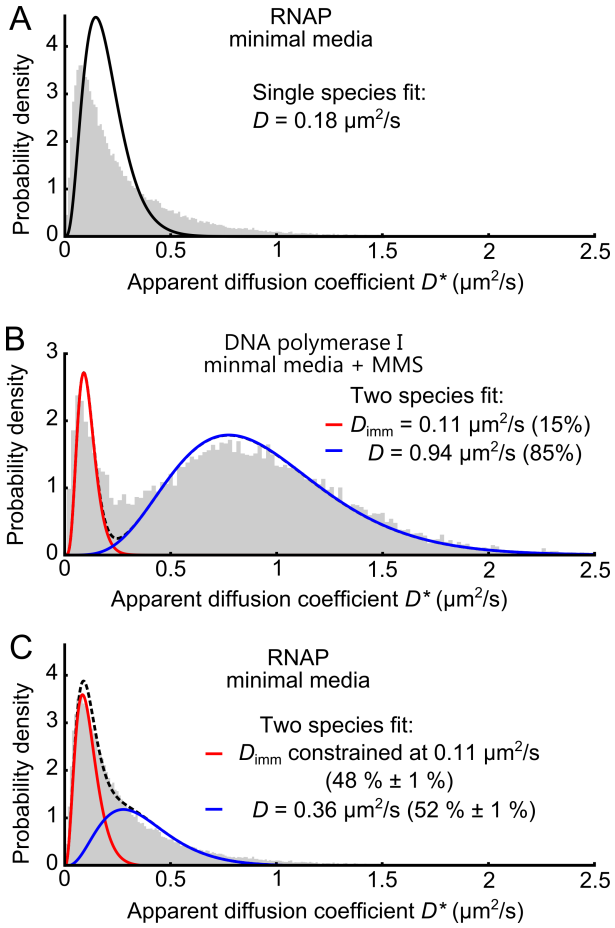


Figure 3.4: **Immobile and mobile RNA polymerase** A) The distribution of RNAP D^* values fits poorly to a single diffusing species, consistent with RNAP being present in two different mobility states in the cell. B) DNA polymerase I (Pol I) was used as a control to determine the D^* value of DNA-bound molecules. When DNA-methylation damage is induced by growing cells for an hour in minimal media supplemented with 100 mM methyl methanesulfonate (MMS) the fraction of bound molecules increases significantly and is clearly resolvable from the pool of mobile molecules (Uphoff et al., 2013). Fitting this distribution with two species model gives D value for molecules specifically bound to DNA of $D_{\text{imm}} = D = 0.11 \pm 0.01\,\mu\text{m}^2/\text{s}$, in good agreement with the estimate for immobile molecules imaged with a localisation error of 40 nm. Range indicates 95 % confidence interval. C) Using this D_{imm} value to constrain one species and allowing a second unconstrained species fits well to the RNAP distribution giving two populations of 48 % ± 1 % immobile and 52 % ± 1 % mobile with $D = D = 0.36 \pm 0.02\,\mu\text{m}^2/\text{s}$.

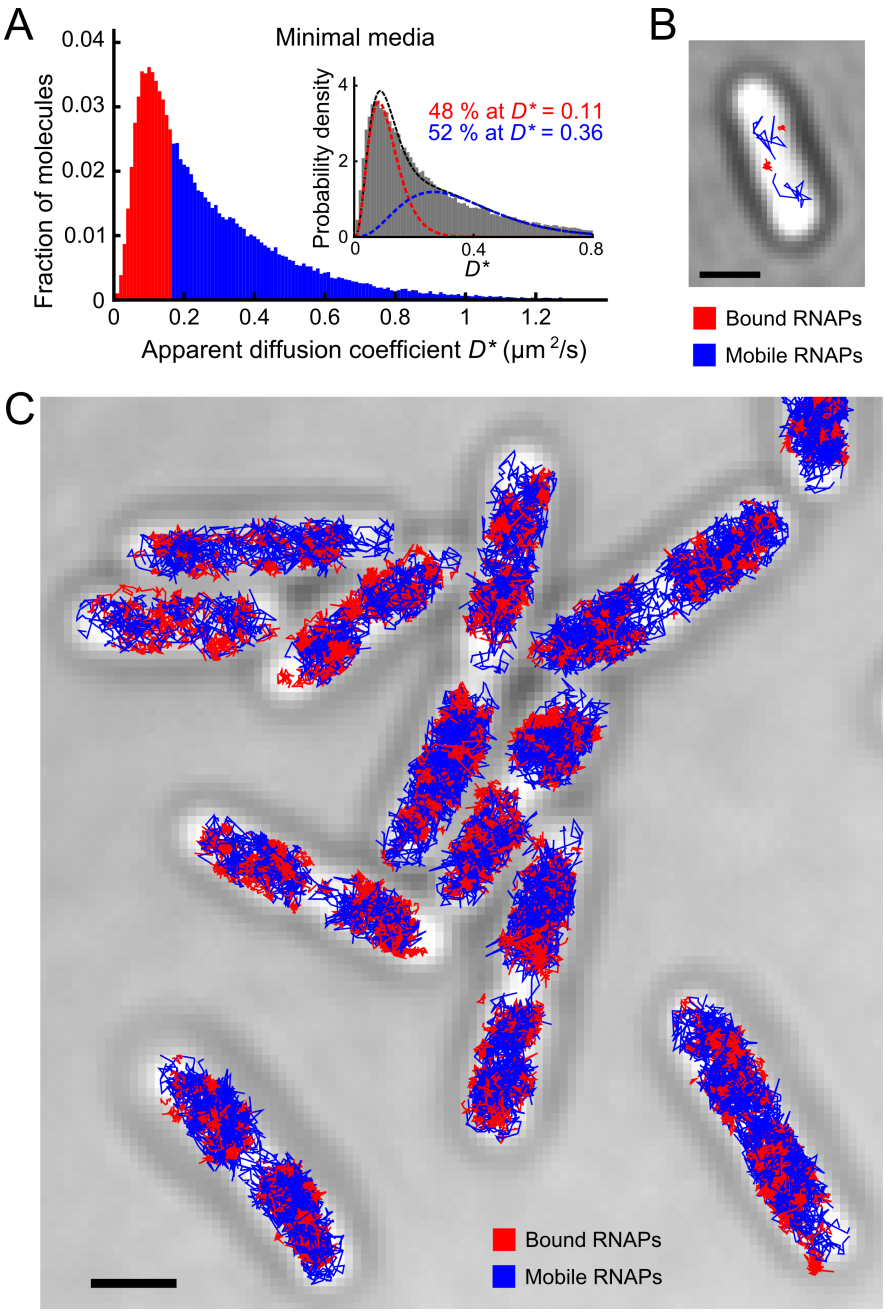


Figure 3.5: **Sorting individual RNAP molecules.** A) Fitting to the distribution of D^* values for RNAP gives a ratio of 48 % bound to 52 % diffusing (inset). A D^* threshold of 0.16 μm^2/s allows individual molecules to be sorted as bound (red) and mobile (blue) with this ratio preserved. B) Example trajectories of individual RNAPs with immobile molecules coloured red and mobile molecules coloured blue. Scale bars: 1 μm. C) An example field of view shows the intracellular spatial distribution of categorised RNAP trajectories.

(Figure 3.6 E and F). Given our estimates for the average RNAP copy number per cell, the bound fraction in the rifampicin-treated cells corresponds to ~430 bound RNAPs per cell. This estimate is in good agreement with results from ChIP experiments, which showed that RNAP remains associated with ~530 promoters per chromosome after rifampicin incubation (Grainger et al., 2005).

The rifampicin treatment also increased the apparent diffusion of mobile RNAPs (D^* increases from 0.36 $\mu m^2/s$ to 0.64 $\mu m^2/s$; Figure 3.6 A and B). This is likely the result of global chromosome decompaction triggered by rifampicin treatment (Cabrera et al., 2009). Chromosome decompaction is expected to increase the confinement volume for RNAP molecules, resulting in mobile RNAPs moving between more distal DNA strands, hence resulting in overall faster apparent diffusion. In agreement with this interpretation, our control protein Pol I also showed increased mobility of its mobile population upon rifampicin treatment (D^* increases from 1.0 $\mu m^2/s$ to 1.4 $\mu m^2/s$; Figure 3.6 E and F).

3.5 Counting protein copy numbers

With PALM imaging and tracking, individual photoactivation events each represent a single molecule, which naturally allows counting protein copy numbers in a cell. Long PALM movies were taken until no further photoactivation was visible (up to 50,000 frames). Over the course of the movie photoactivation intensity was controlled to ensure only single active molecules per cell. Matlab software MicrobeTracker (Sliusarenko et al., 2011) was used to automatically segment cell outlines based on the transmitted light images, and tracking was on localisations falling within this boundary. For RNAP the mean RNAP copy number per cell is 2710 ± 700, which scaled with cell length as expected.

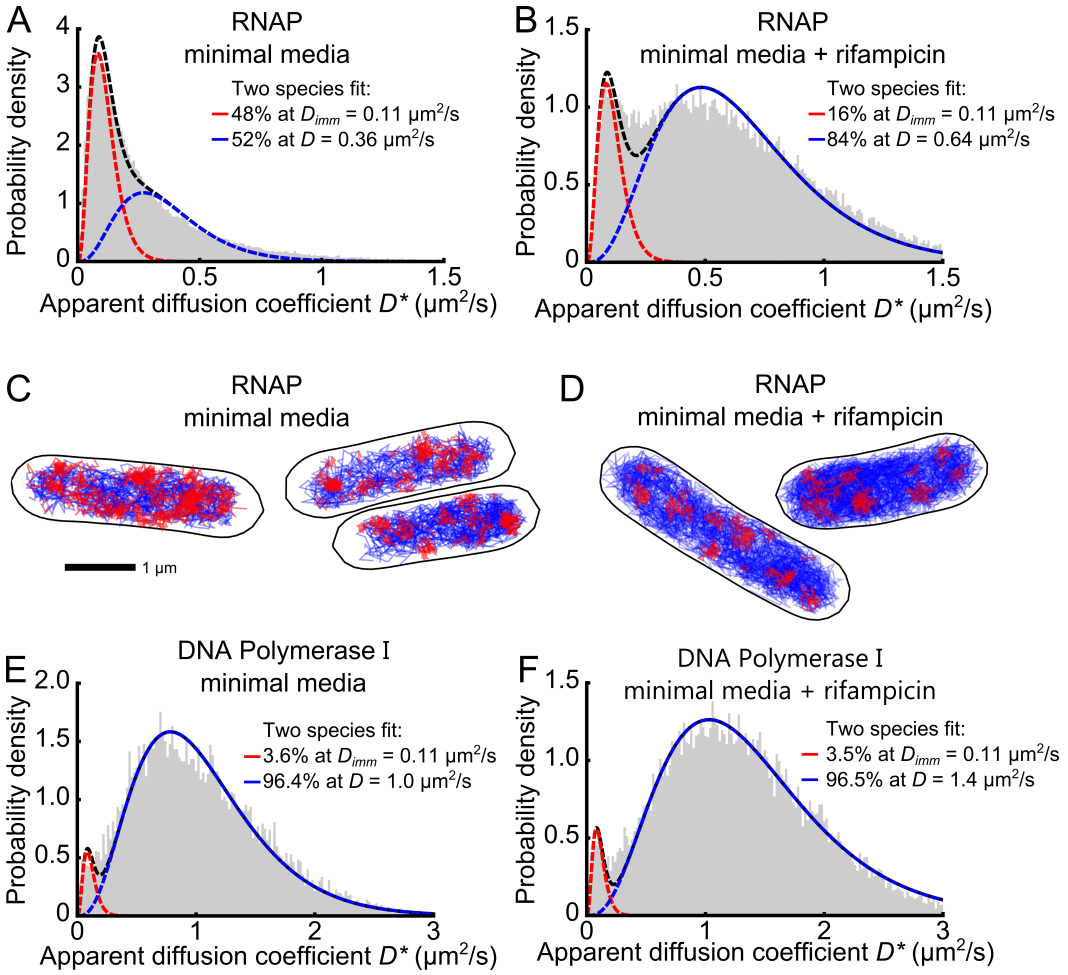


Figure 3.6: **Blocking transcription initiation** A) The fitted distribution of D^* values for RNAP growing in minimal media (from Figure 3.4 C). B) Fitting to the RNAP distribution after rifampicin incubation gives a 3-fold reduction in the fraction of bound molecules to 16 %. The D^* of mobile molecules increases 75 % from $D = 0.36 \pm 0.02\,\mu m^2/s$ to $D = 0.64 \pm 0.03\,\mu m^2/s$. Ranges indicate 95 % confidence intervals. C) Example cells showing immobile RNAP trajectories coloured red and mobile trajectories coloured blue. Scale bar 1 μm. D) Example cells after treatment with rifampicin. E) The fitted distribution of D^* for pol I In unperturbed cells. F) After rifampicin incubation, the bound fraction of Pol I shows negligible change (3.6 % to 3.5 %). The D^* for mobile population of Pol I also increases after rifampicin incubation (40 % increase from $D = 1.05 \pm 0.03\,\mu m^2/s$ to $D = 1.37 \pm 0.03\,\mu m^2/s$.

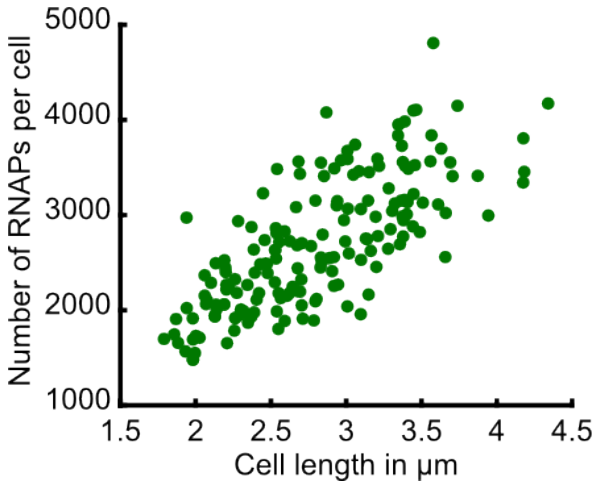


Figure 3.7: **RNAP copy number** The mean RNAP copy number per cell is 2710 ± 700. Copy number calculated from taking long PALM movies (typically 40 000 frames) with low photoactivation levels until all proteins have photobleached.

However, several sources of overcounting and undercounting can effect this number. Photoactivatable proteins may blink after activation; they transition from photoactive to dark states reversibly (Lee et al., 2012). This effect, as well as molecules which transiently disappear as they diffuse in and out of the area of illumination, can cause single molecules to be counted multiple times. We chose to label with PAmCherry as it appeared to show little blinking in comparison to other photoactivatable FPs such as Meos2 (Endesfelder et al., 2013). Nevertheless, to overcome these potential problems, we allowed molecules to be tracked over one dark frame to account for any blinking or transient loss of localisation. On the other hand, sources of undercounting include the presence of unfolded and immature fluorescent proteins. For PAmCherry it has been estimated that up to 50 % of fluorophores do not become fully mature (Durisic et al., 2014), although this characterisation was performed in Xenopus oocytes, and the maturation in bacteria may be very different. We have therefore not applied any correction of undercouting in this research, but we note that the true copy numbers may be up to two-fold higher than those reported here, although relative values should remain unaffected by this. Nevertheless, the

copy number we measured for RNAP is comparable to previous estimates ranging from 1,500 - 5,000 molecules per cell for cells growing in similar growth conditions (Grigorova et al., 2006; Bakshi et al., 2012; Endesfelder et al., 2013).

3.6 Custom analysis software

In this chapter we have demonstrated a method for partitioning a protein into different diffusive states, and determining a threshold for separating individual molecules as DNA-bound or mobile. The tools described in this chapter, as well as several other analysis methods used in this thesis, such as clustering and radial distribution analysis, were compiled into a comprehensive software package (Figure 3.8), which also incorporates single particle tracking code previously developed (Uphoff et al., 2013; Crocker and Grier, 1996), thus allowing these methods to be easily adopted for analysis of other systems. This software package integrates all these features with cell segmentation outlines generated through MicrobeTracker software allowing copy numbers to be easily determined. Moreover, segmentation also allows cells to be binned by length, which is a reasonable proxy for position it the cell cycle, allowing all the various analytical tools to be performed on particular selections of cells. This facilitates much of the analysis in the following chapters.

3.7 Materials and methods

Bacterial strains

The PAmCherry fusion strain of RNAP, expressed from its native promoter in *E. coli* MG1655 is previously characterised by Endesfelder et al. (2013). The DNA polymerase I fusion to PAmCherry is described by Uphoff et al. (2013) and was moved into MG1655 for this work by P1 transduction.

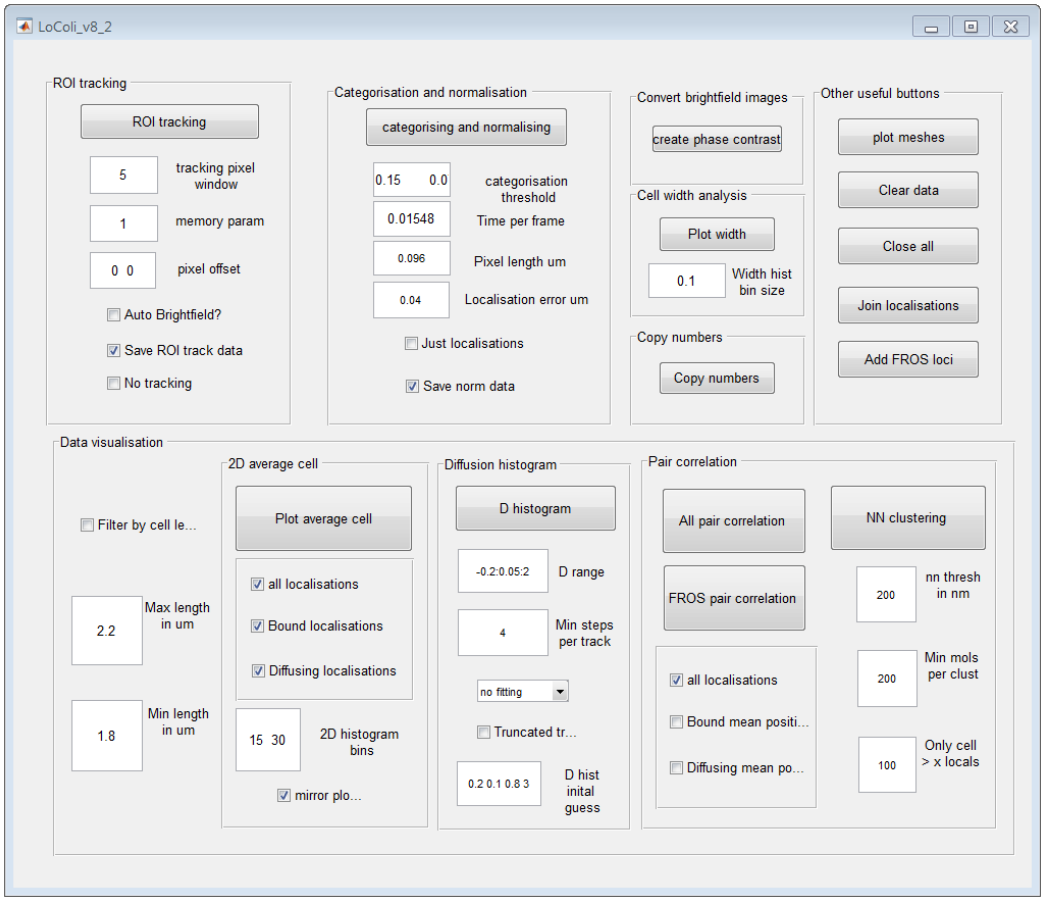


Figure 3.8: **LoColi analysis suite.** Screenshot of the LoColi software developed during this thesis for analysis of single-molecule localisation data in *E. coli*.

Cell preparation

Strains were streaked onto LB plates with 25 µg/ml ampicillin or 50 µg/ml kanamycin. Single colonies were inoculated into LB and grown at 37 °C for 4 h, then diluted into M9 or rich defined medium (RDM) (EZRDM, Teknova) supplemented with 0.2 % glucose and grown overnight at 37 °C. Overnight cultures were diluted and grown for > 2 h at 37 °C to early exponential phase (OD < 0.2). Cells were centrifuged and immobilised on 1 % low fluorescence agarose (Biorad) pads. Glass coverslips (1.5 thickness) were heated to 500 °C in a furnace for 1 h to remove any fluorescent background particles.

Microscopy and analysis

PALM microscopy was performed on the same custom-built TIRF microscope, using the localisation algorithm detailed in Chapter 2. All analysis described in this chapter was written in Matlab (Mathworks).



Chapter 4

RNA polymerase and its role in nucleoid organisation

Despite the fundamental importance of transcription, a comprehensive analysis of RNAP behaviour and its role in the nucleoid organisation *in vivo* is lacking. Here we use the techniques described in the previous chapters to study the localisation and dynamics of the transcription machinery and DNA in live bacterial cells, both at the single-molecule and the population level. Discriminating specifically bound RNAPs shows that low levels of transcription can occur throughout the nucleoid, but clustering analysis and 3D Structured Illumination Microscopy (SIM) showed that dense clusters of transcribing RNAPs form almost exclusively at the nucleoid periphery. In faster growth conditions, the level of RNAP clustering increases, as does the separation between the densest clusters of RNAPs and the densest regions of the nucleoid, showing that transcription can cause spatial reorganisation of the nucleoid.

4.1 Introduction

Unlike eukaryotes, cellular functions in bacteria are not compartmentalised, yet they are still highly organised, with specific subcellular regions associated with specific machineries. In *Escherichia coli*, many of these machineries operate on the bacterial nucleoid, a highly structured and dynamic object segregated from the cytoplasm. Transcription is one of the most abundant processes occurring on the

nucleoid DNA; driven by thousands of RNAP molecules per cell, and plays a crucial role in maintaining both global and local DNA organisation. For example, blocking cellular transcription with the antibiotic rifampicin causes loss of global organisation reflected by nucleoid expansion (Dworsky and Schaechter, 1973; Cabrera et al., 2009). Transcription has also been shown to modify the local DNA topology by unwinding the DNA double helix (Moulin et al., 2005). Despite this, little is known about the spatial distribution of active RNAPs in living bacterial cells and how they affect the organisation of DNA and the nucleoid.

The nucleoid structure depends on growth conditions. Under slow growth conditions, where each cell contains between one and two chromosomes, the nucleoid lacks observable structures and appears diffuse, occupying most of the cellular volume. By contrast, under fast growth conditions, where the cells contain between four and eight chromosome equivalents, the nucleoid is highly structured and displays dramatic variation in local DNA density (Hadizadeh Yazdi et al., 2012). Growth conditions also influence transcriptional activity on different genes; for example, the overall rate of synthesis of ribosomal RNA increases ~40-fold in fast-growth compared to slow-growth conditions (Bremer and Dennis, 1996).This change in the level of expression is reflected in large changes in the spatial distribution of RNAP: under slow growth conditions, RNAP appears to be fairly homogeneously distributed over the diffuse nucleoid (Cabrera and Jin, 2003), whereas under fast growth conditions dense clusters of RNAPs emerge (Cabrera and Jin, 2003; Endesfelder et al., 2013). These dense clusters have been likened to transcription factories in eukaryotic cells, where a single site contains multiple RNAPs active on different genes (Cook, 1999; Jin and Cabrera, 2006); however, they have only been reported in chemically fixed cells and little is known about their presence, dynamics or behaviour in live E. coli.

Transcription and nucleoid structure are also interconnected through the cou-

pling of transcription and translation. In bacteria, ribosomes can bind to nascent mRNA as soon as the ribosome binding site emerges from the transcribing RNAP (Proshkin et al., 2010).This raises an intriguing puzzle, since ribosomes and DNA are spatially separated in *E. coli* (Hobot et al., 1985; Wang et al., 2011; Bakshi et al., 2012), whereas RNAP is strongly associated with DNA (Cabrera and Jin, 2003; Bakshi et al., 2012).To reconcile this with co-transcriptional translation, it has been hypothesised that much of cellular transcription must happen at the surface of the nucleoid, and close to the pool of ribosomes (Jin et al., 2013). However, evidence for this remains weak, since studies under moderate growth conditions did not reveal any RNAP enrichment at the nucleoid periphery (Bakshi et al., 2012).

In the previous chapter we showed that individual RNAP molecules can be categorised based on their mobility, and that the majority of immobile RNAPs are actively transcribing genes. Here, we show that mobile RNAPs are uniformly distributed across the nucleoid, suggesting that all DNA is probed through random non-specific interactions during the promoter search. The distribution of specifically bound RNAPs shows that low levels of transcription can occur throughout the nucleoid. However, clustering analysis and 3D structured-illumination microscopy (3D SIM) show that specifically bound RNAPs are more clustered than the mobile population, and the denser clusters form preferentially at the nucleoid surface, indicating that heavily transcribed genes tend to move out of the bulk of nucleoid DNA. Imaging cells growing in rich media showed that clustering increases, as does the segregation between RNAP and DNA.

4.2 Bound and mobile RNAPs show different spatial distributions

Comparing the spatial location of mobile and bound RNAPs to the distribution of DNA by staining the nucleoid DNA with 488 nm excitable intercalating DNA dye, SYTOX green (Bakshi et al., 2014), showed that the two populations had distinctly different distributions (Figure 4.1 A). The spatial distribution of mobile RNAPs in the cell closely followed that of DNA, with only few excursions of RNAP tracks into the surrounding cytoplasm, indicating that most promoter-searching RNAP molecules remain within the volume of the nucleoid (Figure 4.1 B). In contrast, bound RNAPs were much more heterogeneously distributed, and their distribution did not clearly match to that of DNA (Figure 4.1 C).

To quantify the spatial clustering of mobile and bound RNAPs, we calculated the radial distribution (or pair correlation) function, which gives the relative probability of finding a protein at a given distance away from another protein compared to a random distribution (Sengupta et al., 2011). Only the first localisation in each trajectory was used in the analysis to avoid biasing the results due to the spatial proximity of localisations generated by the same molecule. Cells were segmented from transmitted light images and the pairwise distances for molecules was calculated on a cell by cell basis. Because of the small size of *E. coli*, as the radius r of the pairwise distance increases, much of the area falls outside the cell. To account for this effect, we simulated the randomly distributed points within the same segmented cell outline. See Appendix B for example Matlab code. For each segmented cell, the same number of molecules was simulated as the experimentally observed molecules, and their pairwise distances calculated. This was repeated over all cells and the histogram of pairwise distances was divided by the simulated random pairwise distances to

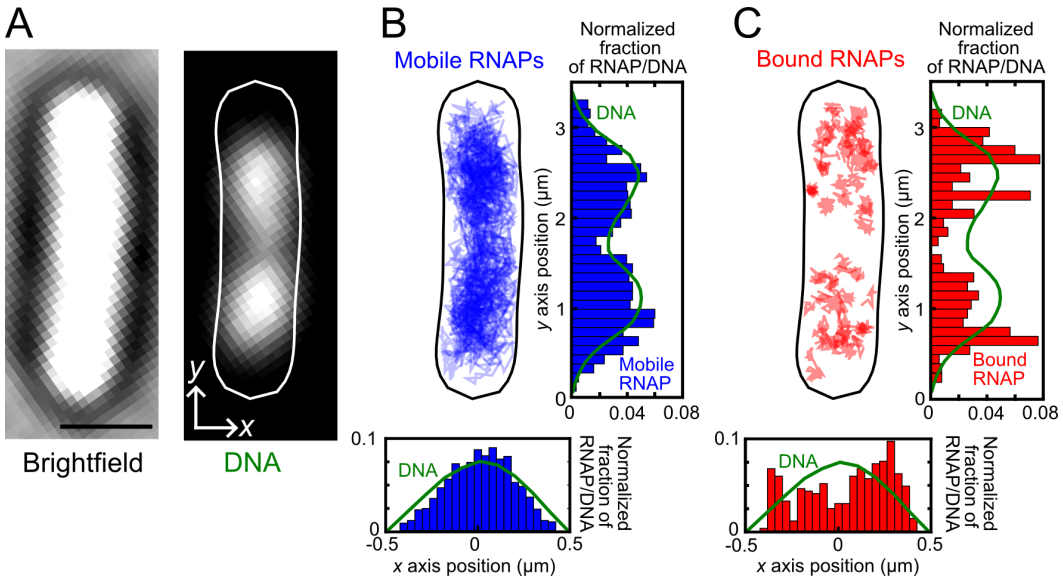


Figure 4.1: **Comparison of categorised RNAPs with the distribution of DNA.** A) A transmission image of an example cell (left; used for cell segmentation), and in a SYTOX fluorescence image (right, showing the nucleoid DNA distribution). x-axis: short cell axis; y-axis: long cell axis. Scale bar: 1 µm. B) The distribution of mobile molecule trajectories (blue lines/bars) closely matched the distribution of DNA (green line) as shown in histogram projections along the x and y cell axes. C) The distribution of bound RNAPs in the same example cell is more heterogeneous, and does not closely follow the DNA distribution.

give the empirical radial distribution function, $g(r)$, with random distribution having a $g(r)$ of one.

The radial distribution function for molecules in > 200 cells shows a clear clustering of bound RNAPs at distances shorter than 150 nm (Figure 4.2), as expected for multiple RNAPs simultaneously bound at highly transcribed genes. This clustering was absent in the case of mobile molecules, pointing to their independent, non-specific RNAP interactions throughout the nucleoid.

As an additional verification that mobile RNAPs remain strongly associated with the nucleoid, we compared the spatial distribution of mobile RNAPs to that of nucleoid-associated Heat Unstable protein (HU), which reports on nucleoid DNA density (Fisher et al., 2013). A stain with an endogenous replacement of HU with

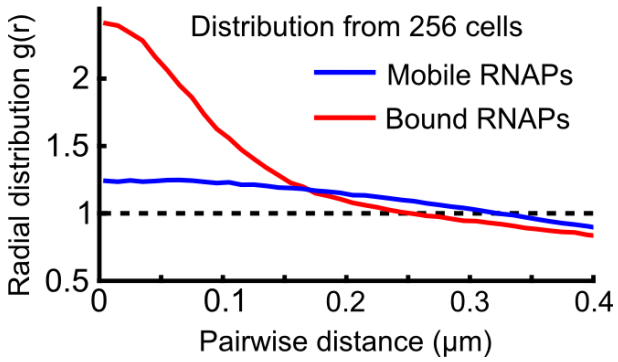


Figure 4.2: **Radial distribution of mobile and bound RNAP.** Radial distribution of bound molecules (from 256 cells) shows a more clustered distribution than mobile molecules.

HU-PAmCherry was imaged in the same way as for RNAP. Cells were segmented from transmission microscopy images using MicrobeTracker, giving a cell outline and a cell midline (Sliusarenko et al., 2011), and the positions of molecule trajectories were determined relative to the cell midline, with the x-axis defined as the cell short axis and the y-axis defined as the cell long axis. The distances were normalised to 1 and -1 at the cell membrane. To determine the probability distribution across the x-axis short cells (1.6 µm to 2.5 µm long) were used as these cells had a single nucleoid located in the centre of the y-axis (Figure 4.3 A). Only molecules falling within the central 30 % of the y-axis were used to exclude molecules towards the cell poles, which may be at the exterior of the nucleoid along the y-axis, but centrally located along the x-axis. The analytical probability of finding a molecule at position x when the molecules are uniformly distributed through a cylindrical volume, $f_{\text{uni}}(x)$ is given by (Sanamrad et al., 2014):

$$f_{\text{uni}}(x) = \frac{\sqrt{\max(1-x^2, 0)}}{\pi} \tag{4.1}$$

This uniform distribution is shown as the grey dashed line in Figure 4.3 B and C. Plotting the probability density of HU molecules across the x-axis, for > 100 cells,

showed that the DNA density is lowest at the cell periphery (Figure 4.3 B). Defining the periphery as the exterior 50 % of the cell width (yellow highlighted areas) gave only 29 % ± 1 % of HU molecules in this region compared to 39 % expected from a uniform distribution. Performing the same analysis on mobile RNAP molecules showed that they matched extremely well to the distribution of HU, and also have 29 % ± 1 % located in the cell periphery rather than uniformly filling the cell volume.

To verify that the distribution of a protein which does not interact with the nucleoid is uniformly distributed across the x-axis, we imaged unconjugated PAmCherry expressed from a plasmid in wildtype cells. Cells were imaged with 1 ms exposure times with 75 mW excitation. The x-axis distribution is in good agreement with the expected uniform distribution with 38 % ± 2 % of molecules located in periphery.

Together these results confirm our SYTOX-based results and strongly suggest that mobile RNAPs explore the entire nucleoid, having access to even the very dense regions of the nucleoid DNA.

4.3 Transcribing RNAPs tend to cluster at the edge of the nucleoid

Simple visual inspection of the spatial distribution of bound RNAPs showed that they were much more frequently located at the periphery of the cell compared to the mobile population (Figure 4.4 A). To quantify this, we plotted the probability distribution of mobile and bound RNAPs across the cell short-axis (x-axis) through the center of the nucleoid, from > 200 cells (Figure 4.4 B). Bound RNAPs showed a broader distribution than the nucleoid (shown by both the HU distribution and mobile RNAP distribution), with 41 % of them located in the cell periphery (yellow

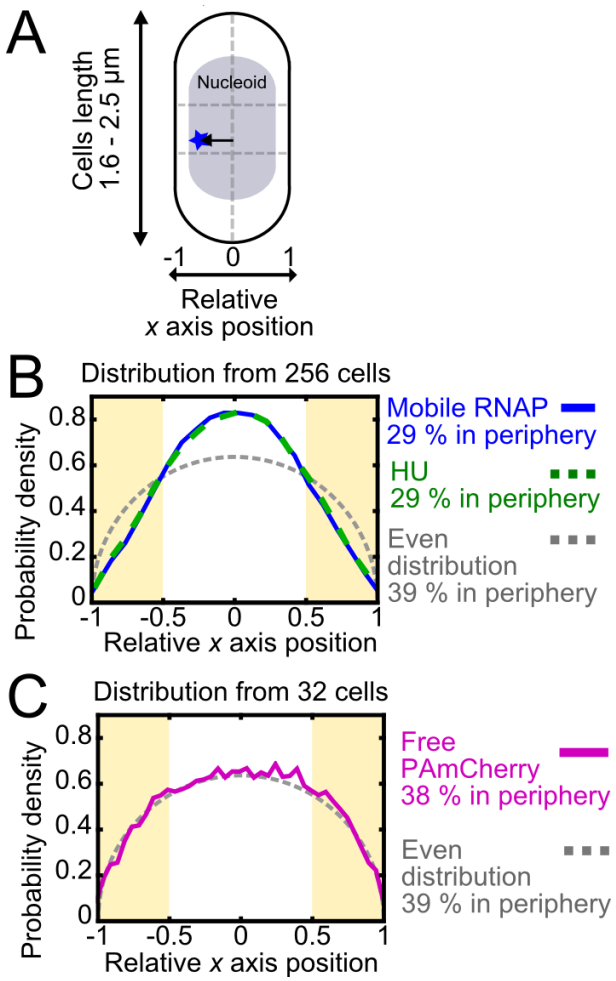


Figure 4.3: **Average spatial distribution of mobile RNAP** A) The average short axis (defined here as the x-axis) distribution of molecules is measured from many cells by taking the relative distance from the cell midline through the centre of the nucleoid, with -1 and 1 representing the cell membrane. Short cells between 1.6 µm to 2.5 µm long were chosen as they have a single nucleoid, centrally located along the y-axis. B) Plot of the x-axis distribution of HU molecules from 256 cells shows the average DNA density is highest at the centre and lowest at the cell periphery, with 29 % ± 1 % of molecules found in the exterior 50 % of the cell width (yellow highlighted area), compared to the expected distribution from molecules evenly occupying the full cylindrical cell volume (39 % expected in cell periphery; dashed grey line). The distribution of mobile RNAP molecules (from 256 cells) matched extremely well to the distribution of HU. C) The probability density distribution of unconjugated PAmCherry across the x-axis for 32 cells. 38 % ± 2 % are located in the cell periphery.

highlighted areas), compared to 29 % for the nucleoid. This broader distribution is evident along the y-axis as well as the x-axis, with bound RNAPs also located at the periphery of the nucleoid towards the cell endcaps, which is highlighted in 2D histograms when the mobile distribution is subtracted from the bound distribution (Figure 4.4 C).

The relative localisation of RNAP and nucleoid DNA was further characterised using 3D Structured Illumination Microscopy (SIM) co-imaging of RNAP-GFP and DAPI-stained DNA in live cells (Figure 4.5 A). Currently, there are no photoactivatable or photoswitchable intercalating DNA dyes for imaging in live cells, making PALM imaging of the nucleoid difficult. Cells can be incubated with Ethynyl deoxyuridine (EdU), a synthetic nucleoside analogue of uridine, which can be incorporated into newly replicated DNA and subsequently click labelled with a fluorophore for super-resolution imaging (Spahn et al., 2014), however this still requires cells to be fixed and permeabilized. SIM, on the other hand, can be used with cell permeable DNA stains to produce a super-resolved visualisation of the nucleoid volume and edges, enabling better assessment of the relative position of the RNAP clusters in three dimensions. Analysis of the SIM images showed that the overall RNAP signal overlaps with the nucleoid; however, the densest RNAP clusters are nearly always located at the nucleoid edge (Figure 4.5 B and C). Projections onto the z-x axis highlight that dense RNAP clusters that appear to be located in the centre of the cell in x-y projections are in fact located at the surface of the nucleoid, where the DNA density is lower.

We confirmed this observation by clustering analysis of DNA-bound RNAPs from PALM data sets. We used a density-based clustering algorithm, DBSCAN (Endesfelder et al., 2013; Ester et al., 1996), which detects clusters based on the local density of points within a search radius, to distinguish non-clustered RNAPs (single and pairs

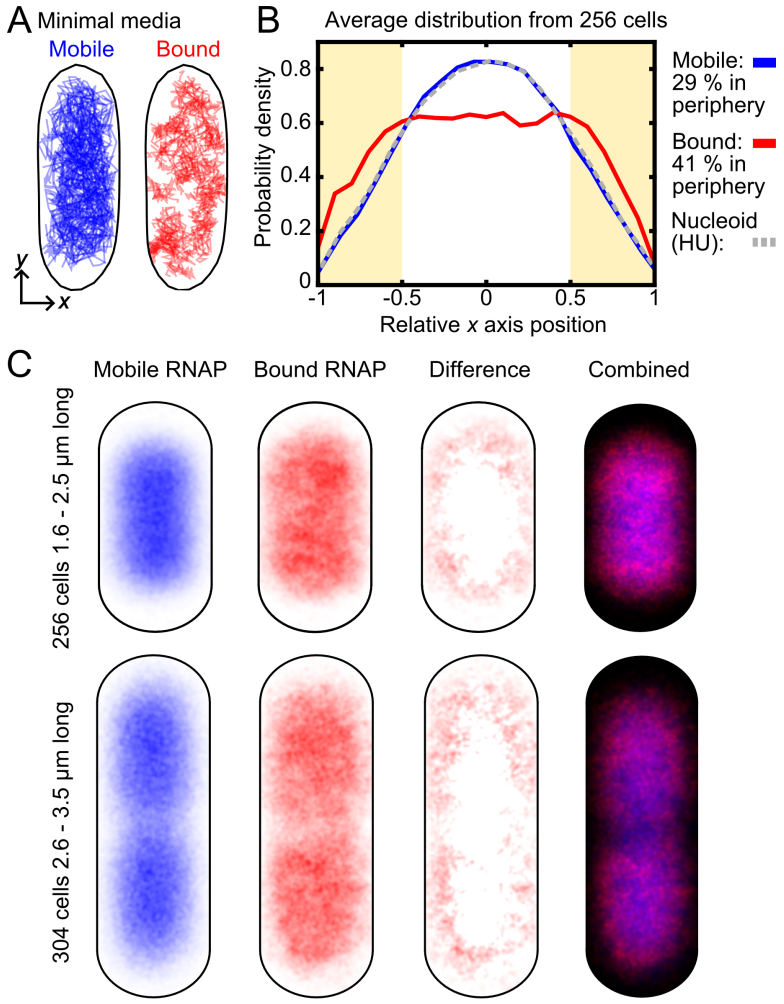


Figure 4.4: **Average spatial distribution of transcribing RNAP** A) An example cell growing in minimal media with trajectories of mobile RNAPs (blue) and bound RNAPs (red). B) Probability density distribution of categorised RNAPs in the nucleoid across the x-axis for 256 cells. Mobile RNAPs (blue line), follow the distribution of the nucleoid (HU distribution; grey line), with 29 % ± 1 % located in the exterior 50 % of the cell width (yellow highlighted area). Transcribing molecules have a significantly broader distribution ($p<0.001$), with 41 % ± 1 % of molecules locating in the periphery. Range indicates standard error of the mean (SEM) of three data sets. C) Normalised 2D histogram plots showing the average spatial distribution of categorised RNAP molecules from many cells. Short cells (1.6 µm to 2.5 µm long) have a single centrally located nucleoid, and longer cells (2.6 µm to 3.5 µm long) have two separate nucleoids. A difference plot showing the distribution of bound RNAPs from which the distribution of mobile RNAPs has been subtracted, highlights the population of bound RNAPs located at the periphery of the nucleoid.

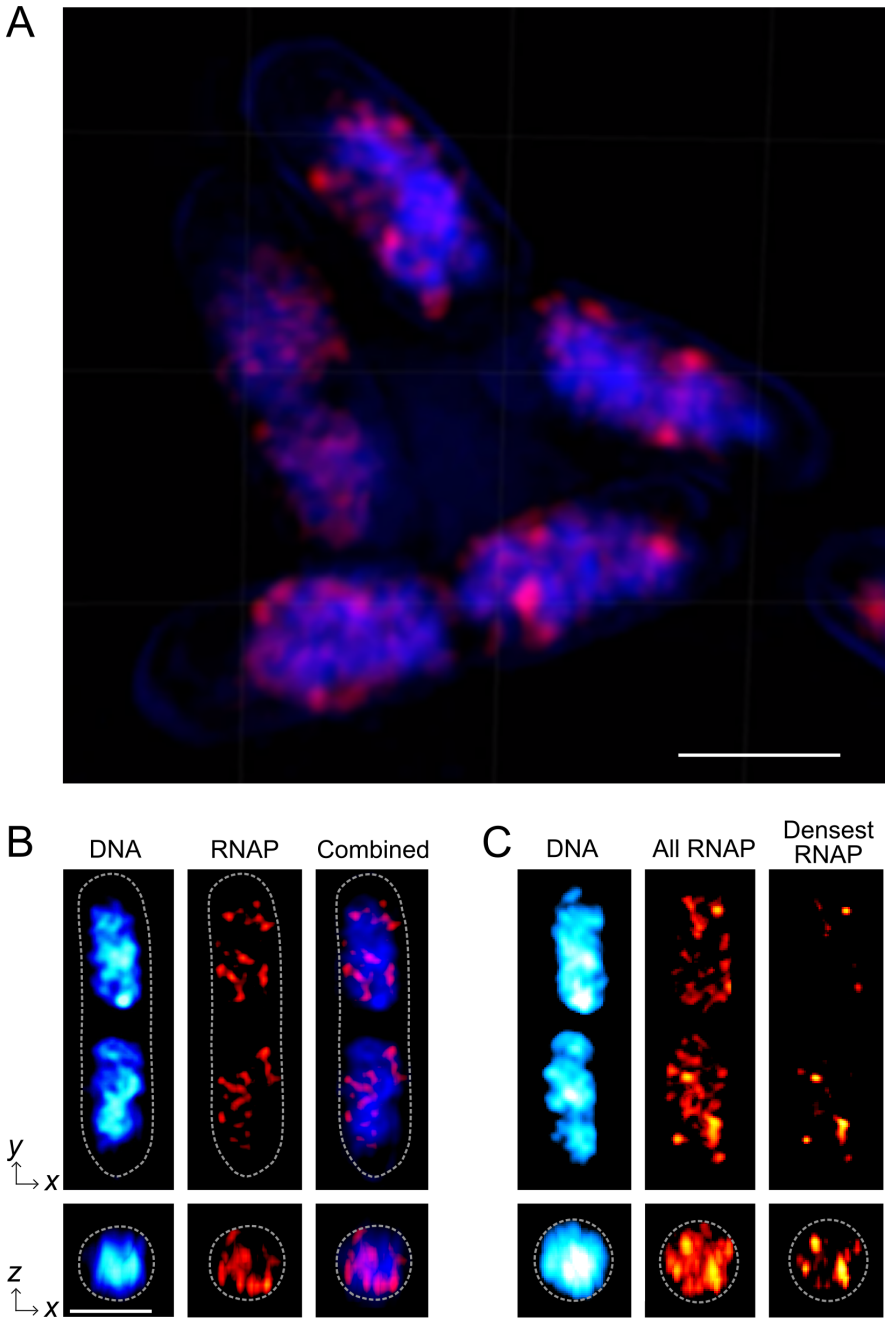


Figure 4.5: **3D structured illumination microscopy images of RNAP and DNA** A) 3D SIM images of live cells in minimal media show that the densest regions of RNAP (red) are located at the edge of the nucleoid (blue). Projections onto the z-x axis highlight that dense regions apparently located in the centre of the cell in x-y projections are in fact above or below the central bulk of the DNA. B) SIM images showing that, while the RNAP distribution frequently overlaps with the DNA distribution, the densest RNAP regions locate at the cell edge.

of RNAPs) from those in dense clusters (> 6 molecules). The DBSCAN algorithm requires two parameters; the neighbourhood radius and the minimum number of points to form a cluster. If an arbitrary molecule contains enough molecules within its neighbourhood radius, a cluster is started. Additional molecules are added if they fall within the neighbourhood radius of the molecules already in the cluster. This process continues until the density-connected cluster is completely found, then the process repeats with a new arbitrary unvisited molecule. To avoid biases in clustering due to variations in density of molecules per cell, the cells were segmented and their density determined from the number of molecules divided by the segmented cell area. Cells below the mean density were not used, and molecules in cells above the mean density were discarded to reduce the density to the mean. It should be noted that the numbers of molecules analysed is only a small subset of the total due to the need to use only trajectories longer than 4 steps for to distinguish immobile molecules, and to remove trajectories for normalising cellular densities. The threshold of > 6 molecules, used to define a cluster RNAPs therefore reflects a much larger group of molecules.

The resulting plot shows that dense clusters of bound RNAPs are clearly located in the cell periphery (71 % in the periphery), whereas non-clustered RNAPs were located throughout the nucleoid (33 % in the periphery) (Figure 4.6 A). This shows that, while low levels of transcription can occur throughout the nucleoid, dense RNAP clusters (i.e. more highly transcribed genes) were located at the nucleoid periphery, exactly at the interface between the DNA and the ribosome-enriched cytoplasm. When we performed the same analysis on another nucleoid associated protein, the histone-like nucleoid structuring (H-NS) protein, a global transcriptional silencer, we saw the opposite trend, with dense clusters located in the nucleoid center (only 12 % in the cell periphery), and small clusters essentially matching the nucleoid density

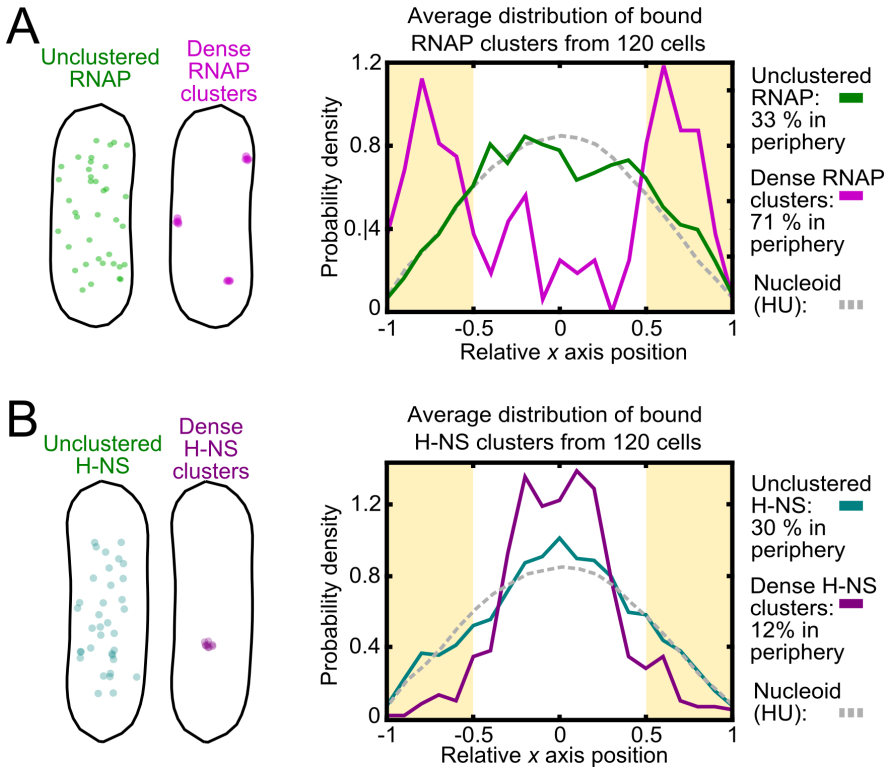


Figure 4.6: **Transcribing RNAPs cluster at the nucleoid periphery** A) An example cell showing clustering of bound RNAP molecules, with clustered RNAPs (>6 molecules) coloured purple, and non-clustered RNAPs (single or pairs of RNAPs) coloured green (left). The x-axis distribution of clustered and non-clustered bound RNAPs in 120 cells shows that non-clustered RNAPs are distributed throughout the nucleoid, whereas dense clusters form at the periphery (71 % ± 9 %)(right). B) An example cell showing the same clustering analysis performed on H-NS (left). The x-axis distribution shows that dense H-NS clusters appear to form preferentially at the centre of nucleoid, with only 12 % ± 5 % forming in the periphery.

(30 % in the periphery)(Figure 4.6 B). This result establishes that the clustering of other nucleoid associated proteins does not necessarily lead to exclusion from the nucleoid, and adds weight to the hypothesis that active transcription drives RNAP segregation from the nucleoid.

Blocking transcription with rifampicin abolished this spatial pattern, as the distribution of bound RNAPs (on promoters, but not transcribing) and mobile RNAPs became almost identical, with 31 % of both distributions located in the cell periph-

ery (Figure 4.7 A and B). Average 2D plots show that cells no longer form distinct nucleoids, but rather fill the cell volume (Figure 4.7 C). When the difference between the average spatial distribution of mobile and bound RNAPs is plotted, it no longer forms any clear pattern. Consistent with this, SIM images show homogeneous, overlapping RNAP and DNA distributions (Figure 4.8). This demonstrates that active transcription is necessary for maintaining this characteristic spatial distribution of RNAP with respect to the nucleoid DNA.

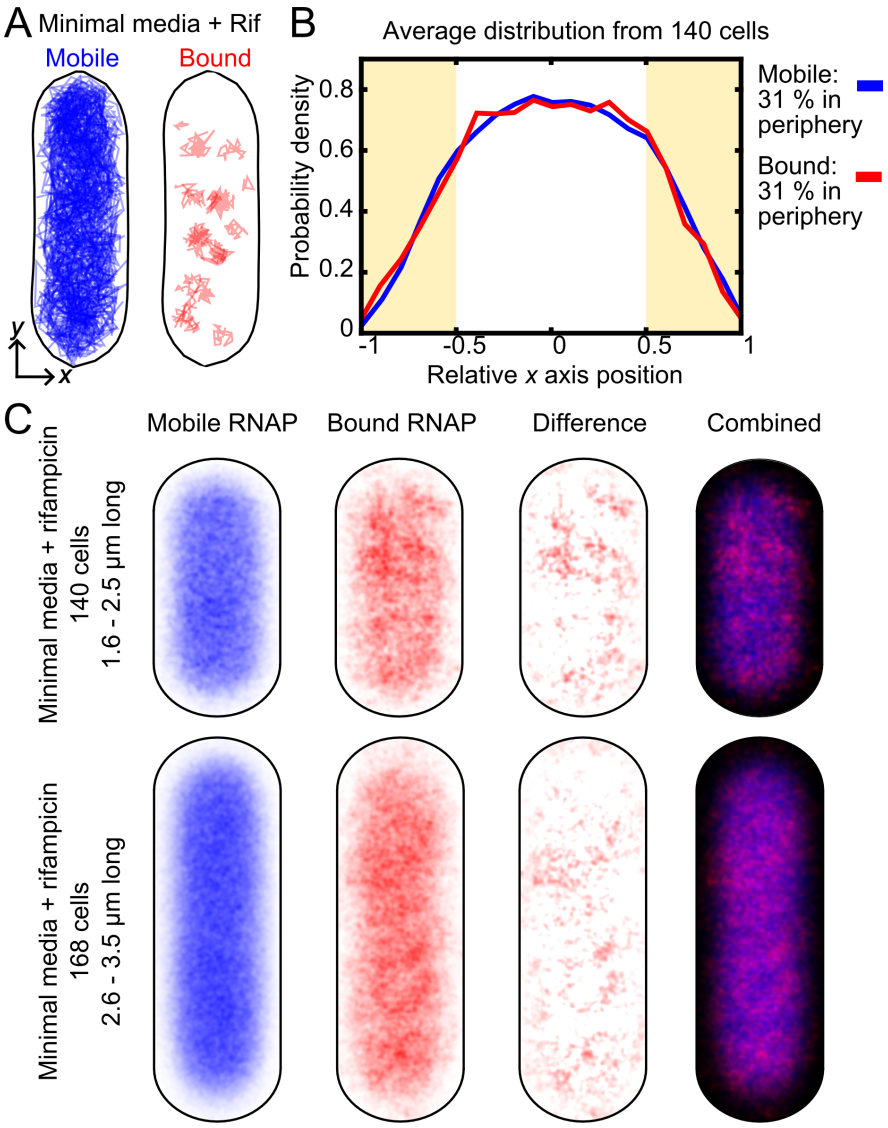


Figure 4.7: **Blocking active transcription**. A) An example cell after incubation with rifampicin. B) After rifampicin incubation, the x-axis distributions of mobile and bound RNAPs become almost identical ($p>0.05$), with 31 % ± 1 % of mobile molecules and 31 % ± 2 % of bound molecules found in the cell periphery. C) Normalised 2D histogram plots showing the average spatial distribution of categorised RNAP molecules from cells after rifampicin incubation. Both short and long cells no longer show clearly defined nucleoids. The difference plot, showing the distribution of bound RNAPs from which the distribution of mobile RNAPs has been subtracted, shows no clear spatial organisation of bound RNAPs compared to mobile RNAPs.

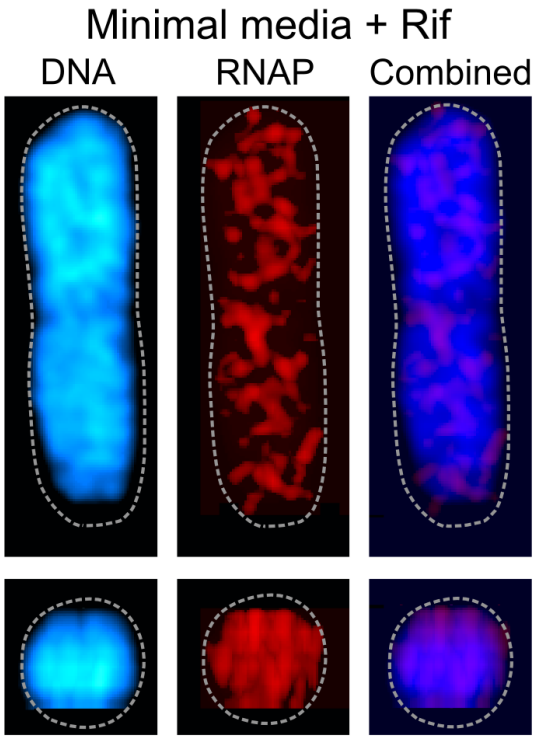


Figure 4.8: **SIM image of an example cell treated with rifampicin.** After rifampicin treatment both DNA and RNAP distributions appear to fill most of the cell volume homogeneously.

We then examined the effect of active translation on the RNAP spatial profile, since co-transcriptional translation has also been proposed to play an important role in the nucleoid structure. According to the transertion model (Woldringh, 2002), transmembrane proteins produced by cotranscriptional translation indirectly attach transcribed genes to the cell membrane, promoting nucleoid decompaction and potentially helping chromosome segregation. To block translation, we used chloramphenicol, an antibiotic that blocks protein chain elongation by inhibiting the peptidyl-transferase activity of the 50S ribosomal subunit. Consistent with previous reports (Jin et al., 2013), SIM imaging showed that chloramphenicol triggers a dramatic compaction of the nucleoid into a ring-like structure that locates at the centre of the cell (Figure 4.9 D); however, RNAP molecules still cluster at the nucleoid

surface. These results are supported by PALM, which show that the nucleoid region condensed along the y-axis, but the periphery bias of bound RNAPs is maintained (Figure 4.9 C); indeed, many cells showed distinct rings of bound RNAPs surrounding the nucleoid (Figure 4.9 A). On average, 46 % and 33 % of bound and mobile RNAPs were located in the periphery, respectively (Figure 4.9 B). These results suggest that transertion indeed promotes chromosome expansion but is not the main driving force for the global spatial localisation of large RNAP clusters to the nucleoid surface.

4.4 Transcription in rich media

As growth rate increases, transcription on most genes is reduced; however, a small number of genes, particularly stable RNAs (such as ribosomal RNAs) become very heavily transcribed (Bremer and Dennis, 1996; Cabrera and Jin, 2003). In cells grown in rich media, and subsequently chemically fixed, RNAP has been shown to form dense clusters, which were not observed in minimal media (Cabrera and Jin, 2003; Endesfelder et al., 2013), however, little is known about their presence, dynamics or behaviour in live E. coli. The observation of these large RNAP clusters, reminiscent of 'transcription factories' in eukaryotic cells (Cook, 1999), raised the question of whether these are static structures that are stably anchored in cells and pull transcribed DNA through them, or whether RNAPs assemble in large numbers on active promoters and simply follow the motion of the segregating chromosome during transcription.

To address this question, we measured the mobility of RNAP clusters using GFP-labelled RNAP in live cells growing in rich media (Cabrera and Jin, 2003). Imaging dense RNAP clusters in live cells has previously proved challenging with conventional fluorescence microscopes. Using the sensitive single-molecule microscope, we clearly resolved bright foci, in GFP labelled RNAP in live cells (Figure 4.10 B), which

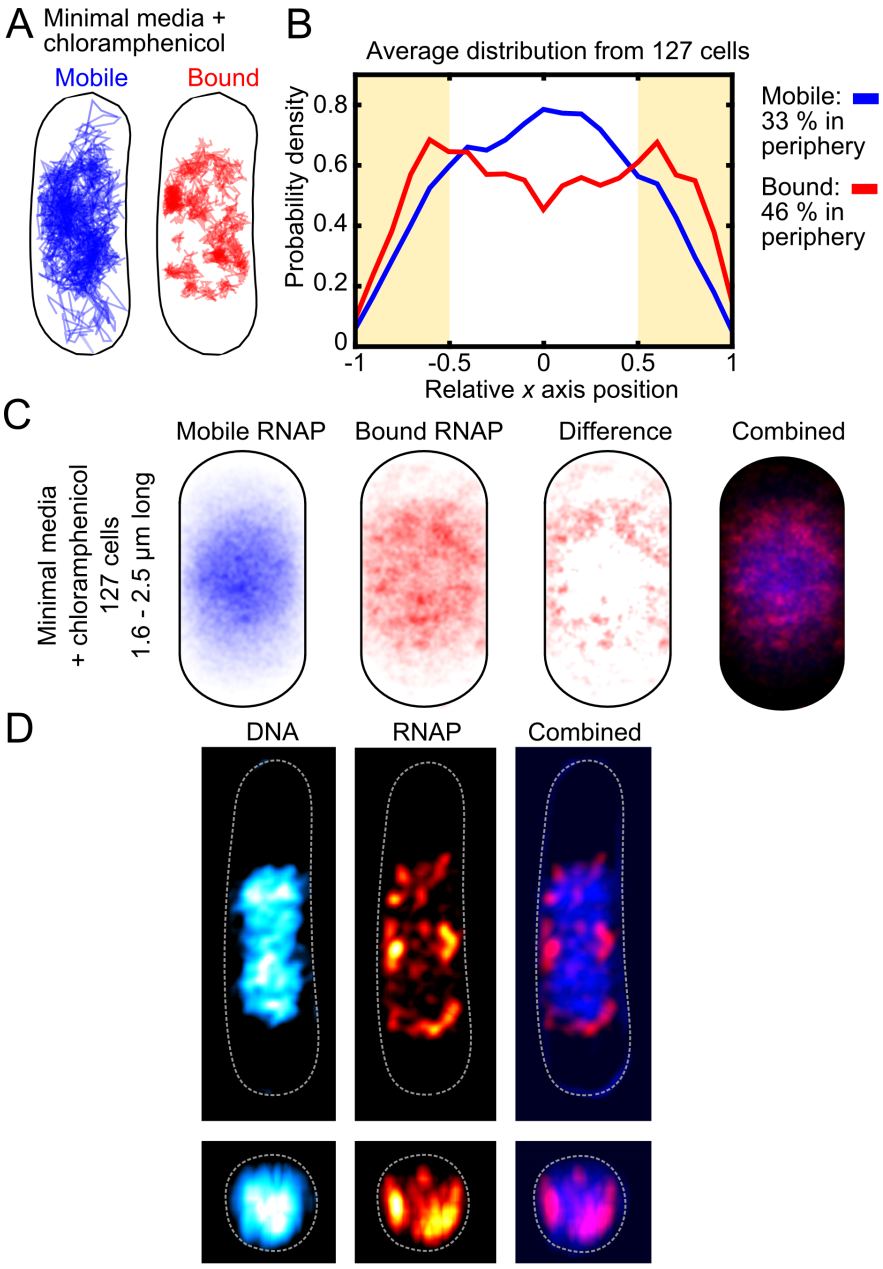


Figure 4.9: **Blocking active translation** A) An example cell incubated with chloramphenicol shows a compacted nucleoid, with bound RNAPs locating to the edge of the bulk of DNA. B) The bias is reflected in the probability distributions showing that 46 % ± 2 % of bound RNAP is found in the periphery compared to 33 % ± 1 % of mobile molecules. C) The compaction of the nucleoid can be seen in 2D average spatial distribution, but the broader distribution of bound RNAP remains. D) 3D SIM image of an example cell incubated with chloramphenicol with a compacted nucleoids, with bound RNAPs locating to the edge of the bulk of DNA.

were not seen in cells grown in minimal media (Figure 4.10 A). To determine the mobility of these foci, we used 2 s exposures, and low excitation power, over a 2 minute time-scale. The centroids of foci were localised using the same Gaussian fitting as that used for single-molecule imaging, with elliptical eccentricity constrained to prevent fitting to the overall RNAP background. Tracking localisations, and plotting the MSD show that these structures are fairly dynamic (Figure 4.10 C). To compare the mobility of these structures with DNA loci, we labelled the origin of replication (*ori*) region of the chromosome using an array of TetO operator sites and chromosomally expressed TetR-YFP fusion. TetR-YFP foci were localised, tracked and analysed as for RNAP-GFP foci. Comparing the MSD for DNA loci to the RNAP clusters showed that they have similar mobility, indicating that these foci were moving along with the transcribed DNA region, and are unlikely to be rigidly tethered to static cellular structures.

To characterise the size of these RNAP clusters in live cells, we performed rapid localisation microscopy with high photoactivation. Typical PALM imaging takes several minutes per super-resolved image, which would blur these RNAP clusters due to their movement over this time-scale. Chemical fixation, on the other hand, distorts biological structures (Schnell et al., 2012). In order to capture sharp super-resolved PALM images in live cells, we localised RNAPs photoactivated at high density and analysed using a crowded-field localisation algorithm (Holden et al., 2011), achieving a 50 nm resolution at 15 s per image.

These snapshots of RNAP localisations confirm that clustering is much more extensive under fast growth conditions compared to minimal media conditions (Figure 4.11 A). To quantify this difference, we calculated the radial distribution of the localisations, which showed an increase in clustering in rich media compared to minimal media (Figure 4.11 C). The most dramatic increase in clustering occurs

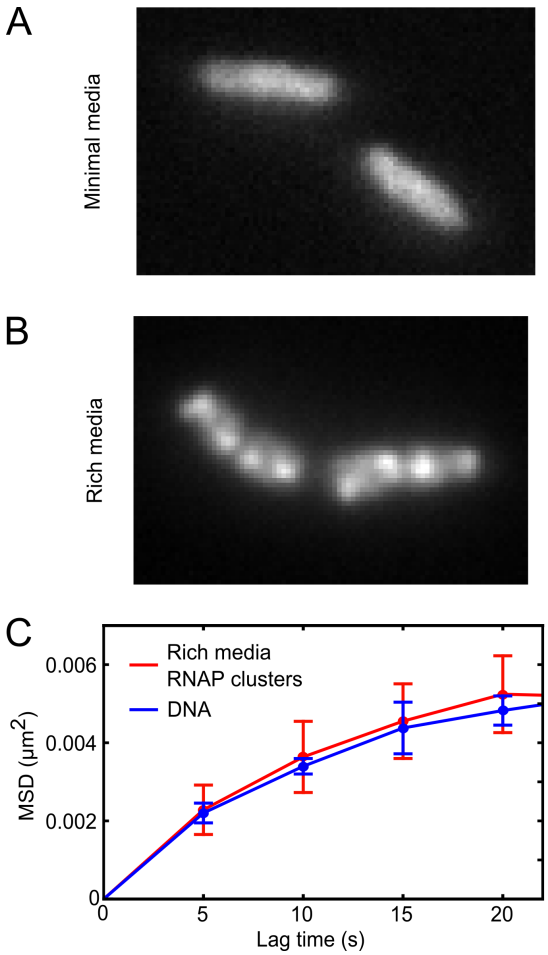


Figure 4.10: **Mobility of RNAP clusters** A) Snapshot image of RNAP-GFP in example live cells growing in minimal media. B) Snapshot image of RNAP-GFP in example live cells growing in rich media. C) Localising dense RNAP foci in cells growing in rich media allows them to be tracked in time-lapse experiments. Mean squared displacements (MSDs) of RNAP foci (red) and labelled DNA loci (blue).

at distances < 150 nm, which indicated the characteristic size of clusters. To confirm that the increased clustering was due to transcribing RNAP molecules, we also acquired imaged cells growing in rich media at normal photoactivation levels and performed the same tracking and categorisation as for cells in minimal media conditions. The radial distribution of bound RNAPs showed a clear increase in clustering for cells in rich media (Figure 4.11 D).

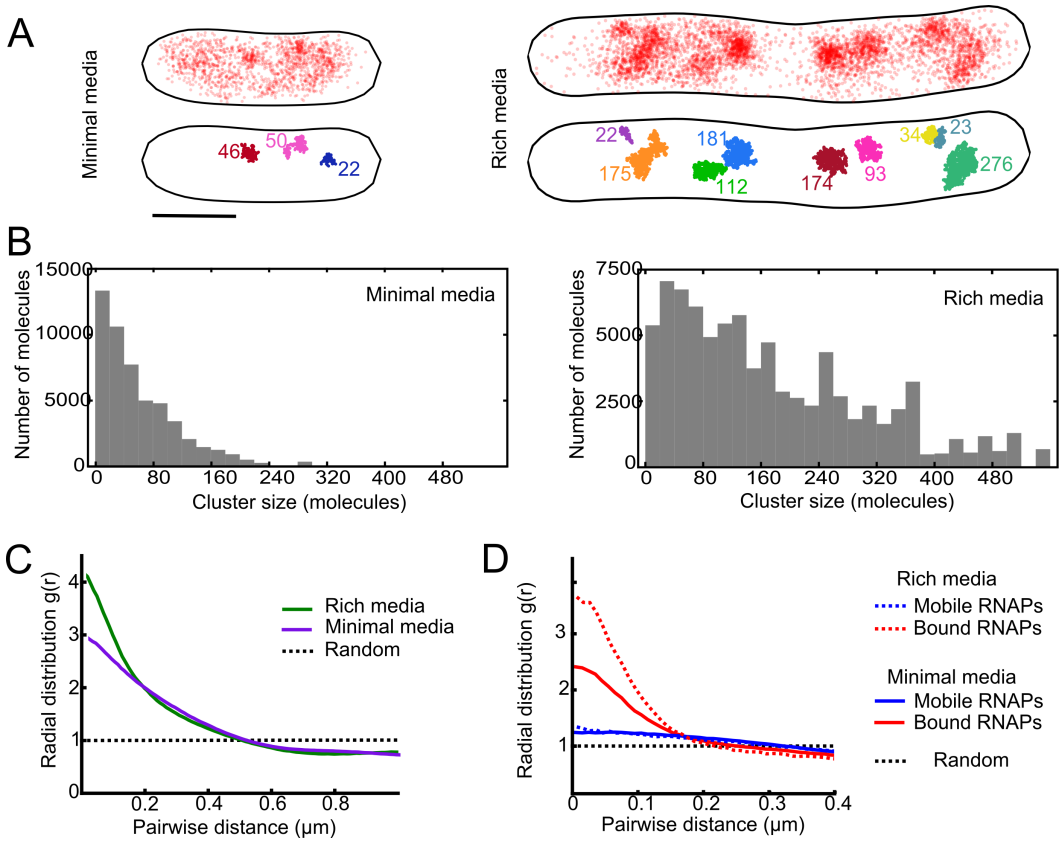


Figure 4.11: **Organisation of transcription in rich media.** A) Representative cells imaged with fast acquisition PALM, showing all localisations (top), and clustered RNAPs, with number of molecules in clusters indicated (bottom). (B) Histogram of the cluster sizes (in number of RNAPs) for 218 cells in rich media and 298 cells in minimal media. C) Radial distribution of all localisations from rapid, high-density PALM shows confirms that clustering increases with growth rate. D) Radial distribution of bound RNAPs (from single-molecule tracking data) showed an increase in clustering for cells in rich media.

The sizes of RNAP clusters were determined by clustering the localisations. We determined the mean number of localisations per molecule (by particle tracking using the same acquisition protocol, but lower photoactivation) and used this to convert localisations to numbers of molecules. The distribution of cluster sizes for the different growth conditions is shown in Figure 4.11 B. Our results showed that the cluster sizes in live cells match broadly to previous estimates in fixed cells (Endesfelder et al., 2013). We observed intermediate-size clusters in minimal media

(~50 molecules), and large clusters (up to 500 molecules) in rich media.

SIM imaging in live cells confirmed that the organisation of the nucleoid also changes as growth rate increases. In minimal media, the nucleoid is relatively homogenous (Figure 4.12 A, left), whereas in rich media nucleoids are more heterogeneous, with distinct structures (Figure 4.12 A, right). In rich media the very dense RNAP clusters clearly form at the nucleoid edge, showing stronger segregation than in minimal media, with very little overlap between RNAP and DNA (Figure 4.12 B). Interestingly, the RNAP clusters form not just at the cell periphery, but also in areas of low-density DNA throughout the cell volume. Comparing rich and minimal media shows that increasing RNAP clustering correlates with increased bias of RNAP to the edge of the nucleoid. In rich media, this effect is more dramatic because, not only is transcription on a few genes increased heavily (leading to increased clustering), but transcription is also reduced on most of the remaining genes, which could allow this DNA to be more tightly compacted.

4.5 Discussion

In this chapter we have used live-cell super-resolution microscopy, single-molecule tracking to shown how RNAP locates promoters distributed throughout the nucleoid, and how active transcription causes spatial reorganisation of the DNA.

In minimal media conditions, approximately half of RNAPs are mobile and diffusing within the nucleoid, engaging in frequent non-specific interactions with DNA as they search for promoters. Contrary to a recent study (Spahn et al., 2015), we see that mobile RNAPs are not excluded from any part of the nucleoid. Indeed, mobile RNAPs show striking agreement with the DNA distribution, with the densest regions of DNA correlating with the densest concentrations of mobile RNAPs. This implies that

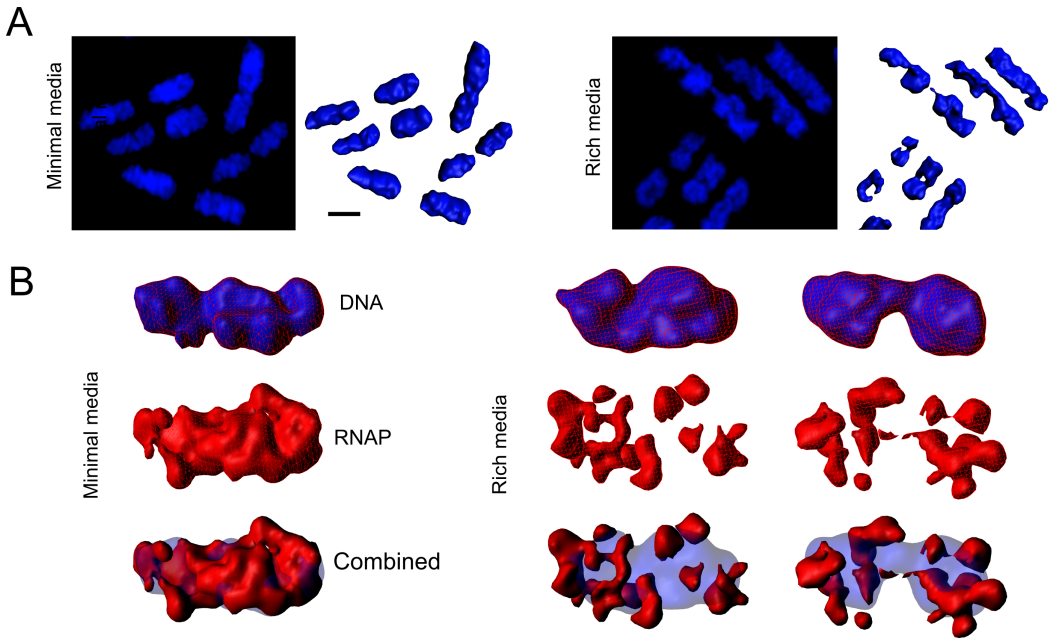


Figure 4.12: **RNAP-DNA segregation in rich media.** A) Example fields of view showing the DAPI-stained nucleoids of live cells imaged with 3D SIM. Surface contour plots highlight the nucleoid structure. B) 3D surface renderings of RNAP-GFP (red) and DNA (DAPI, blue) distributions in example cells grown in minimal (left) and rich media (right).

searching RNAPs can find promoters located throughout the nucleoid (Figure 4.13).

Transcribing RNAPs are more clustered than mobile molecules. Interestingly, non-clustered transcribing RNAPs are located throughout the nucleoid; however, the densest clusters are preferentially located at the edge of the nucleoid where the DNA density is lowest. It appears, therefore, that while low levels of transcription can occur on genes distributed throughout the nucleoid, the most heavily transcribed genes locate to the nucleoid periphery. Incubation with rifampicin shows that active transcription is necessary to maintain these genes at the periphery. These results provide compelling evidence that transcription itself is the driving force for locating active genes to the nucleoid edge, rather than a mechanism whereby regions of the chromosome are inaccessible to RNAP (Spahn et al., 2015), leaving only genes at the periphery available to be transcribed.

One mechanism which could cause highly transcribed genes to locate to the nucleoid periphery is cotranscriptional translation (Jin and Cabrera, 2006; Sanamrad et al., 2014). Ribosomes are strongly segregated from the *E. coli* nucleoid (Hobot et al., 1985; Wang et al., 2011; Bakshi et al., 2012). Statistical models have suggested that this phase separation between the ribosome containing cytoplasm and the DNA can be explained by simple entropic forces: the DNA polymer avoids the walls to maximize conformational entropy, and the polysomes (multiple 70S ribosomes on a single mRNA) occupy the empty space near the walls to maximize translational entropy (Mondal et al., 2011). While mRNA-bound 70S ribosomes are excluded from the nucleoid, it has recently been shown that free 50S and 30S subunits do have access to the nucleoid centre (Sanamrad et al., 2014; Bakshi et al., 2014). These results indicate that both transcription and translation can begin in the centre of the nucleoid (Figure 4.13 A). Transcription initiation and early elongation produce a nascent mRNA that contains a ribosome binding site, which nucleates the recruitment and assembly of one or several ribosomes when translation is initiated. It is plausible that the large DNA-RNAP-ribosome complexes are also excluded from the bulk of the nucleoid and tend to migrate towards the nucleoid periphery, thus leading to the spatial separation of genes depending on their level of expression, with highly expressed genes enriched at the surface of the nucleoid. The inner volume of the nucleoid is consequently less metabolised; potentially allowing it to become more highly compacted. Consistent with this proposal, in minimal media, the nucleoid is less structured, showing less variation in local DNA density due to the fact that genes are expressed throughout the chromosome. In contrast, in rich media, fewer genes are expressed, thus allowing large parts of the chromosome to be inactive, giving rise to more structured and compacted nucleoids.

In rich media, most transcription occurs on the ribosomal RNA operons. Despite

the fact that these transcripts are not translated, there appears to be even stronger segregation between RNAP clusters and nucleoid DNA in rich media compared to minimal media. Coupled transcription and translation is therefore not the only mechanism of locating RNAP clusters to the nucleoid edge. In rich media, the active genes are fewer but are more heavily transcribed: for example, ribosomal operons can have up to 80 transcribing RNAPs on them at any one time (Endesfelder et al., 2013). The large number of RNAPs, which are also be coupled with associated ribosome assembly cofactors (Shajani et al., 2011), may to exert similar bias out of the nucleoid DNA as coupled RNAP-mRNA-ribosome complexes (Figure 4.13 A). Alternatively, the high level of transcription on a section of DNA could cause it to be less interlinked by removing highly abundant DNA bridging proteins due to the translocation of many RNAP molecules. This stretch of DNA may be favoured to locate to the nucleoid surface, compared to the less metabolised, highly interlinked segments of DNA. This mechanism would favour a peripheral location of highly expressed genes encoding stable RNA and mRNA.

Our results also indicate that other potential mechanisms for creating the phase separation between RNAP and the nucleoid have significant shortcomings. For example, because MreB and RNAP co-purify in cell extracts and interact *in vitro* (Kruse et al., 2006), it has been proposed that RNAP clusters could be anchored to the cytoskeleton. While our data do not rule out some interaction between RNAP and the cytoskeleton, we believe it is unlikely for this to be a primary mechanism for locating RNAP clusters to the nucleoid periphery for two reasons. Firstly, tracking of RNAP clusters in rich media showed that they have very similar mobility to DNA, suggesting that they are not rigidly tethered to the cytoskeleton. Secondly, 3D-SIM imaging showed that in rich media, where the nucleoid forms more heterogeneous structures, RNAP clusters form not just at the cell periphery, but also in the central

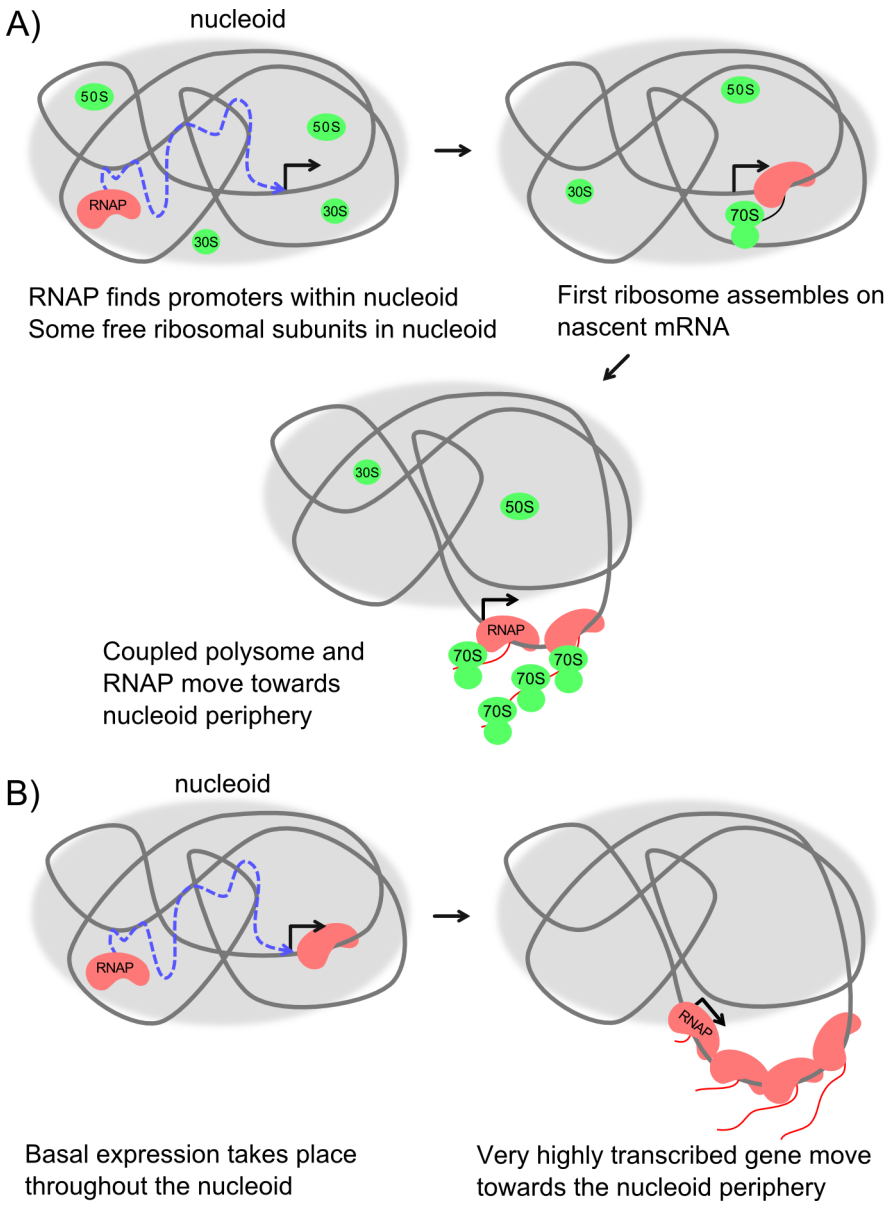


Figure 4.13: **Mechanisms of gene spatial organisation.** A) RNAPs and free ribosome subunits can explore the entire nucleoid in their search for specific nucleic acid sequences. RNAP starts transcribing a gene within the nucleoid, and the first ribosome binds to the emerging mRNA. As the polyribosome grows, and other RNAPs start transcribing, entropic forces may favour movement away from the bulk of the nucleoid. B) In the case of rRNA transcription, there is no coupled translation, but rRNA operons are extremely highly transcribed. Multiple RNAPs on the same gene may also tend to move the gene towards the periphery of the nucleoid.

volume of the cell in low-density regions of the nucleoid that are distal from the cell membrane.

Transertion has also been proposed to play a role in nucleoid decompaction because blocking translation with chloramphenicol causes nucleoid compaction. Indeed, individual genes expressing membrane proteins have been shown to migrate to the membrane when expressed (Libby et al., 2012). However, we showed that clusters of bound RNAPs still remain at the nucleoid edge after chloramphenicol incubation, indicating that transertion might play a role in the nucleoid conformation but is clearly not the driving force behind localisation of large RNAP clusters to the nucleoid surface.

Taken together, our work shows that transcription can cause spatial reorganisation of the nucleoid, with movement of gene loci out of the bulk of DNA as levels of transcription increase. The peripheral localisation of transcription is reminiscent of that of DNA repair (Lesterlin et al., 2013), supporting the view that DNA metabolism activities involving large machineries locate preferentially at the nucleoid periphery. In rich media conditions the number RNAP clusters observed per cell (Cabrera and Jin, 2003), and the number of RNAPs found in each cluster (Endesfelder et al., 2013) has led to the hypothesis that these clusters represent RNAPs active on multiple heavily transcribed genes spatially located to the same site, analogous to transcription factories in eukaryotes (Papantonis and Cook, 2013; Cook, 1999). Similarly, during DNA break repair, distant DNA loci are brought together at the periphery of the nucleoid (Lesterlin et al., 2013). The localisation of heavily transcribed genes outside the bulk of DNA may therefore also facilitate spatial clustering of genes located at distant sites on the chromosome.

4.6 Materials and methods

Bacterial strains

The bacterial strains and plasmids used in this chapter are listed in Table 4.1. All strains are derivatives of the *E. coli* K-12 MG1655 strain. PAmCherry and GFP fusion strains of RNAP, expressed from their native promoter in *E. coli* MG1655, were previously characterised in by Endesfelder et al. (2013). A strain with *ori*1 labelled chromosomal locus was constructed by P1 transduction of *plac-tetR-6aa-ypet, kan* from RRL208 into a strain carrying an array of 240 *tetO* inserted at position 3908.

PAmCherry fusion strains of HUβ and H-NS, expressed from their native promoters, were constructed by Pawel Zawadzki by replacing the genes by their PAmCherry fusion versions. Genes coding for PAmCherry1 and a kanamycin resistance cassette were amplified from the plasmid pROD85 (carrying an 11aa linker preceding PAmCherry and followed by frt flanked kn), using primers with 40-50 nt overhangs homologous to the insertion site at the chromosome. DNA fragments were electroporated in AB1157 cells overexpressing λ-Red proteins (Datsenko and Wanner, 2000). Correct insertion of the fragment into the chromosome was evaluated by PCR using primers flanking the insertion site. Fusions were moved to MG1655 background by P1 transduction. The growth rates in LB and M9 glycerol medium were the same for wild-type and fusion strains. Unconjugated PAmCherry was expressed from a pBad plasmid in wildtype MG1655 cells.

Cell preparation

Cells were prepared as described in Chapter 3. Where indicated, cells were incubated with 50 µg/ml rifampicin or 100 µg/ml of chloramphenicol for 30 min. For nucleoid imaging using SIM imaging 4 µg/ml of 4',6-diamidino-2-phenylindole (DAPI). For

Strains	Genotype
KF26	*rpoC::PAmCherry, amp*; (Endesfelder et al., 2013)
DJ2955	*rpoC::GFP, amp*; (Cabrera and Jin, 2003)
MS002	Δ*galK::plac-tetR-ypet, kan; tetO*240, *gen* at *ori1*
PZ001	*hupB::PAmCherry, kan*
PZ003	*hns::PAmCherry, kan*
MS004	*polA::PAmCherry, kan*
Plasmids	Genotype
pKF01	PAmCherry cloned into pBAD24 using NcoI and XbaI cloning sites

Table 4.1: Bacterial strains and plasmids

nucleoid imaging performed on the PALM microscope, where 405 nm excitation is reserved for photoactivation, SYTOX green was used at 500 nM (in live cells, following Bakshi et al. (2014)).

Microscopy

PALM microscopy was performed on the same custom-built TIRF microscope, using the localisation algorithm detailed in Chapter 2 and with the parameters described Chapter 3. For rapid high-density PALM imaging, data was acquired with a 3 ms frame time with 405 nm photoactivation between 5 nW to 10 nW. Excitation with 561 nm was used at 75 mW Crowded-field localisation was performed using DAOSTORM (Holden et al., 2011).

3D-Structured illumination (SIM) imaging was performed in collaboration with Christian Lesterlin on a DeltaVision OMX V3 (Applied Precision/GE Healthcare) equipped with a Blaze SIM module, a 60X/1.42 oil UPlanSApo objective (Olympus), 405 nm and 488 nm diode lasers and three sCMOS cameras (PCO). Three-dimensional two-colour stacks of RNAP-GFP and DAPI stained DNA were obtained using sequential acquisition of 512 X 512 pixels image stack with 8 to 12 125 nm z-sections. Each z-section results from striped illumination pattern rotated to the

three angles (-60°, 0°, +60°) and shifted in five phase steps. Acquisition settings were as follows: RNAP-GFP, 7 ms to 10 ms exposure with 488 nm laser (100 % transmission) and DAPI, 20 ms exposure with 405 nm laser (100 % transmission). The 3D-SIM raw data was computationally reconstructed with SoftWoRx 6.0 (Applied Precision) using a Wiener filter setting of 0.002 and channel specifically measured optical transfer functions to generate a super-resolution three-dimensional image stacks. Images from the different channels were aligned using parameters obtained from calibration measurements with 120 nm-diameter TetraSpeck beads (Life Technologies) using the OMX Editor software. IMARIS analysis software (BITPLANE) was used to generate 3D rendering of fluorescent signal and the 3D wireframe/surface renderings in Figure 4.12.



Chapter 5

The target search process

In the previous chapter we saw that mobile RNAP molecules were still very highly associated with nucleoid region, and made few excursions into the cell cytoplasm, suggesting a high level of transient interactions with DNA. Here we determine the free 3D diffusion coefficient of RNAP and use it to characterise the fraction of time searching molecules spend bound to non-target DNA. We also establish a universal protocol to partition the search process of any DNA-binding protein and apply it to three different proteins.

5.1 Introduction

The affinity of a protein for DNA has a large electrostatic component, with positively charged functional groups on the surface of the protein interacting with the negatively charged phosphate backbone of DNA. However, for the most part these interactions are sequence independent as the backbone of DNA is largely invariant (Redding and Greene, 2013).

For sequence-specific DNA-binding proteins, such as those involved in transcription, the specificity arises primarily through hydrogen bonds formed with specific bases in the major groove. Sequence-dependent deformations of the DNA helix

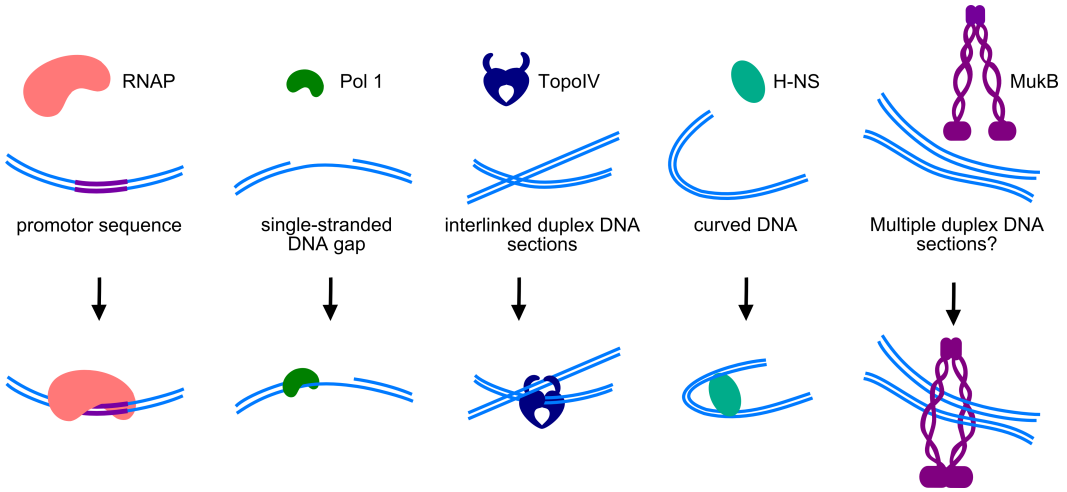


Figure 5.1: **Schematic of different protein-DNA interactions.** The binding behaviour of different DNA-binding proteins. Some proteins, especially those involved in transcription such as RNA polymerase (RNAP), bind to specific DNA sequences. However many other proteins bind to structural motifs rather than specific sequences. For example DNA polymerase I (Pol1) binds to single-stranded DNA gaps. Topoisomerases, such as TopoIV, bind to two duplex sections of DNA in order to remove interlinked, or supercoiled DNA. Other proteins play a role in organising the DNA itself. The histone-like nucleoid structuring protein, H-NS, binds preferentially to curved DNA. MukB, which forms part of the The bacterial structural maintenance of chromosome complex (SMC), may group together multiple DNA strands.

can also play a role (Rohs et al., 2009). However, a huge variety of other proteins recognise specific structural motifs rather than a sequences of base pairs (Figure 5.1). For example, corruption of the information on DNA can be fatal to the cell, hence any damage to DNA must be quickly repaired. RecA, for example, binds to the blunt ends of double-stranded DNA breaks (Lesterlin et al., 2013), DNA polymerase I binds to a gapped section of DNA (Uphoff et al., 2013), and UvrA recognises bulky lesions which distort the DNA helix (Van Houten, 1990).

The entire chromosome must also be completely replicated and segregated in every cell cycle so the information stored in the DNA can be passed on the daughter cell. The single stranded binding protein (SSB), as its name suggests, binds to single stranded DNA prior to synthesis of the complementary strand during DNA replication

(Xie et al., 2008). Type II topoisomerases are responsible for separating interlinked chromosome before they segregate. To do this they bind to two duplex sections of DNA crossing each other at the correct angle (Stone et al., 2003).

Finally, there are also DNA binding proteins whose role is to organise and compact DNA. In eukaryotes histones are one of the most common proteins and are responsible for packaging DNA. Prokaryotes have no direct histone equivalent, however they do have many nucleoid associated proteins (NAPs) (Dorman and Deighan, 2003). These NAPs also bind to double stranded DNA, but many do not recognise a specific sequence. For example the histone-like nucleoid structuring protein (H-NS) binds preferentially curved DNA (Thei Dame et al., 2001), and causes a global effect silencing many genes. The mechanisms by which these diverse proteins locate and recognise their targets are poorly understood.

5.2 RNAP promoter-search

In the previous chapter we showed that mobile RNAP molecules were still very highly associated with nucleoid region, performing few excursions into the cell cytoplasm, suggesting a very high level of transient interactions with DNA as well as 3D diffusion between strands of DNA during the search process. Determining the ratio of non-target binding to free 3D diffusion would shed light on this search mechanism.

Previous studies which have estimated the fraction of free RNAP have been indirect, relying on creating DNA-free minicells, and comparing the cytoplasmic protein concentration in minicells to that in normal unperturbed cells. Using this approach, Shepherd et al. (2001) estimated that the concentration of RNAP found in minicells was 17 % of that found in whole cells. It was inferred that this value represented the fraction of free RNAP in unperturbed cells.

We reasoned that the fraction of time that RNAP spends in binding non-specifically to DNA compared to undergoing 3D diffusion could be determined by obtaining the D of immobile RNAP molecules (D_{imm}), and the D for intracellular RNAP diffusion in the absence of DNA (D_{free}). While obtaining D_{imm} is straightforward (extracted using the control experiments with Pol I in Chapter 3), obtaining D_{free} is more challenging, since virtually all RNAPs in wild-type *E. coli* are located within the nucleoid, complicating the study of freely diffusing RNAP molecules.

We can make a rough estimate of the expected free 3D diffusion of RNAP based on the D values of freely diffusing, unconjugated fluorescent proteins. Estimates of the diffusion coefficient of unconjugated fluorescent proteins diffusing in *E. coli* have been reported between 7 $\mu m^2/s$ for eYFP (Kumar et al., 2010), 8 $\mu m^2/s$ for GFP (Elowitz et al., 1999), 8 $\mu m^2/s$ for Venus (Elf et al., 2007), and 10 $\mu m^2/s$ for Dendra2, (Sanamrad et al., 2014). Given that the diffusion coefficient is inversely proportional to the Stokes radius (see Equation 3.1), and estimated Stokes radii for GFP and RNAP are 2.4 nm and 6.65 nm respectively (Ando and Skolnick, 2010), we would therefore expect free RNAP to diffuse within the cytoplasm with D between 2.5 $\mu m^2/s$ and 3.6 $\mu m^2/s$. In our experiments, mobile molecules appear to have a diffusion coefficient an order of magnitude smaller than this (0.36 $\mu m^2/s$), consistent with our hypothesis that these molecules are undergoing rapid conversion between 3D diffusion and transient binding.

In order to characterise the intracellular diffusion of RNAP without DNA interactions we imaged RNAP-PAmCherry outside the nucleoid, by creating a 'minimal-DNA' strain carrying a temperature-sensitive DnaC mutation (Figure 5.2 A, inset). In order to initiate replication, the duplex strands of DNA at the origin sequence must first be opened to allow the various replication proteins to form the replisome. The initial bubble opening is done by a protein, DnaA, which allows the helicase loader protein

DnaC to bind and facilitate loading of the replicative helicase, DnaB (Sherratt, 2003). At non-permissive temperatures, DnaC(ts) cells are unable to initiate DNA replication but keep elongating, yielding long cells containing a single chromosome and long DNA-free endcaps (Wang and Sherratt, 2010; Maisnier-Patin et al., 2002). We also treated cells with norfloxacin (a DNA gyrase inhibitor) to enhance DNA compaction by preventing supercoil relaxation, and hence increase the relative volume of DNA-free endcaps.

In these minimal-DNA cells, there is a dramatic increase in the mobility of RNAP compared to wildtype cells (Figure 5.2 A). Tracking RNAPs molecules located only in the DNA-free cell ends allowed us to determine the D^* distribution of free RNAP, which was centred at 1.1 $\mu m^2/s$ (Figure 5.2 B). The x-axis distribution of RNAP molecules located in the DNA-free endcaps agreed well with the expected distribution for molecules evenly distributed through across the cell, confirming they were not associated with the nucleoid (Figure 5.2 C). However the MSD of molecules in the endcap still showed sub-diffusion, which we interpreted as confinement due to the small size of bacteria (Figure 5.2 D).

5.3 Monte Carlo simulations of RNAP diffusion

To correct for both localisation error and confinement on the experimentally observed D^* values, and obtain accurate unbiased D values, we used Monte-Carlo simulations of Brownian motion within an average-size cell endcap (Figure 5.3A) (Uphoff et al., 2013; Bakshi et al., 2013). The cell endcap was defined as a cylindrical volume of length 1.2 µm long and 0.85 µm wide with hemispherical endcaps with a radius of 0.85 µm. Each 15 ms frame was split into 100 sub-frames with Gaussian distributed displacements in each sub-frame. Each molecule trajectory was given a random starting time to mimic stochastic photoactivation, and a duration sampled

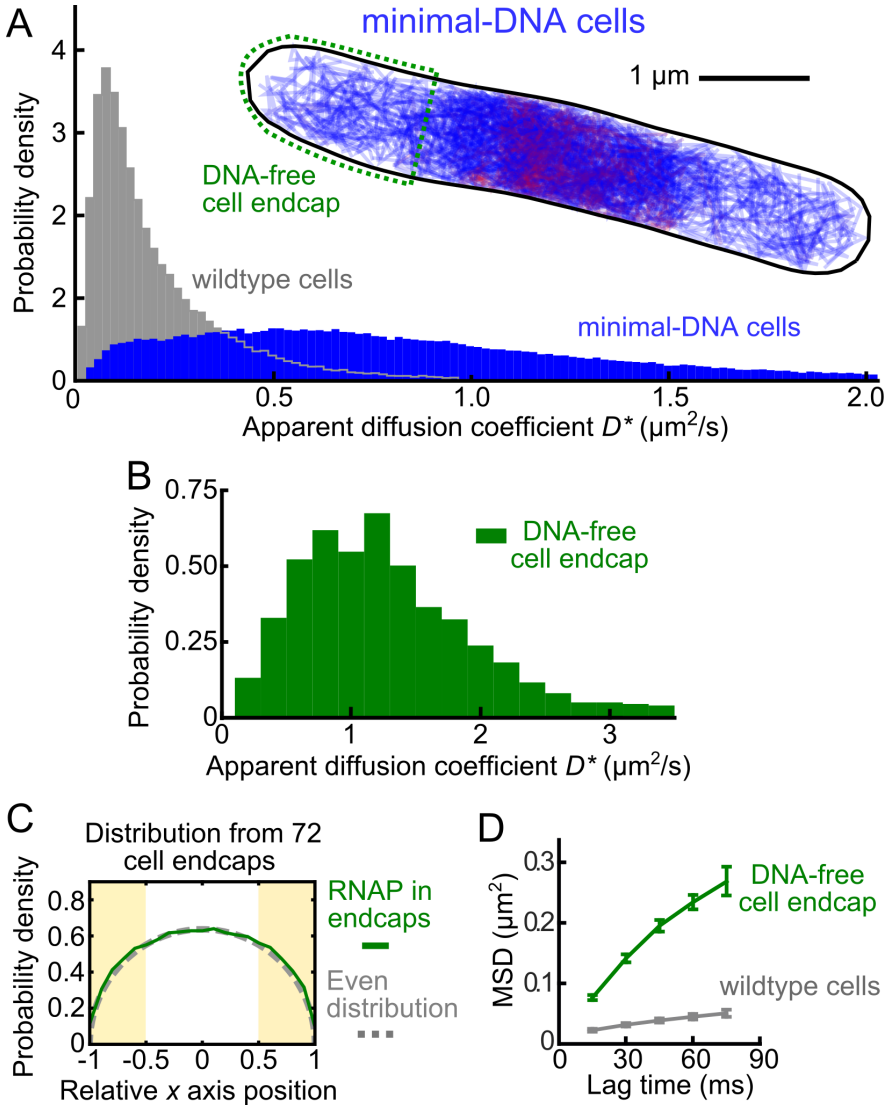


Figure 5.2: **DNA-free diffusion of RNAP.** A) Example 'minimal-DNA' cell (inset); temperature-sensitive DnaC mutant cells are grown at a non-permissive temperature to give long cells with a single centrally located chromosome. Diffusion of RNAP in minimal-DNA cells (blue columns, 97900 molecules) is much faster than normal unperturbed cells (grey columns, 69900 molecules). B) Tracking RNAPs only in the DNA-free cell endcaps (green bars, 2400 molecules) allows the free 3D diffusion to be determined. C) Plot of the x-axis distribution of RNAP molecules from 72 DNA-free cell endcaps. 40 % of these RNAP molecules are found in the exterior 50 % of the cell width (yellow highlighted area). 39 % of molecules evenly occupying the full cylindrical cell volume are expected to be found in the periphery (dashed grey line). D) Mean square displacement for RNAP molecules in normal unperturbed cells (grey line) and in DNA-free cell endcaps (green line).

from an exponential distribution with a mean time equal to our experimentally determined photobleaching lifetime (85 ms). The sub-frame distributions were then averaged to give a position for each frame, and a localisation error sampled from a Gaussian distribution with σ = 40 nm was added. The list of simulated localisations, with their corresponding frame number was then analysed using the same tracking algorithm with the same settings as used for the experimental data. The generated tracks could then be analysed in exactly the same way as experimental data.

By scanning through D values and analysing simulated tracks as we did for experimental tracks, we found that a diffusion coefficient D_{free} of 2.6 $\mu m^2/s$ best matched our data (dashed black line, Figure 5.3B). This diffusion is within the theoretically expected range given the size of RNAP, and noticeably faster than previous estimates (0.7 $\mu m^2/s$; Bakshi et al., 2013).

The interconverting species D_{int} of 0.36 $\mu m^2/s$ determined from unperturbed cells also needs to be corrected for localisation error and confinement. For consistency we performed similar simulations of Brownian motion inside a volume roughly equal to the nucleoid volume, defined as cylindrical volume of length 1.6 µm long and 0.65 µm wide with hemispherical endcaps with a radius of 0.65 µm. The experimental data was best approximated by simulations of molecules diffusing with D = 0.40 $\mu m^2/s$.

The fraction of time, f_{ns}, that the interconverting species diffusing with an effective diffusion coefficient of D_{int}, spends binding non-specifically to DNA can be estimated using (Elf et al., 2007):

$$D_{\text{int}} = f_{ns} D_{\text{imm}} + (1 - f_{ns}) D_{\text{free}} \tag{5.1}$$

After correcting for localisation error, the diffusion coefficient of molecules bound

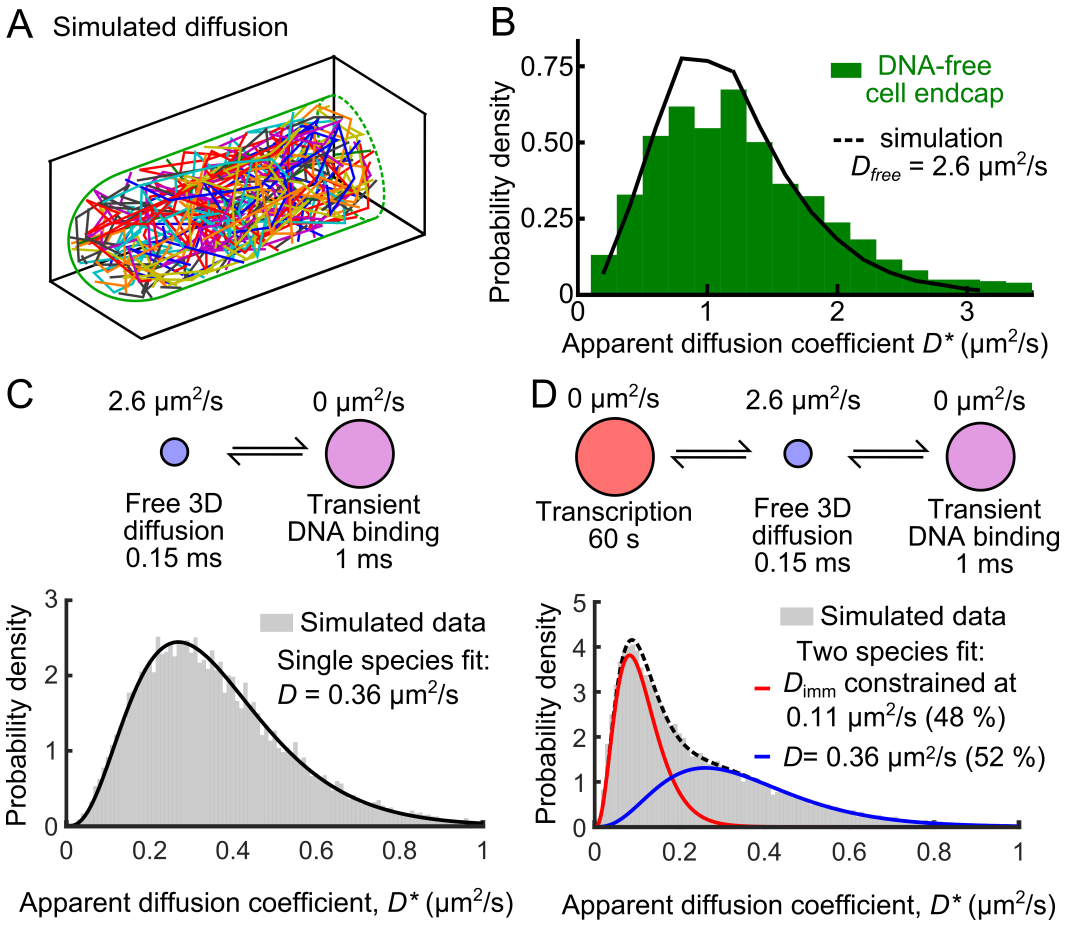


Figure 5.3: **Monte Carlo simulations of RNAP diffusion.** A) Simulated molecular tracks undergoing Brownian diffusion within a confined cell-endcap volume. B) Simulated localisation are tracked and analysed using the same protocols as the experimental data. The best match to the experimental data was found by scanning through simulated D values from 1 μm²/s to 5 μm²/s. A value of D = 2.6 μm²/s best agreed with the data (black dashed line). C) The model for searching RNAP molecules interconverting between free diffusion and transient binding (top). 15,000 simulated interconverting trajectories were tracked and analysed in the same way as experimental data. The distribution of D^* values fits to a single species with the same D value as the mobile RNAP species. D) Three state model for RNAP, with free, non-specifically bound, and specifically bound molecules (top). Simulations show that this model can recapitulate the complete experimental D^* distribution.

to DNA, D_{imm}, is several orders of magnitude less than free molecule (Elmore et al., 2005; Weber et al., 2010). Over the timescales of our observations we therefore assume it is zero. Using the corrected D values to solve the simplified equation, shows that RNAP spends ~85 % of its promoter-search time non-specifically bound to DNA, which is a similar percentage to that seen for the lac repressor (Elf et al., 2007). This high amount of non-specific binding could explain the very high spatial correlation of mobile RNAP with DNA observed in the previous chapter.

To confirm that the normal unperturbed diffusion of RNAP could be recapitulated with the two interconverting species we simulated 15,000 molecules which could rapidly transition between the two states; D_{free} and D_{imm}. The duration of this transient binding, t_{imm}, was randomly sampled from an exponential with a mean of 1 ms. The interconverting species appeared as a single species with reduced mobility, and could be fitted with an analytical equations for a single diffusing species (Figure 5.3 C). We performed the same simulation for various values of f_{ns}, and found that 0.86 best matched the experimentally observed $D = 0.36\,\mu m^2/s$ distribution, in good agreement with the estimate using Equation 5.1.

As a further verification we introduced a second, long lived, immobile species representing transcribing RNAP (Figure 5.3 D). Free molecules could therefore transition into the transient bound state or into a bound state with duration of 60 s. Tracking and fitting the D^* distribution from the simulated trajectories showed that a three state model, with 48 % specifically bound, 44 % non-specifically bound and 8 % free, can accurately recapitulate the experimental D^* distribution.

The fraction of specifically bound molecules, f_{bound}, can tell us about the approximate time-scale for an average RNAP to find a promoter and engage in transcription, t_{mobile} by using (Uphoff et al., 2013):

$$f_{\text{bound}} = \frac{N_{\text{bound}}}{(N_{\text{bound}} + N_{\text{mobile}})} = \frac{t_{\text{bound}}}{(t_{\text{bound}} + t_{\text{mobile}})} \tag{5.2}$$

In minimal media, $f_{\text{bound}} \approx 0.5$, hence the search time, t_{mobile}, is roughly the same as the average transcription event, on the order of 30 s to 120 s considering the approximately timescales for promoter opening, promoter escape and transcriptional elongation, and the length of the average gene.

5.4 Changing environmental conditions

Quantitative partitioning of RNAP into different states provides a new perspective on the expression of unregulated (constitutive) promoters. Transcription of constitutive genes is expected to be dependent on the concentration of free RNAP (Klumpp and Hwa, 2008). Changes in growth rate cause global changes to the expression of genes in *E. coli*. Cells grown in different growth media which support the same growth rate were shown to have the macromolecular composition, indicating that this phenomena is independent of the precise nutrients available to the cells. It has been proposed that growth rate dependent changes in the availability of RNAP forms a 'passive' regulation of gene expression (Klumpp et al., 2009; Gummesson et al., 2009).

When we performed tracking and sorting of RNAP molecules in cells grown in rich media, we found that a significantly higher fraction were bound (63 % ± 1 %), compared to minimal media conditions (48 % ± 1 %) (Figure 5.4 A). This difference indicates a higher fraction of RNAP molecules engaged in transcription, despite the fact that cells in rich media have roughly half the number of active genes relative to minimal media (Tao et al., 1999). This effect is also visible when RNAPs in each cell are sorted as immobile or mobile using threshold based categorisation, which also

gives an indication of the cell-to-cell variability (Figure 5.4 B). The 30 % higher bound fraction also indicates a faster time-scale for an average RNAP to find a promoter. Using Equation 5.2 shows that the average search times lasting only ~60 % of the duration of the average transcription event in rich media.

An unconstrained fit to the mobile population in rich media still resulted in the same D value as for mobile molecules in minimal media, indicating that the ratio of non-specific to free diffusion remains unaffected by increasing growth rate. Partitioning RNAP into the three states suggests that the total fraction of RNAP which is free decreases slightly at faster growth rates, but the effect is not dramatic (Figure 5.4 C). In contrast, in cells incubated with rifampicin (see Figure 3.6) the D value of mobile molecules increases 75 %, suggesting a significant change in the ratio of non-specific to free diffusion. Performing the same simulations of interconverting molecules showed that $f_{ns} = 0.45$.

5.5 A versatile assay for determining the DNA-free diffusion

Partitioning the non-specific binding of a protein based on the free diffusion coefficient, can in principle be applied to any DNA-binding protein. However many DNA binding proteins are present in much lower numbers than RNAP, which makes the approach of imaging in cell endcaps difficult due to the small number of molecules found there, even in minimal-DNA cells. We therefore established a protocol which allows all copies of a protein of interest to be imaged in the absence of DNA by degrading the entire chromosome. This was achieved by inducing two double stranded breaks (DSB) in a strain lacking functional RecA, a key DSB repair protein. DSBs were introduced using the *I-SceI* endonuclease. A strain was created with two 18 bp

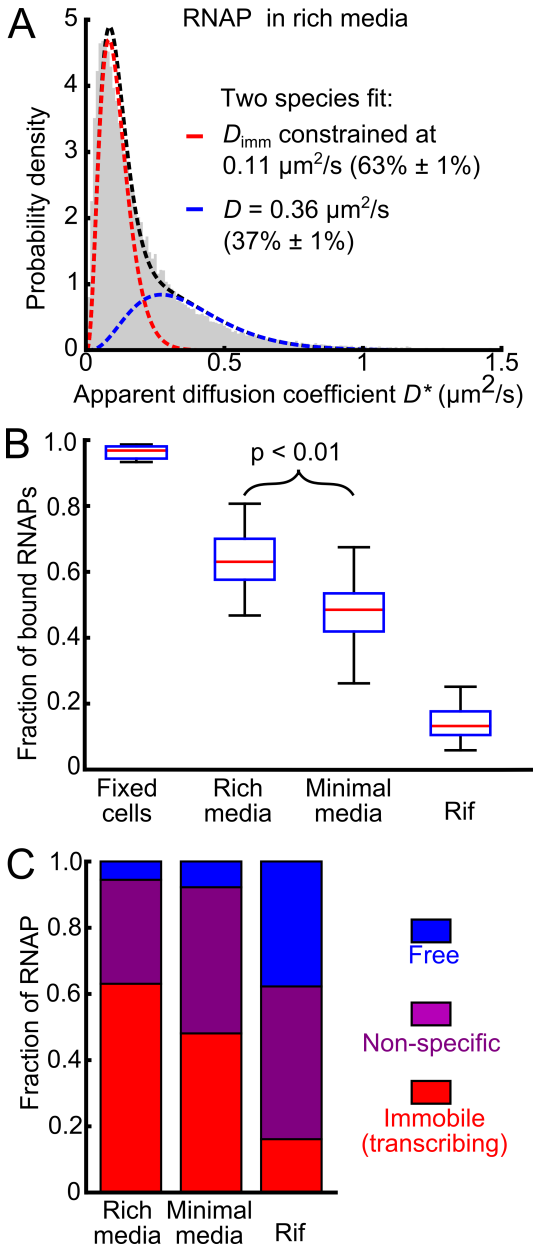


Figure 5.4: **Partitioning RNAP at different growth rates.** A) Two species fitting for cells grown in rich media shows a higher fraction of bound RNAPs (63 % ± 1 %) compared to minimal media. An unconstrained fit to the mobile population still resulted in the same D value as for mobile molecules. B) Plots of the ratio of bound to mobile RNAPs in each cell. Within segmented cells the individual RNAPs are sorted as mobile or immobile using the threshold of $D^* = 0.16\,\mu m^2/s$. The ratio is significantly higher for cells grown in rich media. The same analysis for fixed cells and cells incubated with rifampicin is also shown. D) Comparison of estimates of RNAP partitioning for cells in rich media, minimal media and after rifampicin incubation.

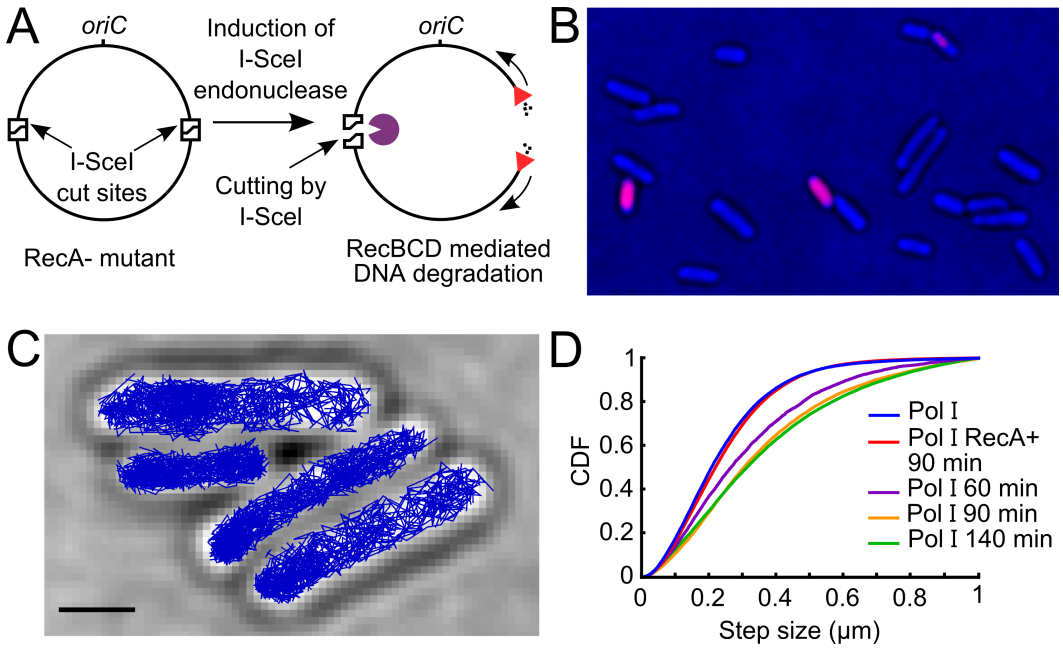


Figure 5.5: **Generating degraded-DNA cells.** A) A cartoon showing the process of DNA degradation. The labelled protein of interest is expressed from its native promoter in a *recA*- strain carrying two *I-SceI* endonuclease cut sites perpendicularly across the chromosome, and an inducible endonuclease on a plasmid. Inducing *I-SceI* expression leads to double stranded cuts and loading of the RecBCD nuclease. Without functional RecA the RecBCD no longer stops at Chi sites leading to eventually digestion of the entire chromosome. B) A representative field of view showing cells 90 mins after inducing cutting. Transmitted light images of cells are overlaid with fluorescence image of DNA stained by syto-16. Between 10 % and 20 % of cells do not fully degrade their chromosome, which can be detected by the DNA stain. C) Example image of PAmCherry labelled Pol I 90 mins after inducing cutting. D) Cumulative distribution of displacements of 15300 Pol I trajectories before inducing DNA cutting (blue curve). After 60 mins of inducing cutting the overall mobility increases (1600 molecules from all cells, purple curve). After 90 mins of induction the majority of cells have fully degraded their chromosomes (2400 molecules, orange curve), and further incubation results in little increase in overall mobility (1200 molecules, green curve). Cells with wildtype RecA functionality successfully repair double stranded breaks, and after 90 mins of inducing cutting there is little change to the overall mobility of Pol I (1600 molecules, red curve).

I-SceI cut sites (*I-SceI*^{CS}) inserted perpendicularly across the chromosome, and *I-SceI* was expressed from an inducible promoter on a plasmid (Lesterlin et al., 2013). The fusion protein of interest was moved into this background and RecA functionality was removed (Figure 5.5 A).

After double stranded cutting is induced the native RecBCD enzyme complex initiates recombination repair at the break, by unwinding the DNA duplex and digesting the individual strands until encountering a Chi site which acts as a recombination hotspot (Singleton et al., 2004). When the complex reaches a Chi site the nuclease activity is attenuated and RecBCD loads RecA onto the 3' tail of the DNA. In the absence of RecA, however, RecBCD continues to digest the DNA until eventually the entire chromosome is degraded.

We chose three diverse DNA-binding proteins as a proof of principle for this assay. Firstly we used RNAP to verify the assay against the values established in minimal-DNA cells. Secondly, we used Pol I which we have previously shown exhibits strikingly different mobility to RNAP in unperturbed cells, with little specific binding and a fast moving mobile species (D = 1.0 μm^2/s). Lastly we used MukB, the *E. coli* Structural Maintenance of Chromosome (SMC) complex. [1] MukB is a large protein with two domains joined together via a ~50 nm long intramolecular coiled-coil. The DNA binding mechanism and the target site of MukB is largely unknown.

After 90 mins of induction 80 % to 90 % of cells show no signal when stained with syto-16, an intercalating DNA stain (Figure 5.5 B). Segmenting cells allows the 10 % to 20 % of cells which fail to completely digest their chromosomes to be filtered out based on the DNA staining (Figure 5.5 B). The remaining cells show tracks which occupy the complete cell volume (Figure 5.5 C).

The overall mobility of Pol I at different time-points after inducing cutting is shown in Figure 5.5 D. Incubation for 140 mins results in only a slight increase in mobility compared to the 90 min time-point. Most cells with wildtype RecA functionality successfully repair double stranded breaks, and after 90 mins of inducing cutting there is little change to the overall mobility of Pol I.

[1] See Chapter 7 for a more in depth introduction to MukB.

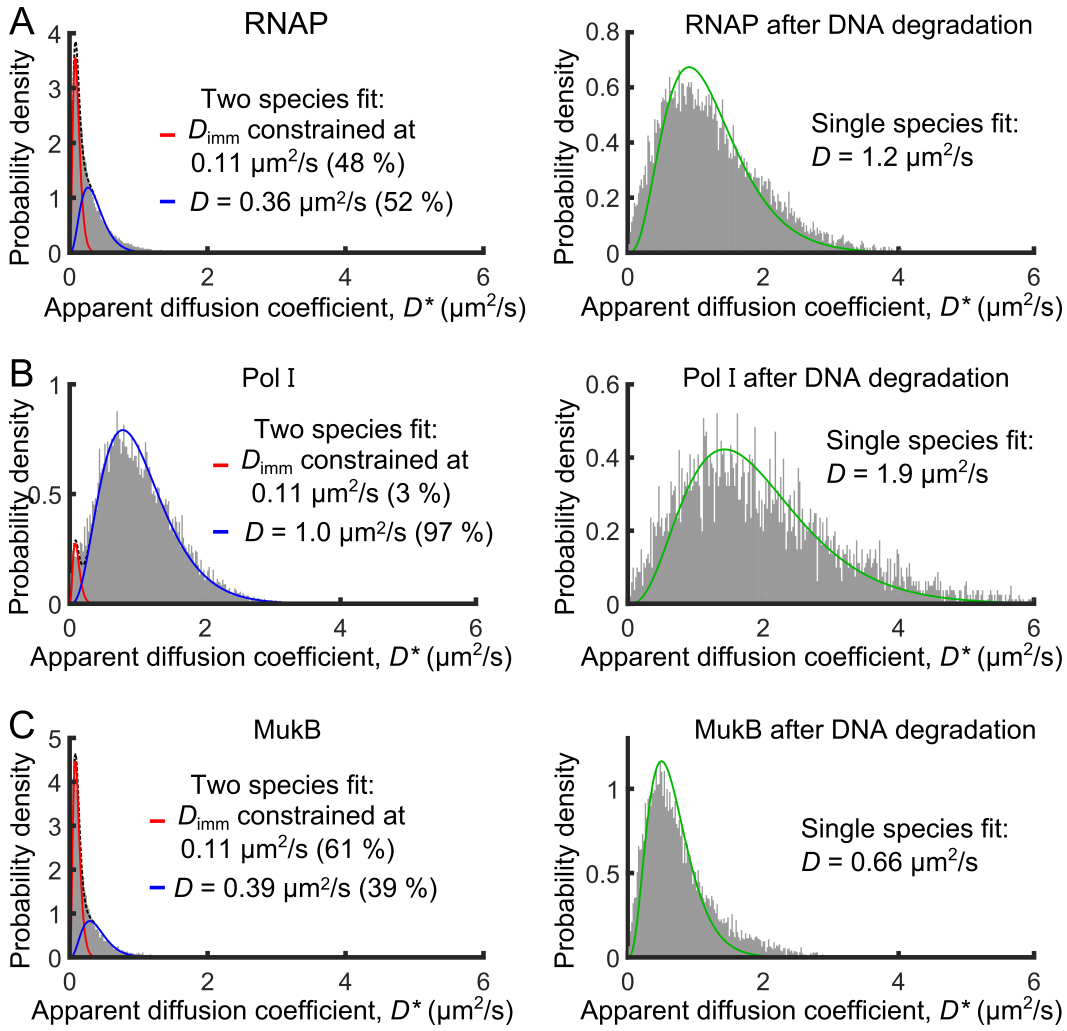


Figure 5.6: **DNA-free diffusion of RNAP, Pol I and MukB** A) The distribution of D^* values for RNAP before (left) and 90 mins after (right) inducing DNA degradation. B) The distribution of D^* values for Pol I before (left) and after (right) DNA degradation. C) The distribution of D^* values for MukB before (left) and after (right) DNA degradation. See Figure 7.4 for a finer scale plot of the unperturbed D^* distribution for MukB.

5.6 Search models of RNAP, Pol I and MukB

The results of complete DNA-degradation are shown in Figure 5.6. After 90 mins of inducing DNA degradation the distribution of D^* values of all three proteins fitted well to a single species model, demonstrating that virtually all DNA binding was lost.

Once again we used simulations to correct for cellular confinement and localisation error. Defining the confinement volume became a simpler task, as the cell membranes could be imaged with transmitted light. We therefore segmented cells and generated 3D volumes from the outline of each cell, by turning each width segment into a circle defining a section across a cylindrical 3D volume (Figure 5.7 A). Diffusing molecules were then simulated inside the 3D volumes using the same number of molecules in each cell as determined experimentally, resulting in the same sampling of the D value as the experimental D^* distributions.

The diffusion of all three proteins in the absence of DNA was significantly faster than the corresponding mobile species in unperturbed cells. For faster diffusion the effect of cellular confinement becomes larger. We therefore also imaged the proteins after DNA degradation at 5 ms exposure time and verified that our corrected D values could also recapitulate the 5 ms experimental D^* distribution by performing simulations of molecules with the same D at both 5 ms and 15 ms (Figure 5.7).

Both the raw and corrected D_{free} value for RNAP were in good agreement with that obtained from the endcaps of minimal-DNA cells. The free diffusion of MukB is very slow, which is is not surprising given its elongated shape. However it still slower than predicted from literature measurements of the hydrodynamic radius (7.6 nm; Melby et al., 1998) which could be due to association with its accessory proteins MukE and F. In agreement with this, the small deviation from the simulation at lager D^* values in Figure 5.7 C, may represent MukB without MukE or F.

We used the D_{free} values for the three proteins to establish the faction time the protein spends undergoing non-specific DNA interactions during its search process using simulations allowing transitions between the bound and free state as before. Scanning through different f_{ns} values, we determined the fraction that could best recapitulate the mobile population in unperturbed cells (Figure 5.8). For

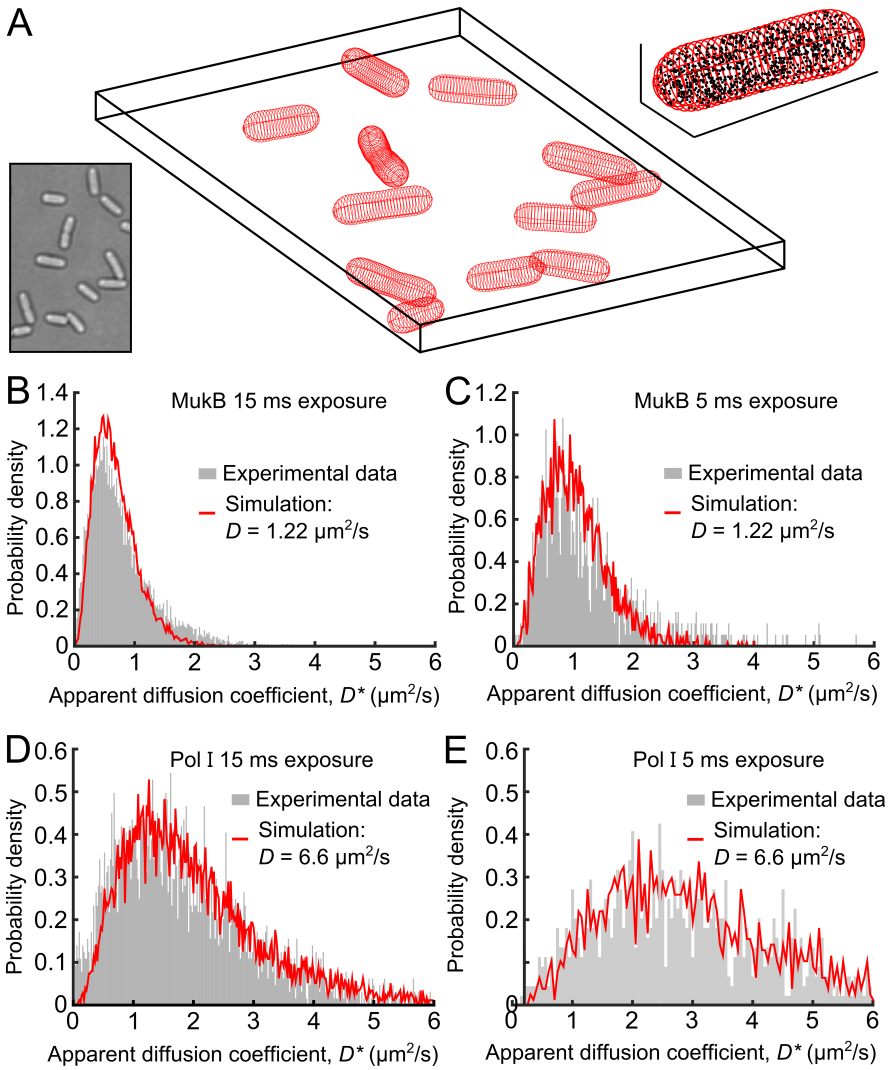


Figure 5.7: **Monte Carlo simulations of diffusion in degraded-DNA cells.** A) 3D volumes of imaged cells were generated by segmenting cells from the transmitted light image (inset, left) using the width of each segment to define the radius of the corresponding cylindrical volume. Monte Carlo diffusion simulations of the same number of molecule determined experimentally were then performed inside the 3D cell volume (inset right). B) The distribution of D^* values for MukB imaged with 15 ms exposure in degraded-DNA cells. The distribution of diffusing molecules with D = 1.22 μm^2/s, simulated in the same cells is shown in red. C) The distribution of experimental and simulated D^* values for MukB imaged with 5 ms exposure in degraded-DNA cells. D) and E) The distribution of experimental and simulated (D = 6.6 μm^2/s) D^* values for Pol I imaged with 15 ms and 5 ms exposures respectively.

RNAP, the model agreed well with the values from our previous assay, with $f_{ns} = 0.87$. Interestingly, both other proteins also spent more than 50 % of their search time associated with DNA ($f_{ns} = 0.61$ for Pol I, $f_{ns} = 0.58$ for MukB). This was especially surprising for Pol I whose diffusion in unperturbed cells is already fast.

5.7 Discussion

In this chapter we have characterised the fraction of time that RNAP, MukB and Pol I spend associated with DNA during their search process. There are several potential diffusion-based mechanisms which might contribute to the target search process. The simplest is 'direct binding', whereby a protein associates directly from solution to its target site via a 3D collision. However, for low copy number proteins, which bind to only few sites on the chromosome, the probability of directly binding its target is very small. In 1970, it was demonstrated *in vitro* that the low copy number lac repressor (LacI) finds its operator apparently faster than the rate limit for a direct binding mechanism (Riggs et al., 1970). Facilitated diffusion was hypothesised to resolve this apparent conundrum. In facilitated diffusion a protein first binds to non-target DNA and then reaches the target by either 1D sliding search, or by 'hopping', with multiple rounds of binding and dissociation to the same section of DNA without re-equilibrating back into solution.

A number of *in vitro* experiments have observed facilitated diffusion of other proteins, and it was generally thought to play a role in the search process of a large fraction of DNA-binding proteins, including RNAP (Ricchetti et al., 1988; Kabata et al., 1993). However, observations *in vitro* do not necessarily translate to *in vivo* behaviour due to non-physiological salt, enzyme, and substrate concentrations as well as the lack of other DNA binding proteins which could act as road blocks.

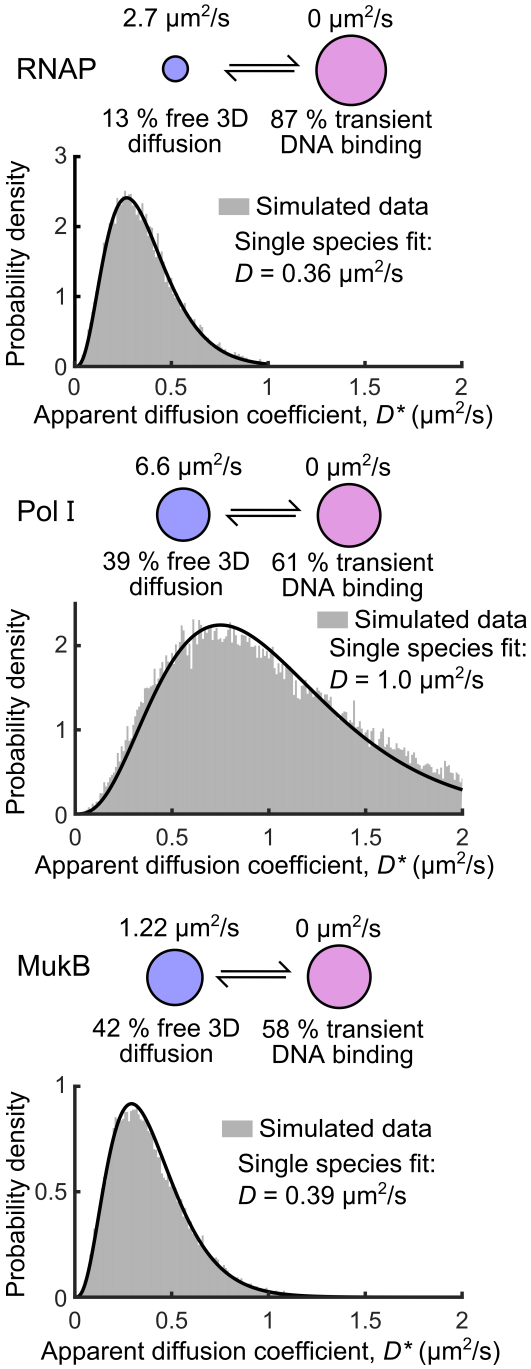


Figure 5.8: **Search models for RNAP, Pol I and MukB.** Simulations of molecules interconverting between free and transiently bound states, using D_{free} values obtained from degraded-DNA cells. Models show the best match to experimentally observed D^* distributions of mobile molecules for RNAP, Pol I, and MukB. Single species fits to the distributions math the experimentally observed mobile species fits.

The first (and only) *in vivo* observation of facilitated diffusion came with LacI, which was shown to slide on DNA for ~45 bp in its approach to its operator (Hammar et al., 2012). In this chapter we show that RNAP also spends a very similar amount of its search process bound non-specifically to DNA when compared to LacI (~85 % for RNAP and ~90 % for LacI, Elf et al., 2007). However, two recent *in vitro* experiments both failed to observe RNAP undergoing 1D sliding, leading to the hypothesis that the search mechanism of RNAP is dominated by 3D diffusion and 'direct binding' (Wang et al., 2013; Friedman et al., 2013). This is likely a consequence of the much higher copy number of RNAP compared to LacI, as increased protein abundance disfavours facilitated searches (Wang et al., 2013). If RNAP does locate its promoters via direct binding, this indicates that a high fraction of non-specific interactions does not necessarily correlate with a facilitated search process.

Despite very different binding targets and clearly different mobility in unperturbed cells, both Pol I and MukB also spend > 50 % of their search process undergoing transient binding. Because the specific DNA interactions of these proteins are likely to be very different from each other, this indicates that the target-independent electrostatic force between the protein and the DNA plays the biggest role in the search process for all three of these proteins. Expanding this analysis may determine if this is true for most DNA-binding proteins.

For RNAP we saw that the average molecule spends from 30 s to 120 s searching for a promoter. We can also perform this calculation for Pol I and MukB using literature values for specific binding dwell times, giving the results in Table 5.1. All three proteins take between 30 s to 120 s to locate their targets. Based on our results we believe that the duration of 3D diffusion between DNA strands is likely to be significantly shorter than 60 ms, the length the trajectories analysed when fitting to D^* distributions. Durations of similar or longer than the observation time would

Protein	f_{bound}	t_{bound}	t_{mobile}
RNAP	0.48	30 s to 120 s*	30 s to 120 s
Pol I	0.03	2.7 s†	~85 s†
MukB	0.61	~50 s‡	~32 s

Table 5.1: The fraction of specifically bound molecules, determined from fitting the D^* distributions for RNAP in rich and minimal media, and for Pol I and MukB in minimal media. Literature values for average binding dwell time allow the average search time for a molecule to be calculated using Equation 5.2. * Values from Bremer and Dennis (1996); † values from Uphoff et al. (2013); ‡ value from Badrinarayanan et al. (2012b).

result in the D_{free} species being observable in unperturbed cells. Taking 60 ms as an upper limit for the duration of free diffusion allows us to estimate a lower limit on the number of non-specific binding events during their target search, with RNAP undergoing at least 50, and Pol I undergoing 550, and MukB 225. In reality, the number of non-target binding events could be much higher for shorter transient binding events. Moving to faster exposure times would allow more accurate values to be determined.

5.8 Materials and methods

Bacterial strains

RNAP-PAmCherry DnaC(ts) cells were constructed by P1 transduction of *dnaC*(Ts) (Wang and Sherratt, 2010) and Tetracycline resistance into RNAP-PAmCherry cells. Colonies were selected with both tet resistance and temperature sensitivity when plated and grown overnight at 42 °C.

Degraded-DNA strains were constructed by Chrisitan Lesterlin. Two different strain background were used with two diametrically opposed *I-SceI*^{CS} cut sites inserted in the chromosome at either *ori* (*ilvA*) and *ter* (*ydeO*), or on the right (*codA*) and

Strains	Genotype
KF26	*rpoC::PAmCherry amp*; (Endesfelder et al., 2013)
MS001	*rpoC::PAmCherry amp*; *dnaC(ts) tet*
ox520	*codA::I-SceI^{CS} frt*; *yeiU::I-SceI^{CS} frt*
ox550	*codA::I-SceI^{CS} frt*; *yeiU::I-SceI^{CS} frt*; *recA- tet*
ox530	*ilvA::I-SceI^{CS} frt*; *ydeO::I-SceI^{CS} frt*
ox540	*ilvA::I-SceI^{CS} frt*; *ydeO::I-SceI^{CS} frt*; *recA- tet*
ox536	*rpoC::PAmCherry amp*; in ox520
ox546	*rpoC::PAmCherry amp*; in ox550
ox535	*polA::PAmCherry kan*; in ox530
ox545	*polA::PAmCherry kan*; in ox540
ox537	*mukB::PAmCherry kan*; in ox530
ox547	*mukB::PAmCherry kan*; in ox540
Plasmids	**Genotype**
pKF01	*PAmCherry1* cloned into pBAD24 using NcoI and XbaI cloning sites
pSN1	pBAD24-*I-SceI amp*
pDL2655	*I-SceI cam*

Table 5.2: Bacterial strains and plasmids

left (*yeiU*) replichores. Protein-PAmCherry fusions were moved into these strains by P1 transduction, followed by RecA inactivatation by P1 transduction of recA-. Finally the *I-SceI* expression plasmids were transformed into recA- strains.

Cell preparation and microscopy

Cells were prepared as described in Chapter 3. Where indicated, cells were incubated with 5 µg/ml norfloxacin for 10 min. For nucleoid staining, 500 nM of SYTOX green (Molecular Probes) was added 10 min prior to imaging. Cutting in DNA-degraded cells was induced with 0.2 % L-Arabinose 90 mins prior to the experiment.

PALM microscopy was performed on the same custom-built TIRF microscope, using the localisation algorithm detailed in Chapter 2 and with the parameters described Chapter 3. Syto-16 stained DNA was imaged with 488 nm laser immediately prior to acquiring PALM movies.



Chapter 6

Single-molecule characterisation of topoisomerase IV

In this chapter we extend the single-molecule techniques and analysis detailed in the previous chapters to characterise the copy number, formation, and behaviour of topoisomerase IV (TopoIV) heterotetramers in live *E. coli*. Imaging both subunits of TopoIV, combined with over-expression of unlabelled subunits, showed that ~60 % of TopoIV subunits were in TopoIV heterotetramers at any time, independent of cell cycle. A replication-dependent increase in the fraction of immobile molecules, together with an apparent catalytic cycle of 1.8 s, indicates that the majority of active TopoIV molecules catalyse decatenation, with only a minority maintaining steady-state DNA supercoiling. The techniques described in this chapter are applicable to many multi-subunit DNA-binding proteins.

6.1 Introduction

During each cell cycle, the newly replicated chromosomes must be successfully segregated before cell division can occur. DNA replication in *E. coli* is performed by two independent replisomes which move in opposite directions around the circular chromosome (Reyes-Lamothe et al., 2008). The double helical structure of DNA means that as each replication fork progresses it rotates interlinking the two sister chromosomes, generating (pre)catenanes behind the replisomes throughout the

chromosome, which all have to be removed. The unwinding of the DNA helix also leads to positive (+ve) supercoils accumulating ahead of the replisome, which must be relaxed to allow replication to continue. Both these crucial topological problems are resolved by the action of type II topoisomerases.

Type II topoisomerases simplify DNA topology by binding to two segments of duplex DNA; firstly a gate (G) segment followed by a transfer (T) segment. Once the T-segment is captured the enzyme introduces a double stranded break in the G-segment and passes the T-segment through, before resealing the break (Figure 6.1). In this process, type II topoisomerases change the linking number of circular DNA by ±2.

In *E. coli*, ~225 000 catalytic events by the type II topoisomerases, DNA gyrase and topoisomerase IV (TopoIV) are required each generation for the segregation of the 4.6 Mb chromosome. DNA gyrase acts preferentially ahead of the replication fork to remove +ve supercoiling (Koster et al., 2010; Vos et al., 2011). *In vitro* TopoIV has been shown to be a versatile protein which acts on topologically different substrates including +ve and negatively -ve supercoiled DNA, knotted DNA and catenated DNA (Postow et al., 2001; Koster et al., 2010; Vos et al., 2011). In theory TopoIV is therefore capable of relaxing supercoils ahead of the replication fork and resolving (pre)catenanes behind the replication fork. Nevertheless, its essential cellular role is in decatenation of newly replicated DNA (Joshi et al., 2013; Wang et al., 2008).

In vivo TopoIV has been shown to act processively on +ve supercoils, with multiple links removed in single long bursts of catalytic activity (Neuman et al., 2009; Stone et al., 2003). It was proposed that multiple T-segments were transported through the cleaved G-segment during a single G-segment capture event (dashed arrow in Figure 6.1). In contrast, the same assay showed that negative supercoils are only relaxed one cycle at a time. The mechanism of how TopoIV recognises and

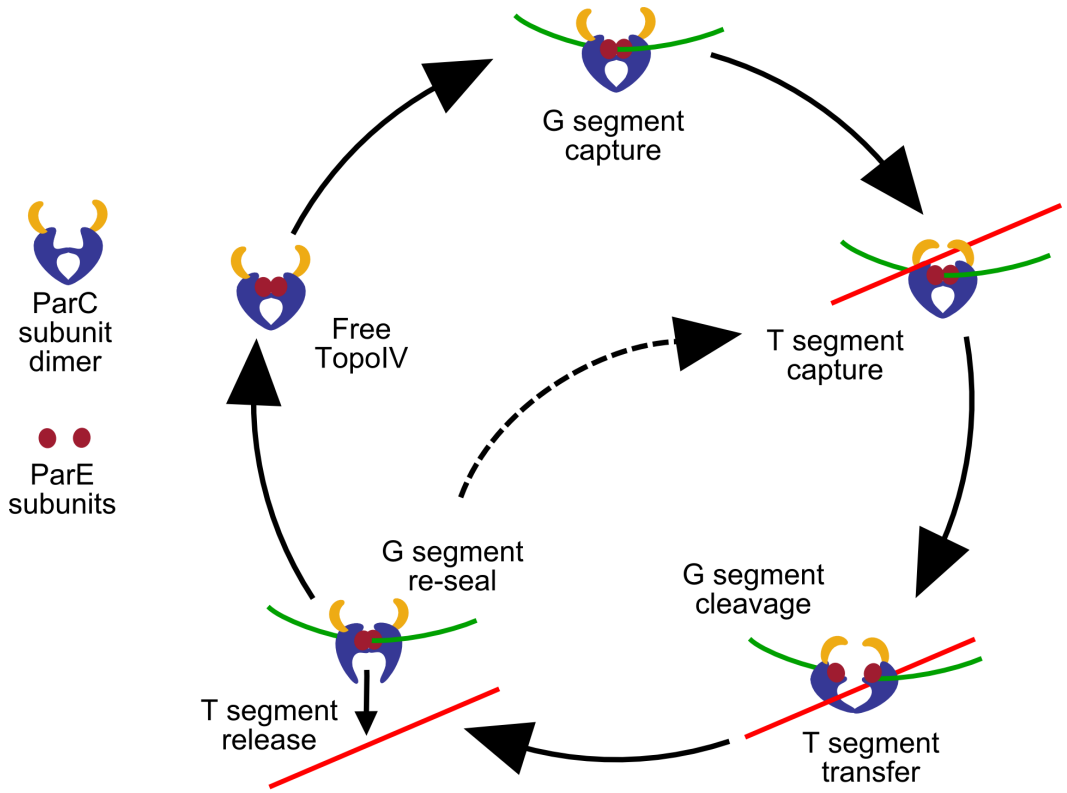


Figure 6.1: **Topoisomerase IV strand passage mechanism.** TopoIV forms a heterotetramer of two subunits; ParC and ParE. TopoIV first binds to the gate (G) segment of duplex DNA (shown in green). It subsequently captures a transfer (T) segment (shown in red), via ATP driven closure of the ParC N-terminal gate (yellow). This is followed by double-strand cleavage of the G-segment and passage of the T-segment across the cleaved DNA. The T-segment is then released as the enzyme dissociates. TopoIV has been shown to act processively on relaxation of +ve supercoils, which is thought to occur by the capture and passage of multiple T-segments during a single G-segment capture event, represented by the dashed arrow.

discriminates its substrates and which substrate is preferred *in vivo* remains to be fully understood (Lee et al., 2013; Vos et al., 2013a).

TopoIV forms a heterotetramer of two subunits; ParC and ParE (Figure 6.1). In this chapter we expand upon the single-molecule tracking techniques and analysis detailed in Chapters 3 and 4 to characterise the copy number, formation, and behaviour of TopoIV heterotetramers in live *E. coli*. Imaging both subunits combined with over-expression of unlabelled subunits allowed the fraction of molecules in

TopoIV heterotetramers to be determined. Imaging TopoIV in non-replicating cells shows that half of immobile molecules are dependent on replication. Finally, detection by immobilisation with long exposure times was used to estimate the specific binding time of TopoIV, showing that most catalysis *in vivo* is not processive.

6.2 Single-molecule tracking of ParC and ParE

Separate strains were created with either ParC or ParE labelled by replacing the endogenous genes with functional C-terminal fusions to the photoactivable fluorophore, PAmCherry (see Appendix A for characterisation of these strains). Photoactivated single-molecule tracking was performed on the same PALM microscope following the same protocols as detailed in Chapter 4. We used long movies photoactivating all molecules to estimate the copy numbers of ParC as 289 ± 34 and ParE as 210 ± 45, normalised to cells of 2.5 µm long.

Intracellular mobility of ParC

Similarly to the DNA-binding proteins in previous chapters, the distribution of ParC D^* values cannot be explained by a single diffusing species (Figure 6.2 A). Following the procedure we established for RNAP, we fitted the D^* distribution with a two-species model with an immobile population of molecules (established from the the PolI control) constrained at $D_{\text{imm}} = 0.11\,\mu\text{m}^2/\text{s}$ and a second unconstrained species. This fitting gave an immobile population accounting for 36 % ± 1 % of the molecules with the remaining 64 % ± 1 % of the population moving slowly at D_{slow} $0.35\,\mu\text{m}^2/\text{s}$ (Figure 6.2 B and C).

The mobile population of molecules had a lower mobility than expected for free 3D diffusion, consistent with them undergoing transient interactions with DNA.

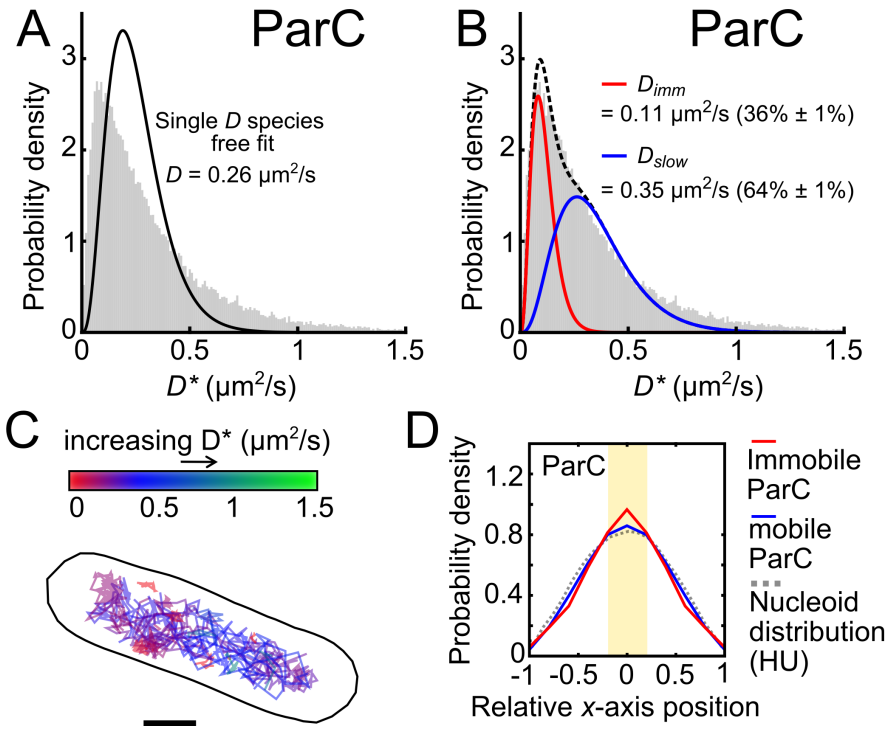


Figure 6.2: **Intracellular mobility of ParC.** A) The distribution of D^* values of 73000 ParC molecules, fitted with a single species model. B) The same distribution fitted with a two-species model with D_{imm} constrained at 0.11 μm^2/s. Ranges indicate 95 % confidence interval. C) An example cell with individual ParC trajectories coloured according to their D^* value. Scale bar = 1 μm. D) The x-axis distribution of immobile and mobile ParC molecules. 39 % ± 2 % of immobile ParC molecules and 34 % ± 1 % of mobile ParC molecules are located in the central region of the x-axis (defined here as the central 20 % shown highlighted in yellow). The distribution of nucleoid associated protein HU (dashed line) shows the average spatial distribution of the nucleoid.

Segmenting cells and plotting the x-axis distribution against the distribution of DNA (as stained by HU) showed that these slowly diffusing mobile ParC molecules were associated with the nucleoid (Figure 6.2 D).

Intracellular mobility of ParE

In contrast to ParC, the ParE D^* distribution showed a third population of molecules with higher mobility in addition to the two populations observed for ParC (Figure 6.3 A). As ParE does not bind DNA (Lee et al., 2013), these fast dif-

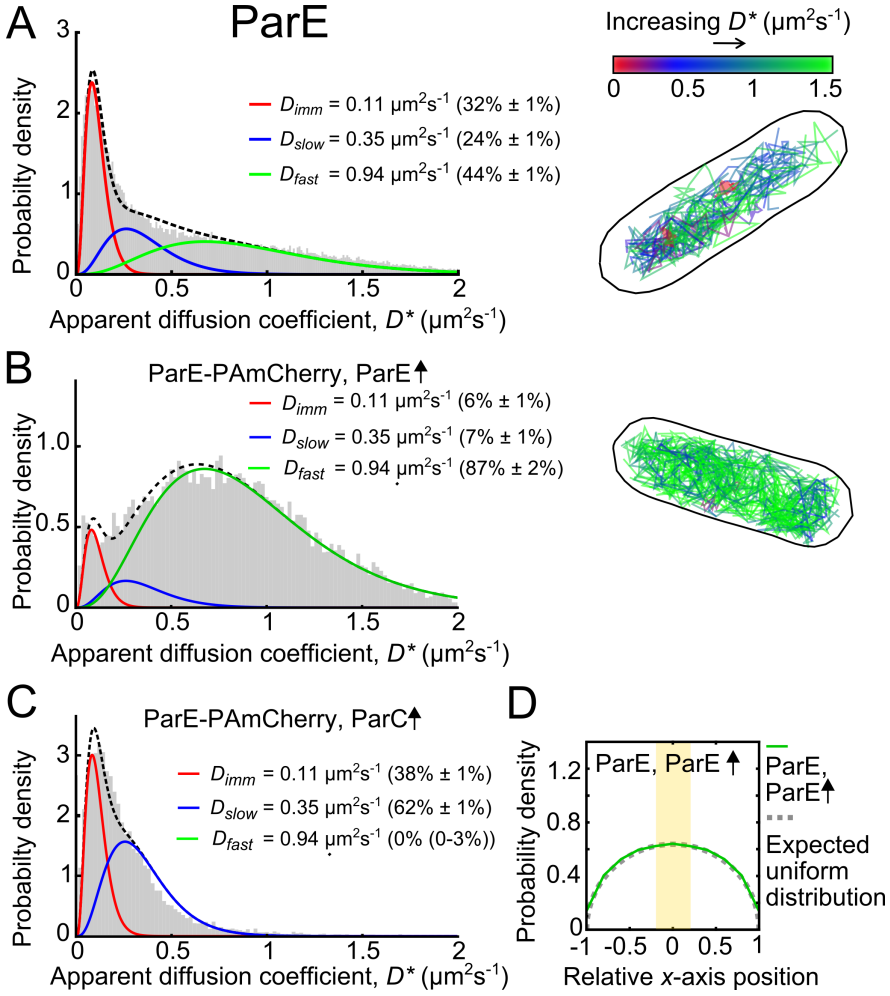


Figure 6.3: **Intracellular mobility of ParE.** A) The distribution of D^* values for 64000 ParE molecules fitted with a three species model (left). An example cell with individual trajectories coloured according to their D^* value (right). B) The distribution of D^* values of 7100 ParE-PAmCherry molecules in a strain overexpressing unlabelled ParE (left). The distribution is fitted to a three species model with two species constrained at $D_{\text{imm}} = 0.11\,\mu\text{m}^2/\text{s}$ and $D_{\text{slow}} = 0.35\,\mu\text{m}^2/\text{s}$. The third unconstrained fit shows the majority of molecules are fast moving, with $D_{\text{fast}} = 0.94\,\mu\text{m}^2/\text{s}$. An example cell with ParE trajectories coloured according to their D^* value. C) The distribution of D^* values of 9500 ParE-PAmCherry molecules in a strain overexpressing unlabelled ParC. The distribution is fitted to a three species model with D_{imm}, D_{slow}, and D_{fast} constrained. C) The x-axis distribution of ParE molecules in a strain over-expressing unlabelled ParE (377 cells). 25 % ± 1 % of molecules are located in the central 20 % region of the x-axis. The expected distribution for molecules uniformly distributed throughout a cylindrical cell volume (dashed grey line) has 25 % of molecules in the central region, agreeing well with the data.

fusing molecules may represent free ParE subunits, whereas the immobile and slow diffusing molecules were in TopoIV heterotetramers. This was tested by imaging ParE-PAmCherry molecules in cells in which unlabelled ParE was overexpressed, out competing labelled ParE in TopoIV heterotetramers. Consistent with the hypothesis that the fast moving population are uncomplexed ParE, ~90 % of ParE-PAmCherry molecules now diffused rapidly (Figure 6.3 B) and were uniformly distributed throughout the cell, showing no bias towards the nucleoid region (Figure 6.3 D). Fitting a three species model to this data (with D_{imm} and D_{slow} constrained) established that $D_{\text{fast}} = 0.94\,\mu\text{m}^2/\text{s}$.

Conversely, imaging ParE-PAmCherry molecules in a strain overexpressing unlabelled ParC resulted in a complete loss of the D_{fast} population (Figure 6.3 C), confirming that ParE molecules in the D_{imm} and D_{slow} states are complexed with ParC in TopoIV heterotetramers. The three species model for the ParE data, with constrained D_{imm}, D_{slow}, and D_{fast} values, showed that 32 % ± 1 % were immobile, 24 % ± 1 % diffusing slowly and 44 % ± 1 % , fast diffusing (Figure 6.3 A).

The copy number estimates (289 ± 34 for ParC and 210 ± 45 for ParE) indicated that there is no excess ParE in the cell. The 44 % of uncomplexed, fast diffusing ParE molecules must therefore reflect a steady state equilibrium of free ParE subunits to TopoIV heterotetramers, with 56 % of ParE subunits (~120 molecules) being present in ~60 TopoIV heterotetramers. Therefore, the ~60 TopoIV molecules present at any time form from a pool of ParC and E molecules sufficient for ~105 TopoIV heterotetramers. The copy numbers are normalised to cells of 2.5 µm long, and can be converted into intracellular concentrations based on the volume of a cell 2.5 µm long and ~0.8 µm wide (~1.1 µm). 1 molecule/cell therefore corresponds to a concentration of roughly ~1 nM.

At equilibrium, the dissociation constant between two proteins is expressed as:

$$K_d = \frac{[A][B]}{[AB]} \tag{6.1}$$

Where $[A]$ and $[B]$ represents the concentrations of protein A and B respectively and $[AB]$ represents the concentration of the complex. To calculate the dissociation constant of TopoIV heterotetramers it was assumed that a dimer of ParC will form a heterotetramer with a dimer of ParE. Using the copy numbers estimated the average concentrations in living cells, of free ParC dimers (85 nM), free ParE pairs (45 nM), and TopoIV (60 nM) yields an *in vivo* dissociation constant of TopoIV heterotetramers of ~65 nM.

Cell cycle independence

The previous results established that there is a steady state equilibrium of TopoIV heterotetramer formation when all cells are included. However, previous publications have proposed that ParC and ParE subunits only form active TopoIV molecules at replication termination as a consequence of replisome disassembly (Espeli et al., 2003). To test this we segmented cells and binned them by cell length (which is a reasonable proxy for their stage in the cell cycle), into three categories; short cells (2 µm to 3 µm long) having a single centrally located nucleoid, intermediate length cells (3 µm to 4 µm long), and longer cells (4 µm to 5 µm long) having replicated most of their chromosome which separate into distinct nucleoids prior to cell division (Figure 6.4 A). Localisations within segmented cell boundaries were tracked and D^* values determined. Fitting the distribution of D^* values for ParC/E in small, intermediate, and long cells showed a similar proportion of heterotetramers in cells at these three different stages of the cell cycle (Figure 6.4 B and C). This is difficult to

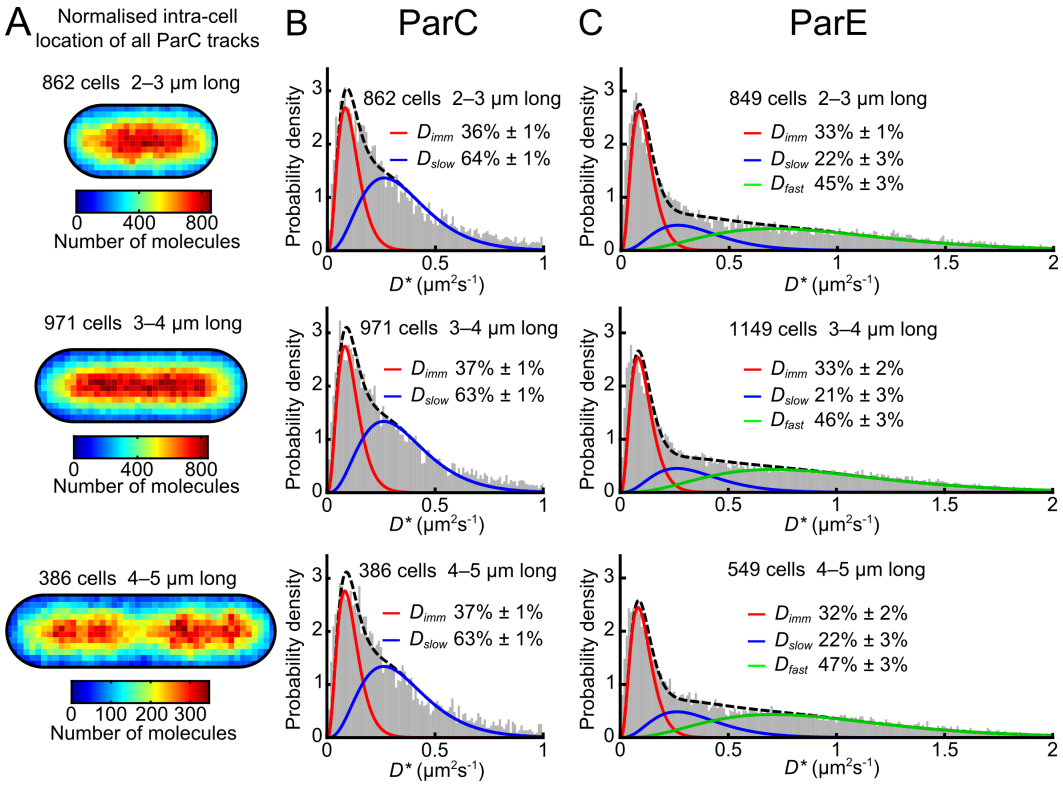


Figure 6.4: **TopoIV diffusion is cell cycle independent.** A) Normalised 2D histograms showing the average spatial distribution of ParC localisations from many cells. Binning by cell length from short to long cells shows progression through the cell cycle, with short cells (2 µm to 3 µm long) having a single centrally located nucleoid, and longer cells (4 µm to 5 µm long) having two clearly separate nucleoids. B) Distribution of ParC D^* values from 862 short cells (top), 971 medium-length cells (middle), and 386 long cells (bottom). C) Distribution of ParE D^* values from 849 short cells (top), 1149 medium-length cells (middle), and 549 long cells (bottom).

reconcile with the hypothesis that active TopoIV molecules form only at replication termination as a consequence of replisome disassembly.

6.3 TopoIV catalysis in non-replicating cells

To address whether TopoIV catalysis occurs in cells lacking (pre)catenanes, we analysed the ~7 % of new-born cells from a steady state population that had not yet initiated DNA replication. To do this ParC-PAmCherry was imaged in a strain with

DnaN labelled with the fluorescent protein mYPet (Figure 6.5 A). DnaN is the β unit of the replisome, which acts as a sliding clamp to keep DNA polymerase III bound to the DNA. The presence of an mYPet foci therefore reports on the formation of the replisome in these cells (Moolman et al., 2014). A snapshot of mYpet fluorescence was taken prior to acquisition of a PALM movie. Cells were segmented and the distribution of D^* values in short cells lacking mYPet-DnaN foci was plotted (Figure 6.5 B).The reduction in the immobile fraction of ParC from 36 % ± 1 % to 21 % ± 2 % in these cells, when compared to the whole population (Figure 6.5 A), indicated that almost half of immobile TopoIV molecules were dependent on replication.

6.4 TopoIV catalytic binding time

A large fraction of the immobile population of TopoIV molecules will be performing catalysis.[1] The strong replication dependence of these catalytically active molecules is consistent with the essential role of TopoIV in decatenation. However, it does not solve the puzzle of whether TopoIV is also significantly involved in relaxing +ve supercoils ahead of the replication fork. Relaxation of positive supercoils by TopoIV has been shown to be processive *in vitro*, with up to 80 left-handed links removed in single long bursts of catalytic activity lasting up to ~40 s (Neuman et al., 2009; Stone et al., 2003). On the other hand, negative supercoils, and presumably replicative catenanes (which have the same right-handed chirality as negatively supercoiled DNA), are relaxed distributively; only one cycle at a time, lasting ~1 s (Neuman et al., 2009; Stone et al., 2003).

[1] Some immobile TopoIV are associated with immobile MukBEF clusters, as detailed in the following chapter, but these are too transient to be detected at the exposure times used in this analysis

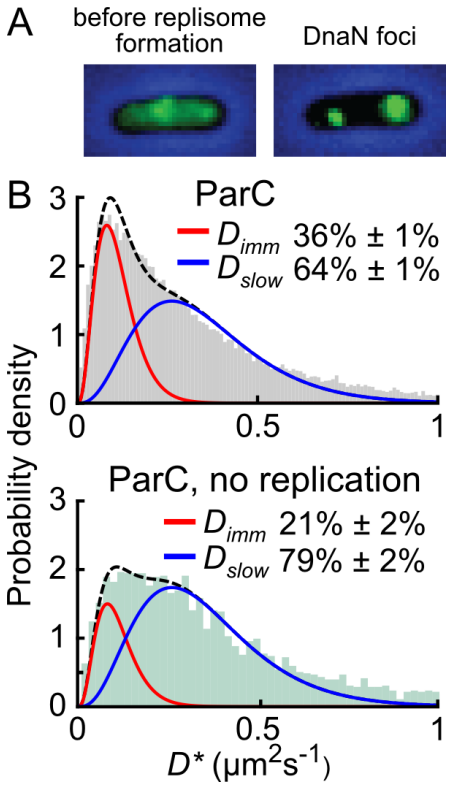


Figure 6.5: **Mobility of ParC in non-replicating cells.** A) Example YPet-DnaN labelled cells before replisome formation (left) and after replisome formation (right). B) Distribution of D^*values for 1930 ParC molecules in non-replicating cells, as assessed by lack of mYPet-DnaN foci prior to PALM data acquisition (bottom). Distribution of D^* values for ParC molecules in untreated cells (top) is reproduced here for comparison. Two species model fits with immobile molecules (red curves; constrained at $D_{\text{imm}} = 0.11\,\mu\text{m}^2/\text{s}$) and slow molecules (blue curves; constrained at $D_{\text{slow}} = 0.35\,\mu\text{m}^2/\text{s}$).

To determine the average duration of catalytic evens we turned to a different technique - detection by immobilisation (Yu et al., 2006; Elf et al., 2007; Xie et al., 2008). In order to increase our observation time we used low excitation intensities, sparse photoactivation, and long (≥ 0.5 s) exposure times. At these exposure times mobile molecules are motion blurred over a large fraction of the cell, whereas immobile molecules appear as point sources, producing a diffraction-limited spot (Figure 6.6 A). Immobile molecules could therefore be distinguished by the width of the elliptical Gaussian fits to the fluorescent spot. Here we refer to the short elliptical

axis as the x-width and the long eleiptical axis as the y-width.

Following the protocol developed by Uphoff et al. (2013), we used thresholds established with Pol1 (with clearly resolvable immobile molecules), of < 160 nm short axis-width and < 200 nm long axis-width to identify immobile molecules (Figure 6.6 B). The distribution of the number of frames which theses molecules were on were fitted with single exponential functions to extract the time constants.

The observed molecule disappearance rate is a sum of the underlying dissociation rate and the photobleaching rate:

$$\frac{1}{t_{\text{on}}} = \frac{1}{t_{\text{bound}}} + \frac{1}{t_{\text{bleach}}} \tag{6.2}$$

The binding-time, t_{bound} can therefore be calculated from the the product of the observed on-time, t_{on} and the bleaching time t_{bleach}:

$$t_{\text{bound}} = \frac{t_{\text{on}} \bullet t_{\text{bleach}}}{t_{\text{bleach}} - t_{\text{on}}} \tag{6.3}$$

The bleaching-time distributions were measured independently using MukB-PAmCherry, which binds DNA in clusters with a dwell time of ~50 s (Badrinarayanan et al., 2012a), with the same acquisition and excitation conditions. We measured ParC on-times at 0.5 s, 0.75 s, and 1.0 s exposure times and corrected the values by measuring the corresponding MukB photobleaching giving the values shown in Table 6.1. The mean binding time for all exposure times measured was 1.8 s ± 0.4 s, which is consistent with distributive relaxation of -ve supercoils seen *in vitro*. Although the presence of a small population of long lasting binding events cannot be ruled out, the results indicate that the majority of catalytic events *in vivo* are not processive, and hence TopoIV is either performing very little +ve supercoil relaxation, or that the

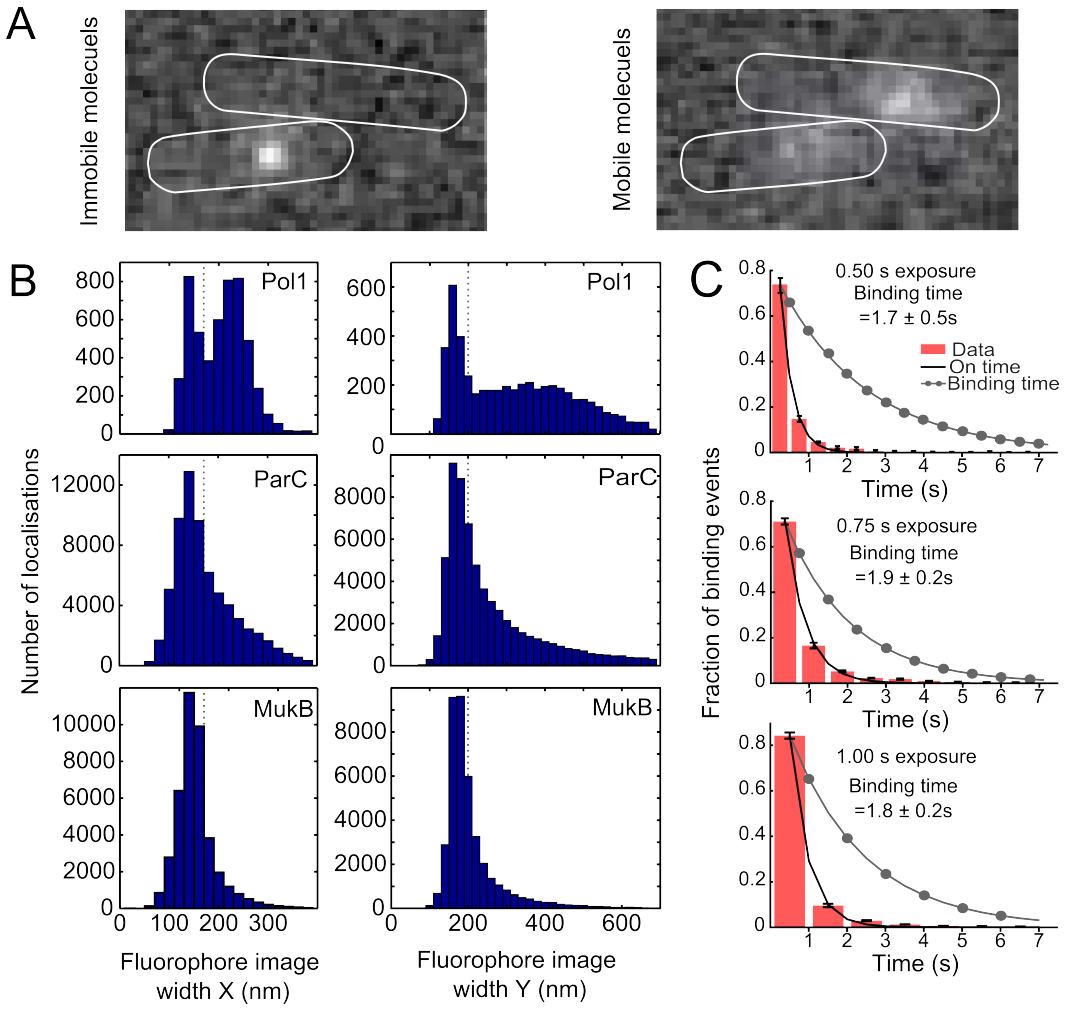


Figure 6.6: **TopoIV binding time.** A) Example 750 ms exposure frames showing cells with an immobile TopoIV molecule (left) and a mobile molecule (right). B) Short axis (x-axis) and long axis (y-axis) elliptical Gaussian fit widths at 750 ms exposure time for Pol I + MMS, ParC, and MukB. We used thresholds of < 160 nm x-axis-width and < 200 nm y-axis-width to identify immobile molecules determined from the Pol I control where immobile and mobile molecules are clearly resolvable (dashed line). C) On-time distributions for bound TopoIV with exponential fit (line) and photobleaching-corrected binding time distribution (line with dots).

relaxation occurs distributively *in vivo*. Because only the presence of precatenanes, and not negative supercoils, are dependent on replication, we concluded that in wildtype cells undergoing replication, the majority of catalytically active TopoIV molecules are performing decatenation.

	500 ms	750 ms	1000 ms
t_{bleach} (s)	0.423	0.647	0.672
t_{on} (s)	0.328 ± 0.020	0.503 ± 0.030	0.484 ± 0.020
t_{bound} (s)	1.68 ± 0.52	1.94 ± 0.22	1.83 ± 0.16

Table 6.1: The exponential time constants for ParC on-times t_{on} at three different exposure times. Errors indicate sem from three experimental repeats. The bleaching time constants for PAmCherry t_{bleach} were determined from the MukB on-times, and used to generate the corrected ParC binding times t_{bound} using Equation 6.3.

6.5 Discussion

This chapter provides a comprehensive understanding of the formation and behaviour of TopoIV molecules in their native unperturbed environment inside living cells. We observed a dynamic equilibrium between free and complexed ParC/E subunits, independent of replication, with ~60 potentially functional TopoIV heterotetramers at birth. The observation that only ~60 % of subunits are in heterotetramers at any given time raises the possibility that more functional enzymes can be assembled if required.

Conventional fluorescence imaging of ParC and ParE saw segregation of the localisation of the two subunits, and led to the hypothesis that active TopoIV molecules form only at replication termination as a consequence of replisome disassembly (Espeli et al., 2003). The results presented in this chapter contradict this hypothesis, showing that there is a similar proportion of heterotetramers in cells of all ages. Furthermore, other results have indicated that TopoIV acts in decatenation at *ori* and the whole chromosome (Joshi et al., 2013; Wang et al., 2008). Taken together these are difficult to reconcile with a model whereby TopoIV formation is highly temporally regulated.

In vitro TopoIV can relax both +ve and -ve supercoiled DNA, as well as catenated DNA (Postow et al., 2001; Koster et al., 2010; Vos et al., 2011). Indeed, *in vivo*, TopoIV

was shown to relax +ve (left handed) supercoils most efficiently, with a single processive burst of catalytic activity lasting tens of seconds that can remove up to 80 links (Crisona, 2000; Stone et al., 2003). In contrast, -ve supercoiled DNA is relaxed distributively, two links at a time. However, *in vivo*, inactivation of TopoIV has been shown to prevent decatenation and segregation of newly replicated sister loci, while replication and transcription remained unaffected (Wang et al., 2008), showing that the essential role of TopoIV is decatenation behind the fork. Consistent with this our results show a reduction of immobile TopoIV molecules in non-replicating cells.

The catalytic cycle of ~1.8 s that we have proposed fits well with a mechanism in which TopoIV acts distributively on right handed replicative catenanes *in vivo*, since right handed supercoils, which have the same chirality, were relaxed distributively *in vitro*. If this processive behaviour were recapitulated *in vivo*, it would lead to catalytically active TopoIV molecules being immobile for tens of seconds, events not detected by PALM. Therefore, we conclude that +ve supercoil relaxation by TopoIV occurs rarely *in vivo*, or it occurs distributively. It is possible that, although TopoIV *in vitro* can relax +ve supercoils, *in vivo* DNA gyrase is responsible for the large majority of +ve supercoil relaxation in front of replication forks and the transcription machinery (Vos et al., 2011), leaving TopoIV free to act in its essential role of decatenation.

6.6 Materials and methods

Bacterial strains

The bacterial strains and plasmids used in this chapter are listed in Table 6.2. All strains are derivatives of the *E. coli* K-12 AB1157 strain(Bachmann, 1972). ParC and ParE PAmCherry fusion strains were generated by Pawel Zawadzki using the λ Red

Strains	Genotype
KK10	*parC::PAmCherry kan*
PZ103	*parE::PAmCherry kan*
PZ124	*parE::PAmCherry kan; dnaN::mYpet frt; Gyr* A^{L83} *tet*
PZ125	*parC::PAmCherry kan; dnaN::mYpet frt; Gyr* A^{L83} *tet*
Plasmids	Genotype
pZ65	*parE* cloned into pBAD24 using NcoI and XbaI cloning sites
pZ66	*parC* cloned into pBAD24 using NcoI and XbaI cloning sites

Table 6.2: TopoIV fusion strains and plasmids

method (Datsenko and Wanner, 2000). For multiple insertions of modified genes, the *kanr* gene was removed using site-specific recombination through expression of the Flp recombinase from the pCP20 plasmid (Datsenko and Wanner, 2000). The fusion strains showed identical growth rates to WT cells, for full characterisation see Appendix A.

Cell preparation

Strains were prepared as detailed in Chapter 3. Cells were grown in M9 growth media supplemented with 0.2 % glycerol. Norfloxacin was used at 5 µg/ml, cells were placed on agarose pad containing the drug 5 minutes prior the experiment.

Microscopy

PALM microscopy was performed on the same custom-built TIRF microscope detailed in Chapter 2 using the acquisition and analysis protocol described in Chapter 3. YFP excitation was provided by a 488 nm laser.



Chapter 7

The action of TopoIV is directed by MukB

TopoIV, which has an essential role in Escherichia coli chromosome decatenation, interacts with MukBEF, an SMC (Structural Maintenance of Chromosomes) complex that acts in chromosome segregation. In this chapter we characterise the interaction between TopoIV molecules and MukBEF clusters *in vivo*. We show that ~15 TopoIV molecules/cell are associated with MukBEF clusters that are preferentially localised to the replication origin region (*ori*), located in the centre of the nucleoid. This interaction is transient, but directs some TopoIV catalytic activity to this region. Finally, we show that the MukB-ParC interaction is crucial for timely decatenation and segregation of newly replicated *ori* DNA.

7.1 Introduction

In the previous chapter we characterised the *in vivo* behaviour of Topoisomerase IV. TopoIV plays a crucial role in DNA segregation as it is responsible for unlinking catenated chromosomes. A second class of proteins, Structural Maintenance of Chromosome (SMC) complexes also play an equally important, but less well understood, role in faithful DNA segregation (Hirano, 2006).

SMC complexes are found in in all three domains of life (Nolivos and Sherratt, 2014), and are essential for successful chromosome segregation and transmission in all organisms. The *E. coli* SMC protein, MukB was identified through the isolation of a *muk*- mutant that generated anucleate cells as a consequence of impaired chromosome segregation (Hiraga et al., 1989). MukB shares little primary amino acid sequence homology with other SMCs, yet it retains the distinctive SMC architecture (Nolivos and Sherratt, 2014); with two domains joined together via a ~50 nm long intramolecular coiled-coil. MukB forms dimers via the hinge domain with both the N and C terminals forming an ATPase head domain at the other end of the coiled coil. The two head domains can be together or separate resulting in an V or O dimer conformation.

Inactivation of the MukB protein or either of the two accessory proteins, MukE and MukF, results in abnormal chromosome organisation and segregation (Danilova et al., 2007; Nolivos and Sherratt, 2014). The circular *E. coli* chromosome is organised by bidirectional replication into two equal left and right arms (replichores). Each arm occupies a separate cell half, with the origin of replication (*ori*) at mid-cell (Danilova et al., 2007). MukBEF forms clusters that associate with *ori* (Badrinarayanan et al., 2012a), and MukBEF impairment has been shown to cause abnormal positioning of *ori* and loss of the separation of replichores into distinct cell halves. It has therefore been proposed that MukBEF initiates the normal bidirectional organisation of the chromosome from the *ori* region (Danilova et al., 2007), however, the mechanism by which MukBEF might bring about this global organisation remains unclear.

In vitro the MukB dimerization hinge has been shown to physically interact with TopoIV via its ParC subunit, which stimulates relaxation of negative supercoils by TopoIV (Hayama and Marians, 2010; Li et al., 2010). *In vivo* an enrichment of ParC molecules in the vicinity of *ori*-associated MukBEF clusters was observed

(Nicolas et al., 2014), but evidence of a direct interaction between TopoIV and MukBEF in living cells and the biological significance of such an interaction have yet to be demonstrated. In this chapter we characterise the interaction between TopoIV molecules and MukBEF clusters and show this interaction directs the location and action of TopoIV.

7.2 Half of immobile TopoIV molecules result from Interaction with MukB

ParC interacts with the MukB dimerization hinge *in vitro* (Hayama and Marians, 2010; Li et al., 2010; Vos et al., 2013b) and shows an enrichment near MukBEF clusters *in vivo* (Nicolas et al., 2014). We tested whether a fraction of the immobile ParC and ParE molecules result from their binding to immobile MukBEF clusters on DNA. Fitting a two species model (with the D_{imm} and D_{slow} populations established in Chapter 6) to the distribution of D^* values for ParC molecules in Δ*mukB* cells showed nearly a 50 % reduction in the immobile fraction of ParC/ParE consistent with these molecules being immobile as a consequence of their interaction with *ori*-associated MukBEF clusters (Figure 7.1 A and B). Using a clustering algorithm to define ParC clusters containing ≥ 25 localisations, showed that ParC formed a median of one cluster per cell and deletion of MukB removed most ParC clustering (Figure 7.1). This was confirmed by the radial distribution analysis of all ParC localisations which showed a strongly clustered distribution, which was reduced ~ four-fold in Δ*mukB* cells (Figure 7.2).

Since Δ*mukB* cells have disorganised chromosomes (Danilova et al., 2007), the reduction in the fraction of immobile ParC/E molecules in these cells could be a consequence of global chromosome organisation rather than a direct interaction

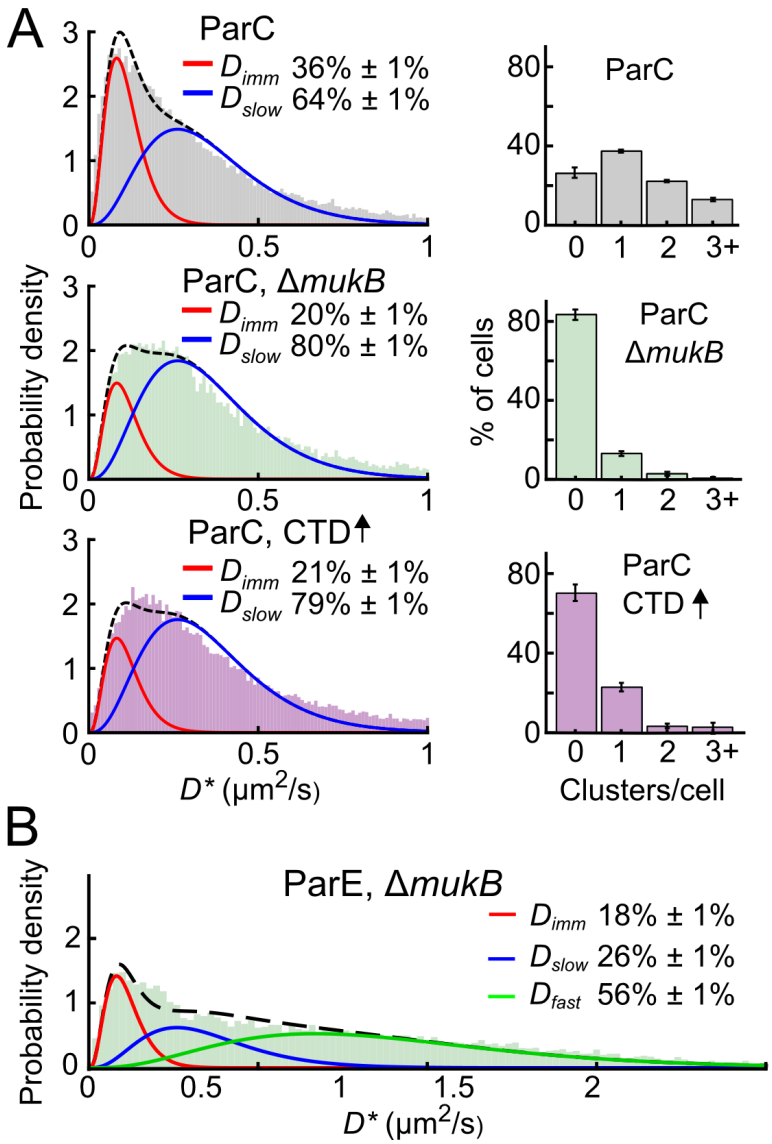


Figure 7.1: **ParC interacts with MukB.** A) Left panels; The distribution of D^* values for ParC molecules fitted with a two species model with immobile molecules (constrained at $D_{\text{imm}} = 0.11\,\mu\text{m}^2/\text{s}$) and slow moving molecules (constrained at $D_{\text{slow}} = 0.35\,\mu\text{m}^2/\text{s}$). Top, ParC molecules in wildtype cells. Middle, 19,000 ParC molecules in Δ*mukB* cells. Bottom, 42,000 ParC molecules after unlabelled ParC-CTD over-expression. Right panels; the number of ParC clusters per cell, determined by clustering all localisations in 2600 wild-type cells, 387 Δ*mukB* cells, and in 214 cell overexpression ParC-CTD. Error bars indicate SEM of three experimental repeats. B) The distribution of D^* values for 16,000 ParE molecules in Δ*mukB* cells fitted with a thee species model with D_{imm}, D_{slow}, and $D_{\text{fast}} = 0.94\,\mu\text{m}^2/\text{s}$ constrained.

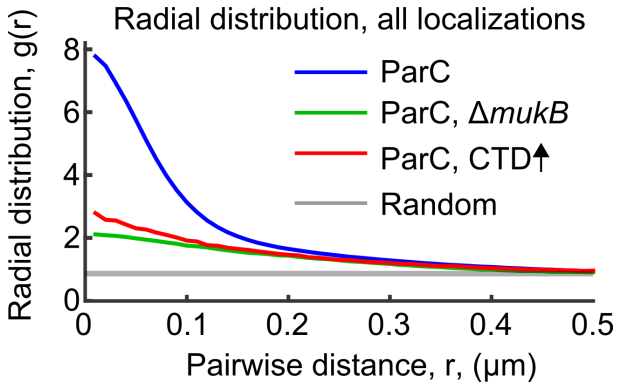


Figure 7.2: **Radial distribution of ParC.** The radial distribution of all ParC localisations for unperturbed cells (blue), Δ*mukB* cells (green), and cells overexpressing ParC-CTD (red).

between TopoIV and MukB. To distinguish these possibilities, the TopoIV-MukB interaction was impaired by over-expressing an unlabelled ParC C-terminal domain (ParC-CTD), which binds MukB (Vos et al., 2013b), thereby out-competing TopoIV binding. Over-expression of ParC-CTD did not significantly affect growth rate, cell length, nucleoid condensation, or formation of anucleate cells (Table A.2), consistent with unperturbed chromosome organisation. Flow cytometry profiles showed a small increase in cells with multiple chromosomes (see Appendix A). Under these conditions, the immobile fraction of ParC was reduced to the level as in Δ*mukB* cells (Figure 7.1 A, bottom), and clustering of ParC was lost, showing that ~half of immobile ParC molecules are dependent on a direct interaction with immobile MukBEF clusters. The loss of clustering was confirmed from the radial distribution of all localisations (Figure 7.2).

To demonstrate that ParC clusters spatially associate with MukBEF clusters, we imaged ParC-PAmCherry and MukB-mYPet in the same cells, by taking an mYPet snapshot immediately prior to PALM acquisition. Calculating the radial distribution function of ParC PALM localisations with respect to the centroid of each MukBEF focus showed a clear enrichment of ParC near MukBEF foci compared to randomly

distributed molecules (Figure 7.3). 20 % of ParC localisations were within 200 nm of MukBEF centroids. This result is consistent with the ~16 % of ParC molecules that were immobile due to a direct interaction with MukB, as judged by the reduction in the fraction of immobile molecules in Δ*mukB* cells.

An intriguing feature of the ParC interaction with MukB is that while ParC clusters are nearly always in close proximity to a MukBEF focus, not all MukBEF foci are associated with a ParC cluster. This is qualitatively highlighted in intensity projections from epifluorescent imaging (Figure 7.4 C). Imaging MukB-PAmCherry with PALM showed that, despite having a similar copy number (195 ± 72 copies/cell, Figure 7.4 A), MukBEF formed ~twice as many clusters per cell as ParC (Figure 7.4 B); indicating an additional level of regulation governing the MukB-ParC interaction yet to be elucidated.

In the previous chapter we estimated the copy number of ParE as 210 ± 45. Roughly 14 % of these molecules are immobile and dependent on MukB (Figure 7.1), indicating that ~15 TopoIV molecules (~30 ParE molecules) are associated with MukB at any given time. This can be verified independently from the 20 % of 289 ParC molecules located within 200 nm of MukB in the radial distribution analysis (~58 ParC molecules; Figure 7.3). Since ~42 % of all ParC molecules were in TopoIV heterotetramers, this gives a value of ~12 TopoIV molecules spatially associated with MukBEF clusters. Converting copy numbers to intracellular concentrations gives between 12 nM and 15 nM of TopoIV-MukB, 45 nM of free TopoIV, and 195 nM of MukB, yielding an *in vivo* dissociation constant of between ParC and MukB of ~0.4 μM, in good agreement with *in vitro* measurements (Li et al., 2010).

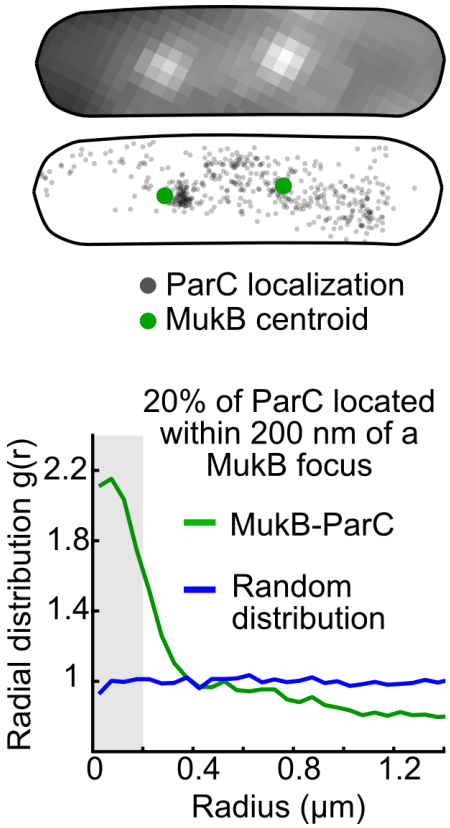


Figure 7.3: **Clustering of ParC near MukB foci.** An example cell with MukB-mYPet foci visualised prior to PALM acquisition (top) and localisations of ParC-PAmCherry molecules (middle). The radial distribution function showing the probability of finding a ParC localisation at distance, r, from a MukB focus compared to random distribution (717 cells; bottom). Grey bar shows localisation within 200 nm.

7.3 MukB directs immobile TopoIV molecules to centre of the nucleoid

To understand further how MukBEF clusters direct the organisation of immobile ParC molecules within the nucleoid, we categorised ParC molecules as immobile or mobile and determined the probability density across the cell x axis, following the methodology described in Chapter 4. The analysis showed that immobile ParC molecules were preferentially located towards the centre of the x axis of the cell (39 % ± 2 % of immobile molecules located in the central 20 % region, yellow highlighted region

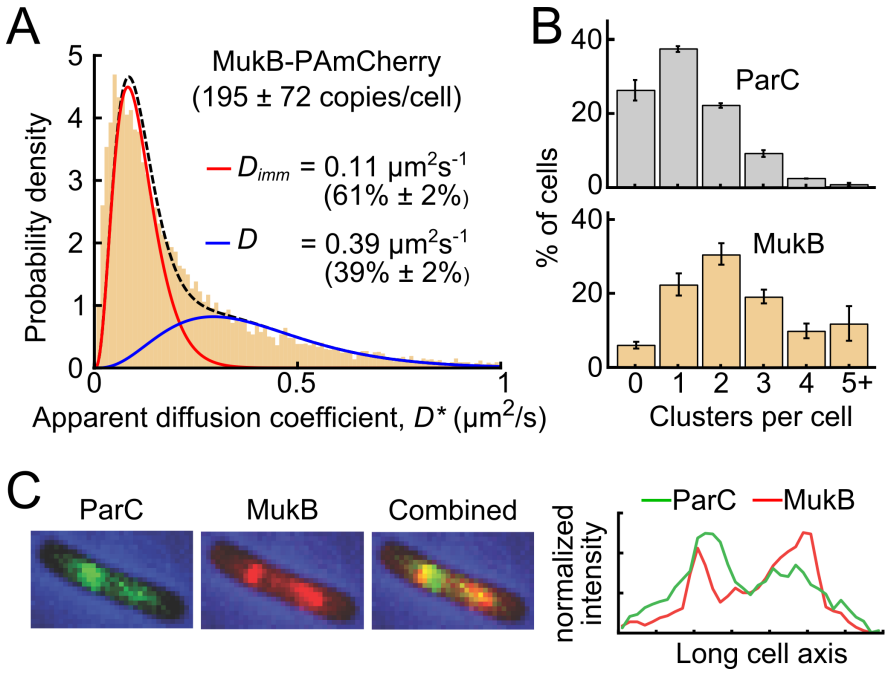


Figure 7.4: **MukB clusters.** A) The distribution of D^* values for 9000 MukB-PAmCherry trajectories. The distribution was fitted to a two species model with D_{imm} constrained at 0.11 μm^2/s and a second unconstrained fit. B) The number of clusters per cell formed by MukB defined by the same clustering parameters used for ParC. For comparison the number of ParC clusters per cell is also shown here. C) Example of a cell imaged in epifluorescence with labelled MukB-mCherry foci (red) and ParC-mYPet foci (green). The long-axis lines-scan of the fluorescence intensity is shown on the right. Frequently we observed only some of MukB foci were associated with distinct ParC clusters.

Figure 7.5 top), indicating their position in the centre of the nucleoid. A similar intracellular positioning was observed for immobile MukB-PAmCherry molecules, with an even stronger bias of immobile molecules towards the centre of the nucleoid (49 % ± 5 % of immobile molecules located in the central region; Figure 7.5 bottom). In Δ*mukB* cells, immobile ParC molecules showed a lower probability of in the centre of the nucleoid, indicating that MukBEF clusters recruit some ParC molecules towards the centre of the nucleoid (32 % ± 4 % of immobile molecules located in the central region; Figure 7.5 , middle). Note that the distribution of mobile molecules (which account for about two thirds of ParC trajectories) varies little in wildtype and

Δ*mukB* cells, highlighting the importance of discriminating immobile and mobile molecules.

7.4 Two Populations of immobile TopoIV Molecules

To dissect TopoIV binding events further we analysed long trajectories of > 10 localisations and applied a 4 frame moving average. Using this average we sorted molecules into three categories: mobile molecules that remained above the $D^* = 0.16\,\mu m^2/s$ threshold for the observation period, immobile molecules which remained below the threshold over the observation period, and molecules which exhibited transitions between these states (Figure 7.6 A). This analysis detected similar fractions of immobile molecules as determined from fits to the D^* distributions. In addition, a fraction of the molecules underwent transitions, consistent with TopoIV molecules being in a dynamic equilibrium between bound and mobile states (Figure 4A; right).

In time-lapse experiments, using 15 ms exposures followed by 35 ms delays, there was a reduction in the population of molecules that remained immobile over the course of the trajectory from 36 % to 14 %, when the same analysis was performed (Figure 7.6 B). This result was also evident in distribution of D^* values (Figure 7.6 C). Molecules in the immobile category in the time-lapse experiments (bound for> 10 localisations with a 50 ms frame time) must remain bound for ≥ 0.5 s, compared to ≥ 0.15 s for the immobile molecules in normal 15 ms frame time experiments (bound for > 10 localisations). The observed reduction in the fraction of immobile molecules shows that (~21 %) of the binding events in Muk+ cells lasted for ≥ 0.5 s.

In contrast, when we performed the same analysis in Δ*mukB* cells, or in cells overexpressing ParC-CTD, the fraction of immobile molecules remained broadly unchanged between normal and time-lapse PALM experiments (Figure 7.6 B and D).

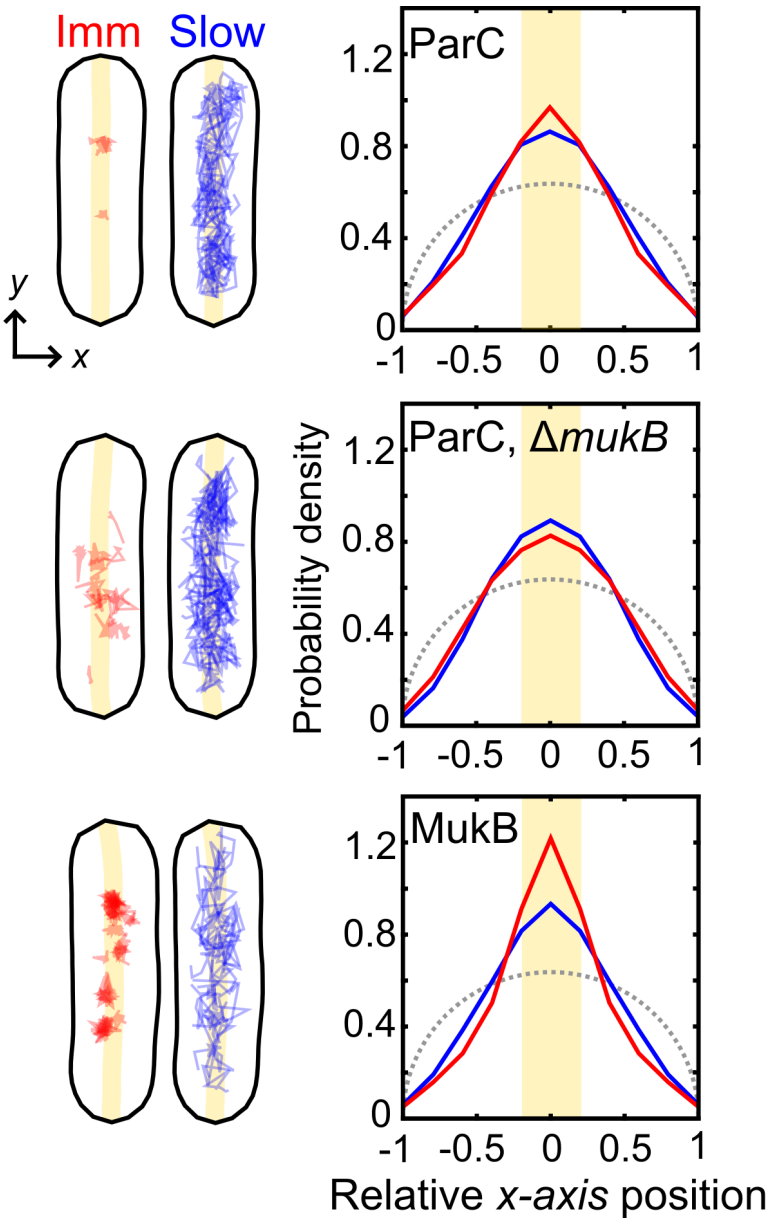


Figure 7.5: **Intracellular organisation of MukB and TopoIV.** Example cells show intracellular location of ParC molecules sorted by their D^* value as immobile (red) and slow (blue)(top left). The probability distribution of immobile and slow moving ParC molecules across the cell x-axis in unperturbed cells (2386 cells, top) and Δ*mukB* cells (253 cells, middle). Also show is the same analysis performed on MukB molecules in wild-type cells (852 cells, bottom). Dotted lines indicate the expected uniform distribution in a cellular cylinder and yellow highlighted regions indicate central 20 % of cell. 39 % ± 2 % of immobile ParC molecules are in the central region in unperturbed cells, which is reduced to 32 % ± 4 % in Δ*mukB* cells. 49 % ± 5 % of immobile MukB are located in the central region

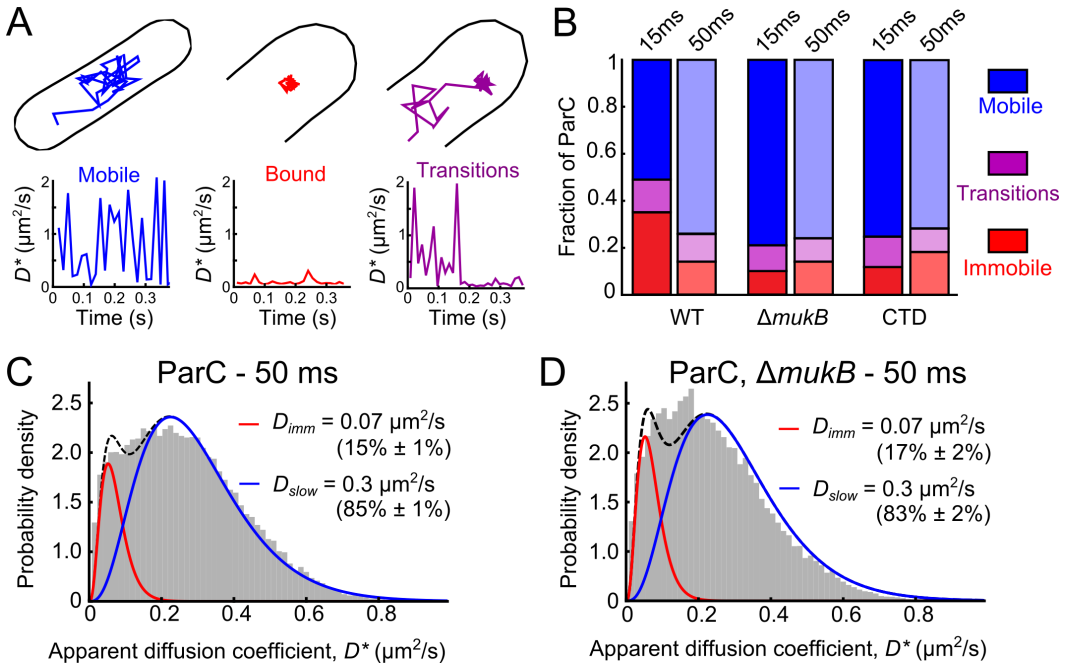


Figure 7.6: **ParC shows a Muk-dependent and Muk-independent binding behaviour.** A) Examples of long ParC trajectories (> 10 localisations) classified according to their D^* transitions. Molecules mobile over observation period (blue), immobile (red), and undergoing transition from one state to another (purple). B) Bar graph of all ParC trajectories for the indicated strains, classified from PALM experiments performed at 15 ms continuous acquisition and time-lapse (15 ms exposure + 35 ms delay). C) The distribution of D^* values for 28,000 ParC molecules imaged with a time-lapse (15 ms exposure + 35 ms delay). D) The distribution of D^* values for 26,000 ParC molecules in Δ*mukB* cells imaged with a time-lapse.

This suggests that in wild-type cells, a population of MukB-dependent transiently immobile TopoIV molecules is present alongside a population of MukB-independent molecules immobile for ≥ 0.5 s. The molecules that remained immobile over the entire trajectory in time-lapse experiments are likely to be molecules undergoing catalysis with a mean binding duration of ~1.8 s as shown in the previous chapter.

Because the underlying binding times are exponentially distributed, they cannot be intuitively extracted from these experiments. We therefore turned to Monte Carlo simulations of molecules interconverting between two diffusive states: D_{free} and D_{imm} (as detailed in Chapter 5) to gain a better estimate of the durations of the

short-lived MukB-dependent binding events. The free diffusion, D_{free}, of TopoIV heterotetramers was calculated based on the free diffusion of ParE (D_{fast}), correcting for their relative sizes. Simulating diffusion inside a confined cell volume showed that the $D_{\text{fast}} = 0.94\,\mu m^2/s$ for ParE was recapitulated with a corrected diffusion coefficient of $D = 2.4\,\mu m^2/s$. The TopoIV heterotetramer contains 2 ParE and 2 ParC molecules. Given this, we estimated that the TopoIV heterotetramer had ~4 x the volume of ParE, and hence a radius ~1.6 times as large. Because the diffusion coefficient for a spherical particle depends linearly on the radius (Equation 3.1), we therefore estimate that the D_{free} for TopoIV is $\sim 1.5\,\mu m^2/s$.

Molecule trajectories were simulated to generate localisations at either 15 ms intervals or 15 ms intervals with 35 ms delays to match normal and time-lapse experiments, respectively. The simulated localisations were analysed with the same tracking and categorising algorithm as used for the experimental data. We simulated interconverting molecules with different exponentially distributed mean binding durations from 0.1 ms to 150 ms (keeping the fraction in each state equal). Plotting the change in the fraction of molecules categorised as bound in time-lapse simulations compared to normal simulations showed that a binding duration of ~30 ms to 70 ms for MukB-dependent TopoIV transient binding events recapitulated the experimentally observed decrease (Figure 7.7). Furthermore, simulations with a binding time much smaller than the 15 ms exposure time showed little difference between continuous and time-lapse simulations, suggesting that non-specific ParC-DNA interactions in the absence of MukB are very transient.

In conclusion, we have shown that TopoIV molecules display two different bound states. In the previous chapter we saw events lasting ~1.8 s, which we hypothesise correspond to molecules preforming catalysis. Here we also see that some immobile molecules are relatively short-lived (~30 ms to 70 ms), and are dependent on a direct

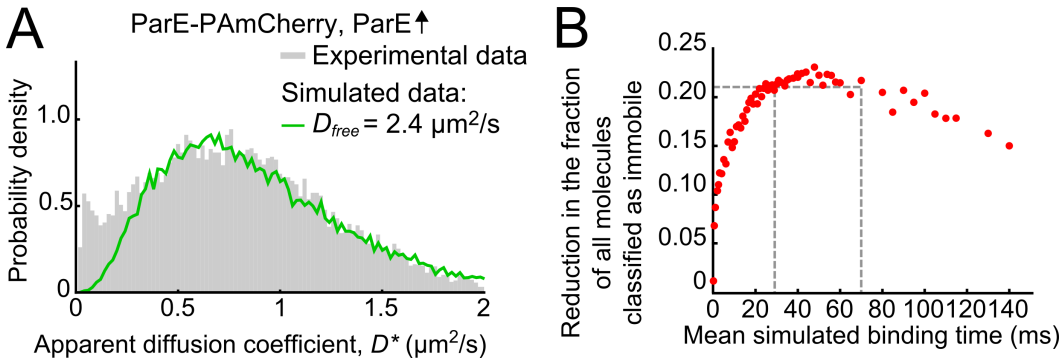


Figure 7.7: **TopoIV simulated diffusion.** A) Comparison between experimental data for ParE-PAmCherry molecules in a strain overexpressing unlabelled ParE and the distributions of the simulated diffusing species. The fast moving population agrees well with simulated molecules with $D_{\text{free}} = 2.4\,\mu\text{m}^2/\text{s}$. B) Transition categorisation analysis of simulated interconverting species with varying binding dwell times. Tracks were simulated with 15 ms exposure times and in time-lapse with 15 ms exposures followed by a 35 ms delay, then categorised in the same way as experimental data in Figure 7.6. The reduction in the total fraction of molecules categorised as immobile between normal and time-lapse simulations was plotted against the binding time. A range of binding times between ~30 ms to 70 ms agree with the observed reduction in immobile ParC in time-lapse experiments (reduced from 36 % to 14 %).

interaction with MukB, leading us to the conclusion that the ParC-MukB complex is short lived.

7.5 Catalytically active TopoIV

To establish if the catalytic activity of TopoIV is also directed by the interaction with MukB, we treated cells with the antibiotic norfloxacin, which specifically blocks type II topoisomerases during their catalytic cycle by covalently binding them to DNA (Khodursky et al., 1995). To prevent norfloxacin acting on DNA gyrase, this was performed in a strain carrying a norfloxacin-resistant gyrase gene. A two-fold increase in the fraction of immobile ParC/E molecules after ~10 min norfloxacin treatment was observed (Figure 7.8 A and B), showing that most TopoIV molecules had performed catalysis during this period; although we cannot exclude the possibility that

norfloxacin captures a fraction of non-productive catalytic events that do not result in topological changes. Longer incubation with norfloxacin did not increase the fraction of immobile molecules, showing that at ~10 min we had reached saturation. In Δ*mukB* cells a similar fraction of immobile molecules was observed after saturating norfloxacin treatment, which agrees with the fact that Δ*mukB* cells can still decatenate their chromosomes (Danilova et al., 2007; Nicolas et al., 2014). Shorter norfloxacin incubation times were not experimentally tractable, hence the overall rate of catalysis in the cell was faster than could be measured in both wildtype and Δ*mukB* cell so any global stimulation of decatenation by the MukB-ParC interaction could not addressed with this assay.

After norfloxacin treatment of wild type cells we observed a modest increase in the number of TopoIV clusters/cell, whereas Δ*mukB* cells showed a similar cluster distribution to wild type cells (Figure 7.9 A), suggesting that TopoIV molecules may undergo catalysis in defined clusters. The enrichment of ParC molecules close to MukBEF clusters, as judged by radial distribution analysis, was retained after norfloxacin treatment (Figure 7.9 A), showing that a fraction of TopoIV molecules underwent catalysis close to MukBEF clusters.

When we analysed the distribution of catalytically active TopoIV molecules across the short cell axis we found that the previous bias towards the cell centre was lost (34 % ± 3 % in the central region after norfloxacin treatment compared to 39 % ± 2 % before treatment), presumably because the additional clusters were not associated with MukBEF clusters (Figure 7.8 C). MukBEF clusters were retained after norfloxacin and ParC colocalised with them (Figure 7.8 B). In the absence of MukBEF clusters, covalently linked DNA-ParC molecules were slightly less likely to be located in the centre of the nucleoid (31 % ± 4 % in the central region), similar to the situation in Δ*mukB* cells before norfloxacin treatment. Taken together, the results indicate

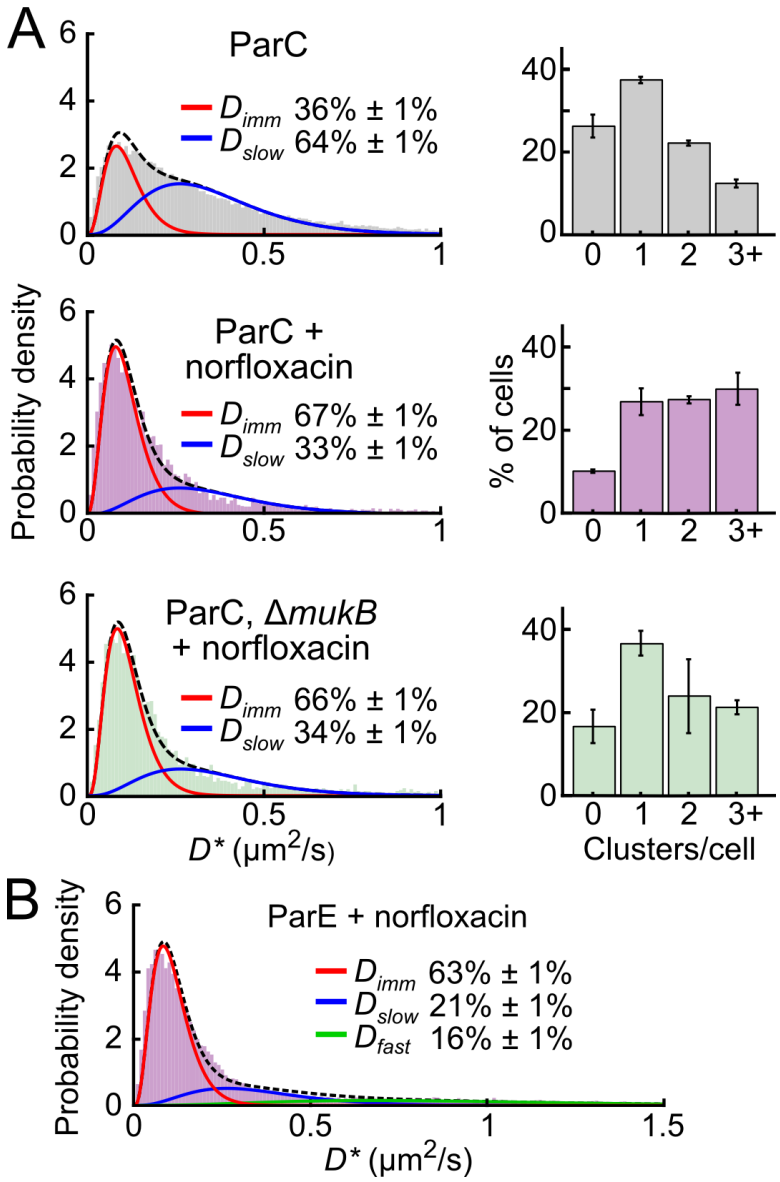


Figure 7.8: **Diffusion of catalytically active TopoIV.** A) Left panels; the distribution of D^* values for ParC molecules in WT cells (middle, 387 cells) and Δ*mukB* (bottom, 214 cells) cells after ~10 min treatment with norfloxacin. Control ParC molecules in untreated cells (top). Two species model fits with immobile molecules (red curves; constrained at $D_{\text{imm}} = 0.11\,\mu\text{m}^2/\text{s}$) and slow molecules (blue curves; constrained at $D_{\text{slow}} = 0.35\,\mu\text{m}^2/\text{s}$). Right panels; number of ParC clusters per cell for steady-state populations of cells. Error bars indicate SEM of three experimental repeats. B) Distribution of D^* values for 17,000 ParE molecules after 10 min treatment with norfloxacin. ParE and ParC had a similar fraction of immobile molecules after norfloxacin treatment.

that MukBEF clusters direct the catalytic activity of some TopoIV molecules to the nucleoid centre, whereas MukBEF-independent catalysis occurs throughout the nucleoid.

7.6 The interaction between ParC and MukB facilitates *ori* decatenation

The observation of a direct interaction between ParC and MukBEF *in vivo*, combined with the evidence in Chapter 6 that the majority of catalysis by TopoIV was decatenation, raised the question if the ParC-MukB interaction had a functional consequence on the decatenation of newly replicated DNA. To complement the single-molecule studies presented here, Kasia Ginda used an epifluorescence assay to assess the time of *ori* separation after replication before and after impairment of the ParC-MukBEF interaction.

Using a strain with a sequence of *tetO* operators inserted at the *ori1* and expressing mYPet labelled Tet repressors shows the location of *ori*. In time-lapse experiments, the time taken for *ori1* loci to separate after replication initiation was measured (Figure 7.10 A and B). Replication initiation was determined by the appearance of a fluorescent mYPet-DnaN focus (as described in Chapter 6). The *ori1* locus replicates < 30 s after initiation at *oriC* (Wang et al., 2008), and a sufficient amount of mYPet-DnaN loads at the forks to visualise it within <2 min of initiation (Moolman et al., 2014). Cells in which the MukBEF-ParC interaction is normal had stably segregated 50 % of the newly replicated *ori1* loci by 17 min after the appearance of mYPet-DnaN. In contrast, cells in which the MukBEF-ParC interaction was impaired by ParC-CTD over-expression showed a ~12 min increase in the time required for 50 % of cells to exhibit *ori*1-segregation (Figure 7.10 C), showing that the MukB-TopoIV interaction

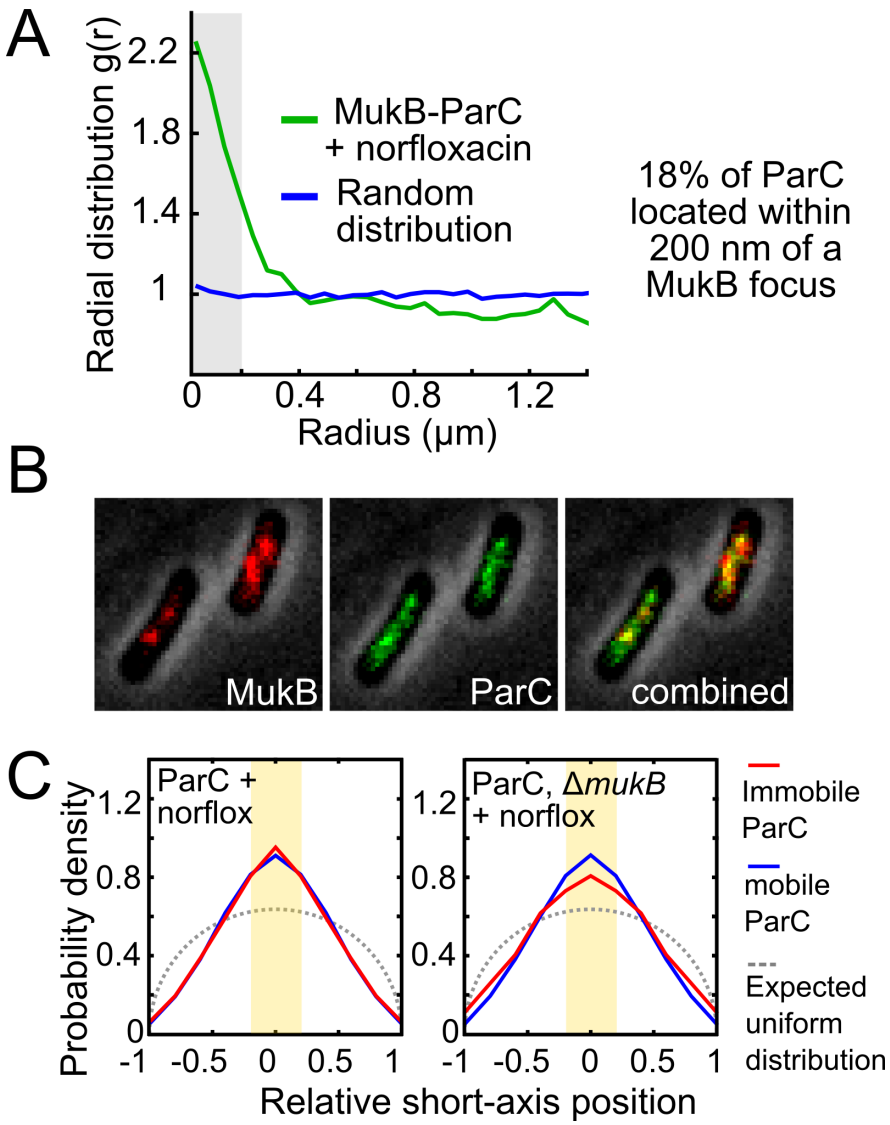


Figure 7.9: **Spatial distribution of TopoIV catalysis.** A) Radial distribution of ParC localisations from each MukB focus in cells treated with norfloxacin, compared to random distribution (726 cells). B) Example cells treated with norfloxacin. Epi-fluorescence microscopy shows MukB-mCherry foci are still present and some of ParC-YPet clusters are associated with these MukB foci. C) The probability density of ParC molecules across the cell x-axis in 858 cells treated with norfloxacin (left). Dashed grey line shows the expected distribution for molecules uniformly distributed throughout a cylindrical cell volume. The mobile and immobile populations have similar distributions, both with 36 % of molecules located in the central 20 % of the width (yellow highlighted area). The distribution of ParC molecules across the cell short-axis in 463 Δ*mukB* cells treated with norfloxacin (right). The mobile population has 36 % of molecules located in the central 20 % of short-axis, whereas the immobile population shows fewer molecules (33 %) in the central region.

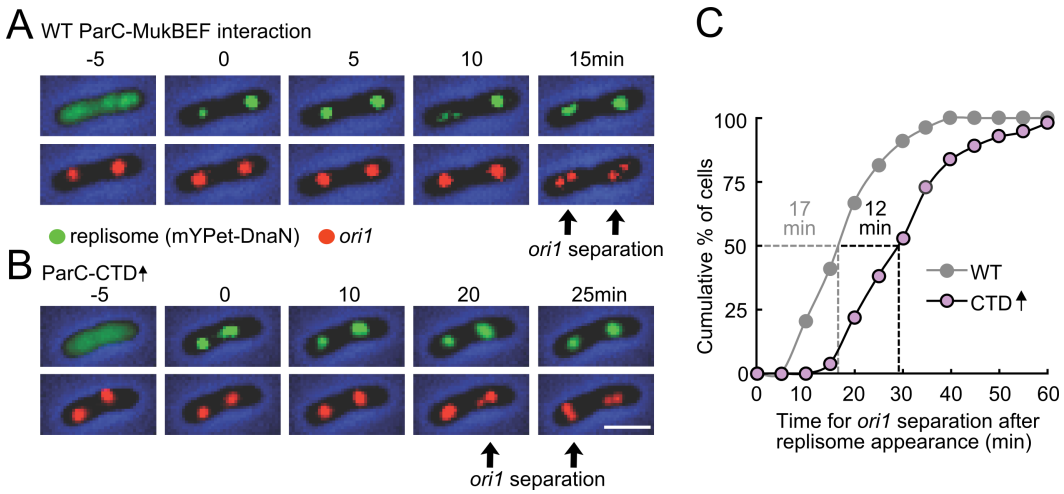


Figure 7.10: **The MukB-ParC interaction stimulates chromosome segregation at *ori*.** A) Representative example cell from time-lapse experiments of wild type cells transformed with empty expression plasmid (pBAD24). Black arrows indicate the time of *ori1* segregation. 0 min time was defined by replisome appearance at *ori1*. B) Representative example cell from the time-lapse experiment where ParC-CTD was over-expression for 3 h prior to the experiment. C) Cumulative distribution of times of *ori1* locus segregation after replication initiation, marked by appearance of mYPet-DnaN foci at *ori1*. The times for 50 % of the population to segregate after initiation are shown.

plays a role in timely decatenation of newly replicated *ori* DNA.

7.7 Discussion

In the previous chapter we characterised the behaviour of TopoIV, showing that catalysis lasted on average ~1.8 s. Mobile ParC molecules, and mobile ParE molecules in TopoIV heterotetramers were slower than expected for free 3D diffusion, and remained associated with the nucleoid region, indicating a high fraction of short lived non-specific binding. We have seen from norfloxacin incubation that TopoIV can mediate catalysis in the absence of MukB, which occurs throughout the nucleoid. This Muk-independent catalytic pathway is highlighted in Figure 7.11.

In this chapter we demonstrated that the interaction seen *in vivo* between ParC

and MukB exists *in vivo*. A fraction of TopoIV molecules appear immobile due to the presence of MukB. Over-expressing the C terminal domain of ParC, which interacts with MukB *in vivo*, demonstrated that this immobile fraction is due to a direct interaction between ParC and MukB. Imaging in time-lapse suggested that the MukB-ParC interaction was relatively short lived, nevertheless it directs the activity of some TopoIV molecules towards MukBEF clusters. The functional role of this interaction was confirmed by the observation of a delay in decatenation of *ori* sisters after the interactions was perturbed. We propose, therefore, that the MukB-dependent pathway guides the activity of some TopoIV molecules towards MukBEF clusters, which in turn promotes decatenation at *ori* (Figure 7.11).

The TopoIV interaction with MukBEF clusters may favour *ori* decatenation due to several different mechanisms, the simplest of which could be that TopoIV-MukB binding causes an increase in local concentration of TopoIV near *ori*. Alternatively the promotion of decatenation could be due to MukB capturing DNA strains in a conformation preferable for TopoIV catalysis. TopoIV must first capture the G-segment, but, before gate cleavage in must also capture the T-segment. Since the MukBEF clusters are relatively stably associated with DNA (Badrinarayanan et al., 2012a), multiple DNA strand held in close proximity by MukB could promote T-segment capture. Finally, the direct interaction with MukB might affect TopoIV substrate specificity to favour decatenation.

The coordination of type II topoisomerases by SMC complexes may not be limited to *E. coli*. In eukaryotes the SMC complex, condensin, has been shown to aid decatenation by the TopoIV counterpart, Topo II (Charbin et al., 2014). Models have also been proposed whereby coordination between condensin and Topo II helps compact the chromosome during mitosis (Baxter and Aragón, 2012). The sequential and coordinated action of TopoIV and MukBEF in the successive steps of decatenation and

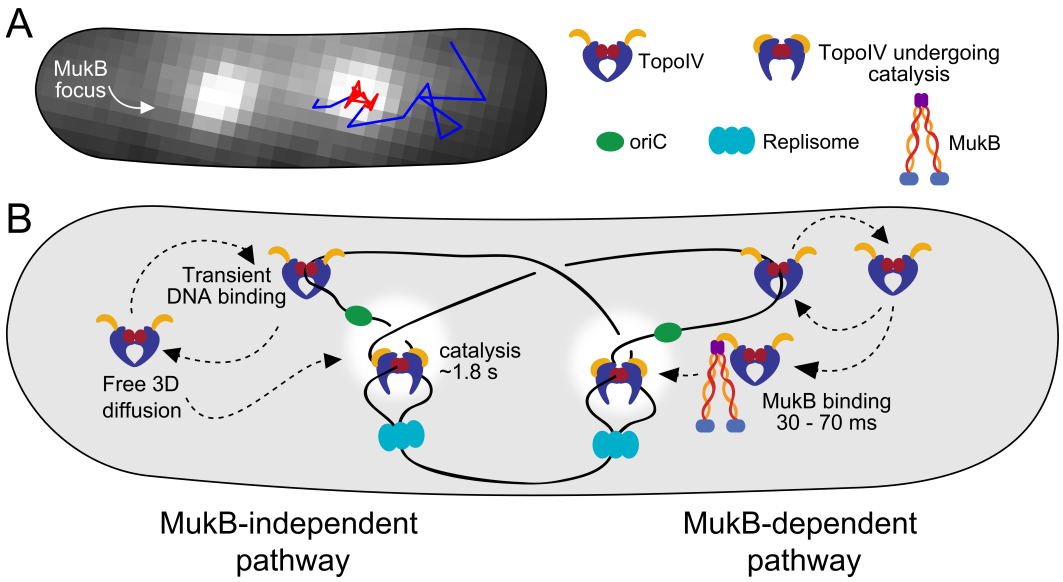


Figure 7.11: **MukB-dependent pathway guides TopoIV activity to the proximity of MukBEF clusters.** A) An example cell showing MukBEF clusters and single TopoIV trajectory showing extended binding in the region of MukBEF cluster. B) Model of MukB-dependent and -independent action of TopoIV. The TopoIV molecules that act independently of MukBEF undergo many unsuccessful attempts before effective capture of G- and T- DNA segments. In contrast, MukBEF clusters direct a fraction of TopoIV molecules to the *ori* region of the chromosome. We propose that the interaction with MukBEF clusters recruits TopoIV and promotes interaction of TopoIV with DNA, thereby enhancing TopoIV catalysis.

chromosome segregation revealed here provides the platform for future mechanistic studies that will reveal whether SMC complexes provide DNA-protein substrates that provide selectivity for topoisomerase action, and precisely how topoisomerase action is coordinated with SMC functions in chromosome processing.

7.8 Materials and methods

The bacterial strains and plasmids used in this chapter are listed in Table 7.1 and were generated by Pawel Zawadzki using the protocols described in Chapter 6. Characterisation of the fusion strains and ParC-CTD overexpression strain are detailed in Appendix A. Strains were prepared as detailed in Chapter 6. Where indicated

Strains	Genotype
PZ120	*parE:: PAmCherry kan*; *mukB::mYPet frt*; $GyrA^{L83}$ *tet*
PZ147	*parC:: PAmCherry kan*; *mukB::mYPet frt*; $GyrA^{L83}$ *tet*
PZ128	*parE:: PAmCherry frt*; Δ*mukB* kan; *dnaN:: mYPet frt*; $GyrA^{L83}$ *tet*
PZ129	*parC:: PAmCherry frt*; Δ*mukB* kan; *dnaN:: mYPet frt*; $GyrA^{L83}$ *tet*
Ab238	*mukB:: PAmCherry kan*; (Badrinarayanan et al., 2012a)
PZ122	*parC::ypet, kan*; *mukB::mCherry frt*; $GyrA^{L83}$ *tet*
Plasmids	Genotype
pZ68	*parC-CTD* (amino acid residues 497-806) cloned into pBAD24

Table 7.1: Bacterial strains and plasmids

norfloxacin was used at 5 μg/ml, cells were placed on agarose pad containing the drug 5 minutes prior the experiment. PALM microscopy was performed on the same custom-built TIRF microscope detailed in Chapter 2. Fluorophore image localisation and tracking was performed as described in Chapter 4.



Chapter 8

Conclusion and outlook

The development of super-resolution microscopy, such as PALM, allows imaging with a resolution below the fundamental diffraction limit of light. But static images can only tell us so much about the underlying mechanisms. When the photoactivation and localisation technique central to PALM imaging is combined with single-particle tracking this technique becomes particularly powerful, giving both increased spatial resolution, but also information on the movement of the protein of interest. This can be complemented by quantitative conventional fluorescence microscopy and ensemble super-resolution techniques. In this thesis we have demonstrated how these can be used to study DNA binding proteins in live *E. coli*, both at the single-molecule and the population level.

8.1 Organisation of enzyme activity in the nucleoid

We first used RNAP as an example to demonstrate how photoactivated single-molecule tracking can be used to discriminate between mobile molecules and

molecules specifically bound to DNA. The distribution of specifically bound RNAP molecules showed that low levels of transcription can occur throughout the nucleoid; however, clustering analysis and 3D SIM showed that dense clusters of transcribing RNAP molecules form preferentially at the nucleoid periphery. In faster growth conditions, the clustering of RNAP increases as does the phase separation between the densest clusters of RNAPs and the densest regions of the nucleoid. These findings suggest that transcription can cause spatial reorganisation of the nucleoid, with movement of gene loci out of the bulk of DNA as levels of transcription increase.

This work provides a global view of the role of transcription in the organisation of the bacterial nucleoid, but it also raises an interesting question. Do other DNA-binding proteins perform their function homogeneously throughout the nucleoid, or like transcription, do they have a preferred location of action? When we look at other results in this thesis we can start to answer this question. We saw that the global transcriptional silencer H-NS has the opposite trend to RNAP, with immobile clusters locating preferentially in the centre of the nucleoid. The SMC complex MukB also tended to be located in the centre of the nucleoid, and directs some of the activity of TopoIV to this region. On the other hand, double stranded DNA break repair has previously been shown to occur at the nucleoid periphery (Lesterlin et al., 2013). It seems clear, therefore, that the activity taking place in the nucleoid is not homogeneously distributed. From the examples here, it seems possible metabolic activity takes place in clusters at the nucleoid surface, with less metabolised DNA and silenced genes in the more compacted centre of the nucleoid. However, we have shown that the action of some TopoIV molecules are directed more centrally in the nucleoid, so this may not be universally true.

The eukaryotic cell cycle allows for chromosome replication, segregation, and cell division to take place in separate stages with checkpoints ensuring completion

of each stage. In contrast, in bacteria such as *E. coli* all these processes happen simultaneously (Cox et al., 2000; Pomerantz and O'Donnell, 2010), yet they are interdependent and must be coordinated. This is especially true at fast growth rates when transcription takes place at the same time as multiple active replication forks, yet RNAP and other NAPs must not prevent the progression of the replisome. Indeed, RNAP clusters and the replisome are thought to generally occupy separate chromosomal territories (Cagliero et al., 2014). However, the mechanism behind this high level of coordination remains poorly understood, but points to the clear need to study this behaviour in its native environment inside living cells. Using the techniques and assays developed in this thesis to perform a more methodical comparison of the spatial distribution of active molecules for a range of DNA-binding proteins could help answer these questions.

8.2 Dissemination of techniques

The wealth of quantitative information available from *in vivo* single-molecules techniques makes them a powerful tool for molecular biology. However, the technical complexity of the microscopy and analysis has limited the uptake of these techniques outside of specialist biophysics labs. As part of this thesis we have designed, built, and maintained a live-cell single-molecule microscope facility, available for any researchers to use. Both the microscope and the analysis software have already been adopted by non-specialist users.

Some of the assays described in this thesis are also be more broadly applicable. For example, as demonstrated in Chapter 6, determining the fraction of subunits of an enzyme in different diffusive states can tell us about its formation and dissociation. Titrating out labelled subunits with unlabelled competitors provides a broadly applicable assay to characterise the formation of a protein complex whose subunits

display significantly different binding behaviour. The mobility based approaches detailed in this thesis are also not limited to studying protein-DNA interactions, but can more generally be applied to any interaction between a diffusing protein and a static, or slowly moving target, such as a small protein interacting with the much larger ribosome (English et al., 2011), or a cytoplasmic protein associating with a membrane protein.

8.3 Future outlook

The focus of this thesis was bacteria, which have served as excellent testing grounds for developing *in vivo* single-molecule techniques. Beyond the importance of understanding bacteria as pathogens and for biotechnological applications, bacterial model systems continue to surprise us with new fundamental biology (Deveau et al., 2010). Nevertheless, there is a clear need to move single-molecule techniques to studying DNA-binding proteins inside the nucleus of eukaryotic cells, despite the technological challenges it entails (e.g., decreasing fluorescence signal at larger depths from the coverslip). Recent developments with reflected light sheet microscopy, where a thin horizontal section of a cell can be illuminated, have shown improvements in signal-to-noise (Gebhardt et al., 2013). This method has recently been used to study the spatial distribution of RNAP II and the dynamics of transcription factors with single molecule accuracy inside mammalian cell nuclei (Zhao et al., 2014; Izeddin et al., 2014). Single-molecule microscopy is not just limited to cell biology, and has already been reported in living vertebrates (Schaaf et al., 2009)

The future promises exciting improvements to the techniques used in this research. Organic dyes hold many advantages over fluorescent proteins; they are smaller, brighter, and more photostable. Labelling with synthetic fluorophores would allow a molecule to be observed over longer times and with better accuracy. Two of

the key obstacles to using these dyes are, firstly, internalising them into live cells, and secondly, directing them to specific targets. Progress is being made on both these fronts; several existing methods are available to direct fluorophores by tag-directed protein labelling, such as SNAP-tags and HaloTags, and enzyme-mediated protein labelling, which uses native or engineered enzymes to ligate a small molecule probe to a recognition site (Fernández-Suárez and Ting, 2008; Hinner and Johnsson, 2010). Unnatural amino acid labelling offers a promising technique to site specifically label proteins in live cells, but the technique remains challenging to implement (Zhang et al., 2003; Plass et al., 2012). Recent work has allowed site-specific labelling of two different fluorophores for in vitro FRET experiments (Wang et al., 2014). An alternative approach involves internalising purified and *in vitro* labelled proteins with electroporation (Crawford et al., 2013). Further development and application of these techniques offers exciting new possibilities for live-cell single-molecule measurements.

The research in this thesis shows how quantitative imaging can bridge the gap between test tubes and living cells, giving information such as, dwell times, enzyme reaction rates, protein copy numbers, and stoichiometries, previously only available from *in vitro* biochemistry assays. *In vivo* single-molecule microscopy is clearly set to become a widely used tool as new methods make completely novel experiments possible, but also as existing techniques and analysis become more accessible to non-specialists users.

References

Ando, T. and Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43): 18457–62, October 2010. ISSN 1091-6490.

Bachmann, B. J. Pedigrees of some mutant strains of Escherichia coli K-12. *Bacteriological reviews*, 36(4): 525–57, December 1972. ISSN 0005-3678.

Badrinarayanan, A., Lesterlin, C., Reyes-Lamothe, R., and Sherratt, D. The Escherichia coli SMC complex, MukBEF, shapes nucleoid organization independently of DNA replication. *Journal of bacteriology*, 194(17):4669–4676, June 2012a. ISSN 1098-5530.

Badrinarayanan, A., Reyes-Lamothe, R., Uphoff, S., Leake, M. C., and Sherratt, D. J. In vivo architecture and action of bacterial structural maintenance of chromosome proteins. *Science (New York, N.Y.)*, 338(6106):528–531, October 2012b. ISSN 1095-9203.

Bakshi, S., Choi, H., Rangarajan, N., Barns, K. J., Bratton, B. P., and Weisshaar, J. C. Nonperturbative imaging of nucleoid morphology in live bacterial cells during an antimicrobial peptide attack. *Applied and environmental microbiology*, 80(16): 4977–4986, August 2014. ISSN 1098-5336.

Bakshi, S., Dalrymple, R. M., Li, W., Choi, H., and Weisshaar, J. C. Partitioning of RNA Polymerase Activity in Live Escherichia coli from Analysis of Single-Molecule Diffusive Trajectories. *Biophysical Journal*, 105(12):2676–2686, 2013.

Bakshi, S., Siryaporn, A., Goulian, M., and Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology*, 85(1):21–38, June 2012. ISSN 0950382X.

Baxter, J. and Aragón, L. A model for chromosome condensation based on the interplay between condensin and topoisomerase II. *Trends in genetics : TIG*, 28(3):110–117, March 2012. ISSN 0168-9525.

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*, 313(5793):1642–1645, September 2006. ISSN 1095-9203.

Bratton, B. P., Mooney, R. A., and Weisshaar, J. C. Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli. *Journal of bacteriology*, 193(19):5138–5146, October 2011. ISSN 1098-5530.

Bremer, H. and Dennis, P. P. *Escherichia coli and Salmonella: cellular and molecular biology*. American Society for Microbiology Press, Washington, DC, 1996.

Cabrera, J. E., Cagliero, C., Quan, S., Squires, C. L., and Jin, D. J. Active transcription of rRNA operons condenses the nucleoid in Escherichia coli: examining the effect of transcription on nucleoid structure in the absence of transertion. *Journal of bacteriology*, 191(13):4180–4185, July 2009. ISSN 1098-5530.

Cabrera, J. E. and Jin, D. J. The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues. *Molecular Microbiology*, 50(5):1493–1505, November 2003. ISSN 0950382X.

Cagliero, C., Zhou, Y. N., and Jin, D. J. Spatial organization of transcription machinery and its segregation from the replisome in fast-growing bacterial cells. *Nucleic Acids Research*, 42(22):13696–13705, November 2014. ISSN 0305-1048.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. Structural

Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell*, 104(6):901–912, March 2001. ISSN 00928674.

Charbin, A., Bouchoux, C., and Uhlmann, F. Condensin aids sister chromatid decatenation by topoisomerase II. *Nucleic acids research*, 42(1):340–348, January 2014. ISSN 1362-4962.

Cheezum, M. K., Walker, W. F., and Guilford, W. H. Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophysical journal*, 81 (4):2378–2388, October 2001. ISSN 0006-3495; 0006-3495.

Cook, P. R. The Organization of Replication and Transcription. *Science*, 284(5421):1790–1795, June 1999. ISSN 00368075.

Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. The importance of repairing stalled replication forks. *Nature*, 404(6773):37–41, March 2000. ISSN 0028-0836.

Crawford, R., Torella, J., Aigrain, L., Plochowietz, A., Gryte, K., Uphoff, S., and Kapanidis, A. Long-Lived Intracellular Single-Molecule Fluorescence Using Electroporated Molecules. *Biophysical Journal*, 105 (11):2439–2450, 2013.

Crisona, N. J. Preferential relaxation of positively supercoiled DNA by E. coli topoisomerase IV in single-molecule and ensemble measurements. *Genes & Development*, 14(22):2881–2892, November 2000. ISSN 08909369.

Crocker, J. C. and Grier, D. G. Methods of Digital Video Microscopy for Colloidal Studies. *Journal of Colloid and Interface Science*, 179(1):298–310, April 1996. ISSN 00219797.

Danilova, O., Reyes-Lamothe, R., Pinskaya, M., Sherratt, D., and Possoz, C. MukB colocalizes with the oriC region and is required for organization of the two Escherichia coli chromosome arms into separate cell halves. *Molecular Microbiology*, 65(6): 1485–1492, September 2007. ISSN 0950-382X.

Datsenko, K. A. and Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6640–6645, June 2000. ISSN 0027-8424.

Deveau, H., Garneau, J. E., and Moineau, S. CRISPR/-Cas system and its role in phage-bacteria interactions. *Annual review of microbiology*, 64:475–493, January 2010. ISSN 1545-3251.

Dorman, C. J. and Deighan, P. Regulation of gene expression by histone-like proteins in bacteria. *Current Opinion in Genetics & Development*, 13(2):179–184, April 2003. ISSN 0959437X.

Durisic, N., Laparra-Cuervo, L., Sandoval-Álvarez, A., Borbely, J. S., and Lakadamyali, M. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nature methods*, 11(2):156–162, January 2014. ISSN 1548-7105.

Dworsky, P. and Schaechter, M. Effect of rifampin on the structure and membrane attachment of the nucleoid of Escherichia coli. *Journal of bacteriology*, 116(3):1364–1374, December 1973. ISSN 0021-9193.

Elf, J., Li, G.-W., and Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, May 2007. ISSN 1095-9203.

Elmore, S., Müller, M., Vischer, N., Odijk, T., and Woldringh, C. L. Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in Escherichia coli. *Journal of structural biology*, 151(3):275–287, September 2005. ISSN 1047-8477.

Elowitz, M. B., Surette, M. G., Wolf, P.-E., Stock, J. B., and Leibler, S. Protein Mobility in the Cytoplasm of Escherichia coli. *J. Bacteriol.*, 181(1):197–203, January 1999.

Endesfelder, U., Finan, K., Holden, S. J., Cook, P. R., Kapanidis, A. N., and Heilemann, M. Multiscale Spatial Organization of RNA Polymerase in Escherichia coli. *Biophysical Journal*, 105(1):172–181, July 2013. ISSN 00063495.

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): E365–373, August 2011. ISSN 1091-6490.

Espeli, O., Levine, C., Hassing, H., and Marians, K. J. Temporal Regulation of Topoisomerase IV Activity in E. coli. *Molecular Cell*, 11(1):189–201, January 2003. ISSN 10972765.

Ester, M., Kriegel, H.-p., and Jörg S, X. X. A density-based algorithm for discovering clusters in large spatial databases with noise. In **S**imoudis, E., Han, J., and Fayyad, U. M., editors, *Second International*

Conference on Knowledge Discovery and Data Mining (KDD-96)., pages 226–231, 1996.

Fernández-Suárez, M. and Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nature reviews. Molecular cell biology*, 9(12):929–943, December 2008. ISSN 1471-0080.

Fisher, J. K., Bourniquel, A., Witz, G., Weiner, B., Prentiss, M., and Kleckner, N. Four-Dimensional Imaging of E. coli Nucleoid Organization and Dynamics in Living Cells. *Cell*, 153(4):882–895, April 2013. ISSN 00928674.

Friedman, L. J., Mumm, J. P., and Gelles, J. RNA polymerase approaches its promoter without long-range sliding along DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 110(24):9740–9745, June 2013. ISSN 1091-6490.

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.

Gebhardt, J. C. M., Suter, D. M., Roy, R., Zhao, Z. W., Chapman, A. R., Basu, S., Maniatis, T., and Xie, X. S. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature methods*, 10(5):421–426, May 2013. ISSN 1548-7105.

Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J., and Busby, S. J. W. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17693–17698, December 2005. ISSN 0027-8424.

Greenfield, D., McEvoy, A. L., Shroff, H., Crooks, G. E., Wingreen, N. S., Betzig, E., and Liphardt, J. Self-organization of the Escherichia coli chemotaxis network imaged with super-resolution light microscopy. *PLoS biology*, 7(6):e1000137, June 2009. ISSN 1545-7885.

Grigorova, I. L., Phleger, N. J., Mutalik, V. K., and Gross, C. A. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(14):5332–5337, May 2006. ISSN 0027-8424.

Gruß mayer, K. S., Kurz, A., and Herten, D.-P. Single-molecule studies on the label number distribution of fluorescent markers. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 15(4):734–742, March 2014. ISSN 1439-7641.

Gummesson, B., Magnusson, L. U., Lovmar, M., Kvint, K., Persson, O., Ballesteros, M., Farewell, A., and Nyström, T. Increased RNA polymerase availability directs resources towards growth at the expense of maintenance. *The EMBO journal*, 28(15):2209–19, August 2009. ISSN 1460-2075.

Hadizadeh Yazdi, N., Guet, C. C., Johnson, R. C., and Marko, J. F. Variation of the folding and dynamics of the Escherichia coli chromosome with growth conditions. *Molecular microbiology*, 86(6):1318–1333, December 2012. ISSN 1365-2958.

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., and Elf, J. The lac repressor displays facilitated diffusion in living cells. *Science*, 336(6088):1595–1598, June 2012. ISSN 1095-9203.

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):405–408, February 2014. ISSN 1546-1718.

Hayama, R. and Marians, K. J. Physical and functional interaction between the condensin MukB and the decatenase topoisomerase IV in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 107(44):18826–18831, November 2010. ISSN 1091-6490.

Heilemann, M., van de Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., and Sauer, M. Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angewandte Chemie*, 47(33):6172–6176, January 2008. ISSN 1521-3773.

Heintzmann, R., Jovin, T. M., and Cremer, C. Saturated patterned excitation microscopy–a concept for optical resolution improvement. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 19(8):1599–1609, August 2002. ISSN 1084-7529.

Hell, S. W. and Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782, June 1994. ISSN 0146-9592.

Hinner, M. J. and Johnsson, K. How to obtain labeled proteins and what to do with them. *Current opinion in biotechnology*, 21(6):766–776, December 2010. ISSN 1879-0429.

Hiraga, S., Niki, H., Ogura, T., Ichinose, C., Mori, H., Ezaki, B., and Jaffé, A. Chromosome partitioning in Escherichia coli: novel mutants producing anucleate cells. *Journal of bacteriology*, 171(3):1496–505, March 1989. ISSN 0021-9193.

Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nature reviews. Molecular cell biology*, 7(5):311–22, May 2006. ISSN 1471-0072.

Hobot, J. A., Villiger, W., Escaig, J., Maeder, M., Ryter, A., and Kellenberger, E. Shape and fine structure of nucleoids observed on sections of ultrarapidly frozen and cryosubstituted bacteria. *Journal of bacteriology*, 162(3):960–971, June 1985. ISSN 0021-9193.

Hohlbein, J., Gryte, K., Heilemann, M., and Kapanidis, A. N. Surfing on a new wave of single-molecule fluorescence methods. *Physical biology*, 7(3):031001, January 2010. ISSN 1478-3975.

Holden, S. J., Uphoff, S., and Kapanidis, A. N. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature methods*, 8(4):279–280, April 2011. ISSN 1548-7105.

Huang, B., Babcock, H., and Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell*, 143(7):1047–1058, December 2010. ISSN 1097-4172.

Izeddin, I., Récamier, V., Bosanac, L., Cissé, I. I., Boudarene, L., Dugast-Darzacq, C., Proux, F., Bénichou, O., Voituriez, R., Bensaude, O., Dahan, M., and Darzacq, X. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife*, 3:e02230, January 2014. ISSN 2050-084X.

Javer, A., Long, Z., Nugent, E., Grisi, M., Siriwatwetchakul, K., Dorfman, K. D., Cicuta, P., and Cosentino Lagomarsino, M. Short-time movement of E. coli chromosomal loci depends on coordinate and subcellular localization. *Nature communications*, 4:3003, January 2013. ISSN 2041-1723.

Jin, D. J. and Cabrera, J. E. Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in Escherichia coli. *Journal of structural biology*, 156 (2):284–291, November 2006. ISSN 1047-8477.

Jin, D. J., Cagliero, C., and Zhou, Y. N. Role of RNA Polymerase and Transcription in the Organization of the Bacterial Nucleoid. *Chemical reviews*, 113 (11):8662–8682, August 2013. ISSN 1520-6890.

Joshi, M. C., Magnan, D., Montminy, T. P., Lies, M., Stepankiw, N., and Bates, D. Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. *PLoS genetics*, 9(8):e1003673, January 2013. ISSN 1553-7404.

Kabata, H., Kurosawa, O., Arai, I., Washizu, M., Margarson, S., Glass, R., and Shimamoto, N. Visualization of single molecules of RNA polymerase sliding along DNA. *Science*, 262(5139):1561–1563, December 1993. ISSN 0036-8075.

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., and Ebright, R. H. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science (New York, N.Y.)*, 314(5802):1144–7, November 2006. ISSN 1095-9203.

Kapanidis, A. N., Margeat, E., Laurence, T. A., Doose, S., Ho, S. O., Mukhopadhyay, J., Kortkhonjia, E., Mekler, V., Ebright, R. H., and Weiss, S. Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. *Molecular cell*, 20(3):347–56, November 2005. ISSN 1097-2765.

Khodursky, A. B., Zechiedrich, E. L., and Cozzarelli, N. R. Topoisomerase IV is a target of quinolones in Escherichia coli. *Proceedings of the National Academy of Sciences*, 92(25):11801–11805, December 1995. ISSN 0027-8424.

Klumpp, S. and Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51):20245–50, December 2008. ISSN 1091-6490.

Klumpp, S., Zhang, Z., and Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–75, December 2009. ISSN 1097-4172.

Koster, D. A., Crut, A., Shuman, S., Bjornsti, M.-A., and Dekker, N. H. Cellular strategies for regulating DNA supercoiling: a single-molecule perspective. *Cell*, 142(4):519–530, August 2010. ISSN 1097-4172.

Kruse, T., Blagoev, B., Lø bner Olesen, A., Wachi, M., Sasaki, K., Iwai, N., Mann, M., and Gerdes, K. Actin homolog MreB and RNA polymerase interact and

are both required for chromosome segregation in Escherichia coli. *Genes & development*, 20(1):113–24, January 2006. ISSN 0890-9369.

Kumar, M., Mommer, M. S., and Sourjik, V. Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli. *Biophysical journal*, 98(4): 552–9, February 2010. ISSN 1542-0086.

Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. Springer US, 3 edition, 2006. ISBN 978-0-387-46312-4.

Lee, I., Dong, K. C., and Berger, J. M. The role of DNA bending in type IIA topoisomerase function. *Nucleic acids research*, 41(10):5444–5456, May 2013. ISSN 1362-4962.

Lee, 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*, October 2012. ISSN 0027-8424.

Lesterlin, C., Ball, G., Schermelleh, L., and Sherratt, D. J. RecA bundles mediate homology pairing between distant sisters during DNA break repair. *Nature*, 506(7487):249–253, December 2013. ISSN 0028-0836.

Li, Y., Stewart, N. K., Berger, A. J., Vos, S., Schoeffler, A. J., Berger, J. M., Chait, B. T., and Oakley, M. G. Escherichia coli condensin MukB stimulates topoisomerase IV activity by a direct physical interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 107(44):18832–18837, November 2010. ISSN 1091-6490.

Libby, E. A., Roggiani, M., and Goulian, M. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(19):7445–7450, May 2012. ISSN 1091-6490.

Lippincott-Schwartz, J., Snapp, E., and Kenworthy, A. Studying protein dynamics in living cells. *Nature reviews. Molecular cell biology*, 2(6):444–456, June 2001. ISSN 1471-0072.

Maisnier-Patin, S., Nordström, K., and Dasgupta, S. Replication arrests during a single round of replication of the Escherichia coli chromosome in the absence of DnaC activity. *Molecular Microbiology*, 42(5):1371–1382, January 2002. ISSN 0950382X.

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, February 2008. ISSN 1548-7105.

McKinney, S. A., Murphy, C. S., Hazelwood, K. L., Davidson, M. W., and Looger, L. L. A bright and photostable photoconvertible fluorescent protein. *Nature methods*, 6(2):131–133, February 2009. ISSN 1548-7105.

Melby, T. E., Ciampaglio, C. N., Briscoe, G., and Erickson, H. P. The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *The Journal of cell biology*, 142(6): 1595–604, September 1998. ISSN 0021-9525.

Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 82(4 Pt 1):041914, October 2010. ISSN 1550-2376.

Mondal, J., Bratton, B. P., Li, Y., Yethiraj, A., and Weisshaar, J. C. Entropy-based mechanism of ribosome-nucleoid segregation in E. coli cells. *Biophysical journal*, 100(11):2605–2613, June 2011. ISSN 1542-0086.

Moolman, M. C., Krishnan, S. T., Kerssemakers, J. W. J., van den Berg, A., Tulinski, P., Depken, M., Reyes-Lamothe, R., Sherratt, D. J., and Dekker, N. H. Slow unloading leads to DNA-bound β2-sliding clamp accumulation in live Escherichia coli cells. *Nature communications*, 5:5820, January 2014. ISSN 2041-1723.

Moulin, L., Rahmouni, A. R., and Boccard, F. Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of Escherichia coli. *Molecular microbiology*, 55(2): 601–610, January 2005. ISSN 0950-382X.

Neuman, K. C., Charvin, G., Bensimon, D., and Croquette, V. Mechanisms of chiral discrimination by topoisomerase IV. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):6986–6991, April 2009. ISSN 1091-6490.

Nicolas, E., Upton, A. L., Uphoff, S., Henry, O., Badrinarayanan, A., and Sherratt, D. The SMC Complex MukBEF Recruits Topoisomerase IV to the Origin of Replication Region in Live Escherichia coli. *mBio*, 5 (1):e01001–13, February 2014. ISSN , 2150-7511.

Nolivos, S. and Sherratt, D. The bacterial chromosome: architecture and action of bacterial SMC and SMC-like complexes. *FEMS microbiology reviews*, 38(3): 380–392, May 2014. ISSN 1574-6976.

Paige, M. F., Bjerneld, E. J., and Moerner, W. E. A Comparison of Through-the-Objective Total Internal Reflection Microscopy and Epifluorescence Microscopy for Single-Molecule Fluorescence Imaging. *Single Molecules*, 2(3):191–201, October 2001. ISSN 1438-5163.

Papantonis, A. and Cook, P. R. Transcription factories: genome organization and gene regulation. *Chemical reviews*, 113(11):8683–8705, November 2013. ISSN 1520-6890.

Patterson, G. H. and Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*, 297(5588):1873–7, October 2002. ISSN 1095-9203.

Persson, F., Lindén, M., Unoson, C., and Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature methods*, 10(3):265–269, February 2013. ISSN 1548-7105.

Plass, T., Milles, S., Koehler, C., SzymaÅĎski, J., Mueller, R., Wiessler, M., Schultz, C., and Lemke, E. A. Amino acids for Diels-Alder reactions in living cells. *Angewandte Chemie*, 51(17):4166–4170, April 2012. ISSN 1521-3773.

Pomerantz, R. T. and O'Donnell, M. What happens when replication and transcription complexes collide? *Cell cycle*, 9(13):2537–2543, July 2010. ISSN 1551-4005.

Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D., and Cozzarelli, N. R. Topological challenges to DNA replication: conformations at the fork. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15):8219–8226, July 2001. ISSN 0027-8424.

Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, 328(5977):504–508, April 2010. ISSN 1095-9203.

Redding, S. and Greene, E. C. How do proteins locate specific targets in DNA? *Chemical physics letters*, 570:1–11, May 2013. ISSN 0009-2614.

Reits, E. A. and Neefjes, J. J. From fixed to FRAP: measuring protein mobility and activity in living cells. *Nature cell biology*, 3(6):E145–E147, June 2001. ISSN 1465-7392.

Reyes-Lamothe, R., Wang, X., and Sherratt, D. Escherichia coli and its chromosome. *Trends in microbiology*, 16(5):238–245, May 2008. ISSN 0966-842X.

Ricchetti, M., Metzger, W., and Heumann, H. One-dimensional diffusion of Escherichia coli DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. *Proceedings of the National Academy of Sciences of the United States of America*, 85(13):4610–4610, July 1988. ISSN 0027-8424.

Riggs, A. D., Bourgeois, S., and Cohn, M. The lac represser-operator interaction. *Journal of Molecular Biology*, 53(3):401–417, November 1970. ISSN 00222836.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. The role of DNA shape in protein-DNA recognition. *Nature*, 461(7268):1248–53, October 2009. ISSN 1476-4687.

Rust, M. J., Bates, M., and Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–795, October 2006. ISSN 1548-7091.

Saitoh, T. and Ishihama, A. Biosynthesis of RNA polymerase in Escherichia coli. *Journal of Molecular Biology*, 115(3):403–416, September 1977. ISSN 00222836.

Sanamrad, A., Persson, F., Lundius, E. G., Fange, D., Gynnå, A. H., and Elf, J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. *Proceedings of the National Academy of Sciences of the United States of America*, 111(31):14113–14118, July 2014. ISSN 1091-6490.

Schaaf, M. J. M., Koopmans, W. J. A., Meckel, T., van Noort, J., Snaar-Jagalska, B. E., Schmidt, T. S., and Spaink, H. P. Single-molecule microscopy reveals membrane microdomain organization of cells in a living vertebrate. *Biophysical journal*, 97(4):1206–1214, August 2009. ISSN 1542-0086.

Schermelleh, L., Carlton, P. M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M. C., Agard, D. A., Gustafsson, M. G. L., Leonhardt, H., and Sedat, J. W. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science*, 320(5881):1332–1336, June 2008. ISSN 1095-9203.

Schnell, U., Dijk, F., Sjollema, K. A., and Giepmans, B. N. G. Immunolabeling artifacts and the need for live-cell imaging. *Nature methods*, 9(2):152–158, February 2012. ISSN 1548-7105.

Schütz, G. J., Schindler, H., and Schmidt, T. Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophysical journal*, 73 (2):1073–1080, August 1997. ISSN 0006-3495.

Sengupta, P., Jovanovic-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):969–975, November 2011. ISSN 1548-7105.

Shajani, Z., Sykes, M. T., and Williamson, J. R. Assembly of bacterial ribosomes. *Annual review of biochemistry*, 80:501–526, January 2011. ISSN 1545-4509.

Shepherd, N., Dennis, P., and Bremer, H. Cytoplasmic RNA Polymerase in Escherichia coli. *Journal of bacteriology*, 183(8):2527–34, April 2001. ISSN 0021-9193.

Sherratt, D. J. Bacterial chromosome dynamics. *Science*, 301(5634):780–785, August 2003. ISSN 1095-9203.

Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C., and Wigley, D. B. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, 432(7014):187–93, November 2004. ISSN 1476-4687.

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):612–27, May 2011. ISSN 1365-2958.

Smith, C. S., Joseph, N., Rieger, B., and Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nature methods*, 7(5): 373–375, May 2010. ISSN 1548-7105.

Spahn, C., Cella-Zannacchi, F., Endesfelder, U., and Heilemann, M. Correlative super-resolution imaging of RNA polymerase distribution and dynamics, bacterial membrane and chromosomal structure in Escherichia coli. *Methods and Applications in Fluorescence*, 3(1):014005, March 2015. ISSN 2050-6120.

Spahn, C., Endesfelder, U., and Heilemann, M. Super-resolution imaging of Escherichia coli nucleoids reveals highly structured and asymmetric segregation during fast growth. *Journal of structural biology*, 185(3):243–249, March 2014. ISSN 1095-8657.

Stone, M. D., Bryant, Z., Crisona, N. J., Smith, S. B., Vologodskii, A., Bustamante, C., and Cozzarelli, N. R. Chirality sensing by Escherichia coli topoisomerase IV and the mechanism of type II topoisomerases. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15):8654–8659, July 2003. ISSN 0027-8424.

Stracy, M., Lesterlin, C., Garza de Leon, F., Uphoff, S., Zawadzki, P., and Kapanidis, A. N. Live-cell super-resolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences*, 112(32): 201507592, July 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):3585–3594, May 2014. ISSN 1873-3468.

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):153–159, February 2009. ISSN 1548-7105.

Tao, H., Bausch, C., Richmond, C., Blattner, F. R., and Conway, T. Functional genomics: expression analysis of Escherichia coli growing on minimal and rich media. *Journal of bacteriology*, 181(20):6425–6440, October 1999. ISSN 0021-9193.

Thei Dame, R., Wyman, C., and Goosen, N. Structural basis for preferential binding of H-NS to curved DNA. *Biochimie*, 83(2):231–234, February 2001. ISSN 03009084.

Thompson, R. E., Larson, D. R., and Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical journal*, 82(5): 2775–2783, May 2002. ISSN 0006-3495.

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, February 2008. ISSN 1548-7105.

Travers, A. A. and Burgess, R. R. Cyclic Re-use of the RNA Polymerase Sigma Factor. *Nature*, 222(5193): 537–540, May 1969. ISSN 0028-0836.

Tsien, R. Y. The green fluorescent protein. *Annual review of biochemistry*, 67:509–44, January 1998. ISSN 0066-4154.

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):8063–8068, April 2013. ISSN 1091-6490.

van de Linde, S., Heilemann, M., and Sauer, M. Live-cell super-resolution imaging with synthetic fluorophores. *Annual review of physical chemistry*, 63: 519–540, January 2012. ISSN 1545-1593.

Van Houten, B. Nucleotide excision repair in Escherichia coli. *Microbiological reviews*, 54(1):18–51, March 1990. ISSN 0146-0749.

Vos, S. M., Lee, I., and Berger, J. M. Distinct regions of the Escherichia coli ParC C-terminal domain are required for substrate discrimination by topoisomerase IV. *Journal of molecular biology*, 425(17): 3029–3045, September 2013a. ISSN 1089-8638.

Vos, S. M., Stewart, N. K., Oakley, M. G., and Berger, J. M. Structural basis for the MukB-topoisomerase IV interaction and its functional implications in vivo. *The EMBO journal*, 32(22):2950–2962, November 2013b. ISSN 1460-2075.

Vos, S. M., Tretter, E. M., Schmidt, B. H., and Berger, J. M. All tangled up: how cells direct, manage and exploit topoisomerase function. *Nature reviews. Molecular cell biology*, 12(12):827–841, December 2011. ISSN 1471-0080.

Wang, F., Redding, S., Finkelstein, I. J., Gorman, J., Reichman, D. R., and Greene, E. C. The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nature structural & molecular biology*, 20(2): 174–181, February 2013. ISSN 1545-9985.

Wang, K., Sachdeva, A., Cox, D. J., Wilf, N. W., Lang, K., Wallace, S., Mehl, R. A., and Chin, J. W. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nature chemistry*, 6(5):393–403, May 2014. ISSN 1755-4349.

Wang, W., Li, 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 1095-9203; 0036-8075.

Wang, X., Reyes-Lamothe, R., and Sherratt, D. J. Modulation of Escherichia coli sister chromosome cohesion by topoisomerase IV. *Genes & development*, 22 (17):2426–2433, September 2008. ISSN 0890-9369.

Wang, X. and Sherratt, D. J. Independent segregation of the two arms of the Escherichia coli ori region requires neither RNA synthesis nor MreB dynamics. *Journal of bacteriology*, 192(23):6143–6153, December 2010. ISSN 1098-5530.

Weber, S. C., Spakowitz, A. J., and Theriot, J. A. Bacterial Chromosomal Loci Move Subdiffusively through a Viscoelastic Cytoplasm. *Physical Review Letters*, 104(23):238102, June 2010. ISSN 0031-9007.

Woldringh, C. L. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Molecular Microbiology*, 45(1):17–29, July 2002. ISSN 0950-382X.

Xie, X. S., Choi, P. J., Li, G.-W., Lee, N. K., and Lia, G. Single-molecule approach to molecular biology in living bacterial cells. *Annual review of biophysics*, 37:417–444, January 2008. ISSN 1936-122X.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767):1600–3, March 2006. ISSN 1095-9203.

Zawadzki, P., Stracy, M., Ginda, K., Zawadzka, K., Lesterlin, C., Kapanidis, A. N., and Sherratt, D. J. The Localization and Action of Topoisomerase IV in Escherichia coli Chromosome Segregation Is Coordinated by the SMC Complex, MukBEF. *Cell reports*, 13(11):2587–2596, December 2015. ISSN 2211-1247.

Zhang, Z., Smith, B. A. C., Wang, L., Brock, A., Cho, C., and Schultz, P. G. A new strategy for the site-specific modification of proteins in vivo. *Biochemistry*, 42 (22):6735–6746, June 2003. ISSN 0006-2960.

Zhao, Z. W., Roy, R., Gebhardt, J. C. M., Suter, D. M., Chapman, A. R., and Xie, X. S. Spatial organization of RNA polymerase II inside a mammalian cell nucleus revealed by reflected light-sheet superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(2):681–686, January 2014. ISSN 1091-6490.



Appendix A

Characterisation of bacterial strains

Strains used in Chapters 6 and 7 were characterised with growth curves, flow cytometry and cell lengths analysis performed by Pawel Zawadzki. Growth curves for ParE-PAmCherry and ParC-PAmCherry showed identical growth to wildtype AB1157 grown in M9-glycerol (Figure A.1).

Cell lengths were determined by segmentation of transmitted light images with MicrobeTracker. Flow cytometry was used to determine the chromosome distribution was perturbed in fusion or overexpression strains (Figure A.2). Cells were grown in M9-glycerol to exponential phase (OD 600 < 0.2). 200 µl of the cultures were fixed with 3.5 ml of ice-cold 74 % ethanol. For the staining procedure, the cells were pelleted and washed twice in 100 µl of cold staining buffer (10 µM Tris pH 7.4, 3 µM $MgCl_2$). The 100 µl samples were mixed with an equal volume of the staining solution Syto-16 (3 µM). 100,000 events were recorded in a Becton Dickinson FACScalibur machine using FL1-H and results were analysed using FlowJo. The cell lengths, generation time, and number or anucleate cells for ParC and ParE fusion strains are

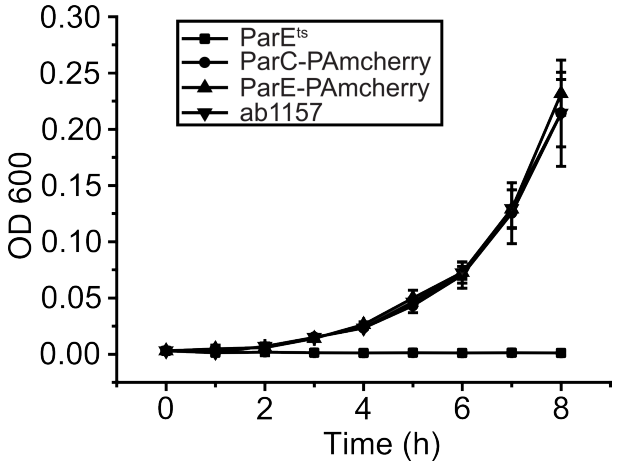


Figure A.1: Growth curves for fusion proteins compared to wildtype AB1157. Cells were grown at 37 °C with shaking. A temperature sensitive ParE(ts) strain serves as a negative control.

	AB1157	ParC-PAmCherry	ParE-PAmCherry
Generation time	86 ± 3 min	85 ± 5 min	81 ± 11 min
Cell length	3.1 ± 0.5 µm	3.3 ± 0.7 µm	3.1 ± 0.6 µm
Anucleate cells		0/86	1/77

Table A.1: Behaviour of ParC and ParE fusion strains.

detailed in Table A.1 and for the ParC-CTD overexpression strain in Table A.2.

	empty pBAD24	ParC-CTD overexpression
Generation time	95 ± 5 min	101 ± 11 min
Cell length	3.7 ± 0.7 µm	3.7 ± 0.6 µm
Anucleate cells	0/80	1/68

Table A.2: Behaviour of ParC-CTD overexpression strains

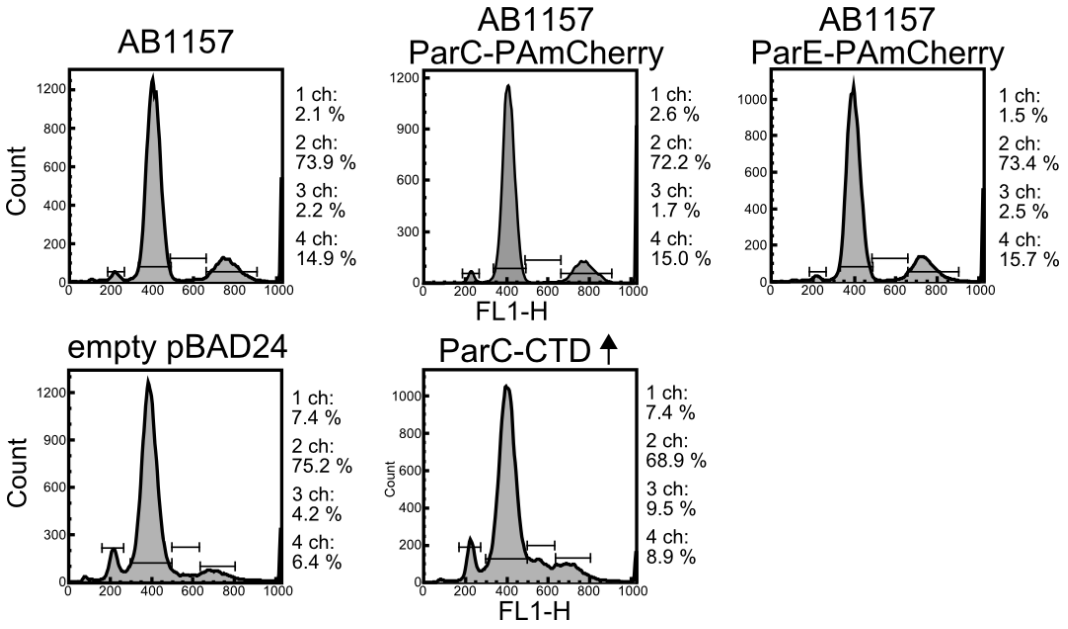


Figure A.2: **Characterisation of fusion and overexpression strains.** Flow cytometry of cells with PAmCherry fusions (derivatives of AB1157) and after ParC-CTD overexpression. Ranges indicate cells with (from the left); one, two, three and four chromosomes.



Appendix B

Example Matlab code

An example of Matlab code used to simulate random positions within a cell outline.

```
%function generates random positions in the cell.
function rand_sample_locals = rand_gen(mesh, ROIVertices, num_locals )
rand_sample_locals = [];

        %use mesh defined from microbeTracker
        plgx = [mesh(1:end-1,1);flipud(mesh(:,3))];
        plgy = [mesh(1:end-1,2);flipud(mesh(:,4))];
        ROIs = [plgx, plgy];

           %first generate equal amount of random simulations to the
           %number of actual localisations in the cell. Some of these will
           %be outsdide the simiulated area
           quantOfRandNum = num_locals;

           lengthOfList = 0; %set lengthList which is updated
```

```
        %loop until the number of random positions equals the number of
        %lacalisations in the cell
        while lengthOfList < quantOfRandNum

            %generate arrays of random positions within a rectangle
            %enclosing the cell area
            numRands = quantOfRandNum - lengthOfList;
            x = min(ROIVertices(:,1)) + ...
                (max(ROIVertices(:,1))-min(ROIVertices(:,1))).*rand(
                    »numRands,1);
            y = min(ROIVertices(:,2)) + ...
                (max(ROIVertices(:,2))-min(ROIVertices(:,2))).*rand(
                    »numRands,1);

            %find the positions that are inside the cell ROI
            xyInPolygon = inpolygon(x,y,ROIs(:,1),ROIs(:,2));

            %Keep the positions inside the cell area and add to the
            %list
            new_locals = [x(xyInPolygon),y(xyInPolygon)];
            rand_sample_locals = [rand_sample_locals; new_locals];
            clear x y new_locals
            lengthOfList =  size(rand_sample_locals,1);
        end
end
```