



Investigating the Communication of Mismatch Recognition and Strand Incision in the DNA Mismatch Repair Pathway

Nehir Banaz

Trinity College



University of Oxford

A thesis presented for the degree of

Doctor of Philosophy

Hillary 2022

Copyright © Nehir Banaz 2022

All Rights Reserved

Acknowledgements

I would like to express my gratitude to my supervisor, Associate Prof. Stephan Uphoff, for his project ideas, continued support and diligent supervision. I have greatly appreciated the environment you created in our group, the scientific discussions we had and your enthusiasm for science. I could not have done this without you as my mentor. I would also like to thank my co-supervisor, Prof. David Sherratt, for his supervision and support.

I further wish to thank all members of the Sherratt group, especially Gemma and Jarno. Gemma, many thanks for teaching me how to purify proteins, even though that project was eventually not included in my thesis due to Covid-19. Jarno, thank you for working with me on the paper and for all your help with writing custom scripts for Matlab.

Working in the Uphoff Lab was a blessing. I thank all members of the group from the bottom of my heart for creating such a collaborative, passionate and constructive environment to work in. My special thanks are extended to life time friends Valentine and Emma. I am lucky to have met you and am grateful for your encouragement along the way.

Finally, I would like to thank the adventure team: Miles, Doğukan, Ines, Laura, Jack and Menelaos. I looked forward to our RPG sessions and the department was not the same after you left. Special thanks to Menelaos, my partner in crime throughout the DPhil. I am lucky to have you in my life. To my Linacre girls, Sapna and Rahel, I know you will achieve great things with your lives. And last but not least, many thanks to Ohad. You make everything more fun, even writing a thesis together.

Abstract

DNA Mismatch Repair (MMR) is crucial for preventing DNA replication errors that escape the proofreading mechanisms of DNA polymerases from becoming permanent mutations. MMR is initiated when MutS recognises a mismatch and recruits MutL to signal the degradation and resynthesis of the nascent DNA. The distinction between the nascent and template DNA strands is a critical aspect of Mismatch Repair in all organisms. In *Escherichia coli*, this problem is solved by MutH endonuclease, which incises the nascent strand specifically at hemimethylated GATC sequences. However, it remains unclear how the MutS-MutL complex communicates with MutH to activate strand incision, given that GATC sites can be located hundreds of bases away from a mismatch. To tackle this long-standing question, we used single-molecule tracking of fluorescently-labeled MutS, MutL, and MutH proteins in live *E. coli* cells. Our quantitative measurements address how MutH is recruited to hemimethylated GATC sites. We find that it is neither dependent on MutS/L nor on the presence of DNA mismatches suggesting that MutH binds to hemimethylated GATC sites independently and becomes activated by interaction with MutL. Furthermore, we show that MutH competes with SeqA and Dam proteins when binding to GATC sites. Hence, there is a brief time window during which MutH has to be activated before the strand discrimination signal is lost. We used Fluorescence Recovery after Photobleaching (FRAP) to measure the stoichiometry and turnover of proteins during Mismatch Repair. Our results complement previous studies, leading to a refined model for the coordination of the *E. coli* Mismatch Repair pathway *in vivo*.

Keywords— Methyl Directed Mismatch Repair - *dam*- *SeqA*- MutH - MutL- MutS - FRAP -Single Molecule Tracking -*Escherichia Coli* - GATC

Table of Contents

1	Introduction	1
1.1	<i>E. coli</i> Cell Cycle	4
1.2	DNA Methylation	7
1.2.1	DNA Adenine Methyltransferase (Dam)	9
1.2.2	SeqA	10
1.3	Methyl-Directed DNA Mismatch Repair	11
1.4	MutS	18
1.5	MutL	22
1.6	MutH	24
1.7	Communication of Mismatch Recognition to Excision Initiation Site	25
2	Methodology	32
2.1	Construction of Halo-tagged Strains Using Lambda Red Recombination	32
2.2	P1 Transduction	34
2.3	Heat-shock Transformation	35
2.4	Growth Conditions	35
2.5	List of Plasmids Used in This Study	35
2.6	List of Primers Used in This Study	36
2.7	List of Strains Used in This Study	38
2.8	Spontaneous Mutation Frequency Assay (Rifampicin Selection)	39
2.9	Labelling of Halo-tagged Proteins	40
2.10	Single Molecule Tracking	42
2.10.1	Histogram Fitting	46

2.11 Fluorescence Recovery after Photobleaching (FRAP)	47
2.12 Statistical Analysis	48
3 Addressing the <i>in vivo</i> Mobility of MMR Proteins	49
3.1 Halo-tagged Mismatch Proteins are Functional	50
3.2 Distribution of Diffusion Coefficients of MMR Proteins in Wild Type Cells	53
3.3 Implications of <i>in vivo</i> Mobilities of Target Search Mechanisms . .	65
4 Investigation of Repair Complexes in MMR	69
4.1 Distribution of Diffusion Coefficients of MMR Proteins after DNA Damaging Treatment	69
4.2 Epifluorescence Imaging of MutS-mYPet, MutL-mYPet and MutH- mYPet	74
4.3 FRAP	78
5 Genetic Perturbations of the MMR System to Investigate the Mecha- nism of MutH Recruitment	83
5.1 Distribution of Diffusion Coefficients of MutH after MutL Deletion	83
5.2 Distribution of Diffusion Coefficients of MutH after Perturbation of the Amount of Dam	85
5.3 Overexpression of the Dam Protein	87
5.4 Deletion of SeqA	91
6 Discussion	94
Appendix A Representative MutH Histograms	118
Appendix B Publication from DPhil Thesis	120

List of Figures

1.1	Schematic summary of the MMR pathway	3
1.2	Replication forks during slow and fast growth	6
1.3	Waves of DNA methylation after replication	12
1.4	Methyl directed mismatch repair in <i>E. coli</i> extracts	14
1.5	Apparent dissociation constants for MutS binding to mismatched pairs (figure from [Modrich, 2016])	15
1.6	Representation of the MMR pathway	17
1.7	MutS Annotated (1E3M) [Lamers et al., 2000]	19
1.8	Overview of MutS-DNA complex (1E3M) [Lamers et al., 2000] . . .	20
1.9	Crystal structure of the trapped MutS-MutL transient complex (5AKB) [Groothuizen et al., 2015] MutS in blue-cyan, MutL-N40 in green .	22
1.10	Crystal structure of <i>E. coli</i> MutH (1AZO) [Ban and Yang, 1998b] . .	25
1.11	Schematic exemplifying looping models	26
1.12	AFM images showing loop binding of MutS to heteroduplex DNA	28
1.13	Schematic exemplifying sliding models	30
1.14	Representative kymographs on DNA curtains	31
2.1	Schematic representation of primer design	34
2.2	Schema explaining spontaneous mutation frequency assay	40
2.3	Schema explaining labelling with Halo-tag	41
2.4	Schematic representation of the optical configuration	45
3.1	Example plates from the spontaneous mutation frequency assay .	51
3.2	Halo-tagged mismatch proteins are functional	53
3.3	Schematic representation of distribution of diffusion coefficients exemplifying bound and mobile molecules	54

3.4	Distribution of diffusion coefficients for MutS-Halo	55
3.5	Different modes of MutS diffusion	55
3.6	Distribution of diffusion coefficients for MutS-Halo with 3 species fit	56
3.7	MutS-Halo diffusion profile in single cells	57
3.8	MutS-Halo diffusion profile showing bound and mobile molecules	58
3.9	Zoomed-in field of view showing MutS-Halo diffusion profile in single cells	59
3.10	Distribution of diffusion coefficients for MutL-Halo	61
3.11	MutL-Halo diffusion profile in single cells	62
3.12	MutL-Halo diffusion profile showing bound and mobile molecules	63
3.13	Distribution of diffusion coefficients for MutH-Halo	64
3.14	MutH-Halo diffusion profile in single cells	66
3.15	MutH-Halo diffusion profile showing bound and mobile molecules	67
3.16	Comparison of distribution of diffusion coefficients for MutS, MutL and MutH	68
4.1	Distribution of diffusion coefficients for MutS-Halo untreated vs treated with 10 mM MMS	70
4.2	Distribution of diffusion coefficients for MutL-Halo untreated vs treated with 10 mM MMS	71
4.3	Distribution of diffusion coefficients for MutH-Halo untreated vs treated with 10 mM MMS	72
4.4	Bar graphs comparing the bound populations of MutS, MutL and MutH, untreated vs treated with 10 mM MMS	72
4.5	MutS and MutL form foci whereas MutH does not	74
4.6	MutS-mYPet forms foci	75
4.7	MutL-mYPet forms foci	76

4.8	MutH-mYPet does not form foci	77
4.9	Fluorescence recovery after photobleaching graph for MutS	81
4.10	Fluorescence recovery after photobleaching graph for MutL	82
5.1	Distribution of diffusion coefficients for MutH-Halo wild type and MutH-Halo $\Delta MutL$	84
5.2	Bar graphs comparing the bound populations of MutH-Halo wild type and MutH-Halo $\Delta MutL$	84
5.3	Distribution of diffusion coefficients for MutH-Halo and Δdam .	85
5.4	Bar graphs comparing the bound populations of MutH-Halo wild type and Δdam MutH-Halo	86
5.5	Brightfield and fluorescent images of MutH-Halo Δdam DnaN- mYPet	86
5.6	SDS-PAGE gel showing overexpression of Dam	88
5.7	Distribution of diffusion coefficients for MutH-Halo with empty pBAD vector vs Dam overexpression	89
5.8	Bar graphs comparing the bound populations of MutH-Halo with empty pBAD vector vs Dam overexpression	90
5.9	Distribution of diffusion coefficients for MutH-Halo and $\Delta SeqA$.	92
5.10	Bar graphs comparing the bound populations of MutH-Halo and $\Delta SeqA$	92
5.11	Example brightfield and fluorescent images of DnaN-mYPet foci in SeqA overexpression	93
6.1	Proposed mutations for MutH	96
6.2	MutH protein purification	98
6.3	Proposed model for mismatch repair mechanism	101
A.1	Distribution of diffusion coefficients for MutH-Halo	119

List of Abbreviations

MMR		Mismatch Repair
DNA		Deoxyribonucleic Acid
<i>E. coli</i>		<i>Escherichia coli</i>
FRAP		Fluorescence Recovery after Photobleaching
TIRF		Total Internal Reflection Fluorescence
TMR		Tetramethylrhodamine
SMC		Structural Maintenance of Chromosomes
Kan		Kanamycin
Amp		Ampicillin
SEM		Standard Error of the Mean
Ori		Origin of Replication
MSD		Mean Squared Displacement
MMS		Methyl MethaneSulfonate
M9		Minimal growth medium
LB		Luria Broth

1 | Introduction

Genetic information of living organisms is encoded in their DNA. DNA consists of two linear polymer strands, coiled around each other to form a double helix, each containing a sequence of nucleotides. These nucleotides are composed of three subunit molecules: a nucleobase, a five-carbon sugar (deoxyribose), and a phosphate group. The 4 bases present in the DNA are: guanine(G), adenine(A), cytosine(C) and thymine(T). DNA strands are held together by hydrogen bonding between the nucleobases, so that the paired bases are on the inside of the polymer and the backbone chains face outwards [Watson and Crick, 1953]. Adenine pairs with thymine *via* two hydrogen bonds, and guanine pairs with cytosine *via* three hydrogen bonds. Cellular DNA is organised into highly compacted chromosomes [Nasmyth, 2017].

In most bacteria genetic information is encoded on a single circular chromosome. Faithful replication of the chromosome is essential for genetic stability. As replication is semi-conservative, replicative polymerases use one DNA strand as a template to synthesize the other. Although this process is highly accurate, sometimes mistakes are made which are called DNA mismatches. The major determinant of this accuracy is a restrictive active site that sterically prevents most mismatches by 'proofreading' them immediately [Johnson, 1993]. The 3' to 5' exonuclease activity of Pol III's proofreading subunit ϵ , (encoded by the *dnaQ* gene), improves accuracy by removing mismatched nucleotides, allowing polymerase to resynthesize the DNA. *In vitro*, proofreading improves the accuracy of DNA synthesis 10- to 100-fold [Bloom et al., 1997]. The mistakes that escape proofreading are corrected by mismatch repair (MMR). This process is central to evolution of species, as mismatches that are not corrected result in mutations, which can be beneficial or detrimental to the survival of

the individual [Kunkel and Bebenek, 2000]. MMR mechanisms are highly conserved and their inactivation has a variety of severe consequences, such as mutator strains in bacteria that build antibiotic resistance or increased susceptibility to cancer in mammals to name a few [Fishel, 1999] [Loeb, 1991].

To correct replication errors, MMR must recognise the mismatched base pairs and identify the new DNA strand that contains the misincorporated nucleotide. In *E. coli* the mismatched base pairs are recognised by a protein called MutS. MutS is a homodimer that exchanges ADP with ATP at its binding site upon mismatch recognition and adopts a sliding clamp conformation that encircles the DNA backbone [Liu et al., 2016]. This modification recruits MutL, a protein matchmaker that binds to MutS and other components of the repair pathway and consequently activates them or recruits them to the DNA. Differentiating between the newly synthesized strand and the template that was used for synthesis is crucial to prevent mismatches from turning into mutations. If the wrong strand is 'corrected' a mutation would be introduced into the genome. [Pukkila et al., 1983] showed that in γ -proteobacteria the methylation state of d(GATC) sites is important for MMR as it dictates strand differentiation. Hemimethylated sites are recognised by a protein called MutH. It is not clear whether MutH finds its substrate on its own or whether it is recruited to these sites by other proteins. However, it is well known that MutH's endonuclease gets activated upon binding to MutL and results in the incision of the unmethylated strand at hemimethylated GATC sites (Figure 1.1).

Hemimethylated GATC sites are also recognised by DNA adenine methyltransferase (Dam) and once recognised they are fully methylated. Once fully methylated MutH can not recognise these sites anymore. As DNA methylation lags behind DNA synthesis [Barras and Marinus, 1989], there is a brief window of opportunity (for about 0.5-3.0 min in cells growing at 30 °C) during which MutH

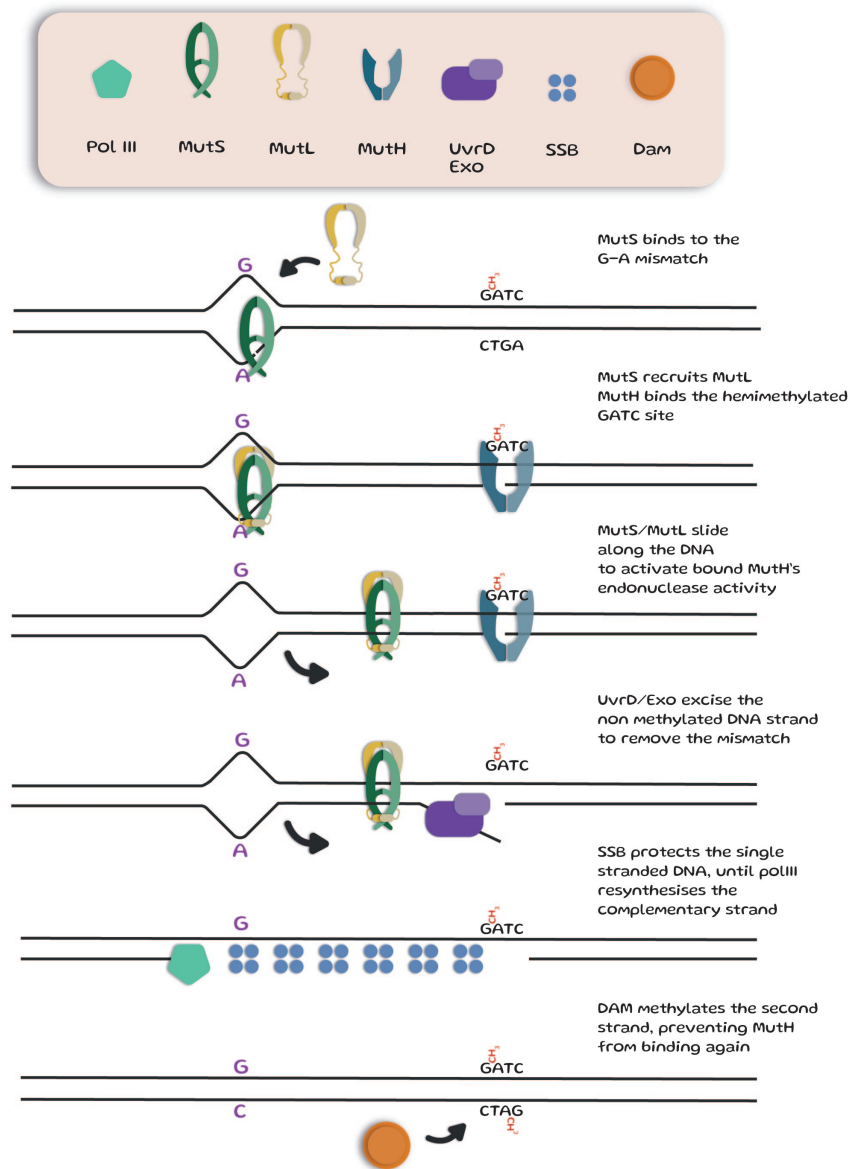


Figure 1.1: Schematic summary of the MMR pathway

Schematic representation of methyl-directed DNA mismatch repair in *E. coli*. MutS recognizes the mismatch and recruits MutL, whereas MutH recognizes hemi-methylated GATC sites. After MutH incises the DNA UvrD unwinds it and exonucleases degrade the nascent strand. SSB protects the single strand and PolIII re-synthesizes the DNA. Finally, ligase seals the gap.

can bind to these sites and use them for strand differentiation. Therefore, transient absence of methylation guides the repair mechanism to the newly synthesized strand.

In other prokaryotes and eukaryotes, the exact mechanism of MMR is not as clear. MutS homologs function similarly by recognising the mismatch and guiding MutL to the DNA. In *E. coli* MutS and MutL are homodimers, whereas in eukaryotes they are heterodimers. Organisms other than γ -proteobacteria lack MutH, therefore MMR is not methyl-directed, instead MutL possesses the endonuclease activity. There are multiple suggestions on the mechanism of strand discrimination, and the exact mechanism remains uncertain. Strand discontinuities associated with replication may serve this purpose *in vivo* while nicks or gaps can direct strand-specific MMR repair activity *in vitro* [Kunkel and Erie, 2005]. Ribonucleotides misincorporated into DNA can also act as strand-discrimination signals in eukaryotic MMR [Ghodgaonkar et al., 2013]. In both eukaryotes and prokaryotes outside of γ -proteobacteria interaction with replisome is observed. In eukaryotes MutS has been shown to interact with proliferating cell nuclear antigen (PCNA) (the DNA clamp that acts as a processivity factor) [Shell et al., 2007] and in bacteria [Liao et al., 2015] visualised MutS in *B. subtilis* with PAmCherry and observed that MutS/replisome interaction is necessary for MutS recruitment and MMR *in vivo*.

1.1 *E. coli* Cell Cycle

To achieve harmonious growth and uniformity in bacterial populations a well-regulated cell cycle is required that responds to a set of internal and external signals. Cells need to achieve a balance of mass, chromosome replication and chromosome segregation. Therefore changes in growth rate must be accom-

panied by changes in the cell cycle to ensure that cell division stays coordinated with mass doubling, chromosome replication and chromosome segregation [Wang and Levin, 2009].

The bacterial cell cycle could be divided into 3 stages:

- B period: The period between cell birth (division) and initiation of chromosome replication
- C period: The period required to replicate a single chromosome
- D period: The period between the end of DNA replication and completion of cell division

In *E. coli* DNA replication begins at a single origin (*oriC*) and proceeds bidirectionally around the circumference of the chromosome, terminating at a region opposite *oriC* called *ter*. Throughout replication, the chromosome is in a condensed ordered structure called the nucleoid [Tolstorukov et al., 2016]. Towards the end of chromosome segregation, dynamic treadmilling filaments are assembled by the tubulin-like protein FtsZ which controls both the location and activity of the associated cell wall synthetic enzymes. FtsAZ filaments treadmill circumferentially around the division ring [Bisson-Filho et al., 2017] and they recruit the division machinery to the division initiation site. They also create discrete sites of cell wall synthesis that circle around the division plane to divide the cell [Yang et al., 2017].

The length of period C and D is dependent on enzyme kinetics and are largely fixed. However, if conditions are favorable; period B becomes shorter, and doubling time can become less than the sum of the C and D periods. This leads to *oriC:terC* ratios that could be as high as 8 or 16. The growth rate of the population is reflected by the copy number ratio of *oriC* to *terC* [Haugan et al., 2018].

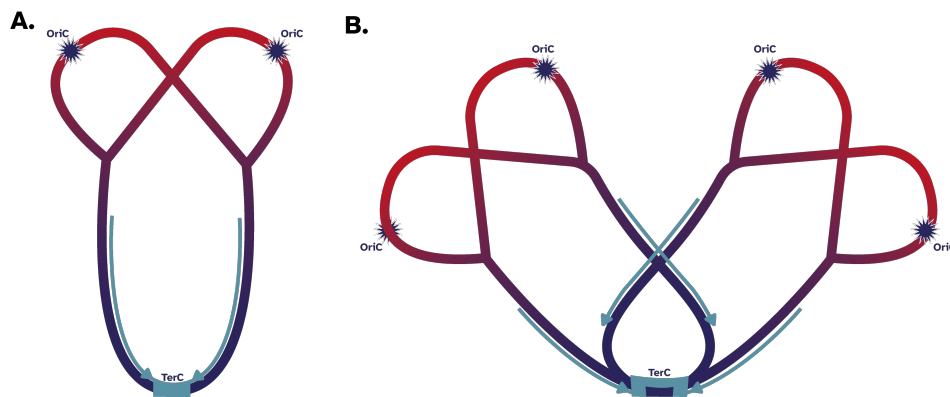


Figure 1.2: Replication forks during slow and fast growth

A) Replication during slow growth: in slow-growing bacterial cells there is a single round of replication per division cycle (doubling time $> C+D$). During slow growth each cell has only two copies of the origin region *oriC* and one copy of the terminus *terC*.

B) Replication during fast growth (multifork replication): in rapidly growing cells (doubling time $< C+D$), each chromosome re-initiates a new round of replication before the first round has terminated. Multifork replication ensures that at least one round of replication is finished before cytokinesis, to guarantee that each daughter cell receives at least one complete genome. During multifork replication cells can have four or more copies of the region proximal to *oriC* and one copy of the region proximal to *terC* (Figure adapted from [Wang and Levin, 2009]).

The number of replication forks in fast growing *E. coli* typically follows 2^n (n = integer) if the replication is synchronous (See Figure 1.2).

The initiation of chromosome replication is dependent on a protein called DnaA. DnaA binds to specific sites at the *oriC* and causes the A-T rich DNA in this region to unwind, which permits the replisome to load [Kaguni, 2006].

Nutrient availability has a dramatic effect on DnaA concentration and directly affects the initiation of chromosome replication [Wang and Levin, 2009]. The intracellular concentration of DnaA increases with the growth rate to ensure that origin firing is triggered more frequently in rapidly growing cells. When cells are starving of amino acids *de novo* DnaA synthesis stops which prevents

new rounds of chromosome replication [[Kaguni, 2006](#)].

1.2 DNA Methylation

DNA methylation is one of the best characterised epigenetic mechanisms in prokaryotes and eukaryotes. *E. coli* has a diverse set of sequence specific DNA methylators (HsdM, YhdJ, Dcm and Dam) [[Marinus and Løbner-Olesen, 2014](#)]. These enzymes methylate the DNA at specific sequences. The post-replicative DNA methylation produced by methylases adds a secondary layer of information to the primary DNA sequence. This information is relevant for transcription, transposition, initiation of chromosome replication, chromosome organisation, restriction modification systems, mRNA utilization, gene regulation and mutation prevention by DNA repair [[Marinus and Løbner-Olesen, 2014](#)], [[Moore et al., 2013](#)] [[Cohen et al., 2016](#)]. In general, methylation can inhibit restriction enzyme cleavage, decrease transformation frequency in certain bacteria and decrease the stability of short direct repeats. Without GATC methylation, MMR becomes deleterious and it results in toxic DNA breaks [[Cohen et al., 2016](#)]. *E. coli* MutH recognizes unmethylated GATC sites and cuts them on both strands of the DNA causing double strand breaks.

Methylation involves addition of a methyl-group to a base by a DNA methyl-transferase. In *E. coli* the most abundant modifications are C^5 -methyl-cytosines and N^6 -methyl-adenines [[Collier, 2009](#)]. N^6 -methyl-adenine modification can facilitate interactions of regulatory proteins with the DNA as it lowers the thermodynamic stability of DNA and alters the curvature. Other DNA binding proteins can also directly recognise the additional methyl group.

E. coli lack de-methylases, therefore its chromosome is normally fully methylated except after the replication fork passes which generates a wave of tran-

siently hemi methylated GATC sites. Most of these sites get methylated a few minutes after the fork passes by Dam protein [Stancheva et al., 1999]. However around the *oriC* and the *dnaA* gene promoter, hemimethylated DNA is much more persistent sometimes lasts throughout the whole cell cycle [Campbell and Kleckner, 1990]. This is possibly due to the high concentration of GATC sites in these regions which are 10- fold over-represented around *OriC*. These GATC sites provide binding sites for multiple proteins such as SeqA, Dam and MutH.

SeqA binding is dependent on the methylation state of the DNA, and thus the position of cells within the cell cycle. When cells are unsynchronized, the largest SeqA binding signal corresponds exactly with the location of *OriC*. Other SeqA binding signals are scattered throughout the genome, and these correspond well to locations where the frequency of GATC sites is higher. A very different SeqA binding profile is found in cells harvested when chromosome replication has been blocked. Little or no SeqA binding is observed at the *OriC*, while clear binding signals are apparent in the *terC* region [Sánchez-Romero et al., 2010].

The DNA of *E. coli* contains on average 19120 N^6 -methyl-adenines [Marinus and Løbner-Olesen, 2014]. These are located in GATC sequences which occur on average, once every 243 nucleotides, which is similar to 1/256 expected for any random combination of 4 nucleotides (4^4 would be the random distribution of nucleotides). d(GATC) methylation in *E. coli* is crucial for regulating multiple mechanisms and 99% of the N^6 -methyl-adenines in *E. coli* is modified by Dam methylase, while the remaining is modified by EcoK enzyme [Marinus and Løbner-Olesen, 2014].

1.2.1 DNA Adenine Methyltransferase (Dam)

DNA adenine methyltransferase (Dam) methylates GATC sites throughout the genome of γ -proteobacteria such as clinically relevant *Vibrio*, *Salmonella* and *Escherichia*. Dam is not an essential protein, however it is crucial for strand determination in correction of replicative errors and controls the frequency of replication initiation at *oriC* and transcription initiation at promoters that contain GATC sequences [Marinus and Løbner-Olesen, 2014]. Throughout the genome, Dam methylation ensures discrimination between the parental DNA strand (fully methylated) and newly synthesized DNA (GATC sites that are not yet modified). *dam* gene is 834 basepairs and encodes for a 32kDa protein. In fast-growing *E. coli* K-12 strains there are 130 Dam molecules per cell [Urig et al., 2002] [Boye et al., 1992]. It is believed that the methylation of GATC sites behind the replication fork of fast-growing *E. coli* takes a few minutes [Campbell and Kleckner, 1990] for the whole chromosome while it takes 2-4 seconds for a shorter plasmid DNA [Stancheva et al., 1999]. *dam* seems to have recently evolved along with *seqA* and *mutH* [Marinus and Løbner-Olesen, 2014].

The *E. coli dam* mutants exhibit a variety of phenotypic traits: Increased number of single-strand and double-strand DNA strand breaks, increased sensitivity to UV light, increased drug-induced mutagenesis and a mutator phenotype to name a few. For Dam function the amount of hemimethylated DNA trailing the replication fork is critical. When the level of hemimethylated DNA is increased by using a Dam overproducing plasmid or decreasing the level with a *dam* mutant, the function is profoundly altered. The mutator phenotype is explained by the loss of strand discrimination signal in Δdam mutants as mutations are introduced into the parental strand 50% of the time. However, it has been observed that mutation rate in Δdam bacteria is much less than 50% that

of $\Delta mutS$, $\Delta mutL$, or $\Delta mutH$ cells [Glickman and Radman, 1980] [McGraw and Marinus, 1980]. This result could be attributed to the fact that mismatch repair becomes lethal and introduces double strand breaks when both strands are unmethylated, which would prevent mistakes being fixed as mutations. *dam* mutants provide further evidence for the role of *dam* in strand discrimination: Contrary to the wild-type where mismatch repair is confined to the newly synthesized strand *dam* mutants can not discriminate between strands, therefore introduce mutations into the parental strand by using the newly (wrongly) synthesized strand as a template. Dam deletions have a specific phenotype, cells are longer, sometimes filamented. Dam mutants are SOS induced due to MutH causing double strand breaks and this phenotype is reversible in MutH deletion strains. When *dam* is overproduced, the high concentration of the protein reduces the transient lifetime of hemimethylated GATCs in the new strand, therefore does not allow enough time for mismatch repair to function hence increasing the mutation rate.

1.2.2 SeqA

The *E. coli* SeqA protein contributes to regulation of chromosome replication by preventing re-initiation at newly replicated origins. SeqA is a homodimer and each dimer binds a pair of hemimethylated GATC sites. SeqA has 181 amino acids and they form two functionally distinct domains [Fujikawa et al., 2003]: The N-terminal domain (residues 1–33) is required for multimerisation, The C-terminal domain (residues 64–181) binds the DNA. Most of the cellular SeqA is found in a complex with the new DNA at the replication forks [Waldminghaus and Skarstad, 2009].

SeqA can bind cooperatively and form a filament through interactions of the N-terminal domain, leading to DNA condensation [Waldminghaus and Skarstad,

2009]. Furthermore as SeqA forms a filament on hemimethylated *oriC*, it binds tighter to this region than Dam, sequestering these sites for an extended period (Figure 1.3) [Guarne et al., 2005] [Kang et al., 2007]. Studies show SeqA clusters bound to hemimethylated DNA behind the replication fork, thereby competing with mismatch repair protein MutH [Brendler et al., 2000] [Hiraga et al., 2000].

Replication initiation synchrony is achieved by SeqA, which negatively regulates initiation of replication by binding to the hemimethylated origins and preventing re-initiation. In *dam* mutants all GATC sites are permanently unmethylated therefore replication origins are not targets for SeqA sequestration. This leads to asynchronous initiation of replication [Løbner-Olesen et al., 1994]. SeqA also has additional binding sites in the chromosome other than around the *oriC* and on the *dnaA* gene promoter. The same phenotype is observed in $\Delta seqA$ cells or Dam over-expressing cells [Boye and Løbner-Olesen, 1990] [Lu et al., 1994]. Microarray studies showed that the transcription profiles of $\Delta seqA$ and Dam-overproducing cells were widely similar [Løbner-Olesen et al., 2003].

1.3 Methyl-Directed DNA Mismatch Repair

Earlier *in vitro* studies using cell extracts discovered that mismatch repair depends on the products of four *E. coli* mutator genes: *mutH*, *mutL*, *mutS*, and *uvrD* [Lu et al., 1984]. Inactivation of any of these genes causes an increase in mutation rates in *E. coli* cells up to 100-fold, indicating that this pathway is important for genetic stability [Glickman and Radman, 1980].

Paul Modrich was one of the pioneers in the field and he was interested in how mismatches might be recognized and how the state of DNA methylation at one site directed the mismatch repair elsewhere on the DNA. They built hemimethylated plasmid DNA and put a G-T mismatch within the recognition site

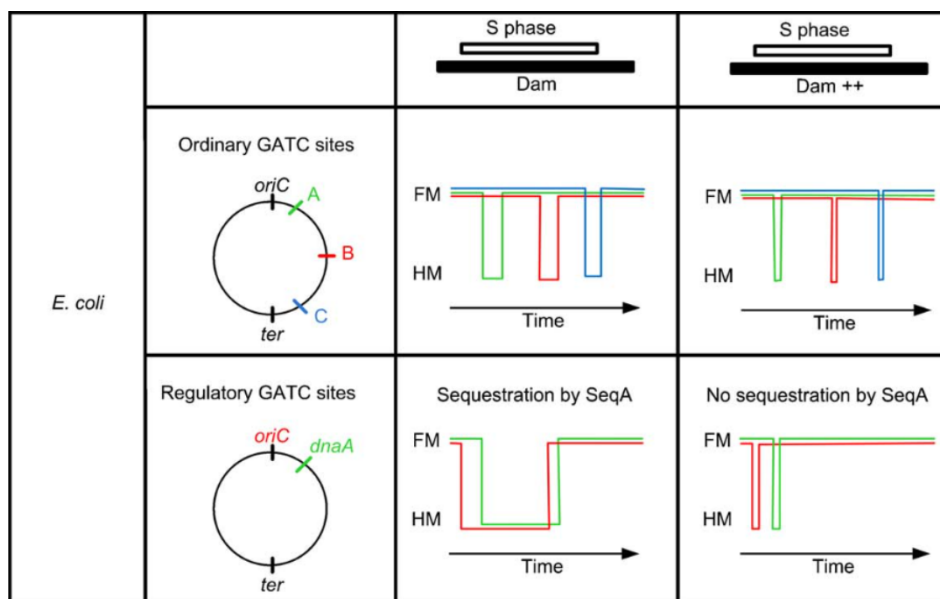


Figure 1.3: Waves of DNA methylation after replication

Methylation of GATC sites lags behind the replication fork. FM stands for fully methylated and HM stands for hemimethylated. Sites close to the *OriC* get replicated faster than sites near *ter*. When Dam is over-expressed methylation occurs more rapidly and SeqA sequestration is lost resulting in asynchronous replication initiation (figure adapted from [Collier, 2009]).

of EcoRI restriction endonuclease (T on the unmethylated strand)(Figure 1.4 A). The mismatch blocks DNA cleavage by EcoRI, therefore digesting this plasmid with EcoRI and BamHI endonucleases yields a full-length linear product (Fig.1.4 B, unrepaired). If repair occurs on the unmethylated strand, the G-T mismatch is corrected to a G-C base-pair restoring EcoRI sensitivity, and digestion with EcoRI and BamHI produces two smaller DNA fragments (Fig.1.4 B, repaired). Variation of this assay showed that all mismatches are repaired by the MMR system except C-C [Su et al., 1988] (Figure 1.5) and at least one hemi-methylated d(GATC) site is necessary for repair [Lahue et al., 1987].

The four proteins described above (Muts,MutL,MutH,UvrD) are not sufficient for mismatch repair. [Lahue et al., 1989] used a biochemical approach similar to that of Figure 1.3 to identify the other components: exonuclease I that hydrolyzes single stranded DNA in 3' to 5' direction, DNA polymerase III holoenzyme that re-synthesizes the DNA after excision, the *E. coli* single strand DNA binding protein SSB and DNA ligase.

In their assay a system comprising 8 proteins MutS, MutL, MutH, UvrD, ExoI, PolIII, SSB and ligase was sufficient to reconstitute methyl-directed mismatch repair *in vitro*. However the presence of only one exonuclease (ExoI) was sufficient only because their PolIII prep was contaminated with ExoVII which has the opposite polarity of ExoI. Later it was discovered that depending on the relative location of mismatch and incision site, an exonuclease with the opposite polarity is required (ExoI or ExoX (3'–5' exonuclease), or ExoVII or RecJ (5'–3' exonuclease) [Li, 2008]).

To elucidate how the pathway works, they isolated MutS, MutL, MutH and UvrD proteins. It was known that the product of *uvrD* gene is DNA helicase II and it is the enzyme responsible for unwinding the DNA in an ATP-dependent fash-

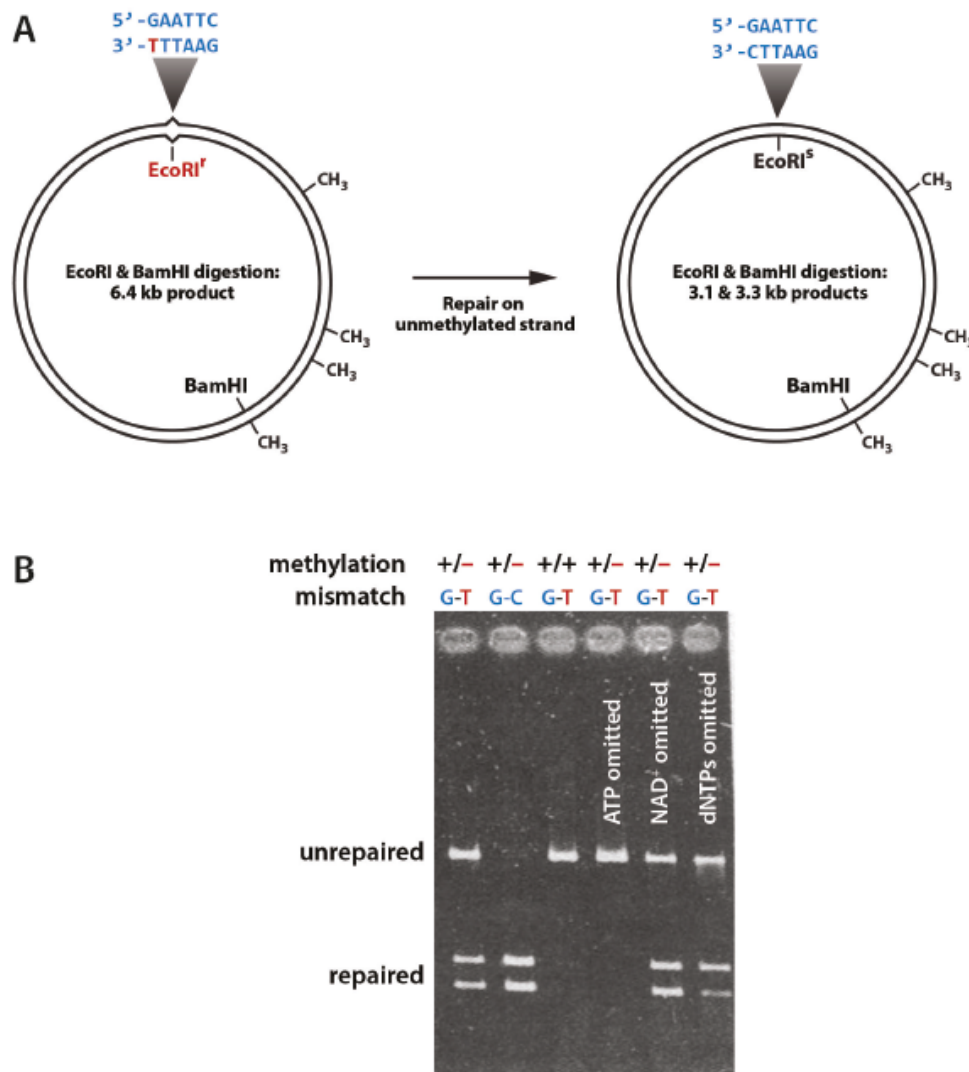


Figure 1.4: Methyl directed mismatch repair in *E. coli* extracts

A) Biochemical Assay for *in vitro* Mismatch Repair. Presence of a G-T mismatch within EcoRI recognition sequence blocks DNA cleavage if unrepaired.

B) *In vitro* repair of above heteroduplex under various conditions adapted from [Lu et al., 1983].

Apparent affinities of mutS protein for base pair mismatches

Mismatch	Apparent dissociation constant
	<i>nM</i>
G-T	39 ± 4
A-C	53 ± 4
A-A	110 ± 7
T-T	140 ± 9
G-G	150 ± 10
A-G	270 ± 30
C-T	370 ± 40
C-C	480 ± 50

Figure 1.5: Apparent dissociation constants for MutS binding to mismatched pairs (figure from [Modrich, 2016])

ion. Later it was discovered that MutS scans the DNA for mismatches and upon recognition of a mismatch exchanges ADP with ATP and becomes a sliding clamp. Once in the sliding clamp conformation MutS recruits MutL, a protein matchmaker, which in turn activates the latent endonuclease activity of MutH. Upon activation MutH incises the DNA on the unmethylated strand and MutL recruits UvrD.

It was shown that the unmethylated strand is repaired through an excision followed by DNA synthesis [Su et al., 1989]. When DNA synthesis is blocked, a single stranded DNA gap formed spanning the shorter path between the mismatch and d(GATC) site on the unmethylated strand. This was true for both strands of the helix, suggesting methyl directed MMR occurs bidirectionally [Cooper et al., 1993].

A study showed that MutL interacts physically with the clamp loader subunits of DNA polymerase III, suggesting that MutL may promote binding of DNA polymerase III to MMR intermediates. Kinetic parameters revealed that the interactions are strong and stable, suggesting that MutL might be involved in the

recruitment of the clamp loader during the resynthesis step in MMR [Li et al., 2008]. This observation suggest that MMR is coupled with DNA replication.

GATC sites are more frequent in translated regions than in non-coding regions, consistent with more frequent mismatch repair surveillance [Marinus and Løbner-Olesen, 2014]. A GATC tetranucleotide is never separated from another GATC sequence by more than 2kb [Marinus and Løbner-Olesen, 2014]. This facilitates *dam*-directed mismatch repair to occur over the entire genome, as it is less efficient at distances larger than 2kb.

The current model of MMR could be summarised as follows: MutS recognizes the mismatches and goes through a conformational change which allows it to recruit MutL. MutL then acts as a protein matchmaker and activates MutH that is bound to GATC sites. MutH incises the newly synthesized strand, creating a gap for MutL recruited UvrD helicase to enter. UvrD unwinds the DNA and exonucleases degrade the newly synthesized strand until right after the mismatch. The single stranded template DNA is protected by SSB and PolIII gets a second chance to correctly synthesize the DNA. Finally, ligase seals the gap (Figure 1.6).

There is little information about the regulation of MMR proteins in the literature. They are in different operons, the transcription units mostly containing the genes themselves. MutS is in position 2,857,093 -> 2,859,654, MutL is in position 4,397,412 -> 4,399,259 and MutH is in position 2,969,662 -> 2,970,351. The copy number of proteins vary depending on the medium. The copy numbers in MOPS medium with 2% glucose are as follows: MutS: 150 molecules/cell, MutL: 77 molecules/cell, MutH: 37 molecules/cell. When grown in rich media (Neidhardt EZ rich defined medium) the copy numbers increase as follows: MutS: 1049 molecules/cell, MutL: 160 molecules/cell, MutH: 119 molecules/cell

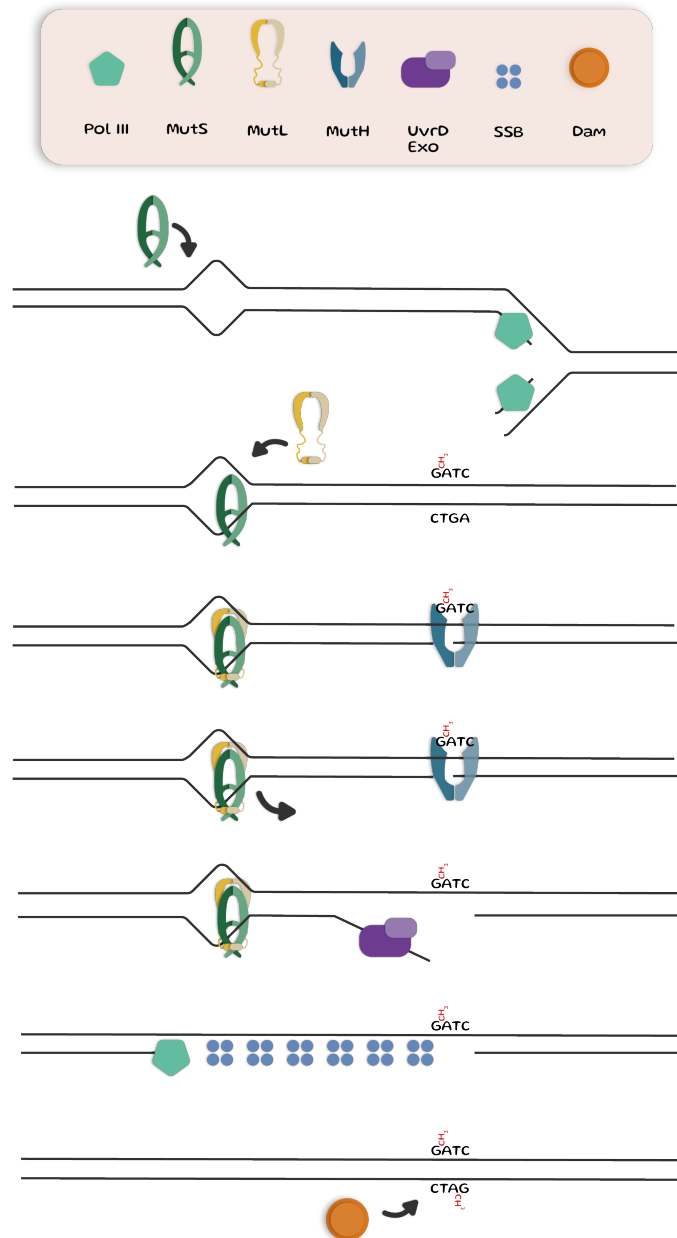


Figure 1.6: Representation of the MMR pathway

MutS recognizes the mismatch and recruits MutL, whereas MutH recognizes hemi-methylated GATC sites. After MutH incises the DNA UvrD unwinds it and exonucleases degrade the nascent strand. SSB protects the single strand and PolIII re-synthesizes the DNA. Finally, ligase seals the gap.

[[Li et al., 2014](#)].

1.4 MutS

MutS is a 95kDA modular protein [[Su and Modrich, 1986](#)] assembled from multiple domains that forms a homodimer and a tetramer. Residues 2–115 form the "thumbs", a six-stranded mixed β -sheet surrounded by three α -helices. Residues 116–266 form the connector domain, which is more regular, comprised of mostly parallel β -sheet, surrounded by four α -helices. The core domain comprises of two sequence regions (residues 267–443 and 504–567), that form a helical bundle. They are called "the levers", two α -helices that extend from within the core domain, embracing but not touching the DNA. The core and connector domains together form the "palm of the hand", sharing a large hydrophilic interface. Residues 444–503 form the clamp domain "the fingertips" which is an insertion at the top of the levers, forming a four-stranded antiparallel β -sheet. The ATPase domain forms "the wrist" (residues 568–765) and is connected to the core domain. The ATPase has a classical Walker A motif and is most similar to the ATPase domain of the ABC transporter histidine permease [[Hung et al., 1998](#)]. Finally, the helix-turn-helix (HTH) domain (residues 766–800) is involved in dimer contacts [[Lamers et al., 2000](#)] (Figure 1.7).

The crystal structure of the *E. coli* MutS dimer bound to a G-T mismatch was solved in the beginning of the millennium confirming that MutS is responsible for mismatch recognition [[Lamers et al., 2000](#)]. MutS dimers look like a pair of hands coming together, almost like a prayer position, with wrists together and fingers touching. Approximate overall dimensions of the complex are 125x90x55 Å [[Lamers et al., 2000](#)]. Almost the entire molecule is negatively charged except for the top of the dimer that binds the DNA. The mismatched

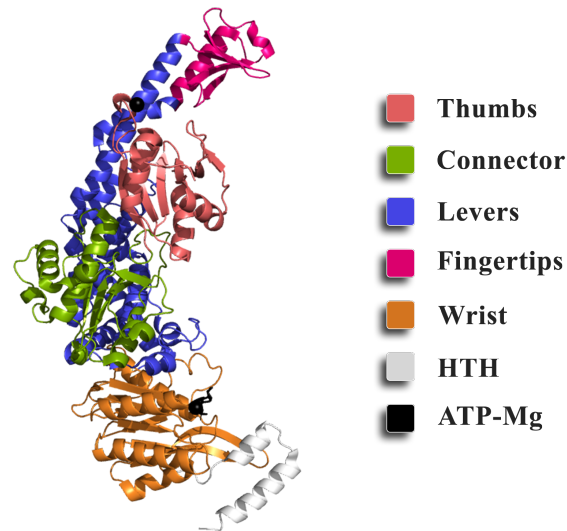


Figure 1.7: MutS Annotated (1E3M) [Lamers et al., 2000]

DNA that it was crystallized with had a strong kink of around 60° at the G-T mismatch. This results in a very wide and shallow minor groove and an associated narrow and deep major groove. This structural change in the DNA is suggested to facilitate recognition of a DNA mismatch and the lack of such kink in C-C mismatches could be the reason that they are poorly recognised by MutS. The clamp domains present a positively charged surface to the DNA, forming several salt bridges with the phosphate backbone and therefore might be responsible for initial contact with the DNA. In the eukaryotic heterodimer only one of the dimers makes contact with the mismatch.

Mismatch recognition is established by stacking the phenylalanine (Phe36) with the mismatched nucleotide. In a normal minor groove these residues would cause electrostatic repulsion of the DNA backbone, but when there is a mismatch the minor groove is widened therefore orientating the mismatch-binding

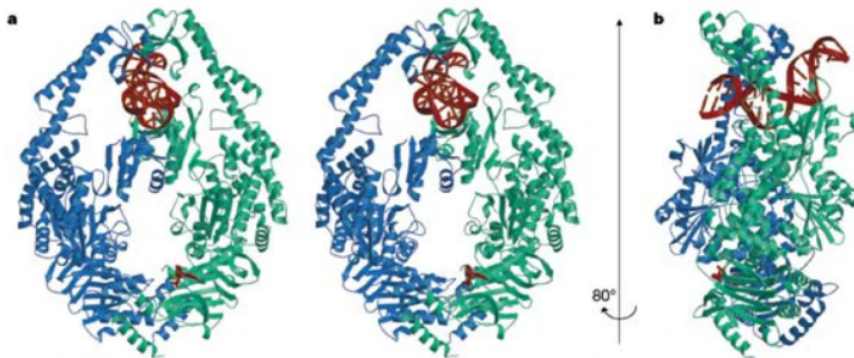


Figure 1.8: Overview of MutS-DNA complex (1E3M) [Lamers et al., 2000]

a) Stereo view from the front. b) Side view, rotated 80° from structure in a. DNA and ADP are coloured red, the mismatch-binding monomer light green, and the second monomer blue.

domain with respect to the DNA backbone. Using dynamic single molecule fluorescence resonance energy transfer (smFRET) [Jeong et al., 2011] determined that MutS scans duplex DNA for mismatched nucleotides by 1D diffusion while in continuous rotational contact with the DNA.

[Cristóvão et al., 2012] demonstrated with single molecule fluorescence and ensemble anisotropy studies in solution that at least two distinct DNA-MutS species exist, with the DNA bent (in which MutS is positioned at the mismatch) and unbent/less bent (in which MutS is positioned at or away from the mismatch). More significantly, they revealed that the MutS dimer binds asymmetric mismatches such as G:T or A:C in a preferred orientation, which explains the overwhelming bias in MutS orientation on DNA in crystal structures of these complexes. They suggested one possible solution to the problem posed by directional mismatch recognition could be release/rebinding of MutS and eventual coverage of all mismatch orientations in multiple rounds of MutS scanning DNA.

When MutS binds the mismatched DNA, it continuously tries to bend the DNA

through the combined action of the mismatch-recognition domain and the clamp domain. The former then checks whether it can modify the DNA conformation using DNA's negative charge to wedge in the phenylalanine. This conformation would only be possible if there is a flexible DNA backbone due to a mismatch or a looped out base, otherwise bend needs to be present owing to a chemical modification by a DNA damaging agent. For example, UV radiation causes two classes of DNA lesions: cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4) [Clancy, 2008]. Both of these lesions distort DNA's structure, introducing bends to the DNA backbone allowing MutS to bind. When there is no flexibility in the DNA backbone, MutS can not stably bind, and disassociates from the DNA [Cristóvão et al., 2012].

Recently, Lamers group determined four cryo-EM structures of MutS bound to ATP, ADP and their analogues in absence of DNA [Borsellini et al., 2021]. Their results shed light into the mechanism of ATP hydrolysis. In all these structures they saw clamp domains were poorly resolved, indicating that in the absence of DNA these positively charged structures are flexible. Like all ABC ATPases, MutS has two ATP binding sites. It needs to bind to ADP on both sites to dimerise and to bind to ATP on both of them to hydrolyze. In the absence of DNA, the two monomers freely open and close by going through cycles of ATP binding and hydrolysis. If MutS associates with homoduplex DNA to scan the DNA for a mismatch, the dimer stays open thus preventing ATP hydrolysis. Only mismatch binding induces a rearrangement that transforms MutS into a sliding clamp allowing both monomers to bind ATP [Hopfner, 2016]. This transformation into the clamp state moves the DNA towards the centre of the dimer while the lever domains cross each other locking the DNA inside the clamp. Now in the clamp state, ATPase sites enable ATP hydrolysis. However with double-stranded DNA keeping the lever domains stuck, the two monomers can not

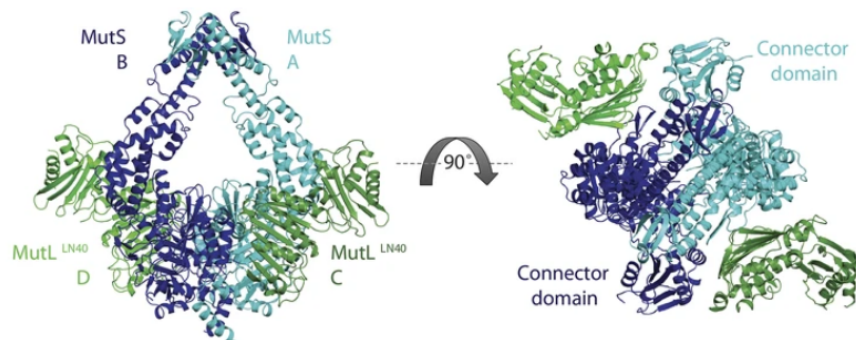


Figure 1.9: Crystal structure of the trapped MutS-MutL transient complex (5AKB) [Groothuizen et al., 2015] MutS in blue-cyan, MutL-N40 in green

open up, preventing the ATPase cycle to continue. Hydrolysis cycle can only continue after UvrD and exonucleases generate a single stranded DNA. When MutS encounters single stranded DNA, it is able to disassociate from the DNA and resume the hydrolysis cycle.

Single molecule studies with *Saccharomyces cerevisiae* show that MutS α can be targeted to mismatched bases through either 1D sliding or 3D diffusion [Gorman et al., 2012b]. [Gorman et al., 2012b] used single-molecule fluorescence imaging to demonstrate eukaryotic MutS and MutL proteins sliding during mismatch search on stretched DNA. They made double-tethered DNA curtains with three tandem G/T mismatches and calculated that MutS located the mismatches either through 1D sliding (42.5% of observed events), with sliding observed over distances up to 3.7 μm (14.6 kbp), or through apparent 3D diffusion (57.5% of observed events).

1.5 MutL

E. coli MutL is a 70 kDa protein formed by an N-terminal 40 kDa fragment (residues 1–349) and a C-terminal 30 kDa fragment (residues 350–615). To date,

the full *E. coli* MutL structure is not crystallized however the crystal structures of a 40 kDa ATPase fragment of *E. coli* MutL (LN40) complexed with a substrate analog ADPnP is published [Ban et al., 1999]. LN40 consists of two domains, residues 20–200 form the first domain and residues 207–331 form the second domain, both of which are of mixed $\alpha\beta$ -folds [Yang, 2000]. The crystal structure of LN40 is strikingly similar to an ATPase fragment from DNA gyrase B subunit (NGyrB). Low ATPase activity ($< 0.5/\text{min}$) has been detected with purified MutL [Ban and Yang, 1998a]. The ATPase activity of the MutL family appears to be essential for the DNA repair function, as the majority of the mutations in human MLH1 and almost all of the mutations in *E. coli* MutL causing the dominant mutator phenotype are in or around the ATP-binding site [Ban et al., 1999]. All members of the MutL family characterized so far form homo or heterodimers *via* the C-terminal region [Yang, 2000].

[Groothuizen et al., 2015] solved the structure of MutS in complex with MutL. As described above, when MutS is activated by a mismatch and ATP it pushes DNA into a new channel, which allows sliding of the protein over DNA. In this new state clamps are crossed over the DNA which makes MutS sliding clamp stable on DNA and enables binding of MutL. The new channel is lined with positive charges which stabilizes the new clamp conformation with the negatively charged DNA. These conformational changes push the connector domain of MutS (shown at figure 1.8) away from the center and on top of the ATPase domains, creating a new docking station for MutL. Therefore, MutS can only bind and recruit MutL onto DNA when activated by a mismatch and while ATP bound. DNA in the channel is also thought to contribute to binding of MutL.

1.6 MutH

E. coli MutH is a 28 Kda, Mg^{2+} dependent monomeric endonuclease [Yang, 2000]. The intrinsic endonuclease activity of MutH is very weak, but it is activated 50-fold in the presence of MutS, MutL, ATP and a mismatch [Au et al., 1992].

The structure of MutH looks like a clamp, divided into two halves by a large cleft. Each half forms a hydrophobic subdomain that has similar structural elements that are connected by three polypeptide linkers. The interface between them is flexible and the two subdomains can pivot relative to each other.

MutH is structurally very similar to PvuII endonuclease and type II restriction endonuclease Sau3AI [Ban and Yang, 1998b]. MutH is a monomer in solution and when crystallized, therefore monomeric MutH is thought to be the active species of the enzyme. This causes it to cleave only one DNA strand in mismatch repair. The DNA binds in the cleft found in between the N and C terminal subdomains, similar to the substrate binding site of PvuII and EcoRV. Located on the entrance of the DNA binding cleft are two loops (named BC and 67) that form a positively charged gateway for the DNA. MutH active site also lies between the N-arm and the C-arm, containing a conserved catalytic sequence motif. MutL binds to the C-terminal tail of MutH. This tail dictates the opening and closing of the cleft, leading to the activation of the endonuclease activity of DNA-bound MutH by MutL [Ban and Yang, 1998b].

Unlike MutS and MutL, no single molecule experiments have been performed with MutH *in vivo* to date.

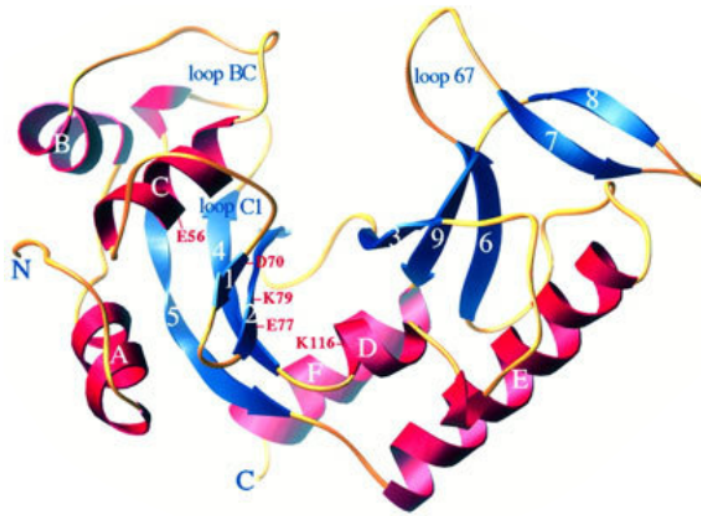


Figure 1.10: Crystal structure of *E. coli* MutH (1AZO) [Ban and Yang, 1998b]

A ribbon diagram of MutH where helices are shown in red and β -sheets in blue. Positively charged loop BC and loop 67 function in orienting the DNA towards the catalytic cleft.

1.7 Communication of Mismatch Recognition to Excision Initiation Site

Mismatch repair is initiated by MutS that scans the DNA continuously and goes through a conformational change upon mismatch recognition by exchanging ADP for ATP. This conformational change creates a DNA sliding clamp and allows MutS to recruit MutL. At a distance, MutH recognizes hemimethylated GATC sites and when activated by MutL incises them. A mismatched base could be as far as 1kb away from the GATC sites that MutH recognizes, so how these distant components of the pathway come together has been a burning question in the field for decades.

4 models have been suggested to explain the communication of mismatch recognition by MutS to a distant GATC incision site recognized by MutH. These mod-

els can be further grouped into two:

- 1) Looping Models
- 2) Sliding Models

The first looping model was suggested by Modrich called Hydrolysis-Dependent Translocation Model. It envisioned the process as follows: MutS and MutL form a stable complex at the mismatch, creating a motor that hydrolyses ATP and uses this energy to pull the DNA into a loop to bring MutS-MutL complex in contact with MutH at the nearest GATC site (Figure 1.11 B) [Modrich, 1989].

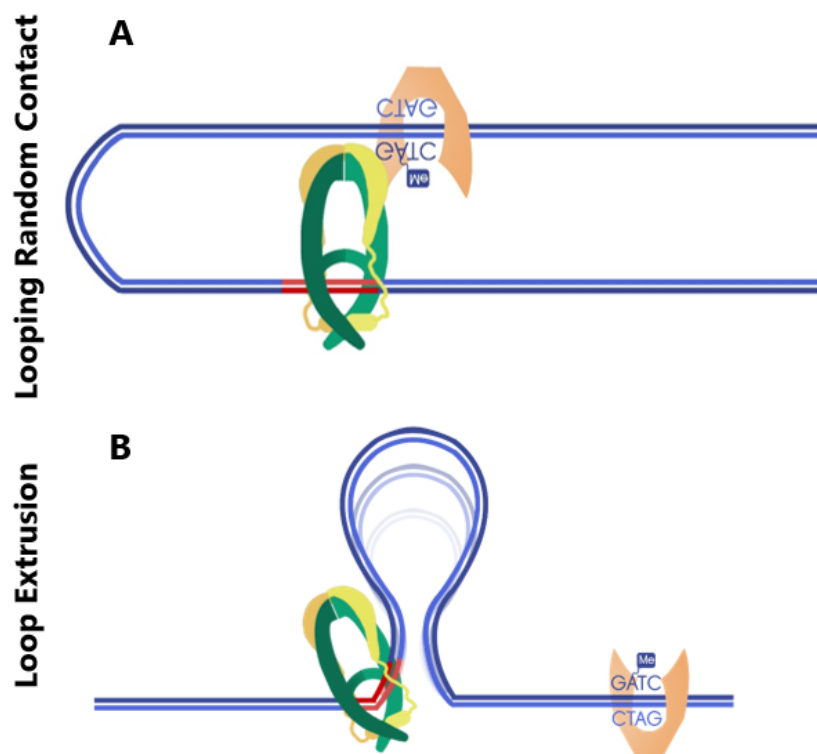


Figure 1.11: Schematic exemplifying looping models

A) MutS-MutL complex can come together with MutH in 3D by Brownian motion of the DNA. B) MutS-MutL complex can actively extrude the DNA into a loop until they reach a MutH bound to a GATC site.

Evidence for this model came from atomic force microscopy (AFM) studies that captured MutS tetramers forming a loop [Iyer et al., 2006], [Jiang and Marszalek, 2011]. (Figure 1.12) In this model ATPase activity is crucial for communication, and further proof to this was presented when AFM showed loops getting bigger in the presence of ATP. It also requires MutS tetramer formation, however, [Mendillo et al., 2007] demonstrated that the mutations that specifically block tetramer formation does not affect MMR *in vivo* and there is no evidence that any other MutS homologue forms tetramers. Therefore, in the schematic representation above, I chose to draw dimers forming a loop.

MutS dimers coming together creates a smaller chamber and a bigger chamber, similar to structural maintenance of chromosomes (SMC) complexes. SMC complexes are ring-shaped ATPases that bind to chromosomes and embrace them topologically [Uhlmann, 2016]. Recently, it has been shown that condensin, an SMC complex, could extrude loops [Ganji et al., 2018] so it is possible to imagine MutS-MutL complex to bring different parts of the DNA together by actively extruding it.

A variation of this model suggested three dimensional (3D) collision of the mismatch recognised by MutS and GATC sites recognised by MutH by Brownian motion of the DNA (Figure 1.11 A). However placing a roadblock made of biotin-streptavidin between the mismatch and strand incision site completely inhibited MMR implicating a DNA translocation mechanism [Pluciennik and Modrich, 2007]. Furthermore, [Mardenborough et al., 2019] showed that their roadblock can be bypassed using the long, disordered linker regions connecting the globular N- and C-terminal domains of MutL. By using functional ensemble and single molecule analyses they showed that shortening the length of the disordered MutL linkers completely abolishes roadblock bypass during MMR strand incision. This finding further suggests that the communication during MMR

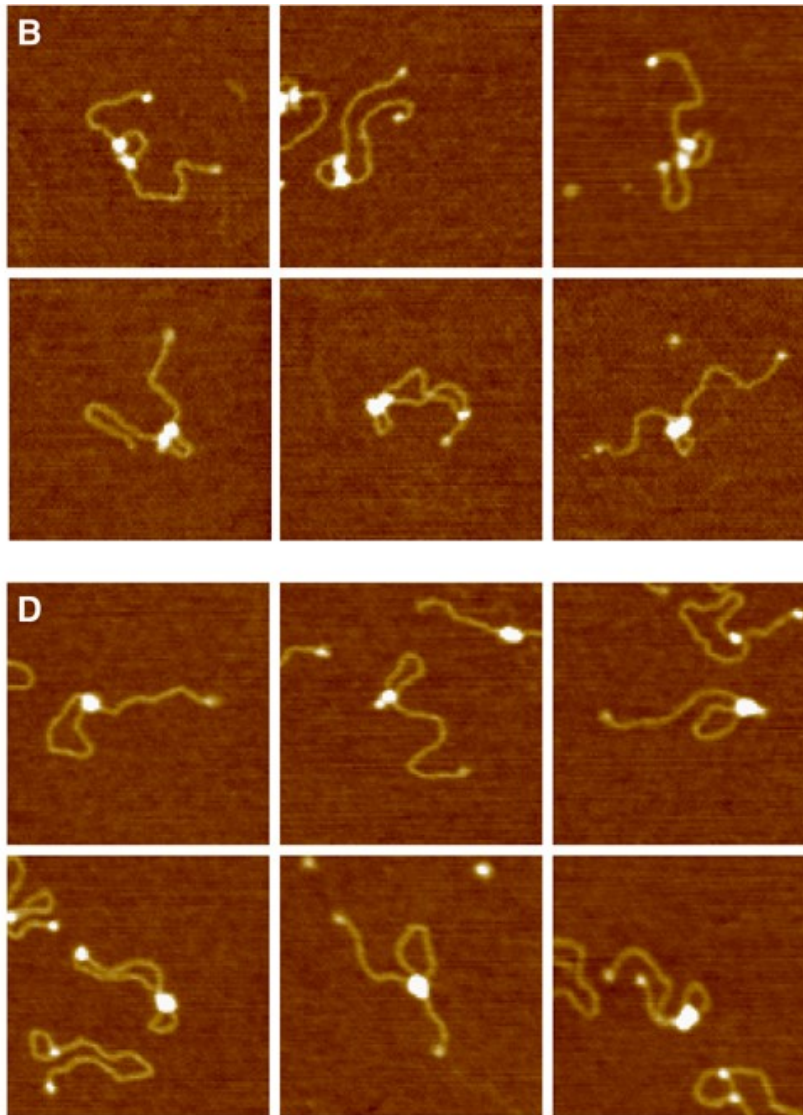


Figure 1.12: AFM images showing loop binding of MutS to heteroduplex DNA
B) Images of MutS loop binding deposited on mica in the absence of nucleotide.
D) Images of MutS loop binding deposited on mica in the presence of 0.5 mM ATP, showing that loops get bigger in the presence of ATP. (Figure adapted from [Jiang and Marszalek, 2011])

initiation occurs along the DNA backbone and not through a 3D looping mechanism which would not have been affected by roadblocks.

The second group of models are based on the sliding hypothesis as numerous studies have visualized mismatch repair proteins diffusing along the DNA. [Liu et al., 2016] visualized *E. coli* MMR and confirmed that MutS forms a stable ATP-bound sliding clamp upon mismatch recognition that randomly diffuse along the DNA with intermittent DNA backbone contact. The MutS sliding clamps then recruit MutL onto the mismatched DNA, forming a MutS-MutL search complex that diffuses along the DNA backbone.

[Junop et al., 2001] demonstrated that MutS-MutL complex can activate MutH in trans, *i.e.* when the DNA mismatch and the GATC incision site are present on separate DNA molecules. They cleaved a 110 bp DNA substrate into two pieces by restriction enzymes so that one DNA fragment contained a T:G mismatch and the other a GATC sequence. They found MutH was activated to nearly the same level as the uncleaved DNA substrate. However, [Wang and Hays, 2007] argued against this finding. They used nicked circular substrates incubated in human nuclear extracts to demonstrate that no nicking was visible when the mismatch and the GATC sites were on different substrates.

There are two possibilities to consider when it comes to sliding models which are intrinsically connected to how MutH finds its substrate. If MutH were to bind to GATC sites independently of the MutS-MutL complex, MutS and MutL sliding clamps would travel the DNA until they found a pre-bound MutH (Figure 1.12 A). If MutH binding is reliant on the mismatch recognition complex, then MutS-MutL and MutH would form a ternary MutSLH sliding complex (Figure 1.12 B). [Liu et al., 2016] used DNA curtains to examine diffusion of MMR proteins on DNA and their data suggest the latter as they see MutS-MutL-MutH

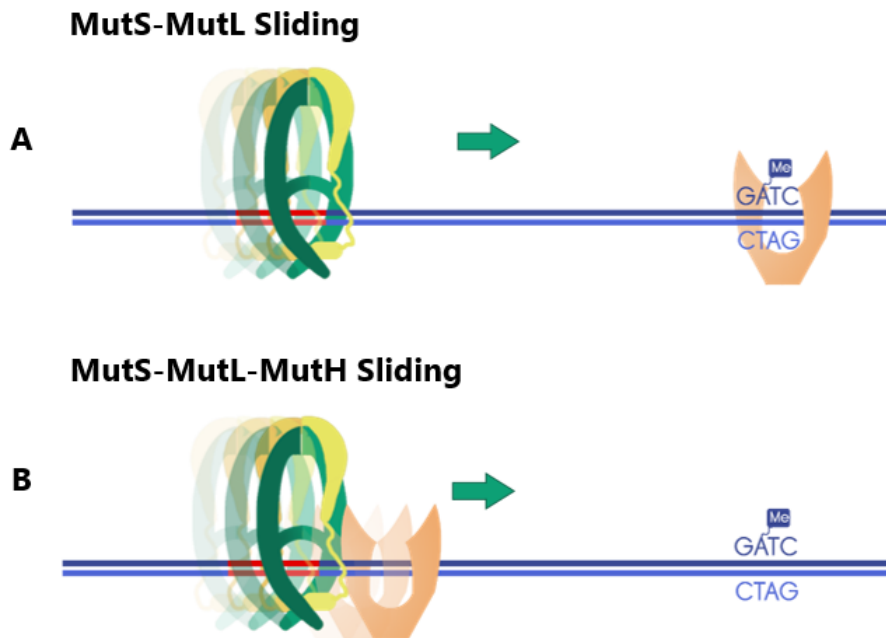


Figure 1.13: Schematic exemplifying sliding models

A) MutS and MutL sliding clamps could travel the DNA until they found a pre-bound MutH. B) MutS-MutL and MutH could form a ternary MutSLH sliding complex, delivering MutH to GATC sites.

sliding together. They only detected MutH bound to the mismatched DNA in <50 mM NaCl, however, when MutS and MutL were present numerous MutH particles were observed on the mismatched DNA at physiological ionic concentration. Furthermore, the majority of MutH (89% of particles) co-localized and displayed coordinated diffusion with MutL(Figure 1.14). If mismatch repair can occur in trans, then MutH would be expected to find its substrate independently of MutS and MutL complexes.

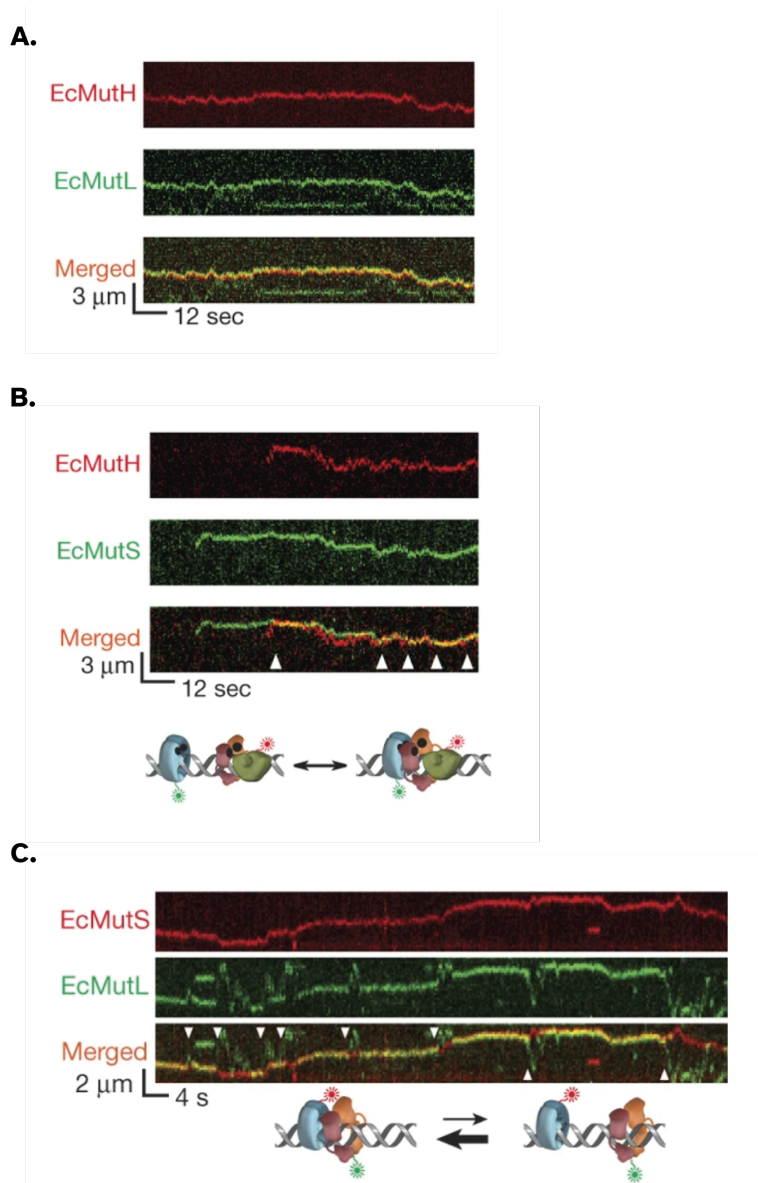


Figure 1.14: Representative kymographs on DNA curtains

A) MutH and MutL are labelled with Alexa647 and Cy-3 respectively (MutS present unlabelled), B) Representative kymographs on DNA curtains for MutH and MutS. MutH and MutS are labelled with Alexa647 and Cy-3 respectively (MutL present unlabelled) C) Representative kymographs on DNA curtains for MutS and MutL. MutS and MutL are labelled with Alexa647 and Cy-3 respectively. All fields are imaged using total internal reflection fluorescence (TIRF) microscopy on doubly-tethered 17.3-kb mismatched DNA. (Figure adapted from [Liu et al., 2016])

2 | Methodology

2.1 Construction of Halo-tagged Strains Using Lambda Red Recombination

C-terminal Halo tagged fusions of DNA binding-proteins, expressed from their native promoter in *E. coli* AB1157 are generated by replacing the endogenous genes using λ -Red recombination. λ -Red Recombination is performed as described before [Murphy, 2016], [Reyes-Lamothe et al., 2008], [Datsenko and Wanner, 2000], [Uphoff et al., 2016]. The forward primer is designed to anneal to 33 nucleotide linker together with the coding region of Halo Tag and contains 50 nucleotide overhang that is homologous to the gene of interest right before its stop codon. The reverse primer is designed to anneal to plasmid pSU005 which contains Halo tag together with a kanamycin cassette, allowing for selection of the chromosomal integration, after an *frt*-kan-*frt* site and also contains 50 nucleotide overhang that is homologous to immediate downstream of the gene of interest.

Frt sites are 34bp sequences that are targeted by the budding yeast recombinase flippase (FLP). FLP binds specifically to the symmetrical elements in the *Frt* sites, and induces an acute bend in the FRT site [Schwartz and Sadowski, 1990] cleaves the FRT site and forms a Holliday-like intermediate by promoting the reciprocal exchange of a pair of strands [Holliday, 1964], [Meyer-Leon et al., 1988], [Parsons et al., 1988] which is then resolved.

dsDNA substrate is generated by high fidelity PCR amplification using pSU005 plasmid and primers listed below. PCR products are incubated overnight with DPN1 restriction endonuclease to degrade the remaining template and subse-

quently gel purified. A bacterial strain containing pKD46 stably called SU083 is used for electroporation to overexpress lambda-Red insertion proteins as described by [Datsenko and Wanner, 2000].

pKD46 carries lambda red genes behind the inducible araBAD promoter, contains an ampicillin resistance cassette and is temperature sensitive (cured when grown at 37 degrees). Recombination is sufficiently induced by adding 0.2% of fresh L-arabinose. The transformed strain is grown at 30 °C to maintain the temperature-sensitive plasmid.

To make the SU083 electrocompetent the strain is grown until the OD600 is approximately 0.6 (0.5 - 0.8 acceptable). The flask is immediately submersed in ice-water slurry while swirling to rapidly cool down and stop metabolic activity, and it's left on ice for 10 min. The ice-cold culture is transferred to a prechilled 50-ml Falcon tube. Next the culture is centrifuged for 10 min at 3000 g and the pellet is resuspended in 35 mL ice cold ddH₂O. This step is repeated once more and after the next centrifugation step the pellet is resuspended in 2 ml ice cold 10 % glycerol. The culture is moved to an eppendorf, centrifuged for 10 min at 3000g and pellet is resuspended in 500 µl ice cold 10% glycerol. 90 microliter of freshly prepared competent cells are immediately electroporated with 100 ng of purified PCR product.

After electroporation bacteria are cured from pKD46 and plated on to selective plates. Successful recombination is validated by using colony PCR with primers flanking the insertion site and demonstrates a shift in agarose gel electrophoresis corresponding to the additional DNA fragment length. After successful electroporation and recombination, strains are moved to wild-type AB1157 background using P1 transduction. Full functionality of the fusion proteins are tested two-fold. The cell growths rates were similar for wild type *E. coli* and fusion

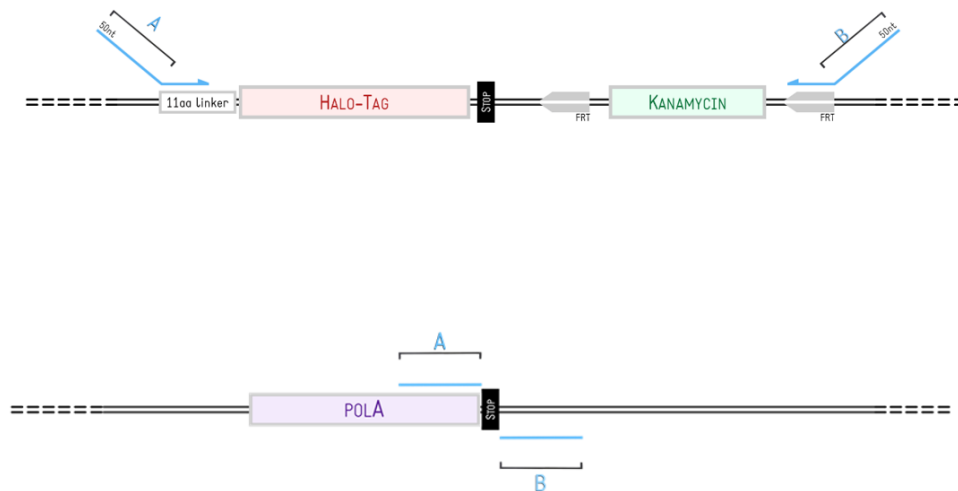


Figure 2.1: Schematic representation of primer design

Plasmid SU005 was used as a template to amplify an 33 nucleotide linker, the Halo tag and a Kanamycin cassette *via* PCR and transformed into bacteria containing recombinase using lambda-red recombination. The forward primer (A) is designed to contain a 50 nucleotide overhang that is homologous to the gene of interest right before its stop codon and anneals to 33 nucleotide linker. The reverse primer (B) anneals to the frt site on plasmid SU005 and contains a 50 nucleotide overhang that is homologous to immediate downstream of the gene of interest.

strains in LB and M9 (minimal) medium. Furthermore, spontaneous mutation frequency assay is performed to confirm the mutation rates are unaffected when Mismatch Repair (MMR) Proteins were tagged.

2.2 P1 Transduction

Strains with chromosomal deletions of genes were obtained from the Keio collection [Baba et al., 2006]. The alleles were moved into the strains with Halo-tag using P1 transduction following the protocol described in [Thomason et al., 2007]. Note that fragments of approximately 100 kb can be transferred by the

P1 bacteriophage and this could lead to further mutations.

2.3 Heat-shock Transformation

Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (QIAGEN) from overnight cultures of NEB5 α -cells (NEB) carrying the plasmid. 60 ng of plasmid DNA was transformed using heatshock (42 °C for 60 seconds) into the relevant strains made competent using 100 mM CaCl₂.

2.4 Growth Conditions

Cells were grown in Lysogeny broth (LB) medium shaking at 220rpm or on LB agar (2%) at 37 °C. Before imaging they are grown in Glucose/glycerol M9 minimal medium containing (1X M9 salts, 0.2% glucose/glycerol, 2mM MgSO₄, 0.1 μ M CaCl₂, 0.5 μ g/ml thiamine, 0.1 mg/ml L-proline and 1x MEM amino acids (Gibco Life Technologies)). For selection, antibiotics were added to plates and growth media to final concentrations of ampicillin 50 μ g/ml, kanamycin 25 μ g/ml. All cultures were grown in well aerated tubes with a 10-fold excess overall volume as Emma Jones from our group noticed that the SOS response was significantly induced in cells that were grown in poorly aerated tubes.

2.5 List of Plasmids Used in This Study

Plasmid Number	Plasmid Name	Made by
PSU026	pBAD24(dam) ::Amp	Nehir Banaz
PSU027	pBAD24(seqA) ::Amp	Nehir Banaz
PSU005	R6kgamma ori 11aa Halo frt kan frt ::Kan	Stephan Uphoff

2.6 List of Primers Used in This Study

Primer Name	Date	Sequence	Made by
SU180 Dam Deletion F1	09/02/2018	CGTTGCGGGAGTTGCCTGAAGCGC	Nehir Banaz
SU181 Dam Deletion R1	09/02/2018	GCG GTA TCT TCA CCC AGG CGG	Nehir Banaz
SU 182 MutS Deletion F1	05/03/2018	AGCATCTTTCCCGGAACCAG	Nehir Banaz
SU 183 MutS Deletion R1	05/03/2018	TCGCCAGTTTTCGAATGCCT	Nehir Banaz
SU 184 MutL Deletion F1	05/03/2018	GAT CGC ACG CTG CCA AAC	Nehir Banaz
SU 185 MutL Deletion R1	05/03/2018	CTT AGG CAG GCT CGC CTT	Nehir Banaz
SU 186 MutH Deletion F1	05/03/2018	TCA GAG AAT TGA ACA ACG CAT G	Nehir Banaz
SU 187 MutH Deletion R1	05/03/2018	GAT CCC AAG AAC GAT CTC CAG	Nehir Banaz
SU188 PolI Halo LRed	08/03/2018	TGCCGTTGCTGGTGGAAAGTGGGGAGT GGCGAAACTGGGATCAGGC GCACTCGGCTGGCTCCGCTGC	Stephan Uphoff
SU189 PolI Halo LRed	08/03/2018	ACGTGACAGCTTATGTT GCTTAC TTACGAAAAA GGCATGTTCA GGCGAATCTATGAATATCCTCCTTAG	Stephan Uphoff
SU217 pBAD rev	05/09/2018	TGAATTCCTCCTGCTAGC	Nehir Banaz
SU218 pBAD fwd	05/09/2018	GGCTGTTTTTGCGGATGAG	Nehir Banaz
SU219 Dam fwd	05/09/2018	gggctagcaggaggaa ttcaATGAAGAAAAA TCGCGCTTTTTTG	Nehir Banaz

Primer Name	Date	Sequence	Made by
SU220 Dam rev	05/09/2018	tctcatccgccaaaacagcc TTATTTTTTCGCGGGTGAAC	Nehir Banaz
SU221 SeqA fwd	05/09/2018	gggctagcaggaggaattca ATGAAAACGATTGAAGTTGATG	Nehir Banaz
SU222 SeqA rev	05/09/2018	gggctagcaggaggaattca ATGAAAACGATTGAAGTTGATG	Nehir Banaz
SU310 MutH Pet28a NdeI	19/11/2019	TATCATCATATGTCCCAACCTCGC	Nehir Banaz
SU311 MutH Pet28a BamHI	19/11/2019	GTACGGATCCCTACTGGATCAGAAAATGACG	Nehir Banaz
SU312 MutH Pet28a loop1	21/11/2019	gatgcagcgaatgcggat GCGCTTACCGAAGCCATT	Nehir Banaz
SU313 MutH Pet28a loop2	21/11/2019	gccatcatcctcatctaa ATTCTCTGGCGTAACCAG	Nehir Banaz
SU314 MutH Pet28a loop3	21/11/2019	ttagatgaggatgatggc TGGATTGGCGTGTTACTG	Nehir Banaz
SU315 MutH Pet28a loop4	21/11/2019	atccgcattcgtgcatc CGGTCGTATCTGTAAATACTC	Nehir Banaz

2.7 List of Strains Used in This Study

Primer Name	Date	Sequence	Made by
SU189	19/04/2016	AB1157, MutS-Halo::kan	Stephan Uphoff
SU190	19/04/2016	AB1157, MutL-Halo::kan	Stephan Uphoff
SU191	19/04/2016	AB1157, MutH-Halo::kan	Stephan Uphoff
SU192	19/04/2016	AB1157, MutH-mYPet::kan	Stephan Uphoff
SU358	06/02/2018	Ab1157 MutS-Halo (frt Kan removed from Su189)	Nehir Banaz
SU359	06/02/2018	Ab1157 MutL-Halo (frt Kan removed from Su190)	Nehir Banaz
SU360	06/02/2018	Ab1157 MutL-Halo (frt Kan removed from Su191)	Nehir Banaz
SU385	15/02/2018	Ab1157 MutH-Halo dam-::Kan (P1 SN203 -> Su360)	Nehir Banaz
SU411	04/04/2018	AB1157 Δ MutL::Kan	Nehir Banaz
SU412	04/04/2018	AB1157 Δ MutH::Kan	Nehir Banaz
SU413	04/04/2018	AB1157 Δ MutL MutH-Halo::Kan	Nehir Banaz
SU417	24/04/2018	AB1157 Pol1-11aa-Halo::Kan	Nehir Banaz
SU472	04/07/2018	AB1157 DnaN-Ypet MutH-Halo::Kan	Nehir Banaz
SU490	23/11/2018	AB1157 Su 472 carrying pBAD(dam)::Kan::Amp	Nehir Banaz
SU503	05/12/2018	AB1157 Su 472 carrying pBAD(SeqA)::Kan::Amp	Nehir Banaz
SU515	04/01/2019	AB1157 Δ MutS::Kan	Nehir Banaz
SU543	26/02/2019	AB1157, MutS-mYPet::kan Δ MutH (P1 SU412 -> SU169)	Nehir Banaz
SU544	26/02/2019	AB1157, MutL-mYPet::kan Δ MutH (P1 SU412 -> SU170)	Nehir Banaz
SU577	10/05/2019	AB1157, DnaN-Ypet MutH-Halo Δ seqA (P1 Su576 -> Su545)	Nehir Banaz

2.8 Spontaneous Mutation Frequency Assay (Rifampicin Selection)

To check whether the fluorescent fusion proteins are as functional as the wild-type, Spontaneous Mutation Frequency Assay is performed. The assay is used to determine the frequency at which mutants emerge when exposed to inhibitory concentrations of rifampicin [Elez et al., 2010]. Rifampicin resistance arises from mutations that map to the *rpoB* gene, encoding the beta subunit of the RNA polymerase [Goldstein, 2014]. These mutations alter the residues of rifampicin binding sites, resulting in decreased affinity for rifampicin. Rifampicin resistance can be used to infer the genomic mutation rates. There are more than 20 mutations identified that confer rifampicin resistance, 17 unique alleles were identified by [Jin and Gross, 1988] (excluding mutants with more than one change) and since then a few other unique RIF mutations have been described in *E. coli*.

The assay is performed with slight modifications to what was described before [Lee et al., 2012]. Strains are streaked on plates containing 25 µg/ml kanamycin and on LB plates without antibiotic for wild-type, then single colonies are inoculated into LB with (25 µg/ml kanamycin) and grown at 37°C for 4–5 h until they reach OD600 0.8. For each strain, 3 colonies are selected. Cultures are then diluted 1/1:000 and grown over-night at 37°C in LB with 25 µg/ml kanamycin.

Next day, the overnight cultures are diluted 1:10 to measure OD600. The culture with the smallest OD600 is determined and the rest of the cultures are adjusted accordingly so each tube contains the same amount of colony forming units (CFU). 100 µl of dilutions of overnight cultures were plated on selective medium (LB containing 100 µg/ml rifampicin) to select rifampicin-resistant colonies

Spontaneous Mutation Frequency Assay

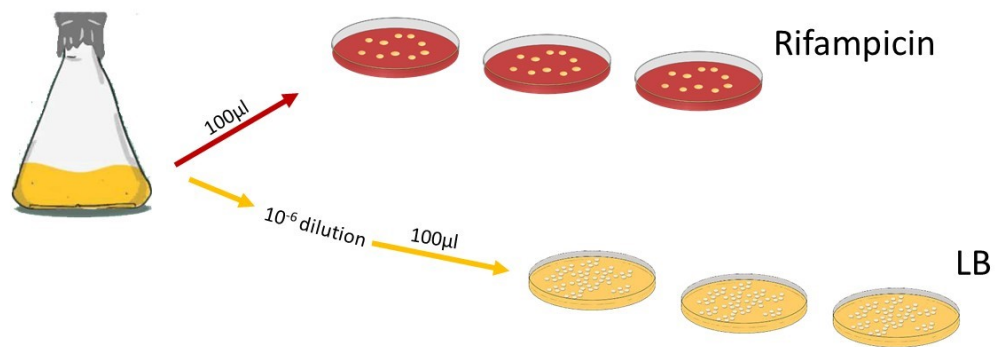


Figure 2.2: Schema explaining spontaneous mutation frequency assay

100 μ l of dilutions of overnight cultures were plated on selective medium (LB containing 100 μ g/ml rifampicin) to select rifampicin-resistant colonies and on LB after a further 10^{-6} dilution to determine the total number of colony forming units. Colonies were scored after overnight incubation at 37°C.

and on LB after a further 10^{-6} dilution to determine the total number of colony forming units. Colonies were scored after overnight incubation at 37°C. The average mutation frequency of each strain was determined from three to six independent experiments performed on different days, with 3 colonies each day. The mutation frequency is estimated by the number of CFU on rifampicin plates normalized to number of CFU on LB plates.

2.9 Labelling of Halo-tagged Proteins

Frozen stock cultures were streaked on to LB agarose with kanamycin selection. A single colony was used to inoculate LB + kanamycin (+ appropriate antibiotic if strain carried a plasmid) and grown for 4 hours until it reached O.D600

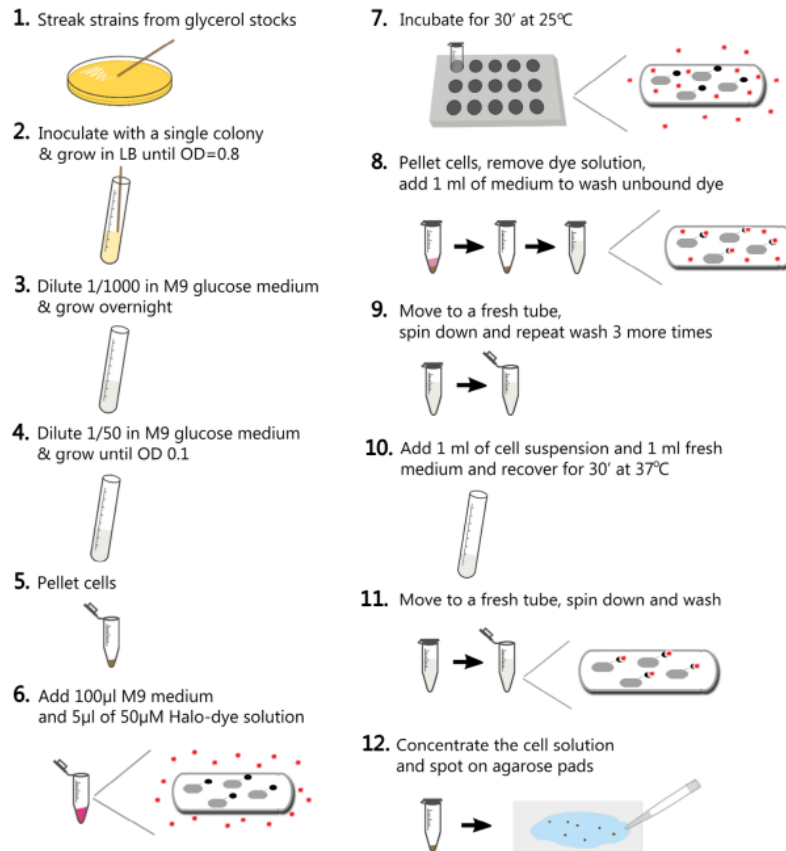


Figure 2.3: Schema explaining labelling with Halo-tag

Frozen stock cultures were streaked on to LB agarose with kanamycin selection. A single colony was used to inoculate LB + kanamycin and grown until O.D₆₀₀ 0.8. The cultures were then diluted 1:1000 into M9 glucose medium and grown overnight to stationary phase. Overnight cultures were further diluted 1:50 into M9 glucose and grown to OD₆₀₀ 0.1-0.2. The cells were pelleted and re-suspended in 100 µl M9 medium. 5 µl of 50 µM Halo-dye solution is added and mixed. The cell-dye solution is incubated at 25 °C for 30 minutes. Cells are pelleted, re-suspended in 1 ml of M9 medium and quickly vortexed then pipetted up and down to wash the unbound dye. The cells are moved to a new tube and the washing step is repeated 3 times. Finally, 1 ml of cell suspension is mixed with 1 ml of fresh medium and left to recover for 30 minutes at 37 °C on a shaker. After recovery, cells are centrifuged and concentrated. The cell concentrate is spotted on agar pads for imaging.

0.8. The cultures were then diluted 1:1000 into M9 glucose medium and grown overnight to stationary phase. In all subsequent steps M9 glucose was used except for strains carrying pBAD plasmids for which M9 glycerol was used to induce expression from the pBAD promoter. Overnight cultures were diluted 1:50 into M9 glucose/glycerol and grown to OD₆₀₀ 0.1-0.2. The cells were pelleted and re-suspended in 100 µl M9 medium. 5 µl of 50 µM Halo-dye solution is added and mixed. The cell-dye solution is incubated at 25 °C for 30 minutes. Cells are pelleted, re-suspended in 1 ml of M9 medium and quickly vortexed then pipetted up and down to wash the unbound dye. The cells are moved to a new tube and the washing step is repeated 3 times. Finally, 1 ml of cell suspension is mixed with 1 ml of fresh medium and left to recover for 30 minutes at 37 °C on a shaker. After recovery, cells are centrifuged and concentrated. The cell concentrate is spotted on agar pads for imaging. Imaging was done using a custom-built total internal reflection fluorescence (TIRF) microscope.

2.10 Single Molecule Tracking

Single Molecule Tracking (SMT) is a technique that allows direct spatial observation of individual molecules as a function of time. The output of the experiment is a video with thousands of frames where the reporter is detected as a visible diffraction pattern with a point spread function (PSF). This PSF is then fit to a 2D spatial Gaussian, to localize the molecules. Once a reporter has been localized in each snapshot, its position is traced across multiple frames tracking the movement of the particle as it was observed over time. From each track with more than 4 steps the mean-square displacement (MSD) is calculated. Tracks

shorter than that are excluded to reduce the statistical uncertainty in the MSD:

$$MSD = \frac{1}{N-1} \sum_{i=1}^{N-1} (X_{i+1} - x_i)^2 + (Y_{i+1} - Y_i)^2 \quad (2.1)$$

N = the number of steps and X and Y are the position coordinates.

$$D^* = \frac{MSD}{4\Delta t} \quad (2.2)$$

For my experiments, I used a custom-built total internal reflection fluorescence microscope [Wegel et al., 2016] under oblique (near TIRF) illumination at 22 °C.

The TIRF microscope for STORM imaging was a custom configuration of ASI Modular Infinity Microscope components (MIM and RAMM, ASI) equipped with an EMCCD camera (iXon 897 Ultra, Andor) and a 100x NA 1.40 oil immersion objective (UPLSAPO, Olympus). A multi laser engine (iChrome, Toptica) with 405 nm (100 mW), 488 nm (100 mW), 561 nm (100 mW), and 640 nm (70 mW) lasers (powers at fiber output) served as excitation light source.

Figure 2.4 is the schematic representation of the optical configuration. The divergent laser beam exiting the optical fiber was collimated using a 150 mm visible achromatic lens (Thorlabs) and focused into the objective back-focal plane using a 200 mm visible achromatic lens (Thorlabs). The fibre mount and lenses were placed on a linear translation mount (Thorlabs) to adjust the angle of excitation. TIRF excitation occurs at excitation angles exceeding a critical angle defined by the refractive index change at the glass-water interphase of the sample. Excitation and emission light were split using a multi-band dichroic mirror (zt405/488/561/640rpc, Chroma) and fluorescence emission filter (zet405/488/561/640m, Chroma), which were supplied pre-aligned in a TIRF filtercube (91041, Olympus). A 300 mm achromatic tube lens module

(ASI) focused the fluorescence image onto the camera chip at a total magnification of 167x, giving an image pixel size of 96 nm.

Samples were mounted on a motorized XY stage (MS2000 closed-loop with linear encoders, ASI) with a Z-piezo insert (PZ-2150, ASI), operated by a stage and focus controller (ASI). The objective was attached to a linear translation stage (LS50, ASI) for coarse focusing. Optional focus stabilization was provided by an autofocus system with an 890 nm IR beam (CRISP, ASI), coupled into the microscope using a longpass dichroic mirror (T800lpxrxt, Chroma). For bright-field imaging, a transillumination kit with an Olympus IX2 condenser and LED light source (ASI) was used.

Laser excitation and image acquisition were controlled by Toptica iChrome and Andor Solis software, respectively.

Before recording, fields of view are pre-bleached by exposure to 561 nm laser illumination at 0.5 kW cm^{-2} for about 30 seconds so there was on average one molecule in fluorescent state per cell. Andor Solis software was used to record movies and snapshots. I recorded movies of 10,000 frames under continuous 561 nm excitation at 0.5 kW cm^{-2} at a rate of 15.48 ms per frame for MutL and MutH. For MutS, I recorded on average 3000 frames with 15 ms exposures and 85 ms dark image collection, with a combined rate of 100 ms per frame. I also recorded a transmitted light image generated by LED illumination for each field of view for the purpose of cell segmentation.

Custom-written MATLAB software (Mathworks) was used for localization [[Martens et al., 2018](#)], tracking [[Crocker and Grier, 1996](#)] and calculation of diffusion coefficients [[Uphoff et al., 2014](#)]. At the moment of writing, we haven't made the whole matlab code available since it is a bundle of different pieces of code (some of which were developed by other labs) and they have already been pub-

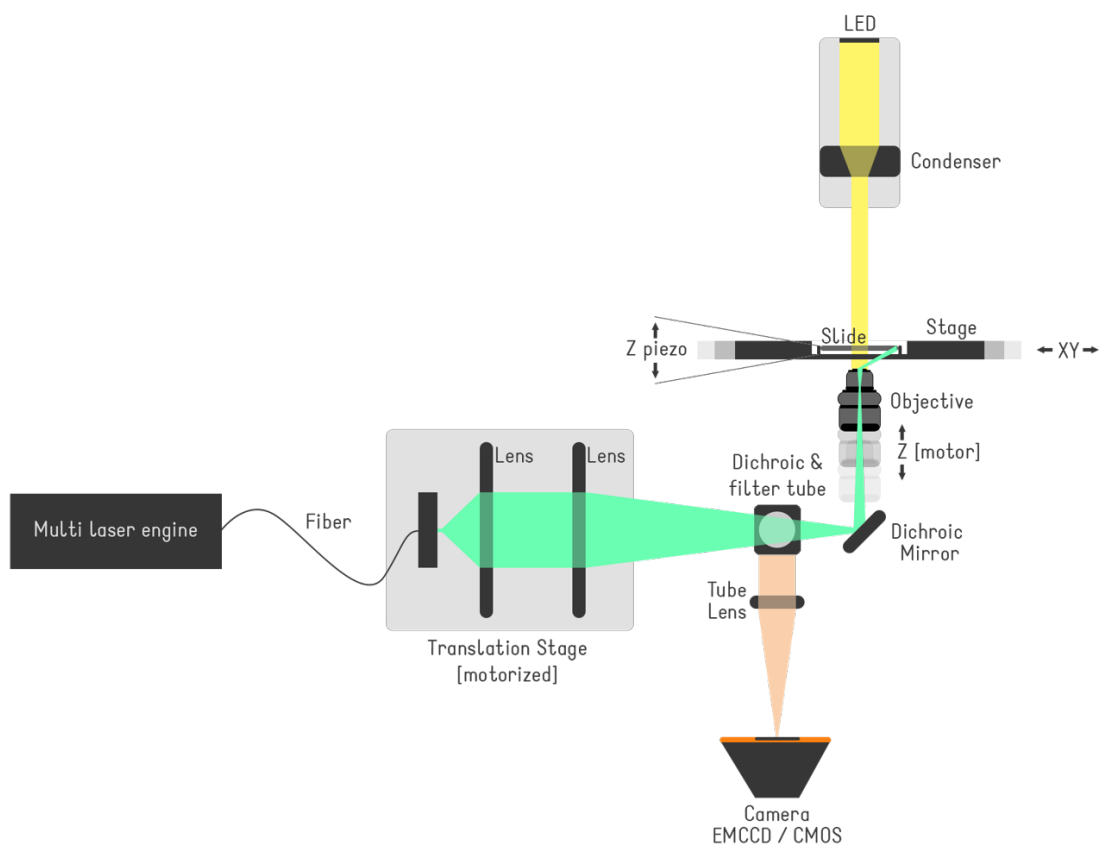


Figure 2.4: Schematic representation of the optical configuration

lished individually.

The code for tracking is published and available here:

<https://site.physics.georgetown.edu/matlab/code.html>

The code for localization analysis is published and available here:

<https://github.com/HohlbeinLab/SMALL-LABS-pSMLM>

When my results will be published we will release the additional code for plotting and fitting histograms, along with the datasets.

Cell outlines were automatically segmented from transmitted light images using a modified version of MicrobeTracker [Sliusarenko et al., 2011] combined with SuperSegger [Stylianidou et al., 2016].

2.10.1 Histogram Fitting

Custom written Matlab software by Matthew Stracy was used in order to estimate the diffusion coefficients of the different species and their relative proportions by fitting the distributions of diffusion coefficients [Stracy et al., 2015]. There is an analytical expression for the distribution of diffusion coefficients when calculated with a fixed number of measured steps. If D denotes the probability density of the diffusion coefficient for the first 4 steps of tracks, diffusion can be modelled by equation 2.3 (where D is the diffusion coefficient and X is the individual molecules).

$$f_D^*(x; D) = \frac{(4/D)^4 x^3 e^{-4x/D}}{6} \quad (2.3)$$

With the same logic multiple species can be fit with different diffusion coefficients (D_1 , D_2 , etc.). Their relative proportions can be estimated as A_1 , A_2 , etc..

Therefore for two species, the combined distribution is modelled by:

$$f_D^*(x; D_1, D_2, A_1, A_2) = \frac{A_1 (4/D_1)^4 x^3 e^{-4x/D_1}}{6} + \frac{A_2 (4/D_2)^4 x^3 e^{-4x/D_2}}{6} \quad (2.4a)$$

$$A_1 + A_2 = 1 \quad (2.4b)$$

2.11 Fluorescence Recovery after Photobleaching (FRAP)

FRAP examines the kinetics of protein binding, diffusion, and transport. To achieve this a small region of fluorescently-labelled molecules in a single cell is photobleached by a high-intensity laser pulse. Afterwards, the return of the fluorescence intensity to a steady state is monitored. The kinetics are elucidated by fitting models that describe the replenishment of fluorescent proteins to the bleached area.

If rapid recovery of fluorescence is observed this signals high protein mobility the rate of which depends on the diffusion and binding kinetics of the molecule being observed while incomplete or slow recovery means proteins bound within or outside the bleached region.

FRAP imaging was done using spinning disk confocal microscope (PerkinElmer UltraVIEW) with an EMCCD (Hamamatsu) camera mounted on an Olympus IX8 microscope with an Olympus 60 1.35 N.A objective. The microscope and the stage was heated to 37 °C. The images were acquired using Volocity software (Perkin Elmer).

Foci FRAP is performed by acquiring 10 prebleach frames at 1 s intervals, and then bleaching a circle 1.2 µm in diameter using the 568 nm laser at 100% power. For MutL, post-bleaching images were acquired at 2 s intervals for 3

min. In most experiments images after 1 min were discarded because they were too bleached to measure. Images were acquired using the 568 nm laser at 300 ms exposure at 15% laser power. For MutS, post-bleaching images were acquired at 1 s intervals for 45 seconds. Images were acquired using the 568 nm laser at 250 ms exposure at 15% laser power. Experiments were performed in at least three biological replicates over 20 foci per session.

To analyse FRAP measurements, first intensity measurements were made using ImageJ. In each cell, fluorescence intensity in the bleached region and a corresponding unbleached region was measured. Afterwards, the intensity of a background region outside of cells was measured (camera background signal) and subtracted from both measurements. The relative intensity of bleached and unbleached regions was calculated by dividing background-corrected unbleached intensity by background-corrected bleached intensity. Then all values are normalised using the pre-bleach measurements achieving that the mean of the relative pre-bleach intensity was 1. The mean of the normalised relative intensity of all repeats are calculated and plotted with error bars indicating standard error of the mean.

2.12 Statistical Analysis

Statistical analysis of results from Spontaneous Mutation Frequency Assay and SMT experiments were performed using student's t-test. Student's t-test is a statistical tool used to compare the means of two samples. P values below 0.05 were considered significant and were denoted 1 star "*". P values below 0.01 were flagged with 2 stars "**". P valued below 0.001 were marked with 3 stars "***".

3 | Addressing the *in vivo* Mobility of MMR Proteins

How proteins, specifically how mismatch repair proteins, find their targets among a vast sea of nonspecific DNA, is a burning question yet to be elucidated. To address this question, among many, and to better understand the *in vivo* mobility of mismatch repair proteins; I first studied their unperturbed behaviour by single molecule tracking microscopy. This method allows us to shine light on the molecular interactions within the cell in real time by examining the movement patterns of molecules. Previously, single-molecule imaging *in vivo* was limited to slowly diffusing membrane proteins tagged with fluorescent markers [Deich et al., 2004] [Sako et al., 2000]. Over the last decade, fluorescence detection sensitivity has improved by the discovery of brighter dyes in combination with better imaging modalities to allow for tracking the movement of single molecules in the cytoplasm.

The diffusion coefficient distribution of a molecule is affected by its size, viscosity of the medium and the temperature [Stracy et al., 2021]. Within the cell under normal circumstances the viscosity of the cytoplasm is homeostatically regulated, therefore the primary parameters that affect the diffusion coefficients of DNA binding proteins were thought to be the size of the protein and its interactions with DNA. Single molecule tracking can distinguish between diffusing molecules and DNA bound molecules. Recently, [Stracy et al., 2021] showed that the intracellular mobility of DNA binding proteins during their target search depends mainly on its interactions with the DNA, rather than the protein's size.

Fluorescent imaging of MMR foci can give limited insight into the dynamics

of MMR. Mismatches are rare, so only a small proportion of cells would contain a foci, therefore the dynamics in majority of cells can only be elucidated with single molecule approaches. For example, the foci that are visible by ensemble fluorescence imaging account for only 10% of cellular MutS in *B.subtilis* [Lenhart et al., 2013]. Until now, only MutS mobility in the cell was measured with single molecule tracking using PAmCherry as a fluorophore [Liao et al., 2015] [Uphoff et al., 2016] but the dynamics of MutL and MutH were yet to be elucidated. All 3 proteins look for different targets. MutS is searching the DNA for distortions which indicate mismatches. It is not known whether MutL is searching for a protein (specifically DNA-bound MutS) or for mismatches directly. It is also unclear whether MutH is looking for hemimethylated GATC sites or DNA-bound MutS-MutL complexes. Therefore, tracking the motion of MMR proteins *in vivo* is crucial to reveal how they function in live cells.

3.1 Halo-tagged Mismatch Proteins are Functional

In order to study the behaviour of mismatch proteins within single cells I obtained Halo tagged versions of MMR proteins that were previously created by Stephan Uphoff. He replaced the endogenous genes with a C-terminal fusion to the Halo-tag using 33 nucleotide linker followed by a kanamycin cassette. Halo tag is a monomeric 33 kDa protein that forms a covalent bond with a synthetic dye (a chloroalkane linker of a ligand) [Banaz et al., 2018]. This reaction is rapid and irreversible under physiological intracellular conditions [Los et al., 2008].

I performed spontaneous mutation frequency assay (explained further in methods) to assess the functionality of Halo-tagged proteins (Figure 3.1). If Halo tag fusions affected the function of MMR proteins, this would increase the frequency of spontaneous mutations genome-wide. This genome-wide increase

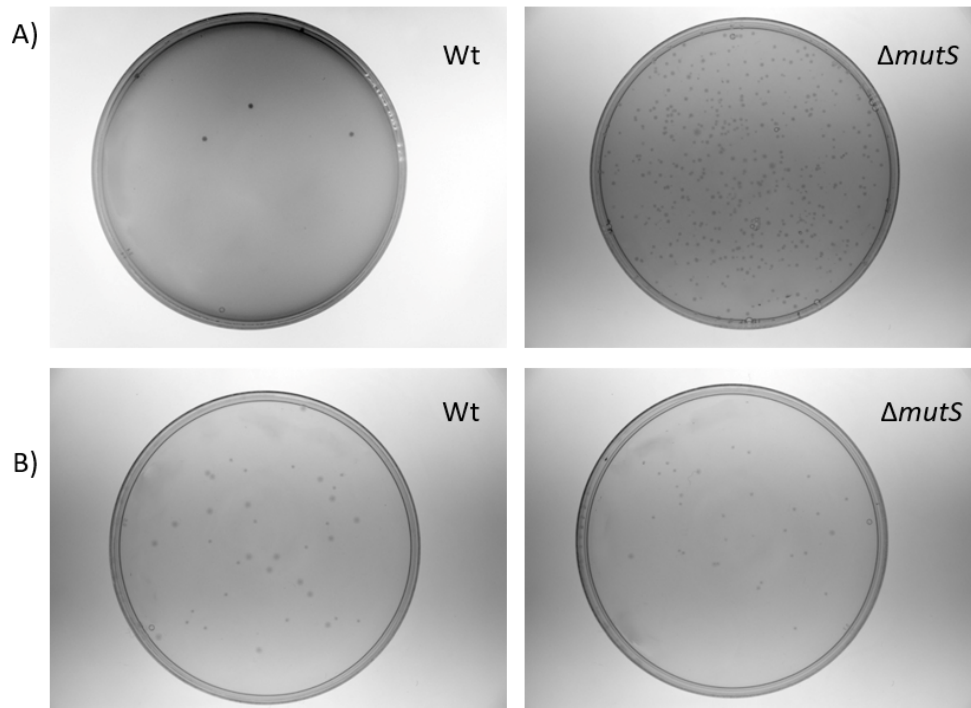


Figure 3.1: Example plates from the spontaneous mutation frequency assay

A) Colonies on rifampicin plates for Wild-type and $\Delta mutS$. Only 3 colonies are visible on the Wild-Type(Wt) rifampicin plate. Deletion of MutS increases the mutation frequency from $0.037 \cdot 10^{-6} \pm 0.012$ to $12.7 \cdot 10^{-6} \pm 2.1$ (343 fold). B) Colonies on LB plates for Wild-type and $\Delta mutS$. As a control, 10^{-6} dilution of the amount plated on the rifampicin plates is plated on the LB plates. The colony number on the LB plates is used to normalize the rifampicin plate colony count.

would lead to an increase of spontaneous resistance to the antibiotic rifampicin.

Wild type cells had a frequency of rifampicin resistance mutations of $0.037 \cdot 10^{-6} \pm 0.012$ (SEM) per cell and Halo tag fusions of MutS, MutL and MutH had mutation frequencies in the same order of magnitude as wild type: $0.031 \cdot 10^{-6} \pm 0.007$ (SEM), $0.048 \cdot 10^{-6} \pm 0.006$ (SEM), $0.028 \cdot 10^{-6} \pm 0.01$ (SEM) respectively as shown in figure 3.2. Moreover deletion of MutS, MutL, MutH resulted in mutation frequencies of $12.7 \cdot 10^{-6} \pm 2.1$ (SEM), $12.9 \cdot 10^{-6} \pm 1.9$ (SEM), $10.8 \cdot 10^{-6} \pm 1.8$ (SEM) respectively. This drastic increase (200 to 300 fold) in mutation frequencies after deletion of MMR is in line with what is described in the literature [Glickman and Radman, 1980]. These results together suggest that tagging MMR proteins with Halo tag does not affect their functionality. Tagging MMR proteins with mYPet also did not result in substantial increase in mutation rates, although caused a slight increase in MutS-mYPet with mutation rates $0.084 \cdot 10^{-6} \pm 0.013$ (SEM) ($p=0.043$), where MutL-mYPet and MutH-mYPet had $0.04 \cdot 10^{-6} \pm 0.012$ (SEM) and $0.054 \cdot 10^{-6} \pm 0.011$ (SEM) respective mutation rates. Tagging the β -sliding clamp DnaN with YPet in MutH-Halo background also did not increase mutation rates significantly, with a rate of $0.04 \cdot 10^{-6} \pm 0.008$ (SEM) allowing us to monitor replicative state of cells while tracking MutH using Halo tag. Tagging MutS with photoactivatable mCherry also resulted in mutation rates comparable to wild type with $0.038 \cdot 10^{-6} \pm 0.007$. Deletion of MutL in MutH-Halo background increased mutation rates drastically to $10.1 \cdot 10^{-6} \pm 2.3$ as expected.

An alternative to the Spontaneous Mutation Frequency Assay I performed could have been measuring the mutation rate using a fluctuation assay. Fluctuation assay would have provided the estimate of the mutation rate per locus per generation. As my mutants have a barely affected growth rate but a very large increase of the mutation frequency, I think in my case measuring mutation frequency is probably good enough.

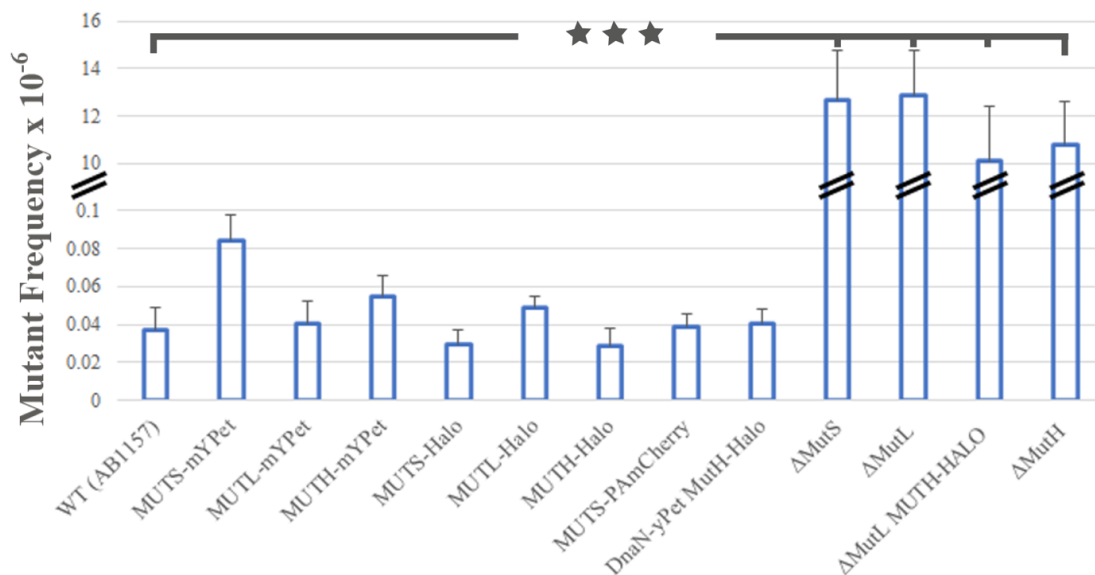


Figure 3.2: Halo-tagged mismatch proteins are functional

Mutation frequencies of strains expressing Halo-tagged mismatch repair proteins, compared to wild-type strain AB1157, and strains with deletions of mismatch repair genes. Wild type cells had the frequency of rifampicin resistance mutations of $0.037 \cdot 10^{-6} \pm 0.012$ (SEM) per cell and Halo tag fusions of MutS, MutL and MutH had mutation frequencies in the same order of magnitude as wild type. Moreover deletion of MutS, MutL, MutH resulted in mutation frequencies of $12.7 \cdot 10^{-6} \pm 2.1$ (SEM), $12.9 \cdot 10^{-6} \pm 1.9$ (SEM), $10.8 \cdot 10^{-6} \pm 1.8$ (SEM) respectively.

Overall, my results suggest that Halo-tagged mismatch proteins are functional and behave as wild type.

3.2 Distribution of Diffusion Coefficients of MMR Proteins in Wild Type Cells

Next, I measured the diffusion coefficients of MMR proteins in live *E. coli* cells. The mobility of tracked MMR molecules provided me with a measure of their activity. DNA bound molecules appear immobile as they are moving very slowly

due to the motion of the DNA therefore this population is called "bound". Diffusing molecules have larger diffusion coefficients and show displacements within successive frames and are called "mobile". One can quantify the differences in mobility by calculating apparent diffusion coefficients (D^*) (Figure 3.3).

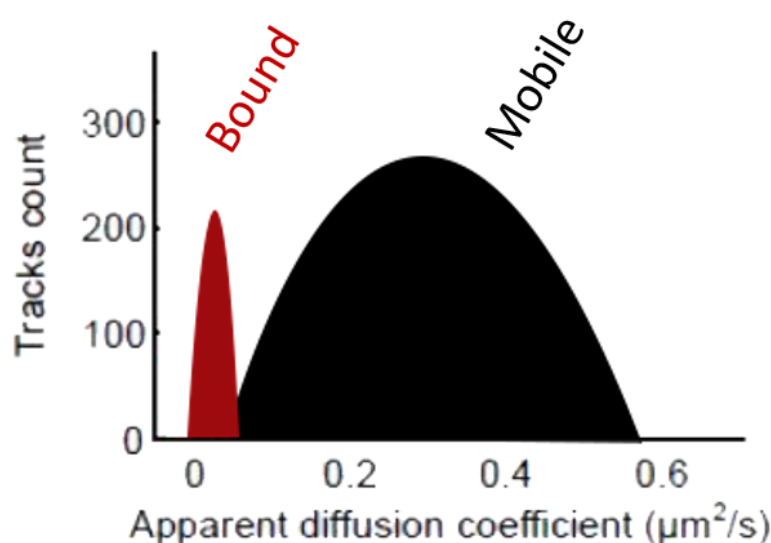


Figure 3.3: Schematic representation of distribution of diffusion coefficients exemplifying bound and mobile molecules

The proportion in red exemplifies the DNA bound molecules whereas the proportion in black are the freely diffusing mobile molecules.

I combined histograms of the apparent diffusion coefficient from 3 independent experiments with Muts-Halo. I observed two distinct populations of MutS, with the majority of MutS molecules slowly diffusing with a mean diffusion coefficient of $0.22 \mu\text{m}^2 \text{s}^{-1}$ with 8.85% of MutS molecules appearing immobile (Figure 3.4). This is much slower than what I measured with Poll, a protein of similar size (125 kDA for MutS+Halo-tag, 139 kDA for Poll+Halo-tag) which has a mean diffusion coefficient of $0.94 \mu\text{m}^2 \text{s}^{-1}$ and 95% of the population mobile [Banaz et al., 2018].

I believe MutS is best represented by two populations, however it is possible

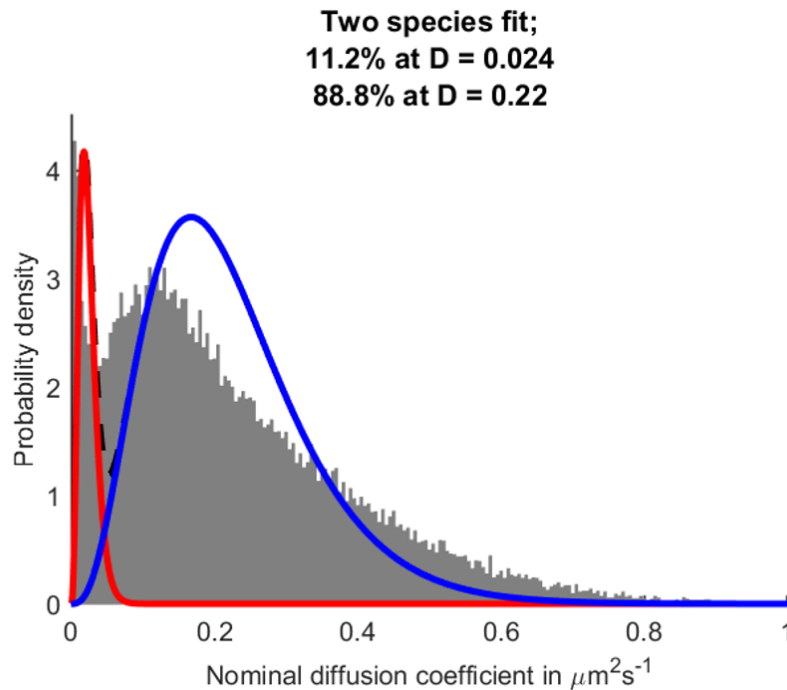


Figure 3.4: Distribution of diffusion coefficients for MutS-Halo

Combined histogram of the distribution of diffusion coefficients of MutS-Halo from 3 independent sets. Two species fit with red denoting the bound population and blue denoting the mobile population.

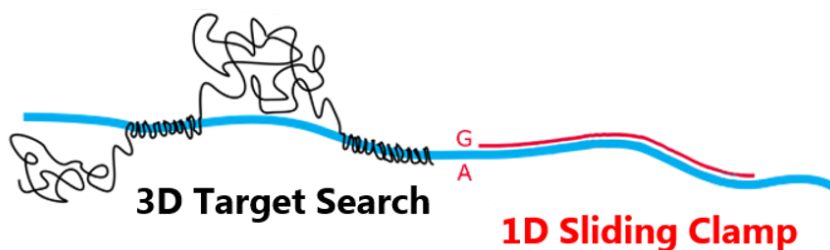


Figure 3.5: Different modes of MutS diffusion

A schematic to illustrate different modes of MutS diffusion. When MutS is searching for mismatches it diffuses in the cytoplasm in 3D. When MutS encounters a mismatch it becomes a stable sliding clamp and diffuses in 1 dimension (which is observed in the bound population).

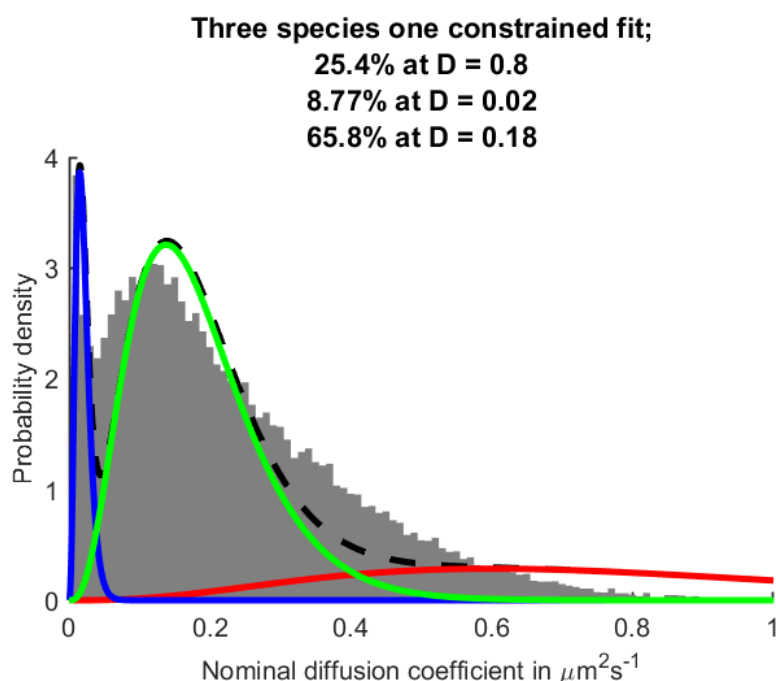


Figure 3.6: Distribution of diffusion coefficients for MutS-Halo with 3 species fit

Combined histogram of the distribution of diffusion coefficients of MutS-Halo from 3 independent sets. Three species fit with blue denoting the bound population and red and green denoting the mobile populations.

two divide the mobile populations into two: A slow species that represent the molecules that spent most of the duration of the track transiently interacting with DNA, and the faster species would represent molecules that spent most of the duration of the track unbound.

The surprisingly low diffusion coefficient could be explained by frequent binding of MutS to DNA. My results suggest that MutS has two different modes of diffusion on the DNA: It engages in a 3D search in the cytoplasm and makes frequent contacts with the DNA to search for mismatches. This frequent contact results in a lower diffusion coefficient for the mobile population. It also occasionally falls off and diffuses in the cytoplasm. Multiple intervals of tran-

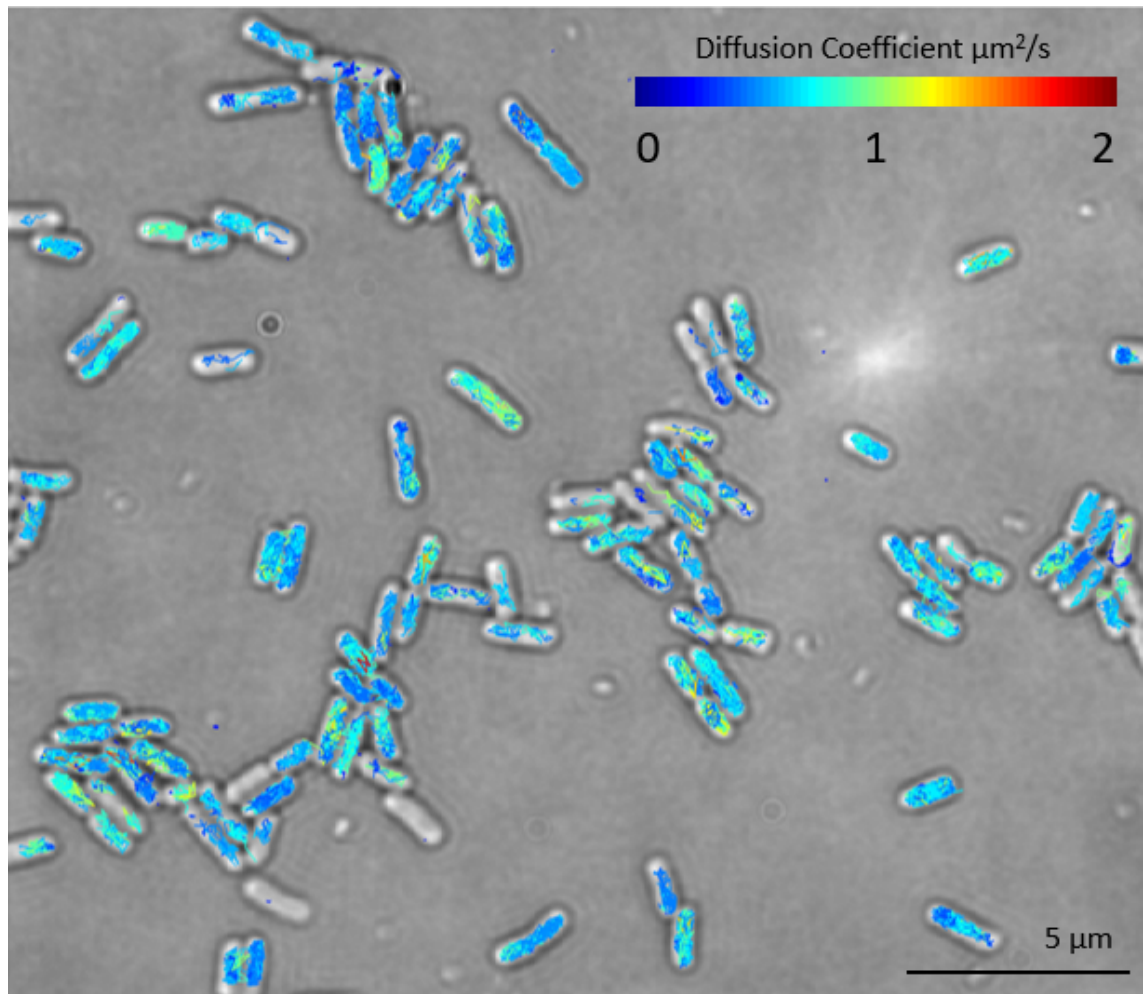


Figure 3.7: MutS-Halo diffusion profile in single cells

Example field of view. Tracks of individual MutS-Halo molecules shown overlaid on a transmitted light image. Track colours correspond to the apparent diffusion coefficient (dark blue = $0 \mu\text{m}^2 \text{s}^{-1}$ to red = $2 \mu\text{m}^2 \text{s}^{-1}$).

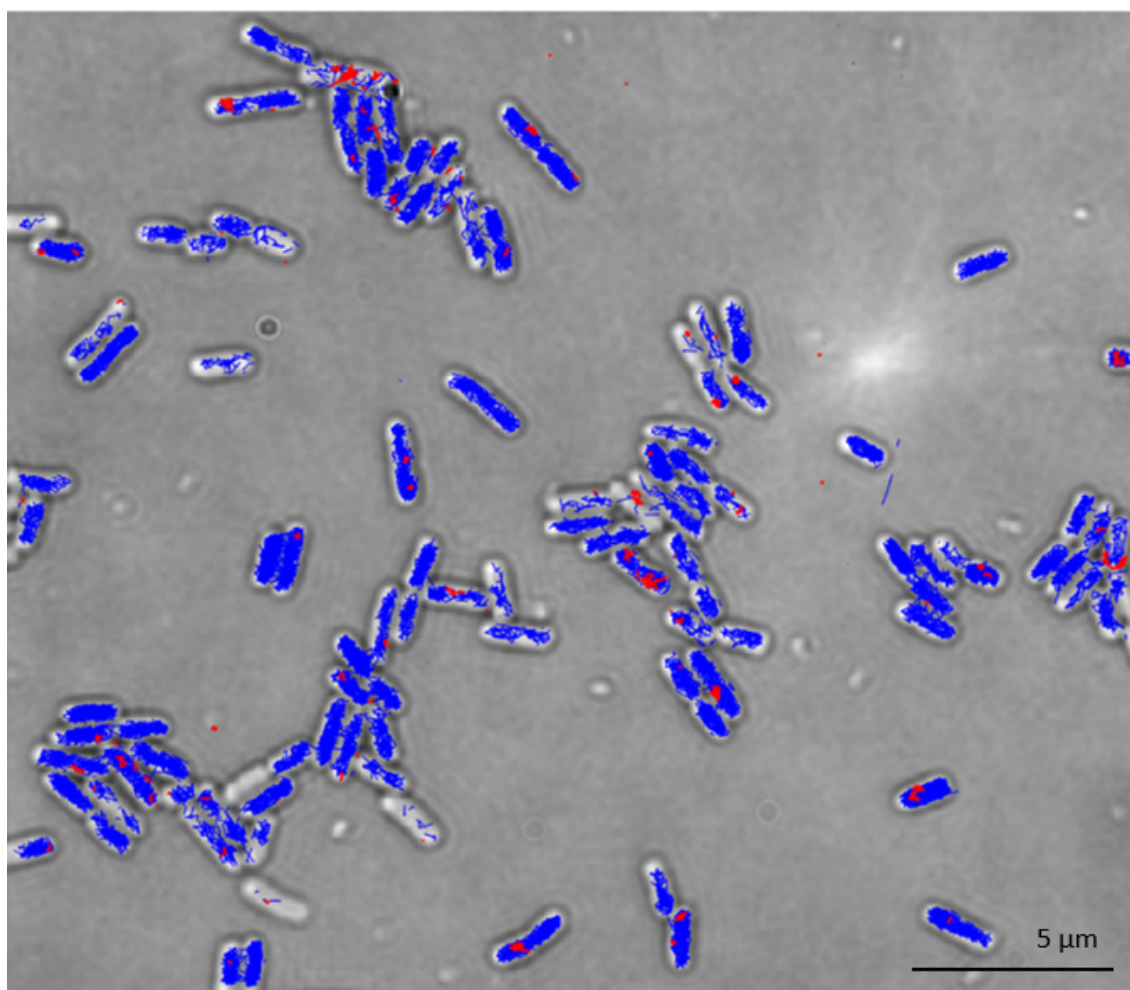


Figure 3.8: MutS-Halo diffusion profile showing bound and mobile molecules

Same data as in Figure 3.7. Tracks of MutS-Halo molecules shown in red (bound molecules) for diffusion coefficient below $0.024 \mu\text{m}^2 \text{s}^{-1}$. Tracks are shown in blue where D is above $0.024 \mu\text{m}^2 \text{s}^{-1}$.

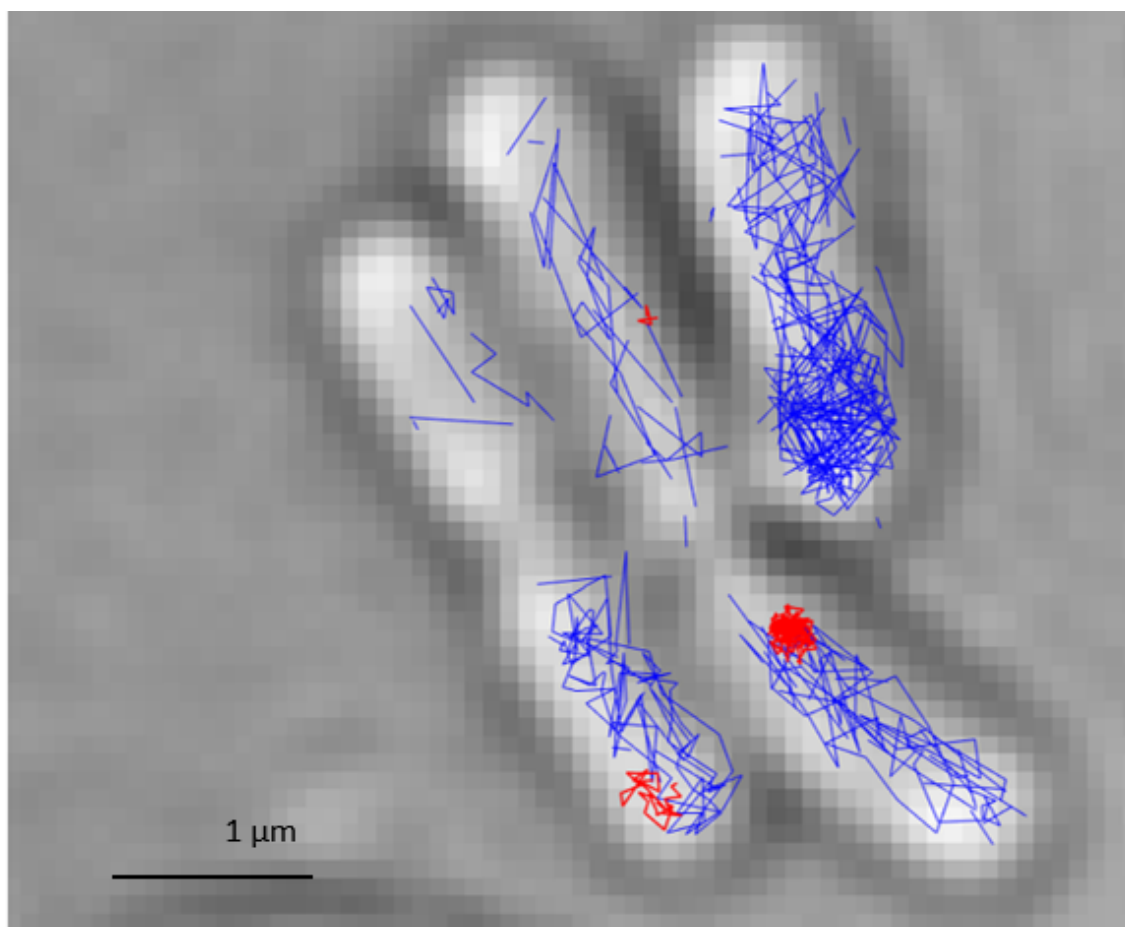


Figure 3.9: Zoomed-in field of view showing MutS-Halo diffusion profile in single cells

Zoomed in view for Figure 3.8. Bound molecules could be seen in red.

sient DNA binding and 3D diffusion could be captured in a track (Figure 3.5) and the MSD of this track would position it in the mobile population. MutS tracks cover the entire cell (Figure 3.7) with only a proportion of cells having bound molecules (Figure 3.8). This observation suggests that there is no direct relationship between MutS binding and the replisome. If that were the case, we would observe two regions of MutS bound molecules (as there are at least two replication forks); however, Figure 3.9 shows only one region or none in each cell.

As known from the literature when MutS encounters a mismatch it goes through a conformational change and becomes a sliding clamp that scans the DNA backbone stably. Our method can not differentiate between a stationary protein on the DNA and one that is sliding along the DNA backbone. Any movement on the DNA would be masked by the observed movement of the DNA itself, making it impossible to differentiate whether the molecule is diffusing one dimensionally on the DNA or whether it is stationary. [Elmore et al., 2005] measured the movement of a point on the DNA to be $4.3 \times 10^{-5} \mu m^2 s^{-1}$ whereas we measured the localization error to be 24 nm for Halo tag [Banaz et al., 2018]. The apparent diffusion coefficient due to this localization error would be $0.04 \mu m^2 s^{-1}$. Therefore, the distances that are covered by the sliding molecules would be small and masked by the localization error.

We quantified immobile and bound populations by fitting distributions with analytical model which assumes Brownian motion for MutL and MutH. This model accurately matched the data whereas for MutS the distribution of the experimental measurements were broader than predicted by the model. This matches our conclusion that MutS motion is frequently interrupted by transient binding to the DNA during target search, such that simple Brownian motion does not completely describe it. The slow diffusion of MutS is consistent

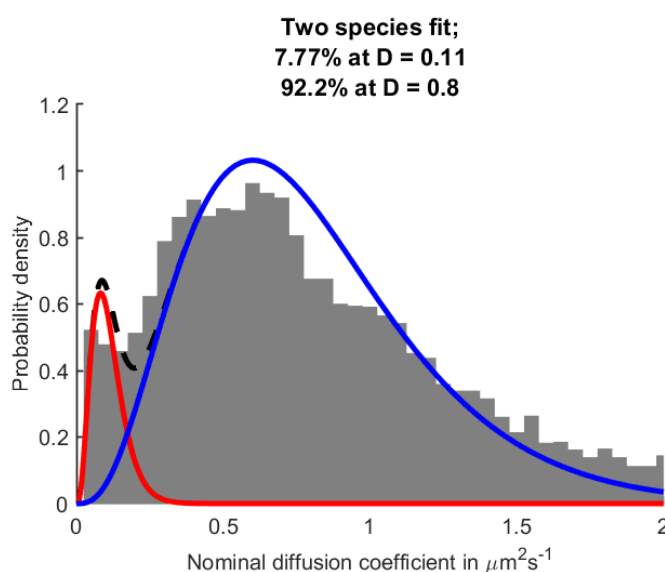


Figure 3.10: Distribution of diffusion coefficients for MutL-Halo

Combined histogram of the distribution of diffusion coefficients of MutL-Halo from 3 independent sets. The histogram could be fit into two species with red denoting the bound population and blue denoting the diffusing population in the cytoplasm.

with [Stracy et al., 2021] that showed that the mobility of DNA-binding proteins is determined by their DNA-interactions rather than their size.

For MutL, I combined histograms from 3 independent sets and observed two distinct populations (Figure 3.10). Interestingly, the mean diffusion coefficient of MutL is higher than MutS with $0.8 \mu\text{m}^2\text{s}^{-1}$ with 7.77% of molecules in an immobile state (Figure 3.11), (Figure 3.12). This is consistent with literature [Liu et al., 2016] where MutL only gets recruited by MutS after mismatch recognition, and acts as a protein matchmaker bringing various components of the MMR pathway together. [Gorman et al., 2012a] observed using *in vitro* single molecule imaging, MutL locating mismatch-bound MutS through 1D hopping and 3D inter-site transfer. MutS-MutL complex is suggested to travel the DNA by rotating 1D diffusion [Gorman et al., 2012a][Liu et al., 2016].

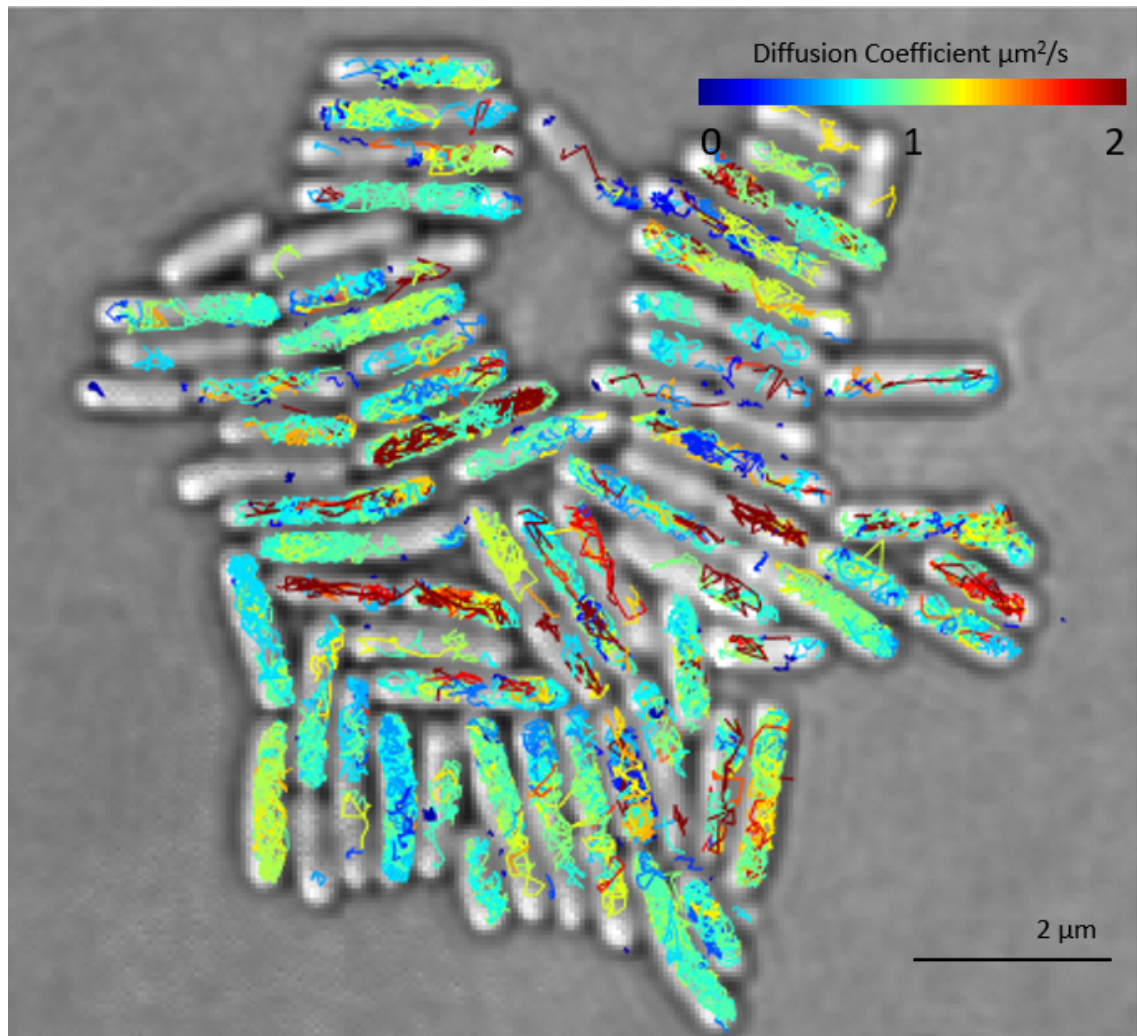


Figure 3.11: MutL-Halo diffusion profile in single cells

Example field of view. Tracks of individual MutL-Halo molecules shown overlaid on a transmitted light image. Track colours correspond to the apparent diffusion coefficient (dark blue = $0 \mu\text{m}^2 \text{s}^{-1}$ to red = $2 \mu\text{m}^2 \text{s}^{-1}$).

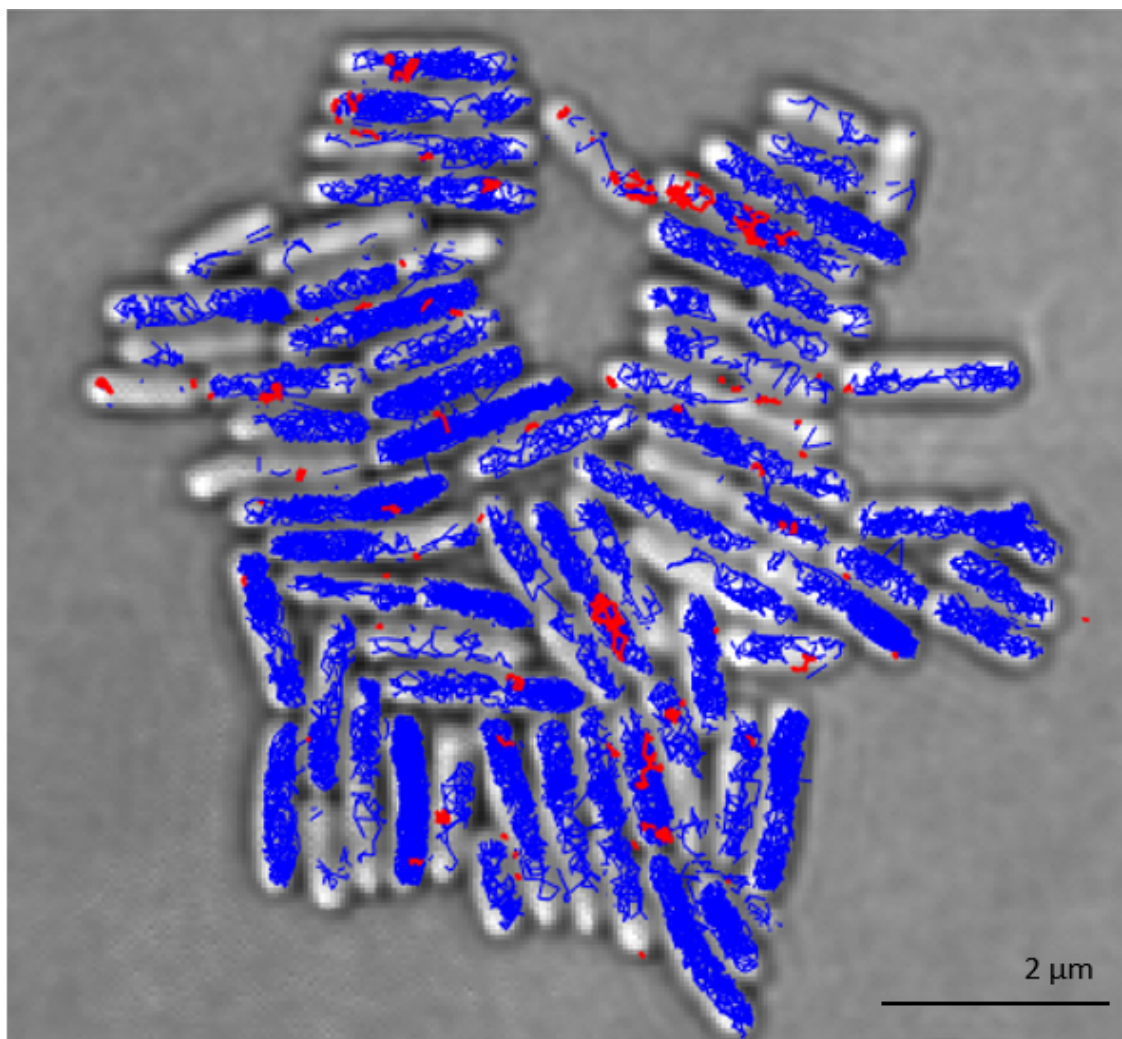


Figure 3.12: MutL-Halo diffusion profile showing bound and mobile molecules

Same data as in Figure 3.10. Tracks of MutL-Halo molecules shown in red (bound molecules) for diffusion coefficient below $0.15 \mu\text{m}^2\text{s}^{-1}$. Tracks are shown in blue where D is above $0.15 \mu\text{m}^2\text{s}^{-1}$.

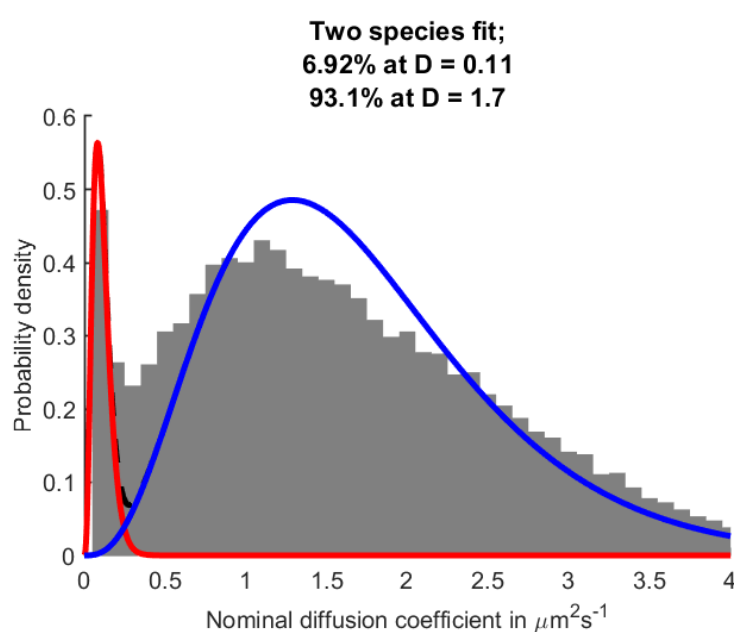


Figure 3.13: Distribution of diffusion coefficients for MutH-Halo

Combined histogram of the distribution of diffusion coefficients of MutH-Halo from 5 independent sets. Two distinct species are visible, with red denoting the bound population and blue denoting the mobile population.

To look at the population dynamics of MutH, I combined histograms from 5 independent sets. To demonstrate the traces are similar and can be combined I added 3 independent repeats to the supplements. As with the other MMR proteins, MutH motion can also be defined by 2 species (an immobile and mobile species) (Figure 3.13). MutH diffuses faster than MutS and MutL with mean diffusion coefficient of $1.7 \mu m^2 s^{-1}$ with 6.92% bound (Figure 3.14), (Figure 3.15). This could be explained by MutH being the smallest of all the three proteins.

3.3 Implications of *in vivo* Mobilities of Target Search Mechanisms

Plotting the diffusion distributions for all 3 MMR proteins together demonstrates that MutS diffuses slower compared to the other two and has a bigger bound population. MutL and MutH has similar bound populations, with MutL diffusing slower than MutH. It is worth noting that MutL diffuses faster than MutS, which is consistent with an hypothesis that it is only recruited to the DNA when MutS recognizes a mismatch and that it spends the rest of the time freely diffusing in the cytoplasm. MutS, on the other hand, is in frequent contact with the DNA which makes its mean diffusion coefficient lower compared to MutL and MutH (Figure 3.16).

It is interesting that the 3 proteins each show different mobilities which demonstrates that although these proteins are expected to form complexes during MMR pathway they search for their targets independently. Any complexes formed between MutS MutL and MutH are likely to be transient. In normally growing cells the majority of the 3 proteins are in mobile state which is consistent with low frequency of DNA replication errors.

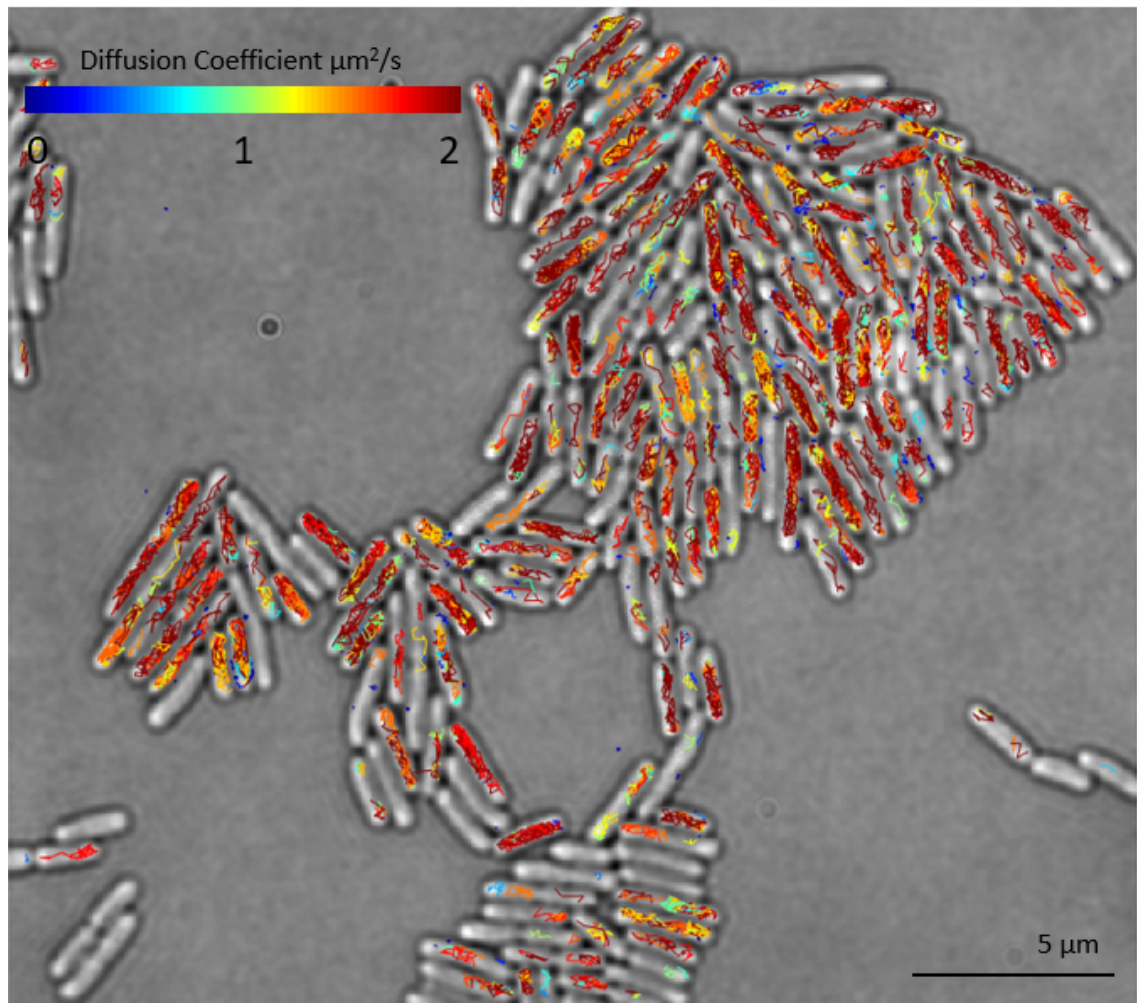


Figure 3.14: MutH-Halo diffusion profile in single cells

Example field of view. Tracks of individual MutH-Halo molecules shown overlaid on a transmitted light image. Track colours correspond to the apparent diffusion coefficient (dark blue = $0 \mu\text{m}^2 \text{s}^{-1}$ to red = $2 \mu\text{m}^2 \text{s}^{-1}$).

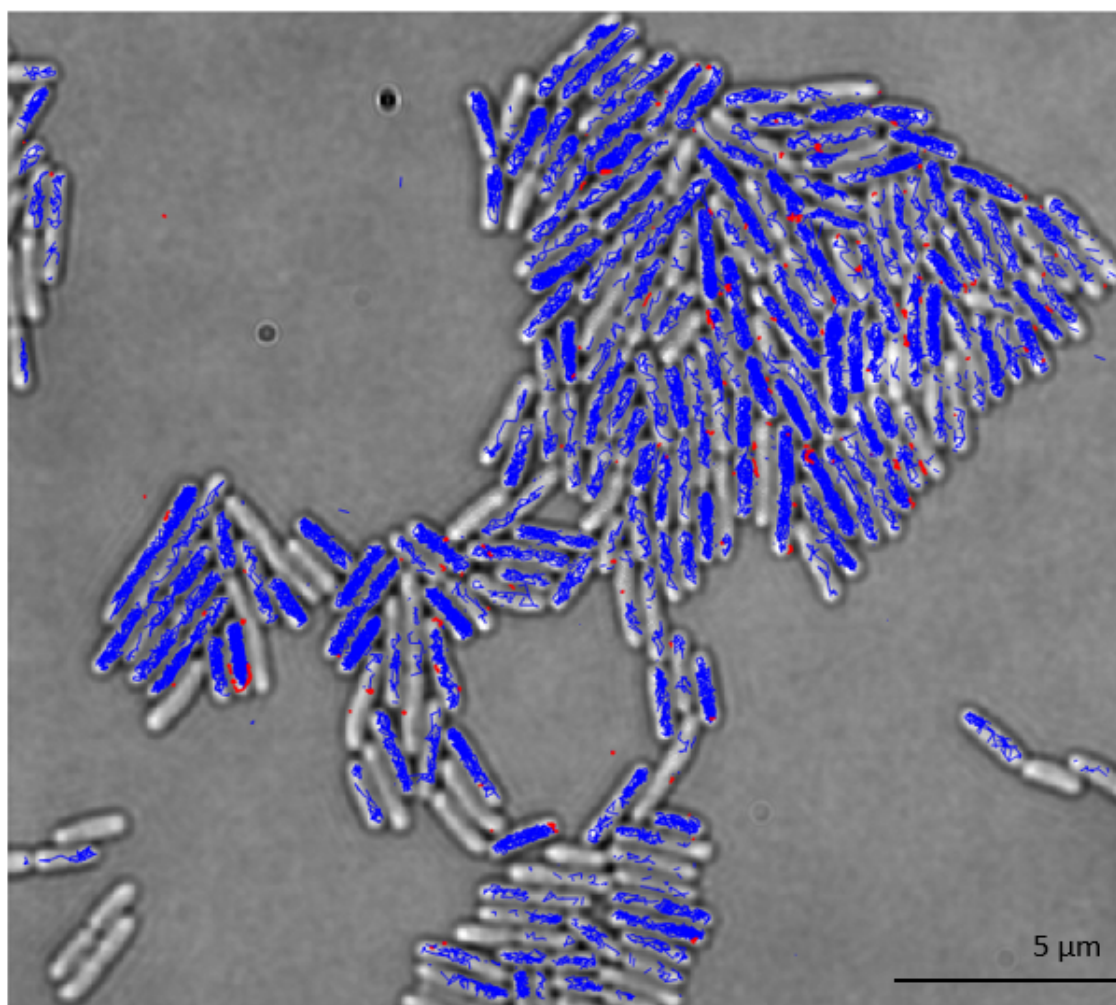


Figure 3.15: MutH-Halo diffusion profile showing bound and mobile molecules

Same data as in Figure 3.12. Tracks of MutH-Halo molecules shown in red (bound molecules) for diffusion coefficient below $0.15 \mu\text{m}^2\text{s}^{-1}$. Tracks are shown in blue where D is above $0.15 \mu\text{m}^2\text{s}^{-1}$.

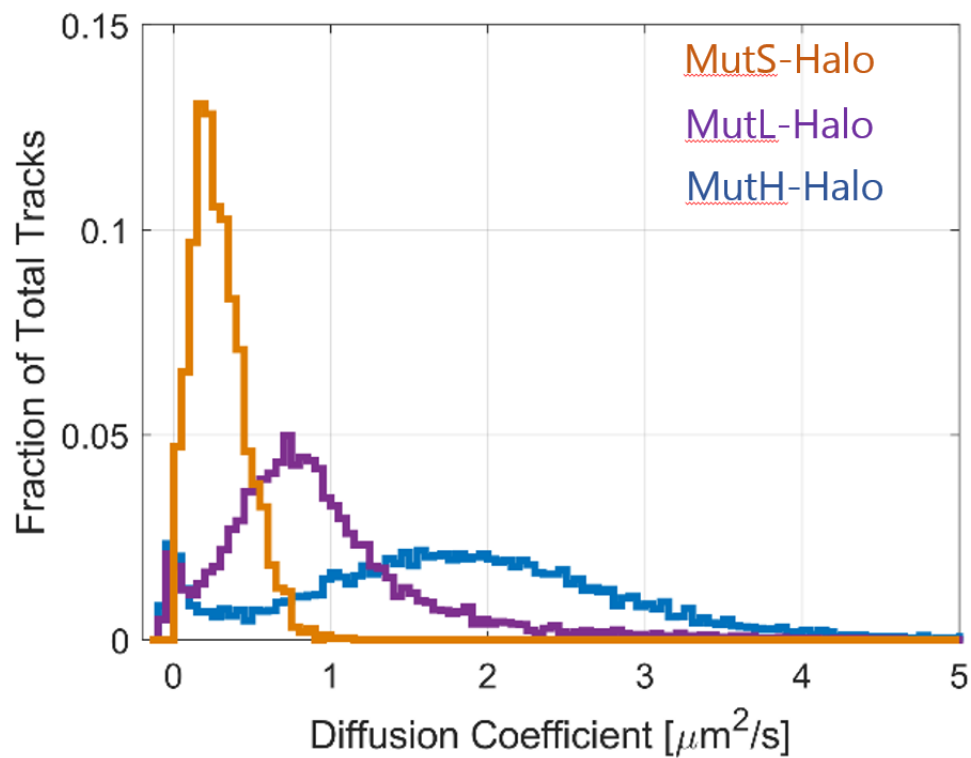


Figure 3.16: Comparison of distribution of diffusion coefficients for MutS, MutL and MutH

The diffusion distributions for all 3 MMR proteins are plotted together. MutS diffuses slower compared to the other two and has a bigger bound population. The majority of the 3 proteins are in mobile state, consistent with low frequency of DNA replication errors.

4 | Investigation of Repair Complexes in MMR

In the previous chapter, I looked at the *in vivo* mobilities of MMR proteins while they were searching for their targets. After MutS recognizes the DNA mismatch, it needs to coordinate with MutL to activate the strand incision by MutH. In this chapter I investigate their coordination after a mismatch is found. At the center of the chapter are the following two questions:

- 1) How is MutH recruited to find its substrate (GATC sites)? Is the recruitment reliant on mismatch recognition by MutS and MutL or does MutH find its substrate independently?
- 2) How is the mismatch recognition by the MutS-MutL complex communicated to trigger strand incision?

To answer these questions, I applied three different live cell approaches:

First, using single molecule tracking, I measured how immobilization of proteins is affected by changes in the frequency of DNA mismatches.

Second, using conventional epifluorescence microscopy, I imaged MutS, MutL and MutH and assessed whether they are forming foci.

Third, using FRAP, I looked at the turnover of MMR proteins at mismatches.

4.1 Distribution of Diffusion Coefficients of MMR Proteins after DNA Damaging Treatment

The model proposed by [Liu et al., 2016] suggests that MutH is recruited to hemimethylated GATC sites by MutS and MutL complex. In order to test this hypothesis I induced DNA damage by using methyl methanesulfonate (MMS)

and observed the effect on the DNA bound populations of MutS, MutL and MutH. MMS is an alkylating agent and a carcinogen. It causes mutagenic DNA lesions which leads to polymerases creating mismatches during DNA replication (specifically O^6 -meG-T) [Sedgwick, 2004]. Methyl methanesulfonate is a methanesulfonate ester resulting from the formal condensation of methanesulfonic acid with methanol. Methanol is known to be toxic to bacterial cells but only at high concentrations.

Instead of using MMS, one could use 2-Aminopurine (2-AP), a purine analog of guanine and adenine. It can pair with thymine as an adenine-analogue, but can also pair with cytosine as a guanine-analogue therefore causes mutations. In our system it did not cause an observable increase in mutation rates therefore I opted to use MMS.

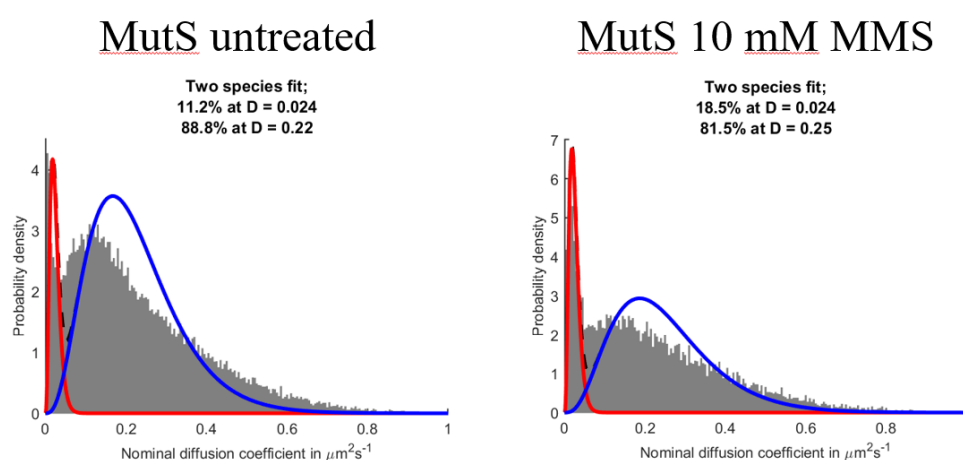


Figure 4.1: Distribution of diffusion coefficients for MutS-Halo untreated vs treated with 10 mM MMS

Combined histogram of the distribution of diffusion coefficients for MutS-Halo from 3 independent sets. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

After treating the cells with 10mM MMS for 15 minutes I observed a significant

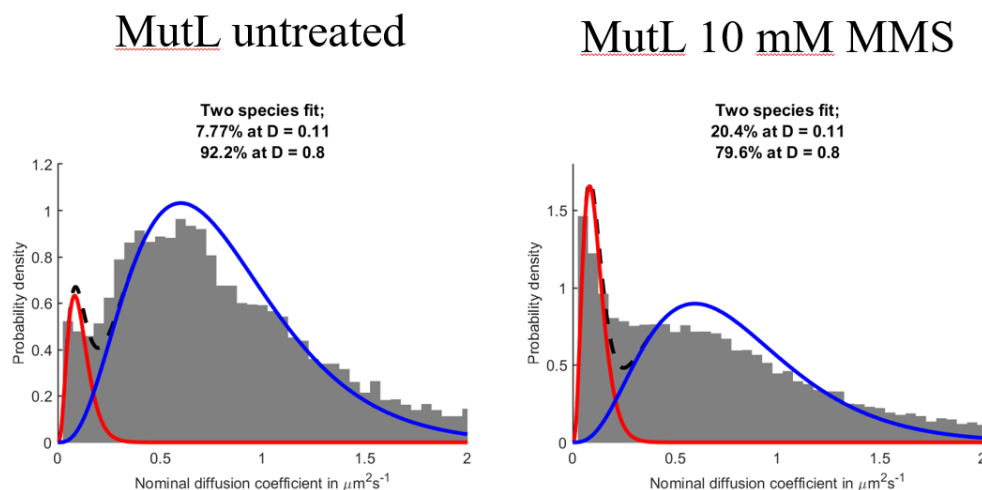


Figure 4.2: Distribution of diffusion coefficients for MutL-Halo untreated vs treated with 10 mM MMS

Combined histogram of the distribution of diffusion coefficients for MutL-Halo from 3 independent sets. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

increase in the immobile population from 11.2% to 18.5% for MutS ($p=0.0044$) (Figure 4.1) and a significant increase from 7.7% to 20.7% for MutL ($p=0.0004$) (Figure 4.2). This increase in the bound population is consistent with MutS and MutL becoming immobilised at DNA mismatches. However MutH bound population remained unchanged with 6.7% before DNA damage treatment and 6.4% after MMS treatment ($p=0.4672$) (Figure 4.3).

If MutH was reliant on MutS and MutL to find the hemi-methylated GATC sites then I would expect an increase in the bound population of MutH after mutagen treatment as the bound populations of MutS and MutL increased. However, I observed no significant change of MutH bound population (Figure 4.4). These results suggest MutH's search for its substrate is independent of the presence of DNA mismatches. In other words, whereas MutS and MutL are recruited to

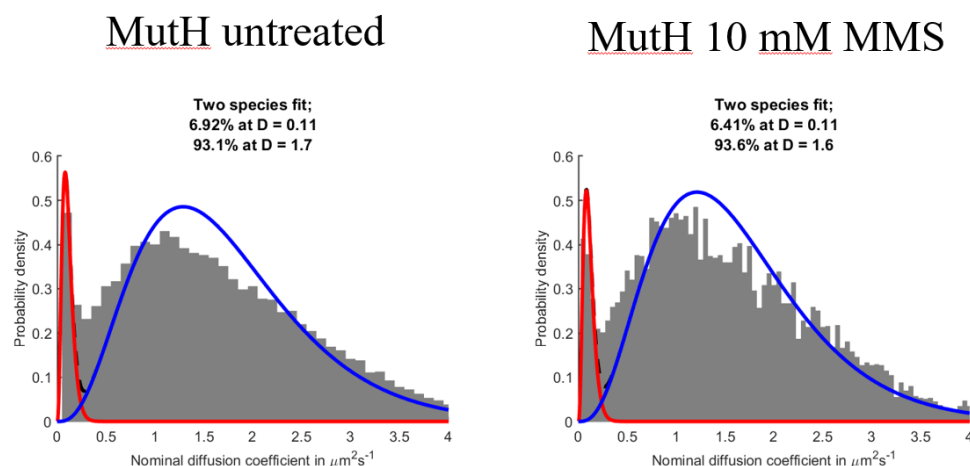


Figure 4.3: Distribution of diffusion coefficients for MutH-Halo untreated vs treated with 10 mM MMS

Combined histogram of the distribution of diffusion coefficients for MutH-Halo from 3 independent sets. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

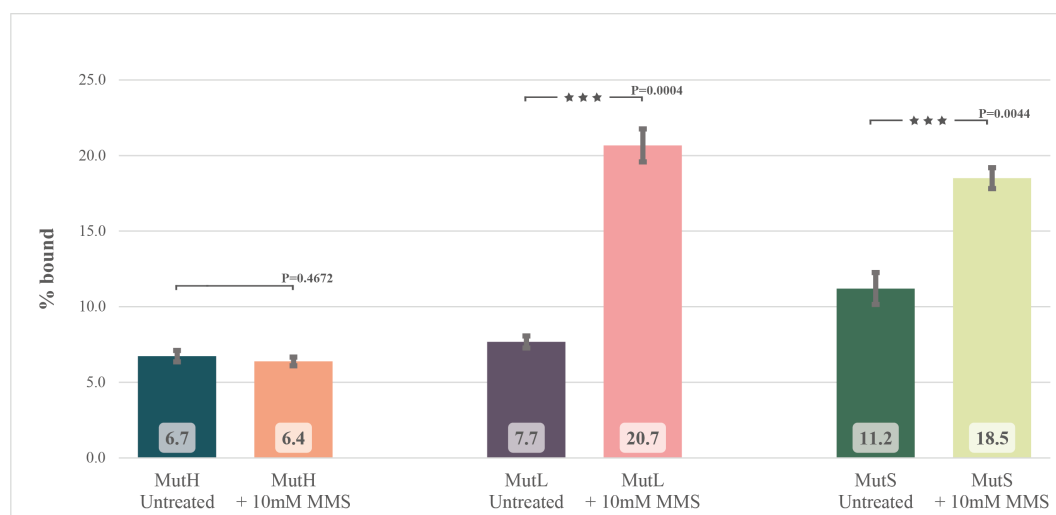


Figure 4.4: Bar graphs comparing the bound populations of MutS, MutL and MutH, untreated vs treated with 10 mM MMS

Bound populations of MutS and MutL increased after mutagen treatment. However, there is no significant change of MutH bound population. Data points represent mean $\pm$ SEM.

DNA in response to DNA mismatch, MutH shows a constant fraction of DNA-bound molecules. This result further indicates that MutH binding is independent of MutS and MutL and I will explore this point further in the following chapter.

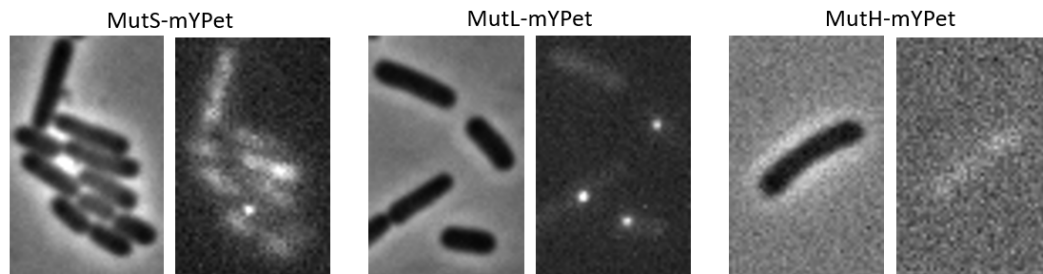


Figure 4.5: MutS and MutL form foci whereas MutH does not
Epifluorescence imaging of untreated MutS-mYPet, MutL-mYPet and MutH-mYPet with Nikon Ti Eclipse inverted fluorescence microscope. MutS and MutL foci formation can be observed however no MutH foci were visible.

4.2 Epifluorescence Imaging of MutS-mYPet, MutL-mYPet and MutH-mYPet

In the previous section, I showed with single molecule tracking that MutS and MutL bound fractions increase upon mutagen treatment. Previous studies have shown that MutS and MutL form foci [Elez et al., 2010], [Elez et al., 2012]. This suggests that there is accumulation of multiple MutS and MutL proteins at a mismatch repair site. MutH has not been imaged in cells to date therefore it was unclear how it is spatially distributed.

I obtained *E. coli* strains expressing MMR proteins labelled with mYPet from Stephan Uphoff and imaged them under a fluorescent microscope (Figure 4.6, Figure 4.7, Figure 4.8). I observed foci formation for MutS-mYPet and MutL-mYPet, suggesting there are multiple proteins at the same spot, however could not see any foci formed for MutH. MutH showed an even distribution across the cytoplasm which is consistent with low fraction of bound molecules I reported in the previous chapter. This finding also challenged previous views in the lit-

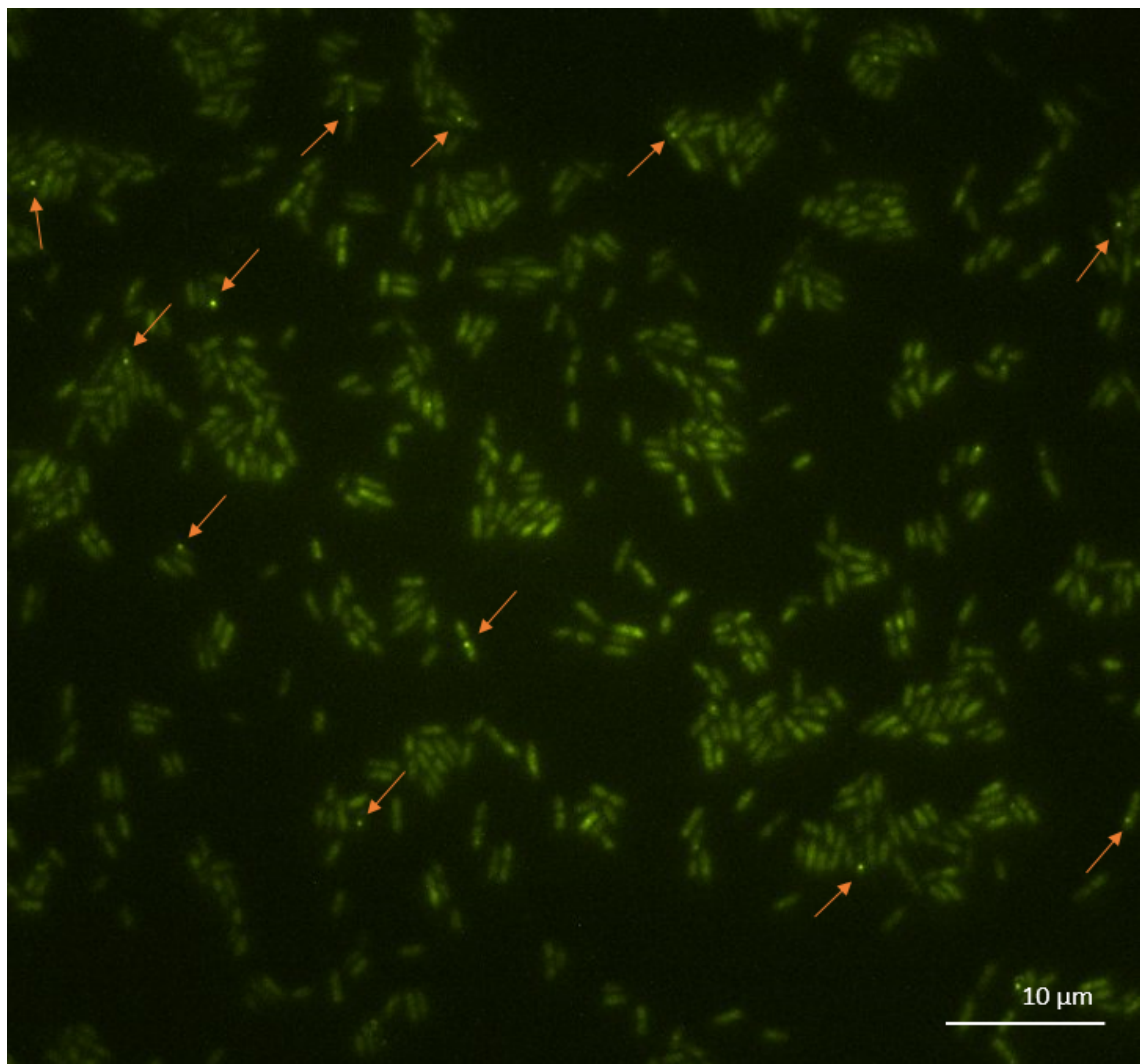


Figure 4.6: MutS-mYPet forms foci

Epifluorescence imaging of untreated MutS-mYPet with Nikon Ti Eclipse inverted fluorescence microscope. MutS foci formation can be observed.

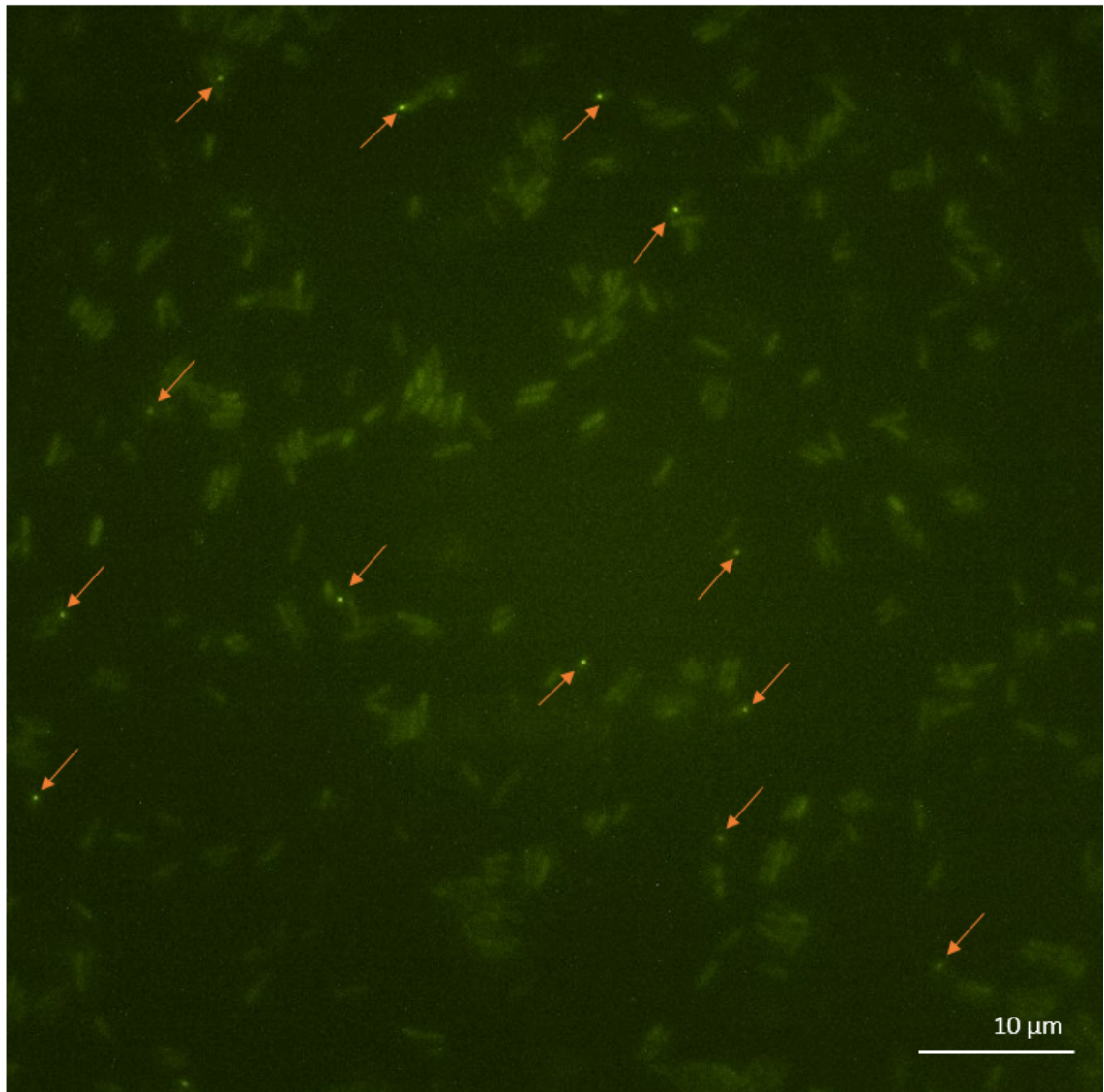


Figure 4.7: MutL-mYPet forms foci

Epifluorescence imaging of untreated MutL-mYPet with Nikon Ti Eclipse inverted fluorescence microscope. MutL foci formation can be observed.

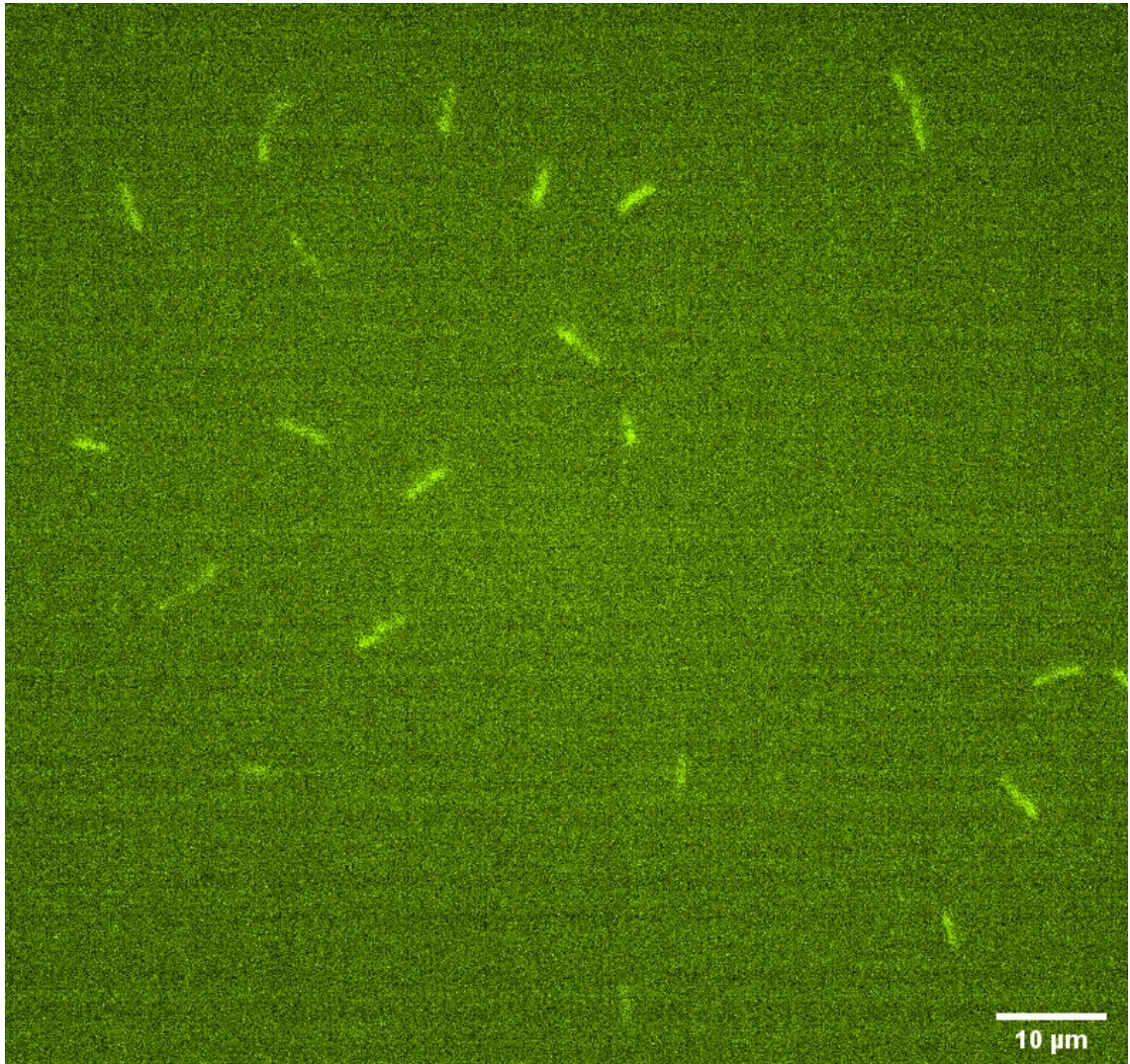


Figure 4.8: MutH-mYPet does not form foci

Epifluorescence imaging of untreated MutH-mYPet with Nikon Ti Eclipse inverted fluorescence microscope. No MutH foci were visible.

erature, If MutS, MutL and MutH formed sliding clamps and moved along the DNA together [Liu et al., 2016], we might have seen MutH foci colocalizing with MutS and MutL foci. Therefore this observation is also consistent with independent binding of MutH to hemimethylated GATC sites. Previously, [Elez et al., 2012] showed MutS and MutL foci colocalizing, while [Uphoff, 2018], [Uphoff et al., 2016] demonstrated MutS and MutL foci frequency increases with mutagen treatment. This is consistent with my result from the previous section that MutS and MutL bound population increases with mutagen treatment.

This experiment could also be performed using Halo-tag. Halo-tag can also be used for epifluorescence and produces images comparable to other fluorescent tags, as we demonstrated in our paper, with MukB-mCherry and MukB-Halo-TMR [Banaz et al., 2018], as well as for quantification of very low-abundant proteins in bacteria [Lepore et al., 2019].

4.3 FRAP

In the previous section, I showed that MutS and MutL form foci. It is, however, unclear whether these foci represent stable protein ensembles that persist from mismatch recognition until the repair is completed. It is also unclear whether there is a turnover of MutS and MutL in the foci. This latter question is inherently related to mechanism of MMR pathway coordination. Looping models require stable complexes to hold the loop together; therefore, low turnover rates should be expected. For looping, whether actively or randomly, it is essential to have a more stable association of MutS and MutL molecules with the DNA at the base of the loop. Sliding models, however, are consistent with high turnover rates, since the sliding away of MutS-MutL complexes that are loaded at mismatches enables loading new molecules.

The turnover of MMR proteins has never been measured *in vivo* to date. To measure the kinetics of protein binding and diffusion, I used FRAP. To perform FRAP, a small region of fluorescently labelled molecules in a single cell is photo-bleached by a high-intensity laser pulse. Afterwards, the return of the fluorescence intensity to a steady state is monitored. Therefore, FRAP could be used to understand the dynamics of MutS and MutL at a mismatch.

Performing FRAP analysis was quite challenging for multiple reasons:

- 1) MutS and MutL foci are rare (present in approximately 1% of cells) leading to low measurement output, which made it difficult to get sufficient statistics.
- 2) Low number of molecules per foci means stochastic recovery kinetics and photobleaching during recovery measurements, which causes low signal-to-noise.
- 3) Foci move and diffuse in and out of focus during recovery measurements.
- 4) Mismatches can be repaired while observing fluorescence recovery, which could lead to the disappearance of foci or incomplete recovery.

To overcome these points, I first scanned a field of view with confocal microscopy, and took snapshots to identify the foci that I will bleach. During analysis, I manually tracked the foci position to ensure all the signal was captured.

To overcome the last point, I performed FRAP analysis in $\Delta mutH$ background, where mismatches can not be repaired. In this strain, MutS and MutL foci form, but they cannot complete repair due to the lack of strand incision. MutS and MutL turnover at such unrepairable mismatches likely still reflects the communication of MutS and MutL in wild type cells, before they contact MutH. I presume that the search of MutS and MutL for MutH would be the same even if the search target is absent.

Figure 4.9 shows the FRAP recovery curve for MutS, normalized to the pre-bleach fluorescence. 3 images were taken before the foci were bleached. It can be observed that the fluorescence intensity decreased by 60% after bleaching and recovered to 80 % in as fast as 5 seconds. This is a striking finding, which suggests that MutS falls off the DNA quite rapidly (high off-rate). The incomplete recovery could be due to the mixture of bleached and unbleached molecules in foci.

Figure 4.10 shows the FRAP recovery curve for MutL, normalized to the pre-bleach fluorescence. It can be observed that MutL also recovers however, the recovery of MutL takes longer (around 30 seconds). This suggests MutL stays on the DNA after MutS disengages. These observations lead to two important conclusions: First, these results are more consistent with a sliding model than with a looping model. Second, this difference in turnover rates challenges the hypothesis that MutS and MutL form a mismatch repair complex with MutH that scans the DNA for GATC sites for MutH to incise.

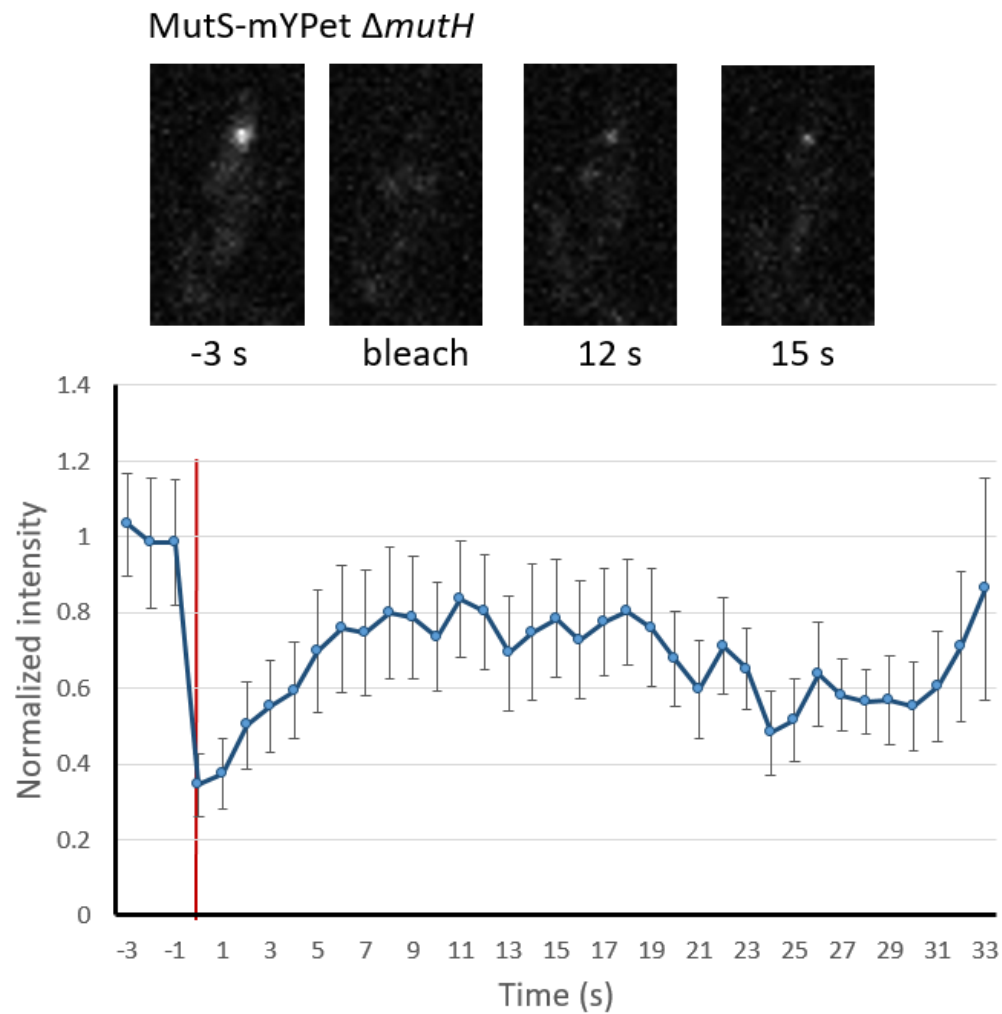


Figure 4.9: Fluorescence recovery after photobleaching graph for MutS
Data is normalized to original fluorescence, prior to the bleach, combined from 43 cells. Data points represent mean $\pm$ SEM.

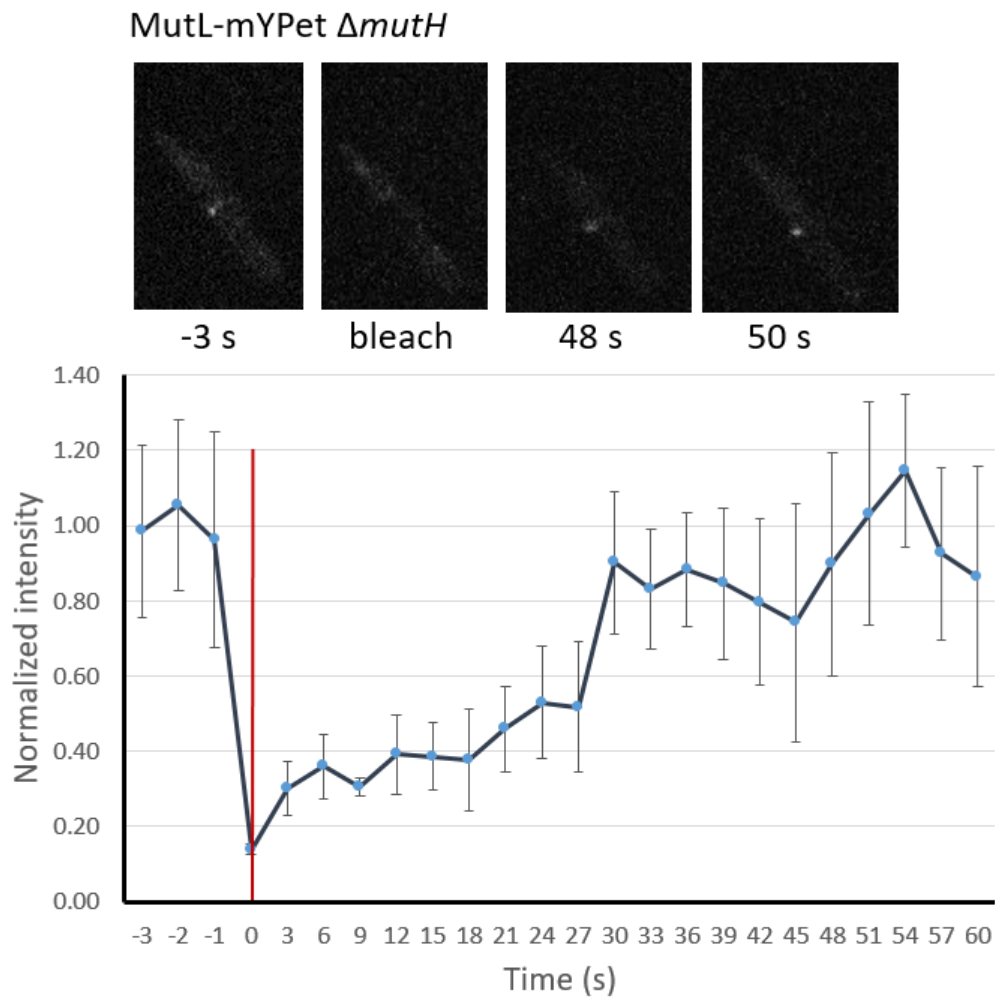


Figure 4.10: Fluorescence recovery after photobleaching graph for MutL
Data is normalized to original fluorescence, prior to the bleach, combined from 37 cells. Data points represent mean $\pm$ SEM.

5 | Genetic Perturbations of the MMR System to Investigate the Mechanism of MutH Recruitment

Chapter 3 and 4 have indicated that the search of MutH for its substrate is independent of the presence of DNA mismatches. It was also evident that its search for hemimethylated GATC sites is not guided by MutS and MutL complexes. In this chapter, I will test these two indications directly.

5.1 Distribution of Diffusion Coefficients of MutH after MutL Deletion

The results of the MMS treatment suggest that MutH is not recruited in response to DNA damage. To test directly if mismatch recognition by MutS-MutL complex is necessary for MutH recruitment during MMR, I deleted MutL in cells where MutH is tagged with Halo and measured the bound population. If mismatch recognition by MutS-MutL complex was indeed necessary for MutH recruitment, deleting MutL should abolish the MutH bound population. However, there was no significant decrease in the bound population of MutH (5.7% vs 6.7% in wild type) ($p=0.1582$) (Figure 5.1), (Figure 5.2). Furthermore, the MutL deletion did not affect the average diffusion coefficient of the mobile population of MutH. This shows that MutH does not bind to MutL in the diffusing state either. The above demonstrates, then, that MutH is not reliant on MutL to find the hemimethylated GATC sites.

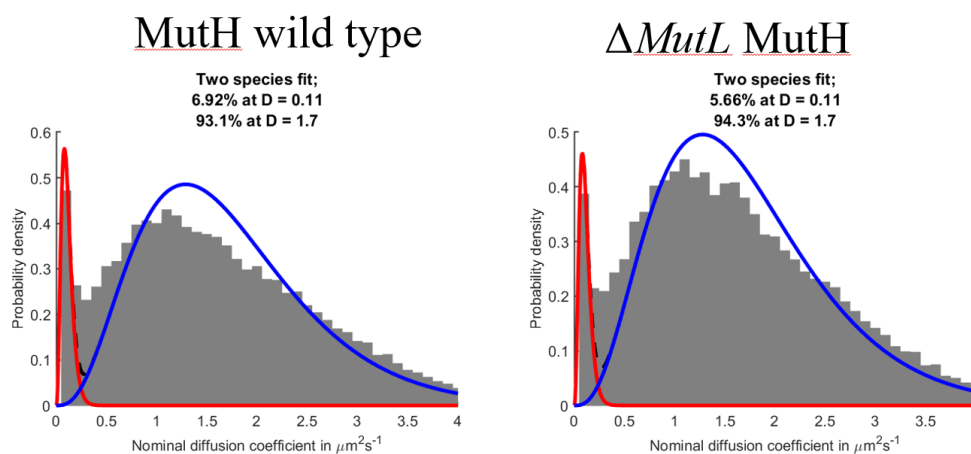


Figure 5.1: Distribution of diffusion coefficients for MutH-Halo wild type and MutH-Halo $\Delta MutL$

Combined histogram of the distribution of diffusion coefficients for MutH-Halo from 3 independent sets. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

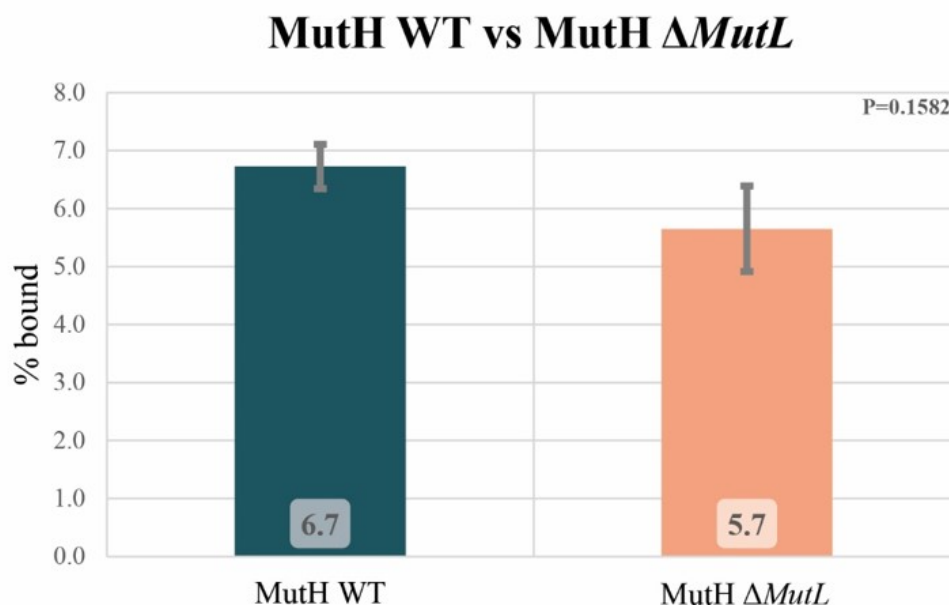


Figure 5.2: Bar graphs comparing the bound populations of MutH-Halo wild type and MutH-Halo $\Delta MutL$

There was no significant change in the bound population of MutH (5.7% vs 6.7% in wild type) ($p=0.1582$). Data points represent mean $\pm$ SEM.

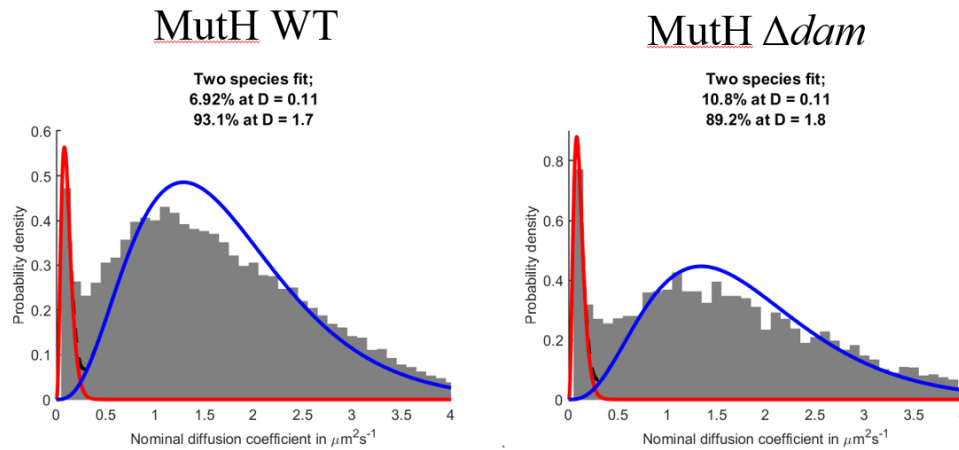


Figure 5.3: Distribution of diffusion coefficients for MutH-Halo and Δdam
Combined histogram of the distribution of diffusion coefficients for MutH-Halo from 3 independent sets. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

5.2 Distribution of Diffusion Coefficients of MutH after Perturbation of the Amount of Dam

The fact that MutH binding relies neither on MutL nor on the presence of DNA mismatches may indicate that it binds to its substrate independently. To address this model, I looked at the binding of MutH when the amount of its competitors is perturbed. MutH competes with SeqA and Dam for hemimethylated GATC sites. If MutH binding is independent of MutS and MutL, I wondered whether the rate limiting factor in MutH binding could be the availability of substrate. I first tested what would happen to the bound population of MutH in *dam* knockout bacteria (Figure 5.3). The results showed a significant increase in Δdam , from 6.7% in wild type to 10.8% ($p=0.0058$) (Figure 5.4).

As perturbations of *dam* affect DNA replication control, I included an internal

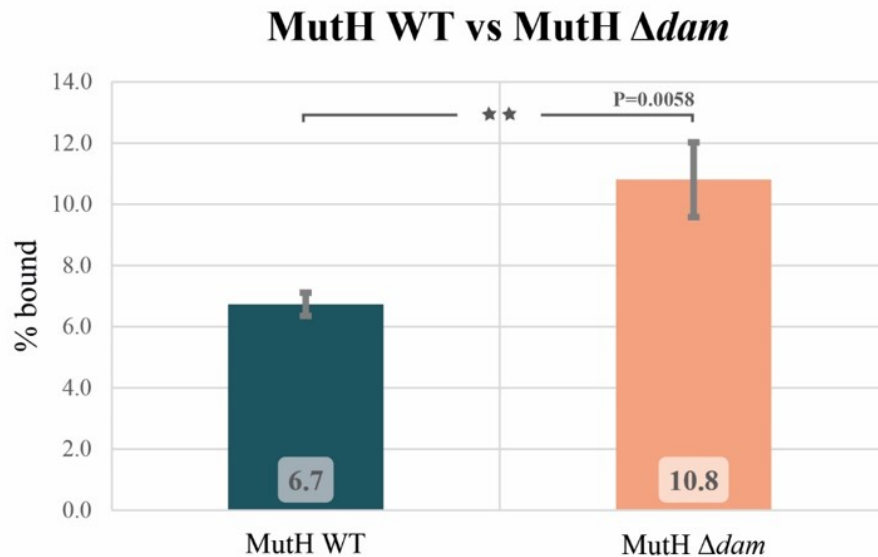


Figure 5.4: Bar graphs comparing the bound populations of MutH-Halo wild type and Δdam MutH-Halo

Bound population of MutH increases significantly in *dam* knockout bacteria (from 6.7% in wild type to 10.8% Δdam) ($p=0.0058$). Data points represent mean $\pm$ SEM.

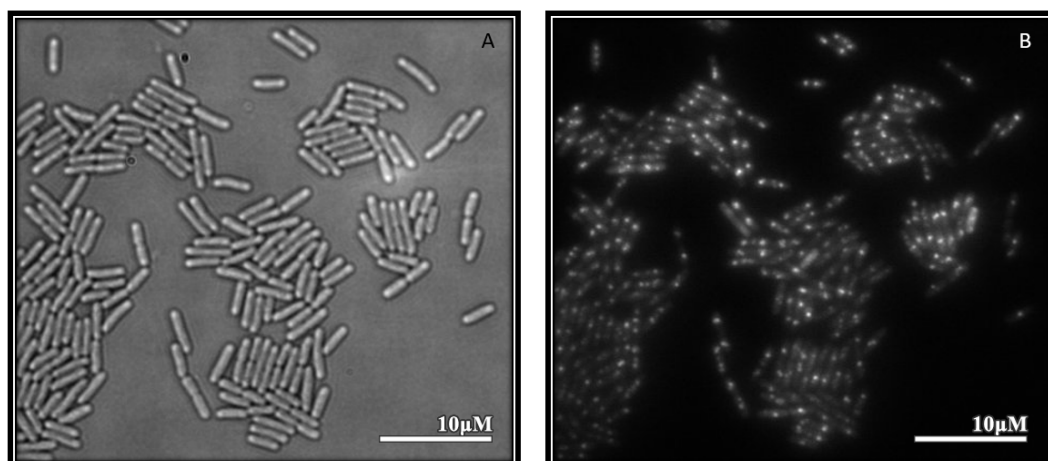


Figure 5.5: Brightfield and fluorescent images of MutH-Halo Δdam DnaN-mYPet

A) Brightfield image of *E. coli* cells B) Fluorescent imaging of DnaN-mYPet that functions as a label on the β -clamp and forms foci at the active DNA replication forks. On average 2 foci per cell is visible. Cells with no foci were removed from analysis.

control to confirm the cells are actively replicating during the single molecule tracking measurements. To this end, I added a DnaN-mYPet allele which functions as a label on the β -clamp that forms foci at the active DNA replication forks [Reyes-Lamothe et al., 2010] [Moolman et al., 2014] (Figure 5.5).

When *dam* is deleted, all GATC sites in the genome become unmethylated. Therefore all GATC sites become a substrate for MutH, as MutH has some affinity for unmethylated GATC sites. An equilibrium ultracentrifugation assay by [Lee et al., 2005] yielded a K_d of 3.5 μ M between MutH and unmethylated GATC sites and 0.9 μ M with hemimethylated GATC sites. A K_d value of 46 μ M is measured for fully methylated sites. The 4 fold decrease in the K_d is offset by 19.120 available unmethylated GATC sites. This affinity of MutH for unmethylated sites explains the increase in the DNA bound population of MutH. My observation is consistent with a model where MutH finds its substrate independent of other MMR components.

5.3 Overexpression of the Dam Protein

To further test whether MutH finds its substrate independent of other MMR components, I overexpressed Dam and quantified the bound population of MutH. To this end, I used the pBAD system which is based on the *araBAD* operon that controls *E. coli* L-arabinose metabolism [Guzman et al., 1995]. I cloned *dam* into the pBAD vector downstream of the *araBAD* promoter, which drives expression of the gene of interest in response to L-arabinose, and is inhibited by glucose. Then, I grew the cells in M9 glucose medium and switched them to M9 supplemented with L-arabinose during recovery phase of Halo tag labelling. I confirmed overexpression by running an SDS-PAGE gel for total proteins (Figure 5.6).

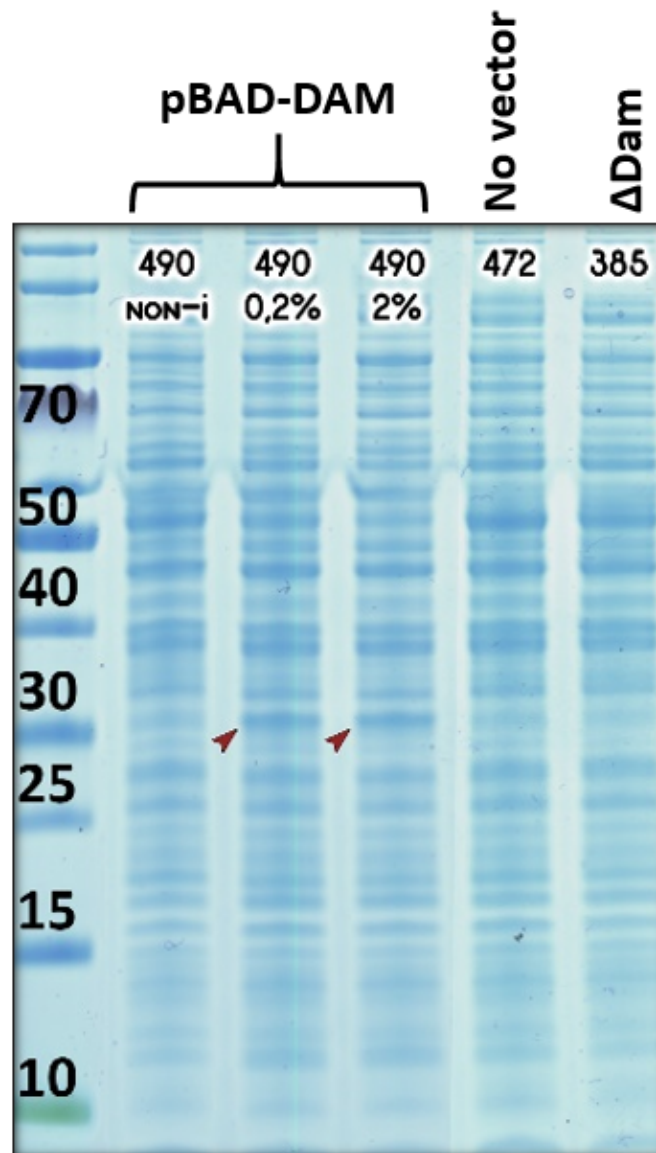
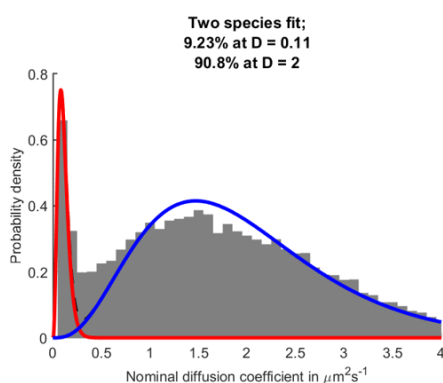


Figure 5.6: SDS-PAGE gel showing overexpression of Dam

SDS-Page gel showing an increase in total protein amounts around 32kDA (the molecular weight of Dam) for strain 490 with 0.2% and 2% arabinose. There is no increase for 490 without arabinose, for strain 472 or strain 385. Strain 472 is AB1157 DnaN-Ypet MutH-Halo::Kan and Strain 490 is 472 transformed with pBAD(Dam). Strain 385 is a Dam knock-out strain.

MutH wild type in arabinose



MutH pBAD(dam)

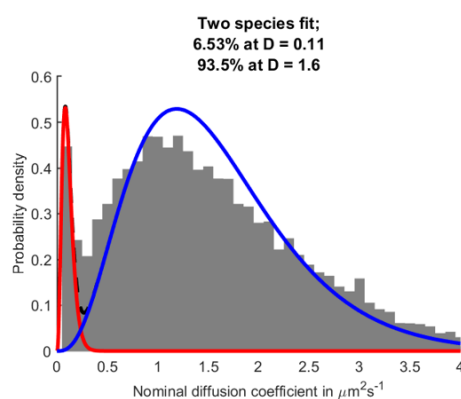


Figure 5.7: Distribution of diffusion coefficients for MutH-Halo with empty pBAD vector vs Dam overexpression

Combined histogram of the distribution of diffusion coefficients from 3 independent sets, Cells were recovered in arabinose for 15 minutes. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

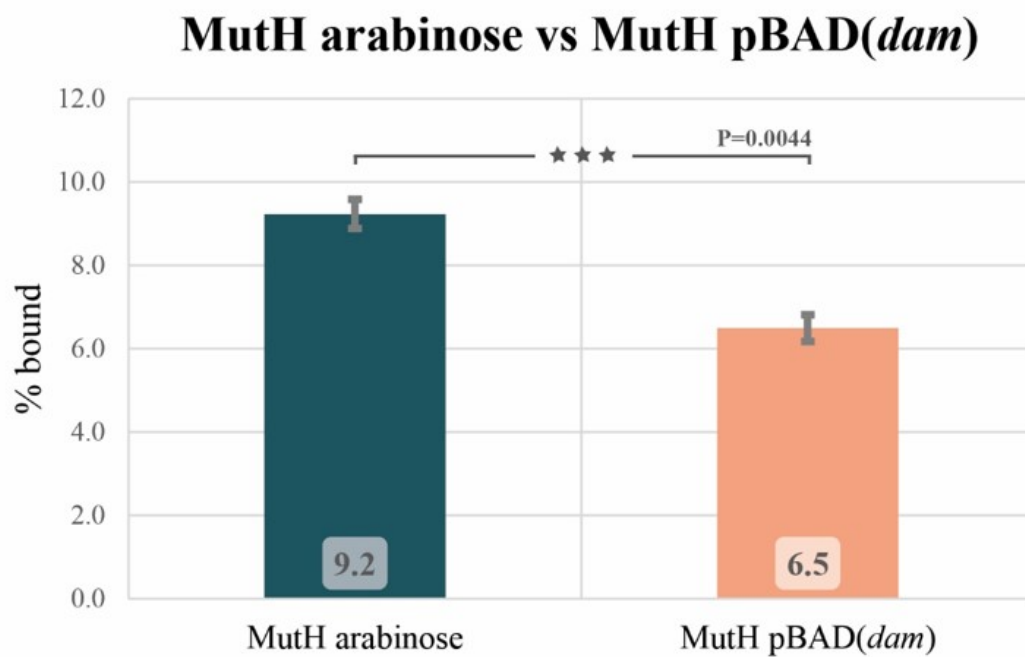


Figure 5.8: Bar graphs comparing the bound populations of MutH-Halo with empty pBAD vector vs Dam overexpression

Overexpression of Dam decreased the bound population of MutH significantly from 9.23% to 6.53% ($p=0.0044$). Data points represent mean $\pm$ SEM.

Overexpression of Dam decreased the bound population of MutH significantly from 9.23% to 6.53% ($p=0.0044$) (Figure 5.7). This decrease is consistent with a model where MutH finds its substrate on its own, therefore is affected by the amount of a competitor for GATC sites. The more proteins present in the cell, the shorter it takes for the hemimethylated GATC sites to get fully methylated, decreasing the chance of MutH finding these sites and binding to them. Note that the bound population of MutH differed in wild-type cells when arabinose was present (Figure 5.8), likely reflecting differences in the number of active replication forks, and hence abundance of hemimethylated GATC sites.

Arabinose induction system is a well-characterized bistable system. [Siegele and Hu, 1997] analyzed population distributions at intermediate sugar levels, and revealed that the *ara* system displays an all-or-nothing expression pattern. They concluded that in uninduced cells the stochastic background expression of the *ara* regulon leads to a wide distribution of *ara* uptake proteins. Addition of arabinose would then lead to different rates of arabinose accumulation, causing heterogeneous timing of gene induction within the population. For this reason, these experiments can be repeated with a different induction system.

5.4 Deletion of SeqA

To further test that MutH binding is dictated by the availability of the hemimethylated GATC sites, I deleted SeqA (Figure 5.9), another competitor of MutH that binds hemimethylated GATC sites, and is known to form foci behind active replication forks [Pedersen et al., 2017]. *SeqA* deletion significantly increased the bound population of MutH from 6.7% to 9.5% ($p=0.001$) (Figure 5.10) mirroring the *dam* deletion (Figure 5.4) and further confirming that the bound

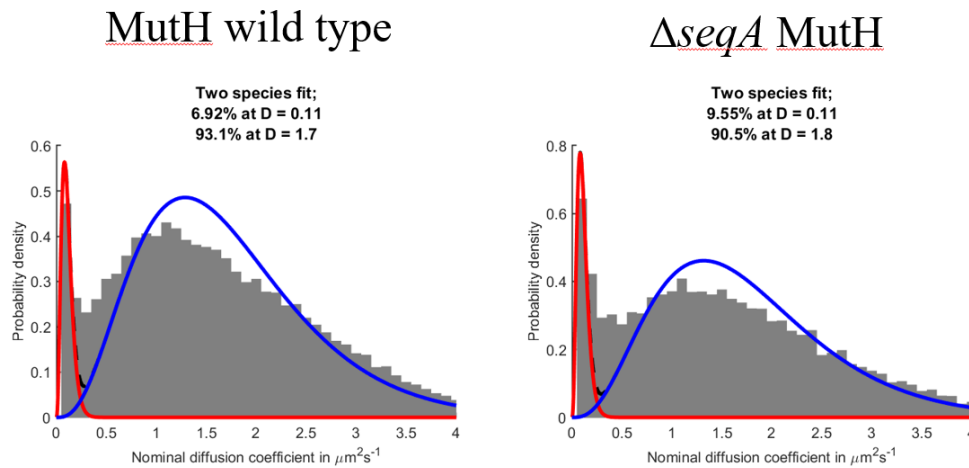


Figure 5.9: Distribution of diffusion coefficients for MutH-Halo and $\Delta SeqA$
Combined histogram of the distribution of diffusion coefficients from 3 independent sets. Two species fit performed with the diffusion coefficient of the bound population fixed and mobile population free.

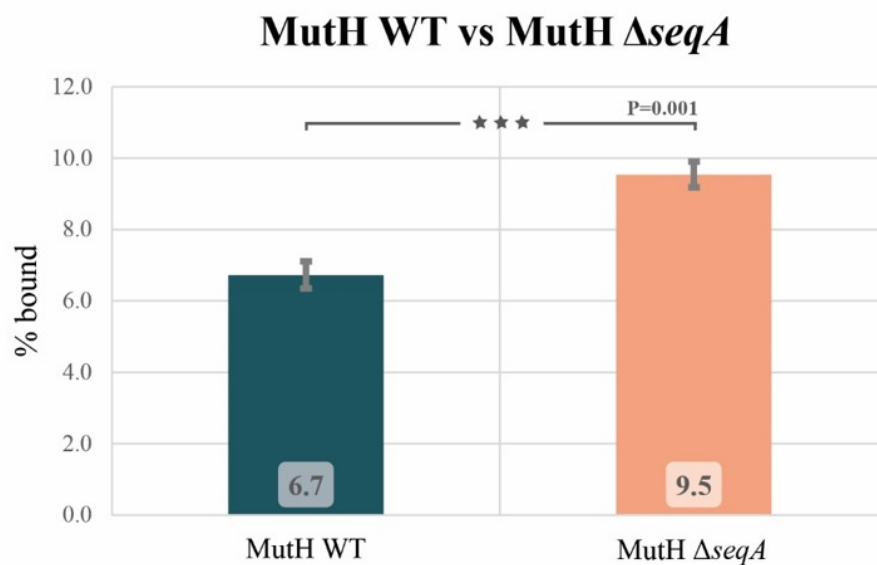


Figure 5.10: Bar graphs comparing the bound populations of MutH-Halo and $\Delta SeqA$

SeqA deletion significantly increased the bound population from 6.7% to 9.5% ($p=0.001$). Data points represent mean $\pm$ SEM.

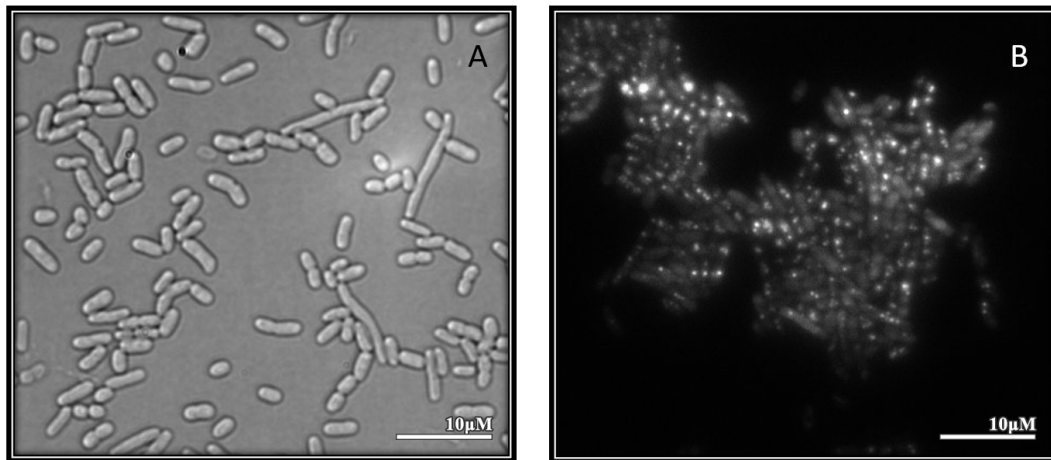


Figure 5.11: Example brightfield and fluorescent images of DnaN-mYPet foci in SeqA overexpression

A) Brightfield image for SeqA overexpression. Note that cells are smaller or larger than their normal size.

B) SeqA overexpression resulted in inconsistent number of replication forks including some cells without any DnaN-mYPet foci.

population of MutH is directly affected by the amount of a competitor for GATC sites. I also tried to overexpress SeqA using a pBAD vector however SeqA overexpression resulted in inconsistent number of replication forks including some cells having stopped replicating therefore making them too unreliable to investigate mismatch repair (Figure 5.11).

6 | Discussion

Accurate replication of DNA is critical for genomic stability. Replication is semi-conservative: replicative polymerases use one strand as a template to synthesize the other. Although this process is highly accurate, sometimes mistakes are made, which are called DNA mismatches. These mismatches are immediately proofread and the ones that escape proofreading are corrected by mismatch repair. Mismatch repair mechanisms are highly conserved and their inactivation has various severe consequences, such as mutator strains in bacteria and increased susceptibility to cancer in mammals.

To correct replication errors, MMR mechanisms must recognise the mismatches and identify the newly synthesized DNA strand that contains the misincorporated nucleotide. In *E. coli*, at the center of this process are 3 proteins called MutS, MutL and MutH. MutS recognizes the mismatches and recruits MutL, which in turn activates MutH. Activated MutH incises the DNA and allows the downstream events in the repair pathway to follow.

In this thesis, I used single molecule tracking to describe the dynamics of the MMR proteins MutS, MutL and MutH. I also used FRAP analysis to understand the kinetics of MutS and MutL in foci in a Δ MutH background, *i.e.*, in the presence of a mismatch without the possibility of repair. Using single molecule tracking, I observed low mobility of MutS within live cells, which indicates non-specific binding and scanning during mismatch search. This is consistent with frequent transient DNA interactions and is in agreement of what is shown for MutS target search in the literature [[Uphoff et al., 2016](#)], [[Liao et al., 2015](#)]. [[Liao et al., 2015](#)] visualised MutS with PAMCherry and observed that MutS tracks explore the entire cell, which I have also observed in *E. coli*. Contrary to their

findings in *B.subtilis*, however, I saw no evidence that there is a significant enrichment of MutS near the replisome in *E. coli*.

The published MutS crystal structures demonstrate that MutS binds to homoduplex DNA and tries to bend it. However the rigid structure of the DNA does not permit bending; therefore, the MutS arms do not rotate and the MutS-DNA interaction is lost. These transient DNA interactions make MutS diffusion slow and MutS residence time on the DNA short, unless it encounters a mismatch. Only in the rare event that MutS encounters a mismatch does it spend a longer time bound on the DNA, going through a conformational change through exchanging ADP with ATP [Liu et al., 2016].

[Gorman et al., 2012a] observed, using budding yeast as a model, that MutS can be targeted to mismatched bases through either 1D sliding or 3D diffusion. In their FRET experiments, they saw asymmetric binding to mismatches, which could result in a substantive fraction of mismatches left unnoticed within a single binding/scanning event (e.g., if the mismatch-binding subunit with the Phe-X-Glu motif is in contact with the G-containing strand for a G:T mismatch). The solution that they suggested for that is a quick release/rebinding of MutS and eventual coverage of all mismatch orientations in multiple rounds of MutS scanning of the DNA. This suggestion of a quick turnover of MutS molecules that they suggest match well my FRAP results, which show recovery in as quick as 5 seconds. One could therefore hypothesize that the foci could be formed by two kinds of MutS molecules: a) MutS molecules that recognise the mismatch and go through ADP-ATP exchange; b) Some MutS molecules that cannot recognise the mismatch and fall off quickly.

Interestingly, I observed a slower turnover for MutL than for MutS. This would suggest that the function of MutS might be to recruit MutL to the DNA, and

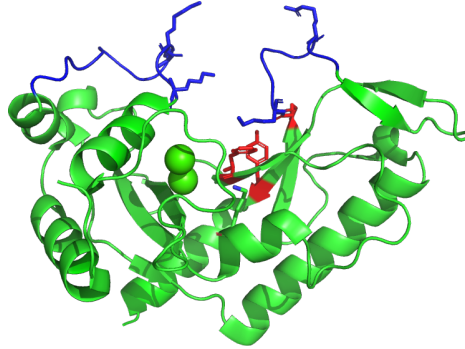


Figure 6.1: Proposed mutations for MutH

Three amino acids F94 Y212 P185 are annotated in red were to be mutated into alanine to perturb recognition of a hemimethylated GATC sequence. Positively charged residues in loop 67 and the BC loop (in blue) were also considered to be mutated into negatively charged amino acids to perturb non-specific DNA-binding.

once that objective is completed the former falls off the DNA. MutL stays on the DNA longer, slides and activates MutH. As I have seen a high off-rate for MutS and MutL using FRAP, sliding models are more plausible than the looping models. Looping models require MutS and MutL to stay stable to hold a loop together, and given the turnover, this does not seem possible.

Before the pandemic, I started a biochemistry project to generate MutH mutants that are defective in DNA-binding and/or recognition of hemimethylated GATC sites to further characterise the bound population of MutH. I planned to change the three amino acids F94 Y212 P185 into alanine to perturb recognition of a hemimethylated GATC sequence (residues annotated in red in figure 6.1). I was also considering changing the positively charged residues in loop 67 and the BC loop (in blue figure 6.1) into negatively charged amino acids to perturb non-specific DNA-binding.

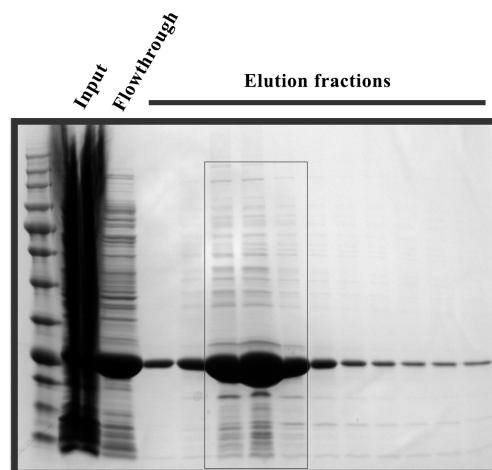
I started the project with purifying the wild-type MutH to learn protein purifi-

cation. I purified the MutH protein from bacteria using the PET expression system and pulling down on a 6-HIS tag introduced to the terminus of the protein (Figure 6.2). The purification included affinity chromatography using Talon beads to bind the 6-HIS tag and subsequently size exclusion chromatography to isolate homogeneous monomeric protein. The next step would have been to introduce the mutations into MutH and purification of the mutants. However, due to the pandemic I wasn't able to complete this project.

Another objective of this thesis has been to understand how MutH finds its substrate. I observed that MutH does not form foci, unlike MutS and MutL. This makes single molecule studies particularly useful, since they were able to detect individual immobilised MutH molecules. By using single molecule tracking, I tested the hypothesis, suggested by [Liu et al., 2016], that MutH is recruited to GATC sites by MutL-MutS complex: having deleted MutL, if MutH recruitment to the DNA was reliant on MutL, MutH binding in MutL knock-out cells would be abolished. Surprisingly, I saw no significant change in the MutH bound population after the deletion. This suggests that MutH finds its substrate independently of MutS-MutL complex. Additionally, the mean diffusion coefficient of mobile MutH molecules remained unchanged, which demonstrates that MutL and MutH do not interact in the cytoplasm.

In addition, I wanted to explore whether MutH binding is reliant on another complex, or whether it is directly affected by the perturbations of the amount of its competitors for the GATC sites. To test this, I deleted *dam* and observed a significant increase in the bound population of MutH. This increase could be attributed to the affinity of MutH for unmethylated DNA: in the absence of Dam, all GATC sites become unmethylated, and thus also a substrate for MutH. Moreover, increasing the amount of Dam resulted in decreased MutH binding: the increased amounts of Dam cause a shorter life-time for hemimethylated

**Talon-HIS
affinity chromatography**



**Size exclusion chromatography
Superdex200**

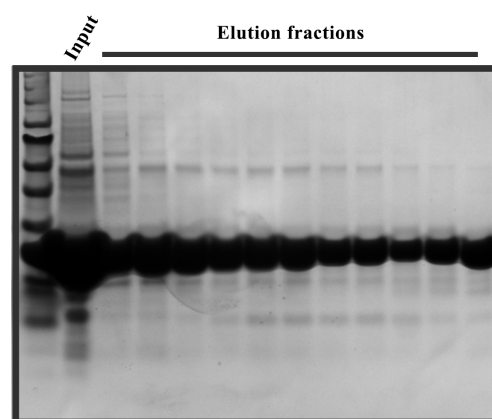


Figure 6.2: MutH protein purification

GATC sites, hence the less available substrate for MutH. Another competitor of MutH is SeqA, which also recognizes hemimethylated GATC sites. Upon deletion of SeqA, I observed a significant increase in the bound population of MutH, too, which lends further support for the suggestion above. I also tested the dependence of MutH binding on DNA damage. Treating the cells with MMS caused a significant increase in the immobile populations of MutS and MutL, indicating that there are more mismatches however MutH bound population remained unchanged.

The above results suggest that the percentage of bound MutH molecules is reliant on (1) the availability of hemimethylated GATC sites and (2) the amount of MutH molecules, but not on MutS or MutL. Furthermore, it is unlikely that MutH binding is reliant on another protein (such as β -clamp of the replicative polymerase), since the amount of competitors have a direct effect on the bound population.

To sum up, my results together with previous reports in the literature yield the following MMR model (Figure 6.3): MutS searches for mismatches on its own while frequently binding and unbinding the DNA. Upon mismatch recognition, it goes through a conformational change, exchanging ADP with ATP, and becomes a sliding clamp on the DNA. This conformational change allows it to recruit MutL, which also takes a sliding clamp shape. MutS has a high off-rate and falls off the DNA soon after recruiting MutL. Meanwhile, MutH finds hemimethylated GATC sites on its own, and upon encountering a DNA-bound MutL it gets activated and incises the DNA. After the incision, MutL recruits UvrD, which unwinds the DNA, and exonucleases degrade the newly synthesized strand until after the mismatch. SSB protects the single stranded DNA while PolIII gets another chance to correctly synthesize the DNA. Finally, ligase seals the gap created by PolIII.

Further experiments could explore the effect of replication on MMR using stationary cells. One could use the FRAP protocol I optimised to look at ATP binding mutants of MutS and MutL, or perform the experiments in WT background. It would be beneficial to repeat the over-expression assays with a different induction system. One could also purify the DNA-binding mutants and further characterise the bound population of MutH.

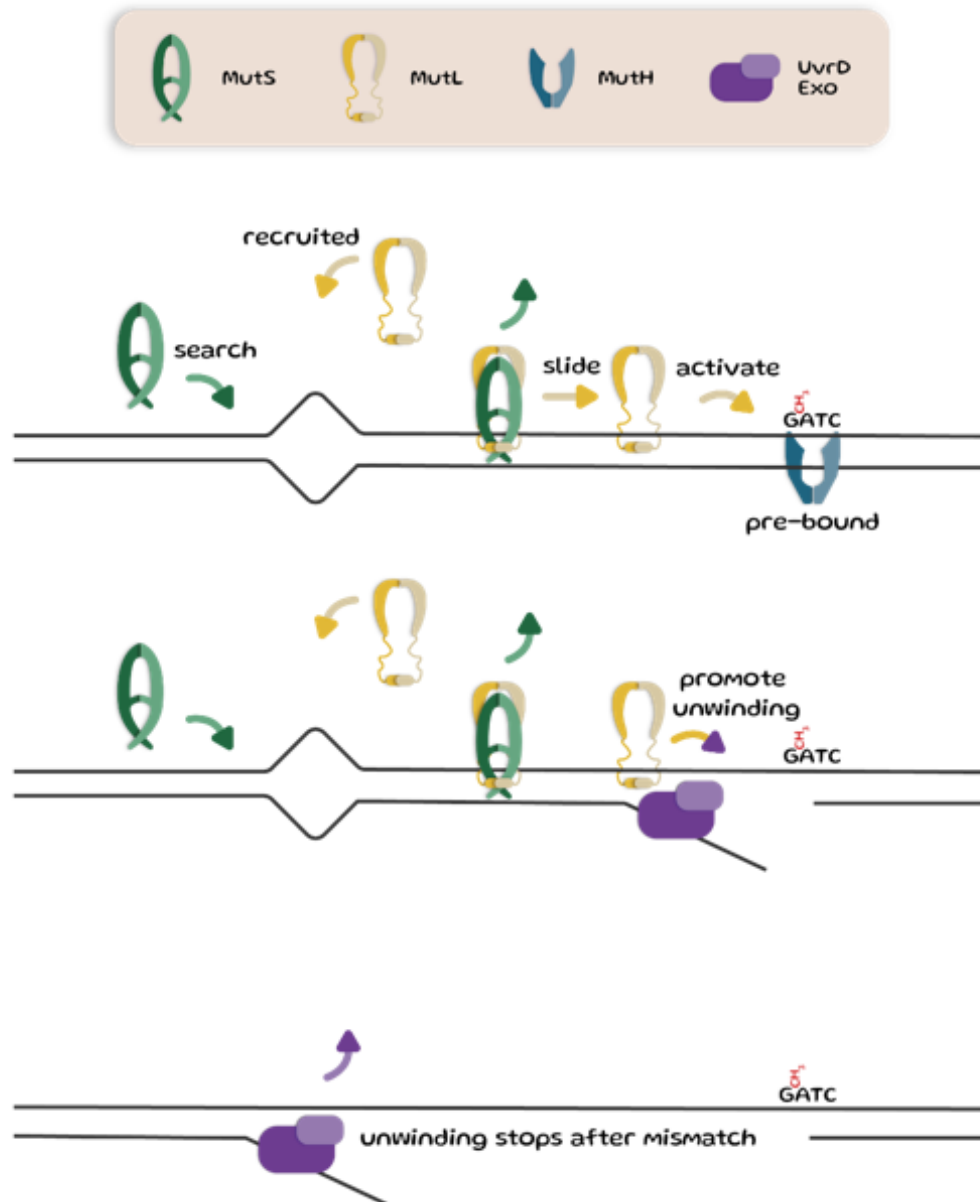


Figure 6.3: Proposed model for mismatch repair mechanism

MutS searches for mismatches. Upon mismatch recognition recruits MutL. MutS and MutL complex slide to MutH that is prebound at a hemimethylated GATC sites. MutS might fall off after recruiting MutL to the DNA. MutL activates MutH's endonuclease activity to incise the DNA. MutL, then, recruits UvrD and promotes unwinding. Unwinding of the DNA stops after the mismatch.

Bibliography

- Karin G Au, Katherine Welsh, and Paul Modrich. Initiation of methyl-directed mismatch repair. *Journal of Biological Chemistry*, 267(17):12142–12148, 1992.
- Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Molecular systems biology*, 2(1):2006–0008, 2006.
- Changill Ban and Wei Yang. Crystal structure and atpase activity of mutl: implications for dna repair and mutagenesis. *Cell*, 95(4):541–552, 1998a.
- Changill Ban and Wei Yang. Structural basis for muth activation in e. coli mismatch repair and relationship of muth to restriction endonucleases. *The EMBO journal*, 17(5):1526–1534, 1998b.
- Changill Ban, Murray Junop, and Wei Yang. Transformation of mutl by atp binding and hydrolysis: a switch in dna mismatch repair. *Cell*, 97(1):85–97, 1999.
- Nehir Banaz, Jarno Mäkelä, and Stephan Uphoff. Choosing the right label for single-molecule tracking in live bacteria: side-by-side comparison of photoactivatable fluorescent protein and halo tag dyes. *Journal of physics D: Applied physics*, 52(6):064002, 2018.
- Frederic Barras and Martin G Marinus. The great gatc: Dna methylation in e. coli. *Trends in Genetics*, 5:139–143, 1989.
- Alexandre W Bisson-Filho, Yen-Pang Hsu, Georgia R Squyres, Erkin Kuru, Fabai Wu, Calum Jukes, Yingjie Sun, Cees Dekker, Seamus Holden, Michael S Van Nieuwenhze, et al. Treadmilling by ftsz filaments drives peptidoglycan synthesis and bacterial cell division. *Science*, 355(6326):739–743, 2017.

- Linda B Bloom, Xiluo Chen, D Kuchnir Fygenson, Jennifer Turner, Mike O'Donnell, and Myron F Goodman. Fidelity of escherichia coli dna polymerase iii holoenzyme: the effects of β , γ complex processivity proteins and ϵ proofreading exonuclease on nucleotide misincorporation efficiencies. *Journal of Biological Chemistry*, 272(44):27919–27930, 1997.
- Alessandro Borsellini, Vladislav Kunetsky, Peter Friedhoff, and Meindert H Lamers. Cryo-em structures reveal how atp and dna binding in muts coordinate the sequential steps of dna mismatch repair. *bioRxiv*, 2021.
- Erik Boye and Anders Løbner-Olesen. The role of dam methyltransferase in the control of dna replication in e. coli. *Cell*, 62(5):981–989, 1990.
- Erik Boye, Martin G Marinus, and A Løbner-Olesen. Quantitation of dam methyltransferase in escherichia coli. *Journal of bacteriology*, 174(5):1682–1685, 1992.
- Therese Brendler, Jim Sawitzke, Kirill Sergueev, and Stuart Austin. A case for sliding seqa tracts at anchored replication forks during escherichia coli chromosome replication and segregation. *The EMBO journal*, 19(22):6249–6258, 2000.
- Joseph L Campbell and Nancy Kleckner. E. coli oric and the dnaa gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell*, 62(5):967–979, 1990.
- Suzanne Clancy. Dna damage & repair: mechanisms for maintaining dna integrity. *Nature Education*, 1(1):103, 2008.
- Nadia R Cohen, Christian A Ross, Saloni Jain, Rebecca S Shapiro, Arnaud Gutierrez, Peter Belenky, Hu Li, and James J Collins. A role for the bacterial gatc methylome in antibiotic stress survival. *Nature genetics*, 48(5):581–586, 2016.

- Justine Collier. Epigenetic regulation of the bacterial cell cycle. *Current opinion in microbiology*, 12(6):722–729, 2009.
- DL Cooper, RS Lahue, and P Modrich. Methyl-directed mismatch repair is bidirectional. *Journal of Biological Chemistry*, 268(16):11823–11829, 1993.
- Michele Cristóvão, Evangelos Sisamakís, Manju M Hingorani, Andreas D Marx, Caroline P Jung, Paul J Rothwell, Claus AM Seidel, and Peter Friedhoff. Single-molecule multiparameter fluorescence spectroscopy reveals directional muts binding to mismatched bases in dna. *Nucleic acids research*, 40(12):5448–5464, 2012.
- John C Crocker and David G Grier. Methods of digital video microscopy for colloidal studies. *Journal of colloid and interface science*, 179(1):298–310, 1996.
- Kirill A Datsenko and Barry L Wanner. One-step inactivation of chromosomal genes in escherichia coli k-12 using pcr products. *Proceedings of the National Academy of Sciences*, 97(12):6640–6645, 2000.
- J Deich, EM Judd, HH McAdams, and WE Moerner. Visualization of the movement of single histidine kinase molecules in live caulobacter cells. *Proceedings of the National Academy of Sciences*, 101(45):15921–15926, 2004.
- Marina Elez, Andrew W Murray, Li-Jun Bi, Xian-En Zhang, Ivan Matic, and Miroslav Radman. Seeing mutations in living cells. *Current Biology*, 20(16):1432–1437, 2010.
- Marina Elez, Miroslav Radman, and Ivan Matic. Stoichiometry of muts and mutl at unrepaired mismatches in vivo suggests a mechanism of repair. *Nucleic acids research*, 40(9):3929–3938, 2012.
- Steven Elmore, Michiel Müller, Norbert Vischer, Theo Odijk, and Conrad L Woldringh. Single-particle tracking of oric-gfp fluorescent spots during chro-

- mosome segregation in escherichia coli. *Journal of structural biology*, 151(3): 275–287, 2005.
- Richard Fishel. Signaling mismatch repair in cancer. *Nature medicine*, 5(11): 1239–1241, 1999.
- Norie Fujikawa, Hitoshi Kurumizaka, Mitsuyoshi Yamazoe, Sota Hiraga, and Shigeyuki Yokoyama. Identification of functional domains of the escherichia coli seqa protein. *Biochemical and biophysical research communications*, 300(3):699–705, 2003.
- Mahipal Ganji, Indra A Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H Haering, and Cees Dekker. Real-time imaging of dna loop extrusion by condensin. *Science*, 360(6384):102–105, 2018.
- Medini Manohar Ghodgaonkar, Federico Lazzaro, Maite Olivera-Pimentel, Mariela Artola-Borán, Petr Cejka, Martin A Reijns, Andrew P Jackson, Paolo Plevani, Marco Muzi-Falconi, and Josef Jiricny. Ribonucleotides misincorporated into dna act as strand-discrimination signals in eukaryotic mismatch repair. *Molecular cell*, 50(3):323–332, 2013.
- Barry W Glickman and Miroslav Radman. Escherichia coli mutator mutants deficient in methylation-instructed dna mismatch correction. *Proceedings of the National Academy of Sciences*, 77(2):1063–1067, 1980.
- Beth P Goldstein. Resistance to rifampicin: a review. *The Journal of antibiotics*, 67(9):625–630, 2014.
- Jason Gorman, Feng Wang, Sy Redding, Aaron J. Plys, Teresa Fazio, Shalom Wind, Eric E. Alani, and Eric C. Greene. Single-molecule imaging reveals target-search mechanisms during dna mismatch repair. *Proceedings of the*

National Academy of Sciences, 109(45):18251–18252, 2012a. ISSN 0027-8424.

URL <https://www.pnas.org/content/109/45/E3074/1>.

Jason Gorman, Feng Wang, Sy Redding, Aaron J Plys, Teresa Fazio, Shalom Wind, Eric E Alani, and Eric C Greene. Single-molecule imaging reveals target-search mechanisms during dna mismatch repair. *Proceedings of the National Academy of Sciences*, 109(45):E3074–E3083, 2012b.

Flora S Groothuizen, Ines Winkler, Michele Cristóvão, Alexander Fish, Herrie HK Winterwerp, Annet Reumer, Andreas D Marx, Nicolaas Hermans, Robert A Nicholls, Garib N Murshudov, et al. Muts/mutl crystal structure reveals that the muts sliding clamp loads mutl onto dna. *Elife*, 4:e06744, 2015.

Alba Guarne, Therese Brendler, Qinghai Zhao, Rodolfo Ghirlando, Stuart Austin, and Wei Yang. Crystal structure of a seqa–n filament: implications for dna replication and chromosome organization. *The EMBO journal*, 24(8):1502–1511, 2005.

Luz-Maria Guzman, Dominique Belin, Michael J Carson, and JON Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *Journal of bacteriology*, 177(14):4121–4130, 1995.

Maria Schei Haugan, Godefroid Charbon, Niels Frimodt-Møller, and Anders Løbner-Olesen. Chromosome replication as a measure of bacterial growth rate during escherichia coli infection in the mouse peritonitis model. *Scientific reports*, 8(1):1–13, 2018.

Sota Hiraga, Chiyome Ichinose, Toshinari Onogi, Hironori Niki, and Mitsuyoshi Yamazoe. Bidirectional migration of seqa-bound hemimethylated dna clus-

- ters and pairing of oric copies in escherichia coli. *Genes to Cells*, 5(5):327–341, 2000.
- Robin Holliday. A mechanism for gene conversion in fungi. *Genetics Research*, 5(2):282–304, 1964.
- Karl-Peter Hopfner. Invited review: Architectures and mechanisms of atp binding cassette proteins. *Biopolymers*, 105(8):492–504, 2016.
- Li-Wei Hung, Iris Xiaoyan Wang, Kishiko Nikaido, Pei-Qi Liu, Giovanna Ferroluzzi Ames, and Sung-Hou Kim. Crystal structure of the atp-binding subunit of an abc transporter. *Nature*, 396(6712):703–707, 1998.
- Ravi R Iyer, Anna Pluciennik, Vickers Burdett, and Paul L Modrich. Dna mismatch repair: functions and mechanisms. *Chemical reviews*, 106(2):302–323, 2006.
- Cherlhyun Jeong, Won-Ki Cho, Kyung-Mi Song, Christopher Cook, Tae-Young Yoon, Changill Ban, Richard Fishel, and Jong-Bong Lee. Muts switches between two fundamentally distinct clamps during mismatch repair. *Nature structural & molecular biology*, 18(3):379–385, 2011.
- Yong Jiang and Piotr E Marszalek. Atomic force microscopy captures muts tetramers initiating dna mismatch repair. *The EMBO journal*, 30(14):2881–2893, 2011.
- Ding Jun Jin and Carol A Gross. Mapping and sequencing of mutations in the escherichia colirpob gene that lead to rifampicin resistance. *Journal of molecular biology*, 202(1):45–58, 1988.
- Kenneth A Johnson. Conformational coupling in dna polymerase fidelity. *Annual review of biochemistry*, 62(1):685–713, 1993.

- Murray S Junop, Galina Obmolova, Kelly Rausch, Peggy Hsieh, and Wei Yang. Composite active site of an abc atpase: Muts uses atp to verify mismatch recognition and authorize dna repair. *Molecular cell*, 7(1):1–12, 2001.
- Jon M Kaguni. Dnaa: controlling the initiation of bacterial dna replication and more. *Annu. Rev. Microbiol.*, 60:351–371, 2006.
- Sukhyun Kang, Joo Seok Han, Sung Ho Kim, Jong Hoon Park, and Deog Su Hwang. Aggregation of seqa protein requires positively charged amino acids in the hinge region. *Biochemical and biophysical research communications*, 360(1):63–69, 2007.
- Thomas A Kunkel and Katarzyna Bebenek. Dna replication fidelity. *Annual review of biochemistry*, 69(1):497–529, 2000.
- Thomas A Kunkel and Dorothy A Erie. Dna mismatch repair. *Annu. Rev. Biochem.*, 74:681–710, 2005.
- Robert S Lahue, Shin-San Su, and Paul Modrich. Requirement for d (gac) sequences in escherichia coli mutHls mismatch correction. *Proceedings of the National Academy of Sciences*, 84(6):1482–1486, 1987.
- RS Lahue, KG Au, and P Modrich. Dna mismatch correction in a defined system. *Science*, 245(4914):160–164, 1989.
- Meindert H Lamers, Anastassis Perrakis, Jacqueline H Enzlin, Herrie HK Winterwerp, Niels de Wind, and Titia K Sixma. The crystal structure of dna mismatch repair protein muts binding to a g·t mismatch. *Nature*, 407(6805):711–717, 2000.
- Heewook Lee, Ellen Popodi, Haixu Tang, and Patricia L Foster. Rate and molecular spectrum of spontaneous mutations in the bacterium escherichia coli

- as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences*, 109(41):E2774–E2783, 2012.
- Jae Young Lee, Judy Chang, Nimesh Joseph, Rodolfo Ghirlando, Desirazu N Rao, and Wei Yang. Muth complexed with hemi-and unmethylated dnas: coupling base recognition and dna cleavage. *Molecular cell*, 20(1):155–166, 2005.
- Justin S Lenhart, Monica C Pillon, Alba Guarné, and Lyle A Simmons. Trapping and visualizing intermediate steps in the mismatch repair pathway in vivo. *Molecular microbiology*, 90(4):680–698, 2013.
- Alessia Lepore, Hannah Taylor, Dirk Landgraf, Burak Okumus, Sebastián Jaramillo-Riveri, Lorna McLaren, Somenath Bakshi, Johan Paulsson, and M El Karoui. Quantification of very low-abundant proteins in bacteria using the halotag and epi-fluorescence microscopy. *Scientific reports*, 9(1):1–9, 2019.
- Feng Li, Qin Liu, Yuan-Yuan Chen, Zi-Niu Yu, Zhi-Ping Zhang, Ya-Feng Zhou, Jiao-Yu Deng, Li-Jun Bi, and Xian-En Zhang. Escherichia coli mismatch repair protein mutl interacts with the clamp loader subunits of dna polymerase iii. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 637(1-2):101–110, 2008.
- Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, 2014.
- Guo-Min Li. Mechanisms and functions of dna mismatch repair. *Cell research*, 18(1):85–98, 2008.
- Yi Liao, Jeremy W Schroeder, Burke Gao, Lyle A Simmons, and Julie S Biteen. Single-molecule motions and interactions in live cells reveal target search dy-

- namics in mismatch repair. *Proceedings of the National Academy of Sciences*, 112(50):E6898–E6906, 2015.
- Jiaquan Liu, Jeungphill Hanne, Brooke M Britton, Jared Bennett, Daehyung Kim, Jong-Bong Lee, and Richard Fishel. Cascading muts and mutl sliding clamps control dna diffusion to activate mismatch repair. *Nature*, 539(7630): 583–587, 2016.
- A Løbner-Olesen, FG Hansen, KV Rasmussen, B Martin, and PL Kuempel. The initiation cascade for chromosome replication in wild-type and dam methyl-transferase deficient escherichia coli cells. *The EMBO Journal*, 13(8):1856–1862, 1994.
- Anders Løbner-Olesen, Martin G Marinus, and Flemming G Hansen. Role of seqa and dam in escherichia coli gene expression: a global/microarray analysis. *Proceedings of the National Academy of Sciences*, 100(8):4672–4677, 2003.
- Lawrence A Loeb. Mutator phenotype may be required for multistage carcinogenesis. *Cancer research*, 51(12):3075–3079, 1991.
- Georgyi V Los, Lance P Encell, Mark G McDougall, Danette D Hartzell, Natasha Karassina, Chad Zimprich, Monika G Wood, Randy Learish, Rachel Friedman Ohana, Marjeta Urh, et al. Halotag: a novel protein labeling technology for cell imaging and protein analysis. *ACS chemical biology*, 3(6):373–382, 2008.
- A-L Lu, K Welsh, S Clark, S-S Su, and P Modrich. Repair of dna base-pair mismatches in extracts of escherichia coli. In *Cold Spring Harbor symposia on quantitative biology*, volume 49, pages 589–596. Cold Spring Harbor Laboratory Press, 1984.
- A-LIEN Lu, Susanna Clark, and Paul Modrich. Methyl-directed repair of dna

- base-pair mismatches in vitro. *Proceedings of the National Academy of Sciences*, 80(15):4639–4643, 1983.
- Min Lu, Joseph L Campbell, Erik Boye, and Nancy Kleckner. SeqA: a negative modulator of replication initiation in *e. coli*. *Cell*, 77(3):413–426, 1994.
- Yannicka SN Mardenborough, Katerina Nitsenko, Charlie Laffeber, Camille Duboc, Enes Sahin, Audrey Quessada-Vial, Herrie HK Winterwerp, Titia K Sixma, Roland Kanaar, Peter Friedhoff, et al. The unstructured linker arms of mutL enable gatC site incision beyond roadblocks during initiation of dna mismatch repair. *Nucleic acids research*, 47(22):11667–11680, 2019.
- MG Marinus and A Løbner-Olesen. Dna methylation. *EcoSal Plus*, 6(1), 2014.
- Koen JA Martens, Arjen N Bader, Sander Baas, Bernd Rieger, and Johannes Hohlbein. Phasor based single-molecule localization microscopy in 3d (psmlm-3d): An algorithm for mhz localization rates using standard cpus. *The Journal of chemical physics*, 148(12):123311, 2018.
- Bonnie R McGraw and Martin G Marinus. Isolation and characterization of dam⁺ revertants and suppressor mutations that modify secondary phenotypes of dam⁻³ strains of *escherichia coli* k-12. *Molecular and General Genetics MGG*, 178(2):309–315, 1980.
- Marc L Mendillo, Christopher D Putnam, and Richard D Kolodner. *Escherichia coli* muts tetramerization domain structure reveals that stable dimers but not tetramers are essential for dna mismatch repair in vivo. *Journal of Biological Chemistry*, 282(22):16345–16354, 2007.
- L Meyer-Leon, LC Huang, SW Umlauf, MM Cox, and RB Inman. Holliday intermediates and reaction by-products in flp protein-promoted site-specific recombination. *Molecular and cellular biology*, 8(9):3784–3796, 1988.

- Paul Modrich. Methyl-directed dna mismatch correction. *Journal of Biological Chemistry*, 264(12):6597–6600, 1989.
- Paul Modrich. Mechanisms in e. coli and human mismatch repair (nobel lecture). *Angewandte Chemie International Edition*, 55(30):8490–8501, 2016.
- M Charl Moolman, Sriram Tiruvadi Krishnan, Jacob WJ Kerssemakers, Aafke Van Den Berg, Pawel Tulinski, Martin Depken, Rodrigo Reyes-Lamothe, David J Sherratt, and Nynke H Dekker. Slow unloading leads to dna-bound β 2-sliding clamp accumulation in live escherichia coli cells. *Nature communications*, 5(1):1–11, 2014.
- Lisa D Moore, Thuc Le, and Guoping Fan. Dna methylation and its basic function. *Neuropsychopharmacology*, 38(1):23–38, 2013.
- Kenan C Murphy. λ recombination and recombineering. *EcoSal Plus*, 7(1), 2016.
- Kim Nasmyth. How are dnas woven into chromosomes? *Science*, 358(6363):589–590, 2017.
- RONALD L Parsons, PARUCHURI V Prasad, RASIKA M Harshey, and MAKKUNI Jayaram. Step-arrest mutants of flp recombinase: implications for the catalytic mechanism of dna recombination. *Molecular and cellular biology*, 8(8):3303–3310, 1988.
- Ida Benedikte Pedersen, Emily Helgesen, Ingvild Flåtten, Solveig Fossum-Raunehaug, and Kirsten Skarstad. Seqa structures behind escherichia coli replication forks affect replication elongation and restart mechanisms. *Nucleic acids research*, 45(11):6471–6485, 2017.
- Anna Pluciennik and Paul Modrich. Protein roadblocks and helix discontinu-

- ities are barriers to the initiation of mismatch repair. *Proceedings of the National Academy of Sciences*, 104(31):12709–12713, 2007.
- Patricia J Pukkila, Janet Peterson, Gail Herman, Paul Modrich, and Matthew Meselson. Effects of high levels of dna adenine methylation on methyl-directed mismatch repair in escherichia coli. *Genetics*, 104(4):571–582, 1983.
- Rodrigo Reyes-Lamothe, Xindan Wang, and David Sherratt. Escherichia coli and its chromosome. *Trends in microbiology*, 16(5):238–245, 2008.
- Rodrigo Reyes-Lamothe, David J Sherratt, and Mark C Leake. Stoichiometry and architecture of active dna replication machinery in escherichia coli. *Science*, 328(5977):498–501, 2010.
- Yasushi Sako, Shigeru Minoghchi, and Toshio Yanagida. Single-molecule imaging of egfr signalling on the surface of living cells. *Nature cell biology*, 2(3):168–172, 2000.
- María Antonia Sánchez-Romero, Stephen JW Busby, Nigel P Dyer, Sascha Ott, Andrew D Millard, and David C Grainger. Dynamic distribution of seqa protein across the chromosome of escherichia coli k-12. *MBio*, 1(1):e00012–10, 2010.
- Carol JE Schwartz and Paul D Sadowski. Flp protein of 2 μ circle plasmid of yeast induces multiple bends in the flp recognition target site. *Journal of molecular biology*, 216(2):289–298, 1990.
- Barbara Sedgwick. Repairing dna-methylation damage. *Nature reviews Molecular cell biology*, 5(2):148–157, 2004.
- Scarlet S Shell, Christopher D Putnam, and Richard D Kolodner. The n terminus of saccharomyces cerevisiae msh6 is an unstructured tether to pcna. *Molecular cell*, 26(4):565–578, 2007.

Deborah A Siegele and James C Hu. Gene expression from plasmids containing the arabid promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences*, 94(15): 8168–8172, 1997.

Oleksii Sliusarenko, Jennifer Heinritz, Thierry Emonet, and Christine Jacobs-Wagner. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Molecular microbiology*, 80(3):612–627, 2011.

Irina Stancheva, Theodor Koller, and Jose M Sogo. Asymmetry of dam remethylation on the leading and lagging arms of plasmid replicative intermediates. *The EMBO journal*, 18(22):6542–6551, 1999.

Mathew Stracy, Christian Lesterlin, Federico Garza De Leon, Stephan Uphoff, Pawel Zawadzki, and Achillefs N Kapanidis. Live-cell superresolution microscopy reveals the organization of rna polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences*, 112(32):E4390–E4399, 2015.

Mathew Stracy, Jakob Schweizer, David J Sherratt, Achillefs N Kapanidis, Stephan Uphoff, and Christian Lesterlin. Transient non-specific dna binding dominates the target search of bacterial dna-binding proteins. *Molecular cell*, 81(7):1499–1514, 2021.

Stella Stylianidou, Connor Brennan, Silas B Nissen, Nathan J Kuwada, and Paul A Wiggins. Supersegger: robust image segmentation, analysis and lineage tracking of bacterial cells. *Molecular microbiology*, 102(4):690–700, 2016.

Shin-San Su and Paul Modrich. Escherichia coli muts-encoded protein binds to

- mismatched dna base pairs. *Proceedings of the National Academy of Sciences*, 83(14):5057–5061, 1986.
- Shin-San Su, Robert S Lahue, KG Au, and P Modrich. Mismatch specificity of methyl-directed dna mismatch correction in vitro. *Journal of Biological Chemistry*, 263(14):6829–6835, 1988.
- Shin-San Su, Michelle Grilley, Randy Thresher, Jack Griffith, and Paul Modrich. Gap formation is associated with methyl-directed mismatch correction under conditions of restricted dna synthesis. *Genome*, 31(1):104–111, 1989.
- Lynn C Thomason, Nina Costantino, and Donald L Court. E. coli genome manipulation by p1 transduction. *Current protocols in molecular biology*, 79(1):1–17, 2007.
- Michael Y Tolstorukov, Konstantin Virnik, Victor B Zhurkin, and Sankar Adhya. Organization of dna in a bacterial nucleoid. *BMC microbiology*, 16(1):1–11, 2016.
- Frank Uhlmann. Smc complexes: from dna to chromosomes. *Nature reviews Molecular cell biology*, 17(7):399–412, 2016.
- Stephan Uphoff. Real-time dynamics of mutagenesis reveal the chronology of dna repair and damage tolerance responses in single cells. *Proceedings of the National Academy of Sciences*, 115(28):E6516–E6525, 2018.
- Stephan Uphoff, David J Sherratt, and Achillefs N Kapanidis. Visualizing protein-dna interactions in live bacterial cells using photoactivated single-molecule tracking. *JoVE (Journal of Visualized Experiments)*, (85):e51177, 2014.
- Stephan Uphoff, Nathan D Lord, Burak Okumus, Laurent Potvin-Trottier, David J Sherratt, and Johan Paulsson. Stochastic activation of a dna dam-

- age response causes cell-to-cell mutation rate variation. *Science*, 351(6277): 1094–1097, 2016.
- Sabine Urig, Humaira Gowher, Andrea Hermann, Carsten Beck, Mehrnaz Fatemi, Andeas Humeny, and Albert Jeltsch. The escherichia coli dam dna methyltransferase modifies dna in a highly processive reaction. *Journal of molecular biology*, 319(5):1085–1096, 2002.
- Torsten Waldminghaus and Kirsten Skarstad. The escherichia coli seqa protein. *Plasmid*, 61(3):141–150, 2009.
- Huixian Wang and John B Hays. Human dna mismatch repair: coupling of mismatch recognition to strand-specific excision. *Nucleic acids research*, 35(20): 6727–6739, 2007.
- Jue D Wang and Petra A Levin. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology*, 7(11):822–827, 2009.
- James D Watson and Francis HC Crick. The structure of dna. In *Cold Spring Harbor symposia on quantitative biology*, volume 18, pages 123–131. Cold Spring Harbor Laboratory Press, 1953.
- Eva Wegel, Antonia Göhler, B Christoffer Lagerholm, Alan Wainman, Stephan Uphoff, Rainer Kaufmann, and Ian M Dobbie. Imaging cellular structures in super-resolution with sim, sted and localisation microscopy: A practical comparison. *Scientific reports*, 6(1):1–13, 2016.
- Wei Yang. Structure and function of mismatch repair proteins. *Mutation Research/DNA Repair*, 460(3-4):245–256, 2000.
- Xinxing Yang, Zhixin Lyu, Amanda Miguel, Ryan McQuillen, Kerwyn Casey Huang, and Jie Xiao. Gtpase activity-coupled treadmilling of the bacterial

tubulin ftsz organizes septal cell wall synthesis. *Science*, 355(6326):744–747, 2017.

A | Representative MutH Histograms

Representative histograms of MutH from 3 independent experiments to demonstrate reproducibility are shown below.

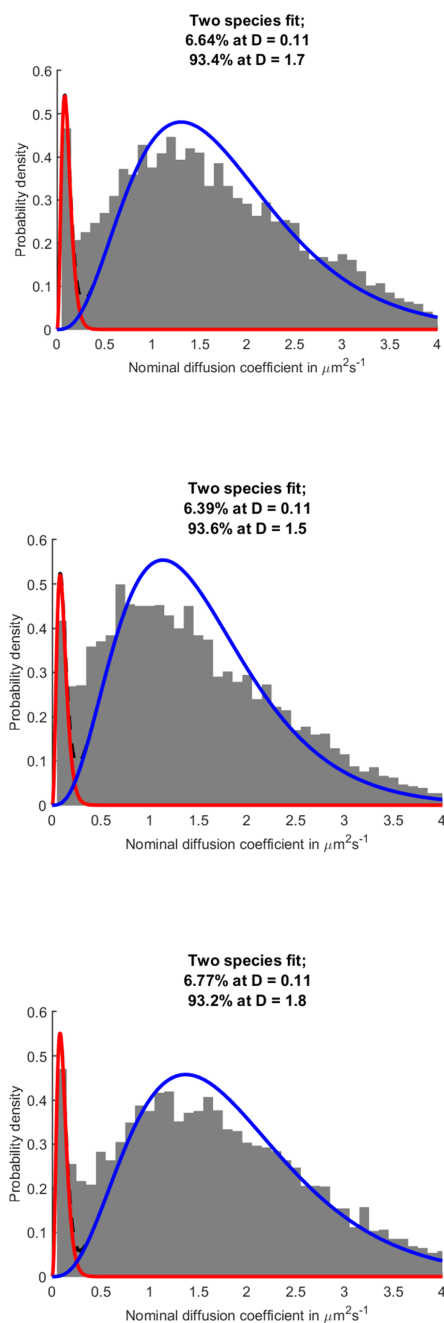


Figure A.1: Distribution of diffusion coefficients for MutH-Halo

B | Publication from DPhil Thesis

PAPER • OPEN ACCESS

Choosing the right label for single-molecule tracking in live bacteria: side-by-side comparison of photoactivatable fluorescent protein and Halo tag dyes

To cite this article: Nehir Banaz *et al* 2019 *J. Phys. D: Appl. Phys.* **52** 064002

View the [article online](#) for updates and enhancements.

You may also like

- [An introduction to optical super-resolution microscopy for the adventurous biologist](#)
J Vangindertael, R Camacho, W Sempels *et al.*
- [Photoactivatable fluorophores for single-molecule localization microscopy of live cells](#)
Yang Zhang and Francisco M Raymo
- [Advantages and limitations of a supernegative GFP in facilitating MyoD intracellular tracking](#)
Lucia Boeri, Emanuela Jacchetti, Monica Soncini *et al.*



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Abstract submission deadline: **April 8, 2022**

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD



Submit your abstract



Choosing the right label for single-molecule tracking in live bacteria: side-by-side comparison of photoactivatable fluorescent protein and Halo tag dyes

Nehir Banaz², Jarno Mäkelä² and Stephan Uphoff¹

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

E-mail: stephan.uphoff@bioch.ox.ac.uk

Received 31 July 2018, revised 7 November 2018

Accepted for publication 20 November 2018

Published 6 December 2018



Abstract

Visualizing and quantifying molecular motion and interactions inside living cells provides crucial insight into the mechanisms underlying cell function. This has been achieved by super-resolution localization microscopy and single-molecule tracking in conjunction with photoactivatable fluorescent proteins (PA-FPs). An alternative labelling approach relies on genetically-encoded protein tags with cell-permeable fluorescent ligands which are brighter and less prone to photobleaching than fluorescent proteins but require a laborious labelling process. Either labelling method is associated with significant advantages and disadvantages that should be taken into consideration depending on the microscopy experiment planned. Here, we describe an optimised procedure for labelling Halo-tagged proteins in live *Escherichia coli* cells. We provide a side-by-side comparison of Halo tag with different fluorescent ligands against the popular photoactivatable fluorescent protein PAmCherry. Using test proteins with different intracellular dynamics, we evaluated fluorescence intensity, background, photostability, and results from single-molecule localization and tracking experiments. Capitalising on the brightness and extended spectral range of fluorescent Halo ligands, we also demonstrate high-speed and dual-colour single-molecule tracking.

Keywords: single-molecule tracking, super-resolution microscopy, Halo tag, photoactivatable fluorescent protein, DNA-binding proteins, *Escherichia coli*, fluorophores

(Some figures may appear in colour only in the online journal)

Introduction

The combination of super-resolution localization microscopy with single-particle tracking has proven very powerful as it allows direct visualization of the activity of thousands

of proteins inside individual living cells. The localizations, mobility, and movement patterns of each molecule report on the molecular interactions and reactions in real-time. These data provide important quantitative information about those interactions, such as the diffusion coefficients of the reactants, dissociation constants, and the spatial distribution of interaction sites relative to other landmarks in the cell. Various implementations of this approach have provided crucial insights into many fundamental molecular processes in prokaryotic and eukaryotic cells [1–7].

Super-resolution localization microscopy is based on the sequential detection and localization of individual fluorescent

¹ Author to whom any correspondence should be addressed.

² These authors contributed equally to the work.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

molecules over the course of a movie [8]. This demands fluorescent labels with certain properties [9–11], most notably (i) means for specific labelling of target proteins ideally in a genetically-encoded manner, (ii) sufficient brightness for detection and accurate localization of single molecules, and (iii) the ability to switch fluorophores between fluorescent and non-fluorescent states so that only a fraction of molecules is visible at any time. For single-molecule tracking, photostability is also an important requirement as bleaching limits the observation time per molecule.

Live-cell localization microscopy was enabled by the development of photoactivatable fluorescent proteins (PA-FPs) such as PA-GFP, PAmCherry, mEOS, and other variants [10–12]. In fixed and permeabilized cells, antibody staining permits the use of synthetic dyes which are generally brighter and more photostable than fluorescent proteins. Fixed samples can be imaged in special buffers that induce reversible photoswitching of many types of synthetic fluorophores [13, 14]. The photophysical properties, cell permeability, and conjugation chemistries of different classes of synthetic dyes have been engineered for microscopy applications. Especially rhodamine, cyanine, and oxazine dye derivatives are commonly used in super-resolution microscopy and single-molecule fluorescence studies [9]. Genetically encoded protein tags have been developed for labelling with synthetic dyes that permeate live cells [15]. Especially Halo tag [16], SNAP tag [17], and CLIP tag [18] are becoming increasingly popular in fluorescence microscopy. In principle, these tags combine the advantages of genetically encoded FPs with the superior brightness and photostability of synthetic dyes. Because of this, they have found many successful applications in super-resolution microscopy [19, 20], single-molecule imaging [21, 22], and single-molecule tracking techniques [5, 7, 23], both in eukaryotic and in prokaryotic cells.

Halo tag is a monomeric 33 kDa protein that was engineered from a bacterial hydrolase enzyme to form a covalent bond with a chloroalkane linker of a ligand. This reaction is rapid and irreversible under physiological intracellular conditions [16]. Promega have commercialised the system and supply a range of fluorescent ligands. These include tetramethylrhodamine (TMR), as well as Oregon Green, and diAcFAM dyes which are less bright than TMR [24]. Alexa fluors are very popular in super-resolution microscopy with fixed cells; however, the Promega Alexa Halo tag ligands do not permeate into live cells. A new palette of bright cell-permeable fluorescent Halo tag ligands has been developed based on the TMR scaffold and termed Janelia Fluors [25]. There are also significant disadvantages and considerations concerning protein tags compared to fluorescent proteins, as the extra labelling step introduces additional experimental complexity and labour, and labelling specificity is not guaranteed. Nevertheless, the separation of protein expression and fluorescence labelling provides another means of experimental control and opportunities for new microscopy-based assays.

Many factors need to be considered when choosing the optimal labelling strategy for any microscopy experiment, and the additional temporal dimension in single-molecule tracking presents further challenges. To facilitate the decision,

we evaluated the Halo tag with different fluorescent ligands in a detailed side-by-side comparison against PAmCherry, a popular PA-FP [12]. PAmCherry has excitation/emission maxima of 564/595 nm, quantum yield of 0.46, and extinction coefficient of $1.8 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ [12]. Based on previous successful applications (e.g. [5, 20, 21]), we first evaluated the Halo ligand TMR, which has excitation/emission maxima of 548/572 nm, quantum yield of 0.41, and extinction coefficient of $7.8 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ [25]. We also tested the dyes JF549 and PA-JF549 which were derived from TMR and have similar spectral characteristics [25, 26]. JF549 has excitation/emission maxima of 549/571 nm, quantum yield of 0.88, and extinction coefficient of $1.0 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ [26]. We also utilized the red-shifted silicon-rhodamine dye JF646 [26, 27] in combination with JF549 for dual-colour imaging. JF646 has excitation/emission maxima of 646/664 nm, quantum yield of 0.54, and extinction coefficient of $1.5 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ [26]. After optimising a protocol for Halo tag labelling, we tested labelling specificity and background, quantified the brightness and photostability of the labels, and performed tracking experiments on several different fusion proteins. We compile these assessments to provide a practical guide for choosing the most suitable label for single-molecule tracking applications in live *E. coli* cells.

Fluorescent labelling of intracellular Halo tag proteins

To evaluate the performance of Halo tag and PAmCherry we chose test proteins that had previously been characterised. MukB is a structural maintenance of chromosomes (SMC) protein that acts in chromosome segregation together with accessory proteins MukE and MukF at a copy number of ~200 MukB dimers per cell [28]. Tracking MukB-PAmCherry revealed the presence of mobile and immobile SMC complexes that accumulated in distinct foci inside cells [29]. These foci appear to position the chromosome replication origin and subsequently aid in the segregation of the replicated origins into the two daughter cells [30]. As a second test protein with different intracellular mobility and localization, we chose DNA polymerase I (PolI), a monomeric protein that is much smaller than the SMC complex and present at ~400 copies per cell [31]. PolI performs DNA synthesis during the processing of Okazaki fragments and in DNA excision repair pathways [32]. Following DNA damage treatment, PolI-PAmCherry molecules become transiently immobilised at DNA repair sites [31]. To complement the existing PAmCherry fusions of MukB and PolI [29, 31], we generated endogenous C-terminal translational Halo tag fusions using the Lambda Red recombination technique [33]. We detected no signs of any growth defects that would indicate impairment of the function of MukB or PolI by the Halo fusions. Single-molecule imaging and tracking was performed on a custom-built total internal reflection fluorescence microscope equipped with an electron-multiplying CCD camera [34]. To image cytoplasmic proteins within ~1 μm thick *E. coli* cells, laser excitation was adjusted to highly inclined illumination mode [35].

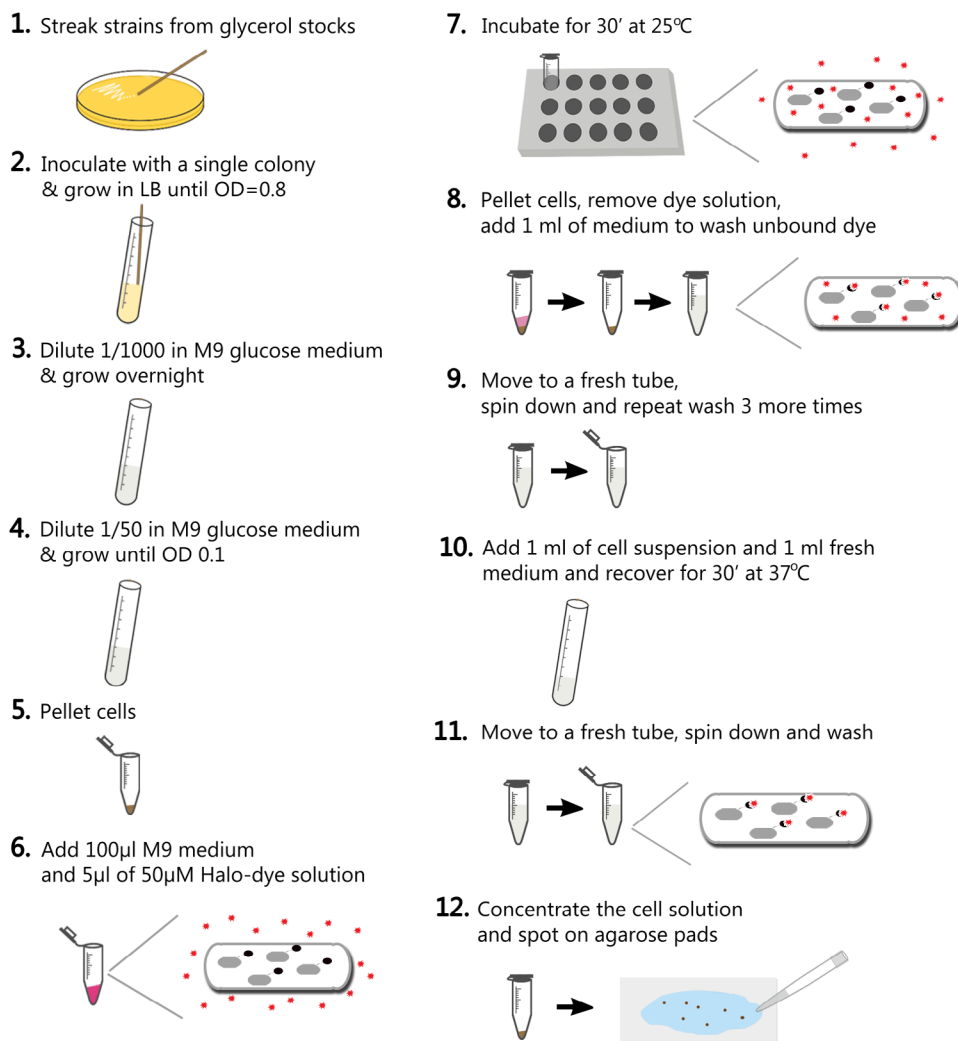


Figure 1. Procedure for labelling Halo tag with a fluorescent ligand in live *E. coli* cells.

We optimized a procedure for labelling Halo tag in live *E. coli* with the commercially available dye TMR (figure 1). Briefly, we grew cell cultures expressing Poll-Halo to early exponential phase at 37 °C in M9 minimal growth medium containing glucose and amino acids. 1 ml of culture was concentrated by centrifugation, resuspended in 100 µl of growth medium and incubated with 2.5 µM TMR dye (5 µl of 50 µM TMR stock solution) at 25 °C for 30 min. Removal of unbound dye was crucial to minimize fluorescent background for robust detection of single molecules inside cells. This is especially important for imaging bacteria directly on the surface of the coverslip where fluorescent particles adhere. We optimized the washing procedure by measuring the background over multiple centrifugation and resuspension cycles (figure 2(a)). After each wash, a small amount of cell suspension was immobilized on an agarose gel pad sandwiched between two microscope coverslips and imaged using 561 nm laser excitation. Intracellular fluorescence became visible above the background noise after 2–3 washes, but substantial background remained both on the coverslip surface and as diffuse background within the agarose pad. Therefore, to facilitate release of unbound dye and to recover the cells from the prolonged centrifugation, we performed a 4th wash and

incubated the diluted labelled culture at 37 °C for 30 min in a shaking incubator, followed by a final 5th wash. This removed the fluorescent background almost entirely. The entire labelling, washing, and recovery procedure takes approximately 90 min. Note that cell growth and division during the recovery dilutes the labelled Halo proteins exponentially while new unlabelled Halo is synthesised.

To quantify how the background affects single-molecule localization microscopy, we acquired movies and performed localization analysis as described below and in our previous work [36]. The frequency of background localizations outside cells decreased with each washing cycle from an average of $4.1 \cdot 10^{-2}$ localizations $\cdot \mu\text{m}^{-2} \cdot \text{frame}^{-1}$ without washing to $1.7 \cdot 10^{-4}$ localizations $\cdot \mu\text{m}^{-2} \cdot \text{frame}^{-1}$ after 5 washes (figure 2(b)). The remaining background after the washing procedure was much lower than the desired signal from Poll-Halo-TMR inside cells (average frequency of $4 \cdot 10^{-2}$ localizations $\cdot \mu\text{m}^{-2} \cdot \text{frame}^{-1}$, figures 2(c) and (d)). To assess the level of non-specific labelling, we measured the localization frequency of TMR inside cells that do not express Halo tag. The localization frequency was approximately 5-fold above the background outside cells, but more than 2 orders of magnitude lower than the specific localizations of Poll-Halo-TMR

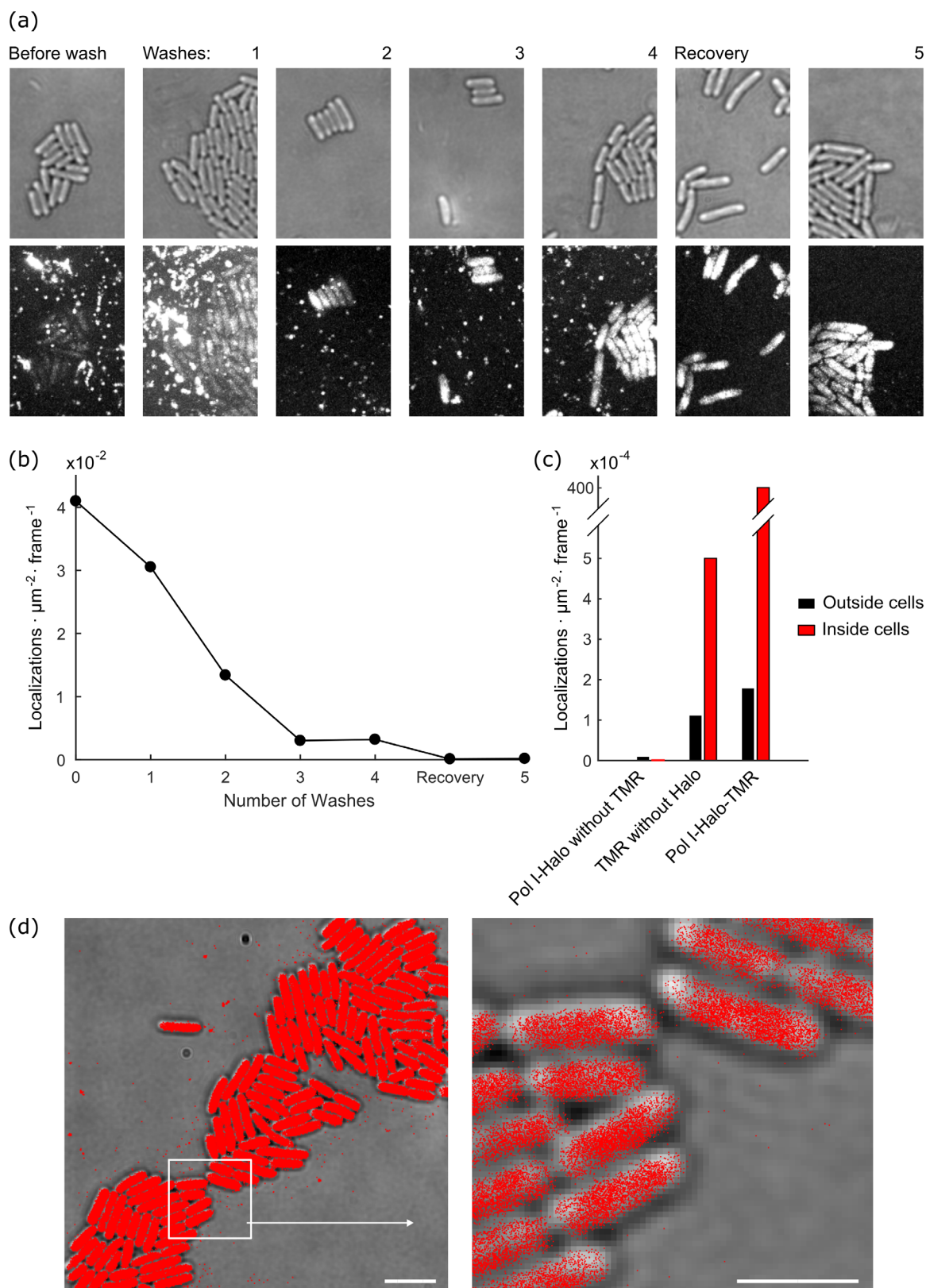


Figure 2. Removal of fluorescent background after Halo tag labelling. (a) *E. coli* expressing PolI-Halo-TMR were labelled and a super-resolution movie of 15 000 frames ($15.48 \text{ ms frame}^{-1}$) was recorded after each washing step to remove unbound dye. Transmitted light images show cell outlines (top). Fluorescence images show maximum intensity projections from the super-resolution movie (bottom, identical grey scaling across images for all washes). (b) Frequency of background localizations outside cells during the washing procedure. The frequency represents the average number of localizations detected per area per frame. (c) Frequency of localizations outside cells and inside cells: (i) for cells without TMR labelling, (ii) for cells that do not express Halo tag but were labelled with TMR (5 washes), (iii) for cells expressing PolI-Halo that were labelled with TMR (5 washes). (d) Map of PolI-Halo-TMR localizations (5 washes) plotted on transmitted light snapshot (scale bar $4 \mu\text{m}$). The boxed area is shown magnified (scale bar $2 \mu\text{m}$).

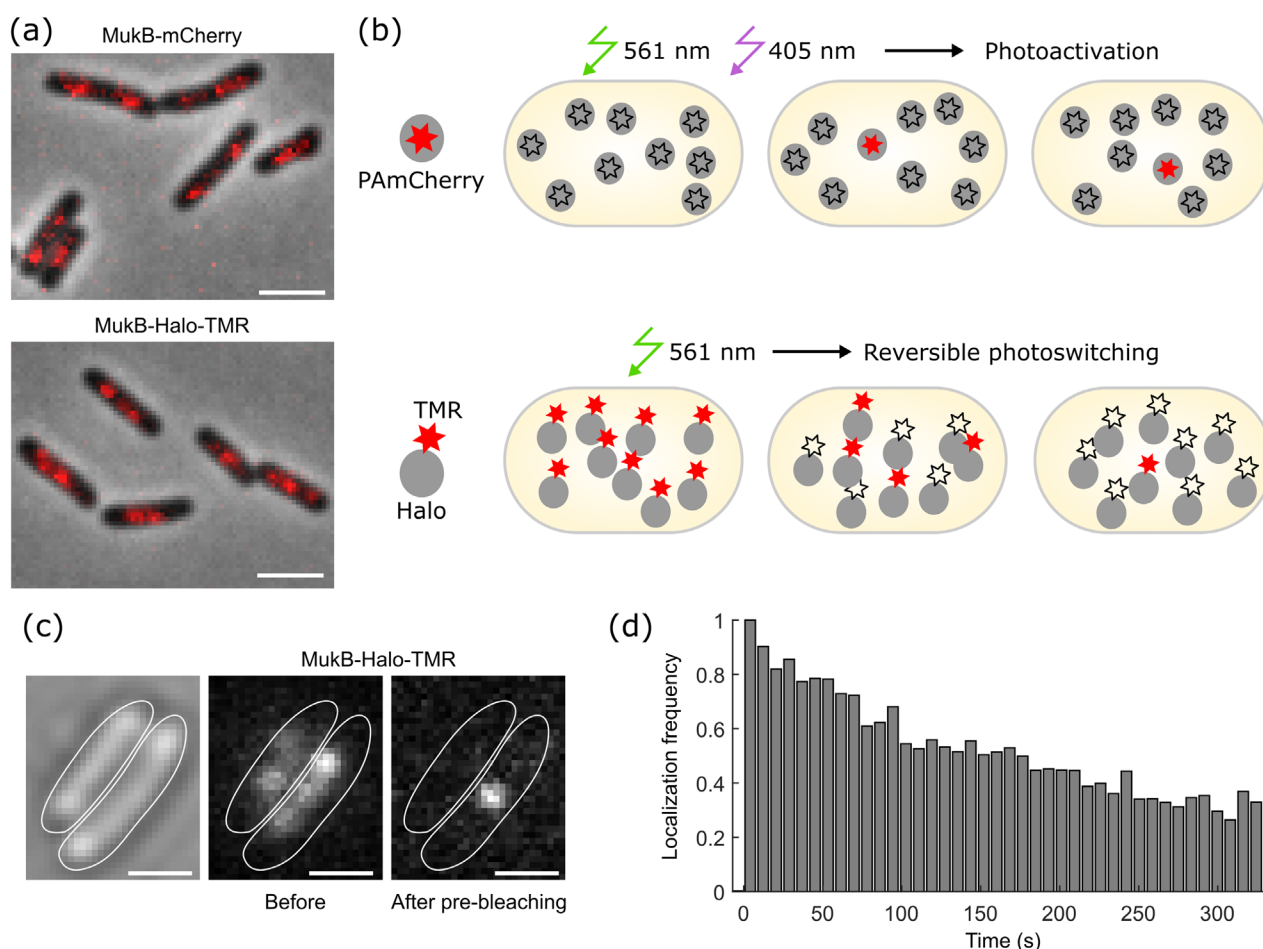


Figure 3. Microscopy data acquisition using Halo tag. (a) Snapshots of MukB-Halo-TMR and MukB-mCherry recorded with identical settings on an epifluorescence microscope (red: fluorescence with 150 ms exposure, grey: phase contrast). Scale bars 2 μm . (b) Principle of data acquisition for single-molecule localization microscopy using PAmCherry or Halo-TMR labels. (c) Snapshots showing MukB-Halo-TMR intensity before and after pre-bleaching (15 ms exposure). Scale bars 1 μm . (d) Frequency of MukB-Halo-TMR localizations over the course of a movie at constant 561 nm excitation (0.2 kW cm^{-2}), normalised to the initial frequency.

(figure 2(c)). Furthermore, the localization frequency inside cells without any TMR labelling was as low as $2 \cdot 10^{-6}$ localizations $\cdot \mu\text{m}^{-2} \cdot \text{frame}^{-1}$ (figure 2(c)). Therefore, the optimised labelling and washing procedure, produces reliable single-molecule localization data for Halo-TMR in live *E. coli* cells with little non-specific background (figure 2(d)).

Single-molecule localization microscopy of Halo tag and PAmCherry fusions

To test the utility of the Halo tag for imaging intracellular protein localization, we recorded snapshots of MukB-Halo-TMR using a standard Nikon epifluorescence microscope. Comparison with a fusion of MukB to the conventional fluorescent protein mCherry shows that both labelling strategies reproduce the characteristic foci of MukB [29] with similar image quality (figure 3(a)). The procedure for recording single-molecule localization or tracking data with Halo-TMR differs from PAmCherry. PAmCherry is initially in a non-fluorescent state and becomes converted by 405 nm illumination to the fluorescent state which can be excited at 561 nm [10, 12]. Consequently, the sample appears dark at the beginning

of the experiment apart from a small number of spontaneously activated molecules (figure 3(b)). On the one hand, this can be a technical inconvenience, as the absence of any signal makes it harder to find regions of interest and to adjust the focus. On the other hand, illumination at 561 nm without 405 nm photoactivation allows bleaching any background fluorescence before data acquisition. Simultaneous 405 nm and 561 nm illumination during the acquisition causes continuous activation and bleaching of PAmCherry for sequential localization of individual molecules (figure 3(b)). Adjusting the 405 nm intensity provides accurate control over the density of fluorescent dyes at any time to facilitate single-molecule detection. Because most PAmCherry molecules bleach irreversibly after one photoactivation cycle, the decay in the density of fluorescent molecules can be counteracted by gradually increasing the 405 nm intensity during an acquisition.

Contrary to PAmCherry, TMR-labelled Halo fusions are initially in the fluorescent state (figures 3(b) and (c)). Therefore, pre-bleaching is required before single-molecule imaging, but this can be useful for capturing a conventional epifluorescence image as a low-resolution reference (figure 3(a)). Under continuous illumination at 561 nm, TMR showed

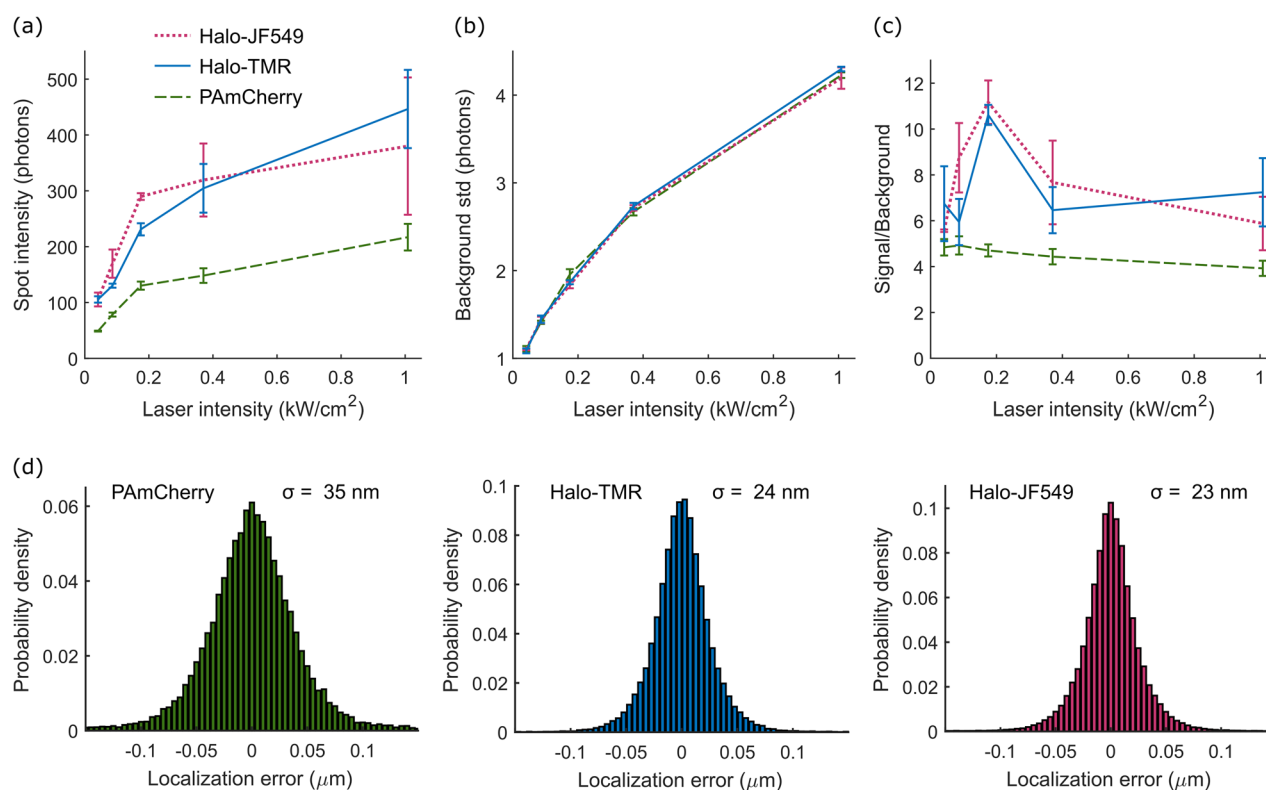


Figure 4. Fluorescence intensity and localization error comparison of PAmCherry, Halo-TMR, and Halo-JF549. (a) Fluorescence spot intensity of single MukB fusion proteins tagged with the indicated fluorophores as a function of 561 nm excitation intensity with 15 ms exposure time. Error bars are $\pm$ SEM from 3 repeats. (b) Standard deviation of the background noise per pixel as a function of 561 nm excitation intensity with 15 ms exposure time. Error bars are $\pm$ SEM from 3 repeats. (c) Signal to background ratio for the data in a-b, calculated from the peak pixel intensity in spots divided by the background std per pixel. Error bars are $\pm$ SEM from 3 repeats. (d) Localization error was determined as the standard deviation of repeated localizations of the same immobile molecules in fixed cells at 15 ms exposure time.

both photobleaching and reversible photoswitching within live cells. Initially, the number of visible fluorophores decayed rapidly due to reversible deactivation of the dye (figure 3(c)), followed by a long period of spontaneous photoswitching (figure 3(d)). The frequency of localisations decreased slowly over the course of a movie because of photobleaching (figure 3(d)). Importantly, imaging was performed in M9 minimal medium containing only supplements for cell growth (glucose, amino acids, and thiamine). There was no requirement for special buffer additives to induce photoswitching (e.g. oxygen scavenging and reducing agents) because the reducing environment inside cells promotes the photoinduced formation of a dark radical anion state that is recovered by molecular oxygen [37]. The lifetime of the reversible dark state of TMR was sufficiently long such that only a sparse subset of molecules appeared fluorescent at any time, which is an important requirement for accurate single-molecule localization and tracking. Compared to PAmCherry, the density of fluorescent TMR molecules cannot be precisely controlled at this stage in the experiment. However, if the density is too high, the initial photobleaching can be prolonged before data acquisition or substoichiometric labelling can be achieved by using a lower concentration of the dye.

We generally detected more localizations from Halo-TMR compared to PAmCherry under the same acquisition settings. This was mostly due to the repeated photoswitching cycles of

TMR as opposed to irreversible activation and bleaching of PAmCherry. Although repeated photoswitching of TMR provides more localization data from the same number of molecules per cell, this may complicate quantitative analysis of the localization data. In particular, blinking of single molecules produces apparent clusters of localizations, which are difficult to distinguish from genuine molecular assemblies [38]. This issue also affects PA-FPs, albeit to a lesser extent than photoswitchable dyes [39]. Several data analysis strategies have been reported to correct for blinking effects in molecule counting and clustering analysis [38–42].

The average intensity of single MukB-Halo-TMR molecules was approximately 2-fold higher than MukB-PAmCherry under the same imaging conditions (figure 4(a)). Notably, fluorescence brightness saturated with increasing 561 nm excitation intensity for both dyes, while the background shot noise continues to increase (figure 4(b)). This creates a peak in the signal to noise ratio at $\sim 0.2 \text{ kW cm}^{-2}$ 561 nm intensity (figure 4(c)). We also tested the Halo ligand JF549, which is a derivative of TMR with improved brightness and photostability [25]. We achieved excellent labelling of MukB-Halo-JF549 using the same protocol as for TMR, and found that it showed similar photoswitching behaviour. JF549 was significantly brighter than TMR at excitation intensities up to 0.2 kW cm^{-2} but there was little difference for higher intensities (figures 4(a)–(c)). To quantify the localization

precision for the different labels, we imaged MukB fusions in cells that were chemically fixed with paraformaldehyde (figure 4(d)). At the optimal excitation intensity of 0.2 kW cm^{-2} we measured localization errors (standard deviation) of $\sigma = 35 \text{ nm}$ (PAmCherry), 24 nm (TMR), and 23 nm (JF549). Photoactivatable Halo tag ligands have also been developed [26]. We tested one such dye, PA-JF549, but found that it did not label MukB-Halo efficiently using our protocol. Only $\sim 1\%$ of cells showed intracellular fluorescence, likely because the dye does not pass the *E. coli* cell wall.

Single-molecule tracking of Halo tag and PAmCherry fusions

The use of photoactivatable or photoswitchable fluorophores generalizes single-particle tracking, such that a very large number of trajectories can be recorded per cell, irrespective of the protein expression level or labelling density [43]. A major appeal of single-molecule tracking experiments is the ability to directly observe and quantify molecular interactions. PA-FPs have been used successfully for this purpose [2, 4, 6, 44]. However, low brightness and photostability of these fluorophores have been major limitations. Fluorophore brightness not only determines the spatial resolution, but also sets a limit on the temporal resolution at which molecules can be tracked. Measuring protein motion in the bacterial cytoplasm requires imaging at high frame rates of typically $10\text{--}1000 \text{ frames s}^{-1}$. Because photobleaching restricts the observation time per molecule, it has been challenging to record long-lived, rare, or transient molecular events. We thus tested whether the enhanced photophysical characteristics of fluorescent Halo tag ligands enable faster and longer observations, which would open avenues for new biological applications.

Using a tracking algorithm, 2D localizations were linked to tracks if they appeared within a radius of 768 nm (8 pixels) in consecutive frames [36]. To reconnect tracks with single missed localizations, we used a memory parameter of 1 frame. An apparent diffusion coefficient $D^* = \text{MSD}/4\Delta t$ was calculated from the mean squared displacement (MSD) between consecutive localizations of individual tracks with $\Delta t = 15.48 \text{ ms}$. For our purposes, D^* serves as a relative measure of mobility but does not represent the accurate diffusion coefficient of a molecule because of several systematic biases, such as motion confinement, motion blurring, and localization error. If required, these biases can be estimated and corrected for according to previously described procedures [31, 36]. For example, the localization error σ adds an offset of $\sigma^2/\Delta t$ to the apparent diffusion coefficient. The distribution of D^* values for MukB-Halo-TMR showed a mixture of immobile and diffusing molecules (figure 5(a)). Stationary MukB molecules were localized in clusters, whereas diffusing molecules moved randomly within the nucleoid (figure 5(b)). These results match our observations for MukB-PAmCherry (figures 5(a) and (b)), and reproduce previous findings [29].

Next, we analysed the duration of single-molecule tracks for MukB-PAmCherry, MukB-Halo-TMR, and MukB-Halo-JF549. For PAmCherry, trajectories are terminated mainly by photobleaching, whereas both reversible

photoswitching and photobleaching occur for TMR and JF549. Track durations were significantly longer for the Halo dyes compared to PAmCherry (figure 5(c)). At optimal excitation intensity (0.2 kW cm^{-2} 561 nm , figure 5(d)), the average track durations were 4.3, 8.7, and 11.5 frames for PAmCherry, TMR, and JF549, respectively. The distribution of track durations shows a long tail with 4.5% and 6.1% of TMR and JF549 molecules lasting for more than 100 frames, compared to 0.4% for PAmCherry. This enables direct observations of protein function within living cells over much longer time scales than previously possible with PA-FPs.

We used the PolI fusions to test if the Halo tag is also suitable for the measurement of transient molecular interactions. PolI is an essential factor for the repair of DNA base damage, filling DNA gaps that are generated by the base excision repair machinery [32]. The majority of PolI-Halo-TMR molecules were diffusing with $D^* \sim 0.7 \mu\text{m}^2 \text{ s}^{-1}$ in cells before DNA damage treatment (figure 6(a)). To monitor the DNA damage response, we first labelled PolI-Halo with TMR, and subsequently treated cells on agarose pads containing the DNA damaging agent methyl methanesulfonate (MMS). We observed immobilization of individual PolI molecules after MMS exposure for PAmCherry and Halo-TMR fusions (figure 6(b)), as previously shown [31].

Halo labelling enables high-speed and dual-colour tracking

To track rapidly diffusing proteins, previous studies relied on stroboscopic illumination patterns in order to reduce blurring of the fluorescent spots due to movement of molecules during the exposure [45, 46]. The superior brightness of fluorescent Halo ligands may enable tracking fast molecular motion with simple continuous illumination. Besides the technical convenience, this also avoids the use of very high excitation intensities required for stroboscopic illumination pulses (i.e. hundreds of kW cm^{-2} [46]). To test the limits of high-speed single-molecule tracking, we measured unconjugated Halo molecules that were expressed from a plasmid and labelled with TMR (figure 7(a)). With continuous 561 nm illumination (0.4 kW cm^{-2}) and a frame rate of $134 \text{ frames s}^{-1}$ ($7.48 \text{ ms frame}^{-1}$), fluorescent spots were sufficiently bright for robust localization and tracking despite residual motion blurring (figures 7(b)–(d)). This was in stark contrast to unconjugated PAmCherry molecules, which were almost indistinguishable from the background noise at such short exposure times. Because any protein fused to Halo should have a lower mobility than the Halo tag alone, this result demonstrates the technical feasibility of tracking even the most rapidly diffusing fusion proteins with simple continuous illumination. Stable non-specific binding of unconjugated Halo molecules was reported in eukaryotic nuclei [47]. This complicates the interpretation of the tracking data and may perturb the function of fusion proteins. Here, we observed only a small fraction of apparently immobile Halo-TMR molecules in *E. coli* (1% of all tracks, figure 7(d)), which was likely due to residual fluorescent background particles (as quantified in figure 2).

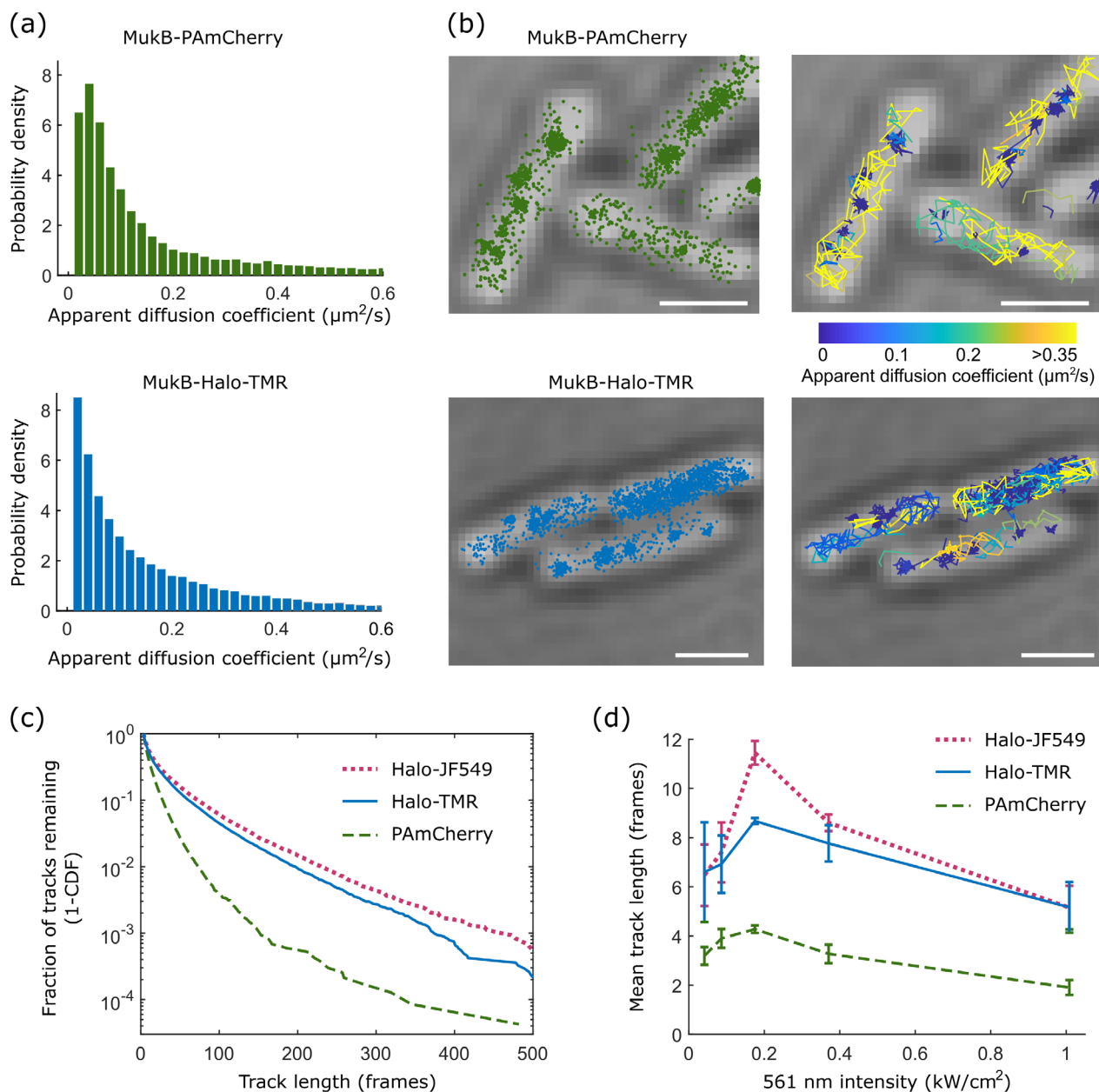


Figure 5. Application of PAmCherry and Halo tag labels for single-molecule tracking of MukB. All data recorded at $15.48 \text{ ms frame}^{-1}$. (a) Distributions of the apparent diffusion coefficient for MukB-PAmCherry and MukB-Halo-TMR. (b) Localisations and tracks of MukB-PAmCherry and MukB-Halo-TMR. Track colours correspond to the apparent diffusion coefficient. Scale bars 1 μm . (c) Track length distributions at 0.2 kW cm^{-2} 561 nm excitation intensity for MukB labelled with PAmCherry, Halo-TMR, and Halo-JF549. (d) Mean track length as a function of 561 nm excitation intensity. Error bars are $\pm$ SEM from 3 repeats.

To choose appropriate imaging conditions for high-speed tracking experiments, first the exposure time per frame should be sufficiently short such that motion blurring does not prevent single-molecule localization. This may require reducing the size of the field of view to increase the readout speed of the camera. On the other hand, molecular motion should yield significant displacements beyond the localization error between successive frames to accurately measure mobility. After determining the exposure time and frame rate, the excitation intensity should be adjusted to optimise localization precision and track duration as shown (figures 4 and 5).

Especially for rapidly diffusing proteins, it is crucial to maintain a low density of fluorescent spots during data acquisition to avoid tracking errors.

The ability to use Halo tag with a range of different ligands is a further advantage of this labelling approach. It facilitates multi-colour imaging and enables the use of fluorophores in the far-red spectrum where cellular auto-fluorescence and phototoxicity are less pronounced. We tested dual-colour tracking using Halo labelling with JF549 and the red-shifted dye JF646 [25]. As a proof-of-principle, we first labelled MukB-Halo with JF646 according to the protocol in figure 1. Following

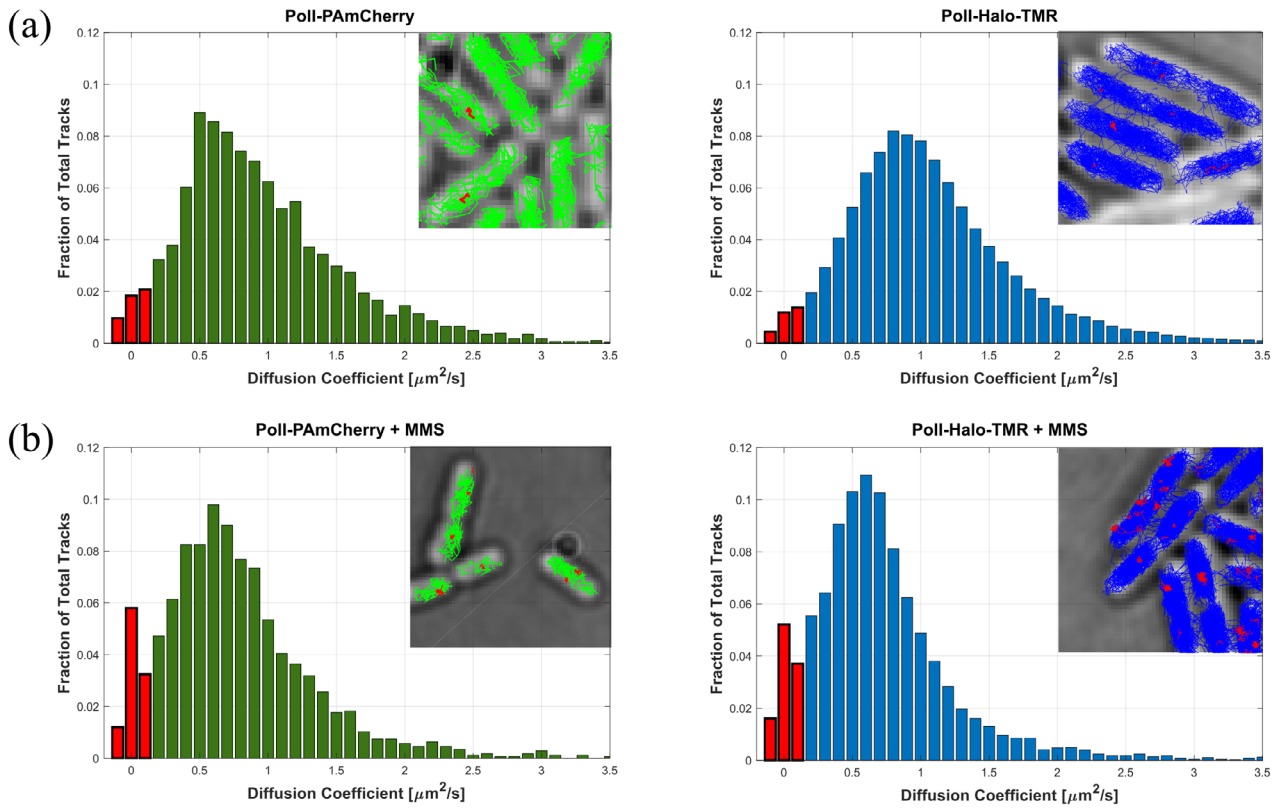


Figure 6. Tracking DNA repair activity of PolI. Cells with PolI-PAmCherry (left) or PolI-Halo-TMR (right) were imaged before (a) and after treatment with 100 mM MMS for 15 min. Immobile molecules with $D^* < 0.15 \mu\text{m}^2 \text{s}^{-1}$ are shown in red. All data recorded at $15.48 \text{ ms frame}^{-1}$.

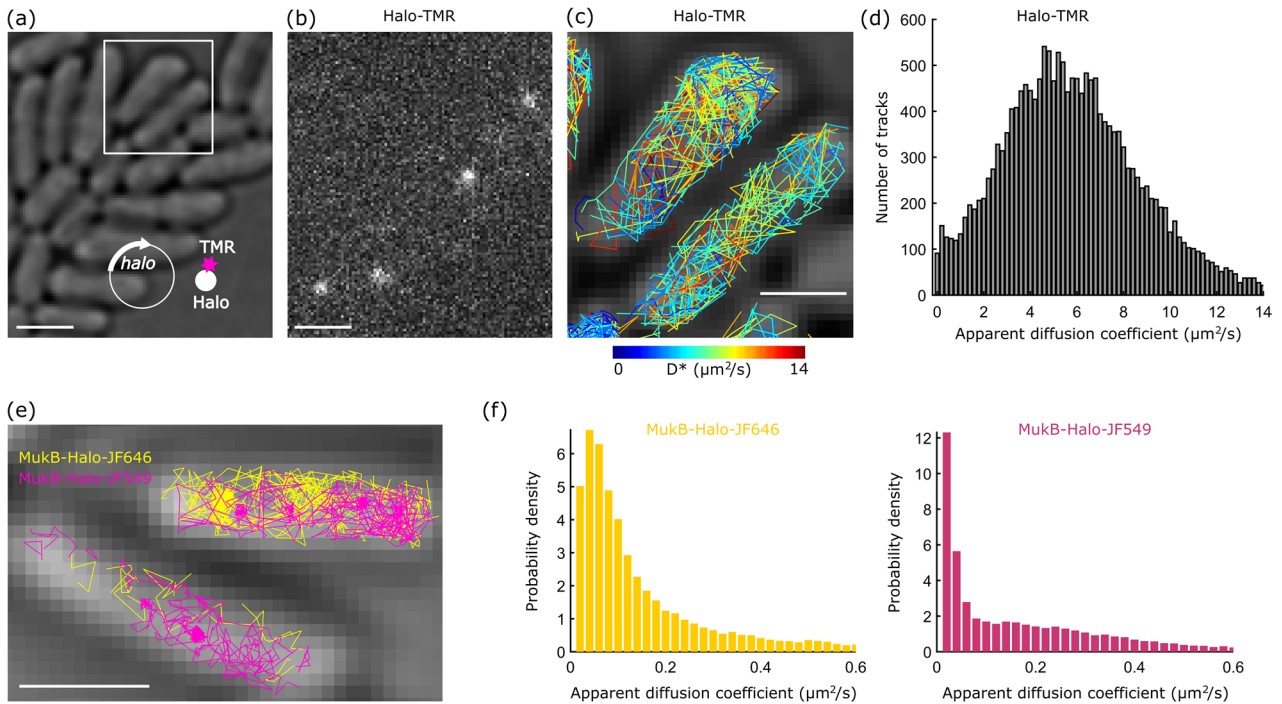


Figure 7. High-speed and dual-colour single-molecule tracking using Halo tag. (a) Cells expressing unconjugated Halo tag from a plasmid. Scale bar $2 \mu\text{m}$. (b) Snapshot from a super-resolution movie showing Halo-TMR molecules at 7.48 ms exposure time. Scale bar $2 \mu\text{m}$. (c) Halo-TMR tracks from the boxed area in panel a, shown in colours corresponding to the apparent diffusion coefficient. Scale bar $1 \mu\text{m}$. (d) Histogram of apparent diffusion coefficients for unconjugated Halo-TMR. (e) Dual-colour tracking of MukB-Halo-JF646 (yellow) and MukB-Halo-JF549 (pink), recorded at $15.48 \text{ ms frame}^{-1}$. Scale bar $1 \mu\text{m}$. (f) Histogram of apparent diffusion coefficients from dual-colour tracking of MukB-Halo-JF646 and MukB-Halo-JF549.

Table 1. Practical considerations regarding labelling approaches for single-molecule localization and tracking experiments in live bacterial cells.

	Advantages	Disadvantages
Photoactivatable fluorescent protein (e.g. PAmCherry)	• Genetically encoded fluorophore (no additional labelling required, labelling specificity guaranteed) • Independently controlled photoactivation with 405 nm excitation • Very low background noise (background can be bleached before photoactivation) • Cells can be imaged during continuous growth • Protein synthesis during cell growth replenishes fluorescence • Established and well-characterised labels	• Low brightness limits spatial and temporal resolution • Observation time per molecule is limited by rapid photobleaching • Irreversible bleaching prohibits repeated measurements of the same molecules • Limited spectral range • Phototoxicity of 405 nm illumination for photoactivation • Multi-colour tracking is limited because most PA-FPs are activated by near-UV light • Quantitative localization analysis complicated by incomplete photoactivation and long maturation times • Aggregation artifacts have been reported
Protein tag labelled with synthetic fluorophore (e.g. Halo-TMR)	• Superior brightness provides enhanced localization precision and temporal resolution • Reduced photobleaching increases observation time per molecule • Available dyes span visible to far-red spectrum • Compatible with conventional fluorescence imaging • Multi-colour tracking possible using orthogonal labelling of Halo, SNAP, and CLIP tags with different dyes • Reversible blinking allows repeated measurements of the same molecules • Ability to perform sequential labelling (e.g. before and after a cell perturbation)	• Additional labelling step (laborious, may interfere with other cell treatments, time-course experiments, etc) • Removal of unbound dye requires extensive cell washing and centrifugation • Labelled protein is diluted with each cell division, causing loss of signal during prolonged time-lapse imaging • Pre-bleaching step required before single-molecule imaging • Limited control over photoswitching kinetics • Reversible blinking produces apparent clustering artifacts • Quantitative localization analysis complicated by reversible blinking, unknown labelling efficiency, and dilution of labelled proteins during cell growth • Not all Halo tag dyes are cell permeable • Not all fluorescent Halo ligands are commercially available or can be expensive

2 h of cell recovery, we labelled the newly synthesised MukB-Halo with JF549 using the same protocol. We imaged the two labels sequentially, recording 21 000 frames under 640 nm excitation (0.24 kW cm^{-2}) followed by 21 000 frames under 561 nm excitation (0.2 kW cm^{-2}). The dual-colour tracking data shows the expected diffusing and immobile MukB molecules for both labels (figures 7(e) and (f)). Under our imaging conditions, JF646 had a lower intensity (116 photons/frame) compared to JF549 (290 photons/frame). As a result, the mean apparent diffusion coefficient was marginally higher for MukB-JF646 ($0.190 \pm 0.005 \mu\text{m}^2 \text{ s}^{-1}$, SEM) compared to MukB-JF549 ($0.182 \pm 0.007 \mu\text{m}^2 \text{ s}^{-1}$, SEM) due to the increased localization error of JF646.

Choosing the right label for single-molecule tracking experiments

Applications that require maximum fluorophore brightness for fast tracking or high localization precision benefit from the use of protein tags with synthetic dyes. The high excitation intensities required for super-resolution microscopy and single-molecule tracking cause significant phototoxicity in cells [48, 49]. Especially near-UV light required for photoactivation of PA-FPs is problematic [48]. Ongoing development

of synthetic fluorophores will further improve their photo-physical characteristics, cell permeability, and availability. However, several important limitations need to be taken into account. The requirement for labelling and extensive washing of the fluorescent ligand may complicate applications that require precisely controlled growth conditions or cell perturbations. This may be overcome using certain fluorophores such as JF646 whose fluorescence increases upon conjugation to a protein tag, thereby reducing the need for removal of unconjugated dye [25]. Prolonged time-lapse imaging benefits from the continuous synthesis and replenishment of fluorescent proteins. Improved fluorescent proteins are also reported, with extended spectral range [50], beneficial photoactivation/conversion/switching properties, reduced maturation time [51]. We close with a table showing what we consider the most important practical considerations in favour and against PA-FPs and synthetic dyes for protein tags (table 1).

Acknowledgments

We thank Luke Lavis for providing the Halo ligand dyes JF549, PA-JF549, and JF646, and David J Sherratt for discussions. Imaging in the Micron Advanced Microscopy Facility was funded by Wellcome Trust Strategic Award 107457/Z/15/Z.

This work was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant No. 206159/Z/17/Z) and a Wellcome-Beit Prize (206159/Z/17/B) to SU. SU holds a Hugh Price Fellowship at Jesus College Oxford.

Author contributions

NB and JM contributed equally to the work. SU initiated the study. NB, JM, SU performed experiments and analysed the data. NB, JM, SU prepared manuscript figures. SU wrote the manuscript with input from all authors.

ORCID iDs

Nehir Banaz  <https://orcid.org/0000-0003-2384-0857>
Stephan Uphoff  <https://orcid.org/0000-0002-3579-0888>

References

- [1] Saxton M J and Jacobson K 1997 Single-particle tracking: applications to membrane dynamics *Annu. Rev. Biophys. Biomol. Struct.* **26** 373–99
- [2] Gahlmann A and Moerner W E 2013 Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging *Nat. Rev. Microbiol.* **12** 9–22
- [3] Kapanidis A N, Uphoff S and Stracy M 2018 Understanding protein mobility in bacteria by tracking single molecules *J. Mol. Biol.* **430** 4443–55
- [4] Uphoff S and Sherratt D J 2017 Single-molecule analysis of bacterial DNA repair and mutagenesis *Annu. Rev. Biophys.* **46** 411–32
- [5] Chen J et al 2014 Single-molecule dynamics of enhanceosome assembly in embryonic stem cells *Cell* **156** 1274–85
- [6] Li Y, Schroeder J W, Simmons L A and Biteen J S 2018 Visualizing bacterial DNA replication and repair with molecular resolution *Curr. Opin. Microbiol.* **43** 38–45
- [7] Knight S C et al 2015 Dynamics of CRISPR-Cas9 genome interrogation in living cells *Science* **350** 823–6
- [8] Huang B, Bates M and Zhuang X 2009 Super-resolution fluorescence microscopy *Annu. Rev. Biochem.* **78** 993–1016
- [9] van de Linde S, Heilemann M and Sauer M 2012 Live-cell super-resolution imaging with synthetic fluorophores *Annu. Rev. Phys. Chem.* **63** 519–40
- [10] Lippincott-Schwartz J and Patterson G H 2009 Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging *Trends Cell Biol.* **19** 555–65
- [11] Wang S, Moffitt J R, Dempsey G T, Xie X S and Zhuang X 2014 Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging *Proc. Natl Acad. Sci.* **111** 8452–7
- [12] Subach F V, Patterson G H, Manley S, Gillette J M, Lippincott-Schwartz J and Verkhusha V V 2009 Photoactivatable mCherry for high-resolution two-color fluorescence microscopy *Nat. Methods* **6** 153–9
- [13] Heilemann M, Dedeker P, Hofkens J and Sauer M 2009 Photoswitches: key molecules for subdiffraction-resolution fluorescence imaging and molecular quantification *Laser Photonics Rev.* **3** 180–202
- [14] Vogelsang J, Steinhauer C, Forthmann C, Stein I H, Person-Skegro B, Cordes T and Tinnefeld P 2010 Make them blink: probes for super-resolution microscopy *ChemPhysChem* **11** 2475–90
- [15] Crivat G and Taraska J W 2012 Imaging proteins inside cells with fluorescent tags *Trends Biotechnol.* **30** 8–16
- [16] Los G V et al 2008 Halo tag: a novel protein labeling technology for cell imaging and protein analysis *ACS Chem. Biol.* **3** 373–82
- [17] Juillerat A, Gronemeyer T, Keppler A, Gendreizig S, Pick H, Vogel H and Johnsson K 2003 Directed evolution of O⁶-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules *in vivo Chem. Biol.* **10** 313–7
- [18] Gautier A, Juillerat A, Heinis C, Corrêa I R Jr, Kindermann M, Beaufils F and Johnsson K 2008 An engineered protein tag for multiprotein labeling in living cells *Chem. Biol.* **15** 128–36
- [19] Hein B, Willig K I, Wurm C A, Westphal V, Jakobs S and Hell S W 2010 Stimulated emission depletion nanoscopy of living cells using SNAP-tag fusion proteins *Biophys. J.* **98** 158–63
- [20] Barlag B et al 2016 Single molecule super-resolution imaging of proteins in living *Salmonella enterica* using self-labelling enzymes *Sci. Rep.* **6** 31601
- [21] Ke N, Landgraf D, Paulsson J and Berkmen M 2016 Visualization of periplasmic and cytoplasmic proteins with a self-labeling protein tag *J. Bacteriol.* **198** 1035–43
- [22] Talyor H, Lepore A, Landgraf D, Okumus B, Jaramillo-Riveri S, McLaren L, Paulsson J and Karoui M E 2017 Quantification of very low-abundant proteins in bacteria using the HaloTag and epi-fluorescence microscopy 237248 (bioRxiv) (<https://doi.org/10.1101/237248>)
- [23] Rhodes J D, Mazza D, Nasmyth K A and Uphoff S 2017 Scc2/Nipbl hops between chromosomal cohesin rings after loading *eLife* **6** e30000
- [24] Norris S R, Núñez M F and Verhey K J 2015 Influence of fluorescent tag on the motility properties of kinesin-1 in single-molecule assays *Biophys. J.* **108** 1133–43
- [25] Grimm J B et al 2015 A general method to improve fluorophores for live-cell and single-molecule microscopy *Nat. Methods* **12** 244–50
- [26] Grimm J B et al 2016 Bright photoactivatable fluorophores for single-molecule imaging *Nat. Methods* **13** 985–8
- [27] Lukinavičius G et al 2013 A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins *Nat. Chem.* **5** 132–9
- [28] Rybenkov V V, Herrera V, Petrushenko Z M and Zhao H 2014 MukBEF, a chromosomal organizer *J. Mol. Microbiol. Biotechnol.* **24** 371–83
- [29] Badrinarayanan A, Reyes-Lamothe R, Uphoff S, Leake M C and Sherratt D J 2012 *In vivo* architecture and action of bacterial structural maintenance of chromosome proteins *Science* **338** 528–31
- [30] Reyes-Lamothe R, Nicolas E and Sherratt D J 2012 Chromosome replication and segregation in bacteria *Annu. Rev. Genet.* **46** 121–43
- [31] Uphoff S, Reyes-Lamothe R, de Leon F G, Sherratt D J and Kapanidis A N 2013 Single-molecule DNA repair in live bacteria *Proc. Natl Acad. Sci. USA* **110** 8063–8
- [32] Friedberg E C 2006 *DNA Repair and Mutagenesis* 2nd edn (Washington, DC: American Society for Microbiology)
- [33] Datsenko K A and Wanner B L 2000 One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products *Proc. Natl Acad. Sci.* **97** 6640–5
- [34] Wegel E, Göhler A, Lagerholm B C, Wainman A, Uphoff S, Kaufmann R and Dobbie I M 2016 Imaging cellular structures in super-resolution with SIM, STED and localisation microscopy: a practical comparison *Sci. Rep.* **6** 27290
- [35] Tokunaga M, Imamoto N and Sakata-Sogawa K 2008 Highly inclined thin illumination enables clear single-molecule imaging in cells *Nat. Methods* **5** 159–61

- [36] Uphoff S, Sherratt D J and Kapanidis A N 2014 Visualizing protein-DNA interactions in live bacterial cells using photoactivated single-molecule tracking *J. Vis. Exp.* **85** e51177
- [37] van de Linde S, Krstić I, Prisner T, Dooze S, Heilemann M and Sauer M 2011 Photoinduced formation of reversible dye radicals and their impact on super-resolution imaging *Photochem. Photobiol. Sci.* **10** 499–506
- [38] Annibale P, Vanni S, Scarselli M, Rothlisberger U and Radenovic A 2011 Quantitative photo activated localization microscopy: unraveling the effects of photoblinking *PLoS One* **6** e22678
- [39] Lee S-H, Shin J Y, Lee A and Bustamante C 2012 Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM) *Proc. Natl Acad. Sci.* **109** 17436–41
- [40] Coltharp C, Kessler R P and Xiao J 2012 Accurate construction of photoactivated localization microscopy (PALM) images for quantitative measurements *PLoS One* **7** e51725
- [41] Sengupta P, Jovanovic-Talisman T and Lippincott-Schwartz J 2013 Quantifying spatial organization in point-localization superresolution images using pair correlation analysis *Nat. Protocols* **8** 345–54
- [42] Fricke F, Beaudouin J, Eils R and Heilemann M 2015 One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localization microscopy *Sci. Rep.* **5** 14072
- [43] Manley S, Gillette J M, Patterson G H, Shroff H, Hess H F, Betzig E and Lippincott-Schwartz J 2008 High-density mapping of single-molecule trajectories with photoactivated localization microscopy *Nat. Methods* **5** 155–7
- [44] Dersch S and Graumann P L 2018 The ultimate picture-the combination of live cell superresolution microscopy and single molecule tracking yields highest spatio-temporal resolution *Curr. Opin. Microbiol.* **43** 55–61
- [45] Elf J, Li G-W and Xie X S 2007 Probing transcription factor dynamics at the single-molecule level in a living cell *Science* **316** 1191–4
- [46] English B P, Hauryliuk V, Sanamrad A, Tankov S, Dekker N H and Elf J 2011 Single-molecule investigations of the stringent response machinery in living bacterial cells *Proc. Natl Acad. Sci.* **108** E365–73
- [47] Gebhardt J C M, Suter D M, Roy R, Zhao Z W, Chapman A R, Basu S, Maniatis T and Xie X S 2013 Single-molecule imaging of transcription factor binding to DNA in live mammalian cells *Nat. Methods* **10** 421–6
- [48] Wäldchen S, Lehmann J, Klein T, van de Linde S and Sauer M 2015 Light-induced cell damage in live-cell super-resolution microscopy *Sci. Rep.* **5** 15348
- [49] Icha J, Weber M, Waters J C and Norden C 2017 Phototoxicity in live fluorescence microscopy, and how to avoid it *BioEssays* **39** 1700003
- [50] Shcherbakova D M, Baloban M, Emelyanov A V, Brenowitz M, Guo P and Verkhusha V V 2016 Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging *Nat. Commun.* **7** 12405
- [51] Balleza E, Kim J M and Cluzel P 2018 Systematic characterization of maturation time of fluorescent proteins in living cells *Nat. Methods* **15** 47–51