THE UNIVERSITY
of EDINBURGH

Quantitative Detection of Low Abundance Gene Expression Products in Individual *E. coli* Cells

Hannah L. Taylor

Thesis presented for the degree of Doctor of Philosophy

Institute of Cell Biology

The University of Edinburgh

August 2017

Declaration

I hereby declare that this thesis was composed by me, and that the research presented is my own, except where otherwise stated. This work has not been submitted for any other degree or professional qualification.

Hannah L. Taylor

August 2017

Acknowledgements

I would like to thank Meriem El Karoui for giving me the opportunity to carry out the research for my Ph.D. in her lab and under her supervision. Her guidance and support were indispensable. Alongside Meriem, I would like to thank Lorna McLaren, Alessia Lepore, Sebastian Jaramillo-Riveri and Xavier Zaoui for the help, encouragement and advice they have given me, and for making the lab such a fun place to work.

Outside the lab, my parents, Ed and Doug Taylor, my sister Robyn Taylor and her partner Asaad Baksh have been a constant source of support and encouragement and their contribution has been invaluable. The same goes for my wonderful partner Nathaniel Melman.

Finally I would like to thank my excellent flatmates, Colette Bush and Sarah Noonan, who were somehow able to bear with me throughout.

Abstract

Stochastic fluctuations in mRNA and protein copy number between cells are inevitable during the process gene expression, even when cells carry identical chromosomes. Such fluctuations are able to impact the phenotypic fate of the cell, and are known to have greater impact when the copy number of the molecule involved is low. Additionally, up to 50% of proteins in *Escherichia coli* are present in the cell at a level of 10 molecules per cell or fewer (Taniguchi *et al.* 2010). As such, quantification of low copy number gene expression products and their distribution in cellular populations is key in understanding the process of gene expression. Currently, there are few techniques that allow investigation with the single cell and single molecule resolution required to study low copy number gene expression products. This work presents a novel method for protein quantification at the single molecule level, Quantitative HaloTag-TMR labelling, and uses the technique to quantify the absolute numbers of the low copy number RecB, RecC and RecD subunits of the bacterial DNA repair enzyme RecBCD, finding each subunit is present at between two and eight molecules per cell with mean numbers per cell of 4.9, 4.7 and 4.5 respectively. Additionally single molecule mRNA FISH was used to quantify the mRNA levels of *recB* and *recD* within cells, with means of 0.21 and 0.31 mRNA per cell being observed respectively. Finally this work presents a new method for use detecting both mRNA and protein simultaneously in individual cells by combining the HaloTag and FISH protocols to give HaloFISH. This work introduces two novel techniques that allow for single cell examination of gene expression, and investigates RecBCD expression at the single molecule level.

Abbreviations

bp	Base pair
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
FISH	Fluorescent *in situ* hybridisation
FP	Fluorescent protein
G	Gram
kDa	Kilodalton
l	Litre
M	Molar
mg	Milligram
ml	Millilitre
mM	Millimolar
mRNA	Messenger RNA
Nal	Naladixic acid
ng	Nanogram
nm	Nanometre
nt	Nucleotide
OD_{600}	Optical density at 600 nm
RecBHalo	RecB protein fused to HaloTag
RecBHcoH1	RecB protein fused to one HaloTag and one codon optimised HaloTag, joined with a 6 amino acid linker
RecBHcoH2	RecB protein fused to one HaloTag and one codon optimised HaloTag, joined with a 31 amino acid linker
RecCHalo	RecC protein fused to HaloTag
RecDHalo	RecD protein fused to HaloTag
RNA	Ribonucleic acid

rRNA	Ribosomal ribonucleic acid
smFISH	Single molecule FISH
TMR	Tetramethylrhodamine
tRNA	Transfer RNA
µg	Microgram
µl	Microliter
µm	Micrometre
µM	Micromolar
UTR	Untranslated region

Table of Contents

Table of Contents…… VI

Chapter 1: Introduction…… 1

1.1 Bacterial gene expression…….. 1
1.2 Stochastic fluctuations in living systems……... 5
1.3 Stochasticity and low level expression…….. 9
1.4 Measuring stochastic fluctuations……. 11
1.4.1 Current methods in mRNA detection…… 12
1.4.1.1 Population methods for mRNA detection and quantification……. 12
1.4.1.2 Single cell methods for mRNA detection and quantification……. 15
1.4.1.2.1 RT-qPCR……... 15
1.4.1.2.2 Single cell RNA-Seq……. 17
1.4.1.2.3 MS2 and PP7 aptamers…….…… 20
1.4.1.2.4 Fluorescent RNA aptamers…… 24
1.4.1.2.5 mRNA FISH……..…… 26
1.4.1.2.6 mRNA detection summary…… 28
1.4.2 Current methods in protein detection…… 29
1.4.2.1 Population methods for protein detection and quantification……. 30
1.4.2.2 Single cell methods for protein detection and quantification……..… 34
1.4.2.2.1 Flow cytometry and microfluidic flow cytometry…….…… 34
1.4.2.2.2 Separation based methods……..… 35
1.4.2.2.3 Genetic and chemical probes…….…… 36
1.4.2.2.4 Simultaneous detection of mRNA and protein in single cells…… 41
1.4.2.2.5 Protein detection summary…… 42
1.5 RecBCD copy number……...…… 43
1.5.1 RecBCD expression……..…… 44
1.5.2 RecBCD structure and function……...…… 46
1.6 Scope of this thesis…… 49

Chapter 2: Materials and Methods…….. 51

2.1 Materials…… 51
2.1.1 Stock solutions……... 51
2.1.2 Culture media……... 53

2.1.3 Imaging media……... 54
2.1.4 Buffers……... 55
2.1.5 Fluorophores…….. 56
2.2 Methods……. 64
2.2.1 Bacterial methods…….. 64
2.2.2 DNA cloning techniques……... 68
2.2.3 Phenotypic testing……. 73
2.2.4 Protein and mRNA detection techniques……. 75
2.2.4.1 Protein detection by western blot…….. 75
2.2.4.2 Quantitative labelling with HaloTag-TMR……... 77
2.2.4.3 Labelling of mRNA with single molecule fluorescence in situ hybridisation (smFISH)…….. 78
2.2.4.4 Fluorescent in situ hybridisation combined with HaloTag labelling……. 80
2.2.4.5 Overexpression of mRNA or protein with arabinose…… 82
2.2.5 Microscopy…… 83
2.2.5.1 Preparation of 2% agarose pad for imaging…….. 83
2.2.5.2 Imaging conditions..... 84
2.2.5.3 Image analysis……... 85
2.2.6 DNA primers, DNA sequences and gBlocks, bacterial strains and plasmids……... 88
2.2.6.1 DNA primers……. 88
2.6.2.2 DNA sequences and gBlocks…… 92
2.2.6.3 *E. coli* strains……. 95
2.2.6.4 Plasmids……. 95

Chapter 3: Developing Quantitative single cell HaloTag-TMR labelling in E. coli and application of the method to each subunit of the bacterial DNA repair protein RecBCD…… 97

3.1 Introduction……... 97
3.2 Characterization of the method……... 98
3.2.1 Specific detection of HaloTag-TMR…….. 99
3.2.2 Detection of a protein across a wide range of expression levels…….. 96
3.2.3 Detection of low copy number protein at single molecule level……... 99
3.2.4 Assessment of HaloTag-TMR labelling efficiency…….. 107
3.2.5 Assessment of HaloTag-TMR labelling reproducibility…. 118
3.3 Quantitative detection of RecB, RecC and RecD……. 119
3.3.1 Quantification of RecBHalo……... 119
3.3.2 Phenotypic and viability testing of RecDHalo and RecCHalo…….. 121
3.3.3 Quantification of RecDHalo…….. 123
3.3.4 Quantification of RecCHalo…….. 125

3.4 Conclusions and future work……. 127
3.4.1 Conclusions……... 127
3.4.2 Future work…….. 128

Chapter 4: Establishing single molecule mRNA FISH for quantitative mRNA detection, applying the method to *rec*D and *rec*B mRNA and introducing combined HaloFISH…… 129

4.1 Introduction…….. 129
4.2 Establishing mRNA detection with smFISH……. 130
4.2.1 mRNA FISH is able to detect r*ecD* mRNA specifically…….. 131
4.3 Quantitative detection of *recB* and *rec*D mRNA……. 138
4.3.1 Quantification of *recD* transcript number…… 138
4.3.2 Quantification of *recB* transcript number…… 142
4.3.3 Comparison of *recD* and *recB* transcript number…….. 144
4.4 Combining HaloTag-TMR and smFISH to label protein and mRNA simultaneously within cells …….. 145
4.4.1 Protocol combination…… 145
4.4.2 Proof of principle…….. 147
4.5 Conclusions and further work…….. 150
4.5.1 smFISH…….. 150
4.5.2 HaloFISH…….. 151

Chapter 5: Discussion…… 153

5.1 HaloTag-TMR labelling of RecB, RecC and RecD protein……. 153
5.1.1 Quantitative HaloTag-TMR labelling of low copy number protein…… 153
5.1.2 Quantification of RecB, RecC and RecD protein…… 156
5.2 smFISH and HaloFISH labelling…… 159
5.2.1 smFISH labelling of mRNA…… 159
5.2.2 Quantification of *rec*B and *rec*D mRNA…… 161
5.3 HaloFISH detection of mRNA and protein…….. 162
5.4 Conclusion…… 163

References …… 165
Appendix …… 174

Table of Figures

Chapter 1:

Figure 1.1. Transcription elongation…… 4

Figure 1.2. Living systems are subject to stochastic fluctuations…… 6

Figure 1.3. HipA overexpression causes the cell to enter a state of growth arrest……..... 8

Figure 1.4. MS2 tagging of mRNA……..... 21

Figure 1.5. PP7 tagging of mRNA……..... 23

Figure 1.6. mRNA labelling with the spinach aptamer……..... 25

Figure 1.7. GFP β-barrel structure……..... 37

Figure 1.8. HaloTag, SNAP and CLIP tags expressed with protein bind to fluorogenic ligands……..... 40

Figure 1.9. Chromosomal structure of *recC*, *recB* and *recD*……..... 45

Figure 1.10. Model of RecBCD enzyme mechanism..... 49

Chapter 2:

Figure 2.1. Plasmid mediated gene replacement…….. 67

Figure 2.2. Polymerase chain reaction for gibson assembly…… 69

Figure 2.3. Gibson assembly reaction…….. 71

Figure 2.4. Labelling of protein with the HaloTag……... 78

Figure 2.5. Labelling of mRNA with mRNA FISH……. 80

Figure 2.6. Labelling of protein and mRNA with HaloFISH…… 82

Chapter 3:

Figure 3.1. Specific detection of TMR bound to induced HaloTag protein…… 99

Figure 3.2. Mean fluorescence detection per cell following induction at

varying concentrations of arabinose induction……..... 100

Figure 3.3. Culture-averaged fluorescence (fluorescence/OD_{600}) of *E. coli* cultures containing promoters of different strengths for *ara*E……... 101

Figure 3.4. RecBHalo strain function and viability……..... 103

Figure 3.5. Specific detection of TMR bound to low copy number RecBHalo…….. 105

Figure 3.6. The HaloTag-TMR protocol allows visualization of single TMR molecules which can then be quantified…… 107

Figure 3.7. RecB double HaloTag constructs and experimental design……. 109

Figure 3.8. Distributions of foci in RecBHalo/RecBHcoH1……..... 111

Figure 3.9. Distributions of foci in RecBHalo/RecBHcoH2……..... 113

Figure 3.10. Distributions of foci detected with 0.05µM, 0.5µM and 5µM TMR in single and double HaloTag strains……..... 115

Figure 3.11. Distributions of RecB foci when labelled with HaloTag-TMR and GFP……..... 118

Figure 3.12. Distributions of RecB foci when observing cells <3.5µm and cells of all lengths……..... 120

Figure 3.13. Phenotypic and viability assays for RecCHalo and RecBHalo…….. 122

Figure 3.14. Distributions of RecD foci when labelled with HaloTag-TMR…….. 124

Figure 3.15. Distributions of RecC foci when labelled with HaloTag-TMR…….. 126

Chapter 4:

Figure 4.1. Phenotypic and viability assays for Δ*rec*DMG…….. 132

Figure 4.2. mRNA FISH detects *rec*D mRNA specifically in single cells…… 133

Figure 4.3. mRNA FISH detects *rec*D mRNA specifically across the population……..... 134

Figure 4.4. *rec*D overexpression is detected in 87% of cells……... 135

Figure 4.5. mRNA FISH can detect changes in *rec*D mRNA concentration across a wide range of induction levels…….... 137

Figure 4.6. *rec*D quantification with smFISH……..... 141

Figure 4.7. *recB* quantification with smFISH……..... 143

Figure 4.8. Distributions of *rec*D and *recB* mRNA……..... 145

Figure 4.9. Simultaneous labelling of highly expressed mRNA and protein is possible with HaloFISH……..... 148

Figure 4.10. Signal is detected on induction of mRNA and protein expression with arabinose, and the amount of signal increases as the level of induction increases……. 149

Appendix:

Figure A.1. Phenotypic tests for double HaloTag strains…… 171

Figure A.2. Western blot and growth curves for RecCHalo and RecDHalo……..... 172

Table of Tables

Chapter 1:

Table 1.1. Methods for RNA detection……. 29

Table 1.2. Methods for protein detection…….. 43

Chapter 2:

Table 2.1. DNA oligonucleotide probes for *recB* mRNA……. 57

Table 2.2. DNA oligonucleotide probes for recD mRNA……. 60

Table 2.3. DNA oligonucleotide probes for *HaloTag* mRNA……... 62

Table 2.4. DNA Primers……..... 88

Table 2.5. HaloTag, codon optimized HaloTag and large and small linkers……..... 92

Table 2.6. gBlocks…….. 93

Table 2.7. *E. coli* strains……..... 95

Table 2.8. Plasmids……. 96

Chapter 1

Introduction

1.1 Bacterial gene expression

The circular *Escherichia coli* (*E. coli*) chromosome consists of almost 4.7 million base pairs and the organism is undoubtedly one of the best-studied models in genetics (Kohara *et al*. 1987). There are, however, many unanswered questions surrounding the chromosome and the expression of the approximately 4000 genes it contains (Blattner *et al*. 1997). One such question concerns the observation that cells coming from isogenic populations can express their identical genomes differently. Such differences arise due in part to unavoidable stochasticity that occurs during the processes of transcription and translation. Transcription is the process responsible for reading the nucleotide sequence of protein coding DNA and producing mRNA, and translation is the process by which this mRNA is read and an amino acid chain is polymerized into protein which can be utilized by the cell. The central dogma of molecular biology can be defined as the passage of information through this process, from DNA to protein and the inability of information to pass backwards (Crick 1970).

The first stage of gene expression in bacteria is the transcription of the information contained within deoxyribonucleic acid (DNA) into ribonucleic acid (RNA). The process of transcription is catalysed in all organisms by RNA polymerase (RNAP), a conserved multi-subunit complex (Murakami 2015). In

eukaryotic cells there are three distinct RNA polymerases, each transcribing a different class of RNA while in prokaryotes the core RNAP (made up of two α subunits, one β subunit, one β′ subunit and one ω subunit) links transiently to a number of different σ factors to form a holoenzyme. The binding of specific σ factors allows the enzyme to bind to promoter elements to initiate the transcription of protein coding genes (Cooper & Hausman 2008; Hurwitz 2005). After RNAP binds to the promoter the polymerase unwinds ~15 bps of DNA around the initiation site to form an open promoter complex in which single stranded DNA is available as a template for transcription. Transcription is initiated by the joining of two free nucleotide triphosphates (NTPs), and the σ factor is released after about 10 NTPs have been added to the nascent chain. RNAP continues move along the DNA, unwinding the base pairs in front of it and rewinding those behind, giving a 17 bp transcription bubble. RNA synthesis continues until RNAP encounters a termination sequence, when this happens a RNAP dissociates from the DNA template and mRNA transcript is produced (Cooper & Hausman 2008).

As bacteria do not possess a membrane enclosed nucleus, transcription and translation are coupled and can occur in direct sequence (Gowrishankar & Harinarayanan 2004). Translation is carried out by the 70S ribosome, which is composed of two subunits – the smaller 30S subunit and the larger 50S subunit. The 30S subunit consists of 16s rRNA and 21 ribosomal proteins (S1-S21) while the 50S subunit is composed of a 5S and a 23S rRNA and 33 ribosomal proteins (L1-35) (Starosta *et al*. 2014; Schmeing & Ramakrishnan 2009). Once assembled the 70S subunit allows the composition of polypeptide chains. The 30S subunit is responsible

for ensuring the mRNA is correctly positioned and establishing the reading frame of the protein. The 50S subunit contains the site of peptide bond formation – the peptidyl transferase centre. There are 3 binding sites in the 70S ribosome, these are the A-, P-, and E-sites. The A-site receives the aminoacyl-tRNA, and the P-site holds the peptidyl tRNA for peptide bond formation. The E-site forms the exit for uncharged or deacytlated tRNAs to move through before exiting the ribosome (Starosta *et al*. 2014). Translation initiation involves the assembly of the 70S subunit and the positioning of the fMet-tRNA (which will become the first amino acid in the nascent peptide) with the mRNA start codon at the P-site of the ribosome (subsequent aminoacyl tRNAs will enter in the A site as described above). Translation initiation is facilitated with the assistance of initiation factors 1, 2 and 3. Elongation follows initiation (a schematic summary of translational elongation can be seen in Fig.1.1). During the first round of elongation, elongation factor EF-Tu-GTP delivers the correct aminoacyl-tRNA to the A site of the ribosome, where interaction with the mRNA codon triggers hydrolysis of the GTP to GDP by EF-Tu and EF-Tu-GDP dissociates from the ribosome. The aminoacyl-tRNA moves into the peptidyl transferase centre and undergoes peptide bond formation with the fMet-tRNA in the P-site. This results in a deacytlated tRNA in the P-site and a peptidyl-tRNA in the A site. Binding of a second elongation factor, EF-G-GTP is hydrolysed to EF-G-GDP to catalyse the translocation of the ribosome, which moves one frame along the DNA moving the deacytlated tRNA into the E-site and the peptidyl-tRNA into the P-site. The E-G-GDP dissociates from the ribosome and a new aminoacylt-tRNA is able to move into the A-site. The appearance of a stop codon (recognised by

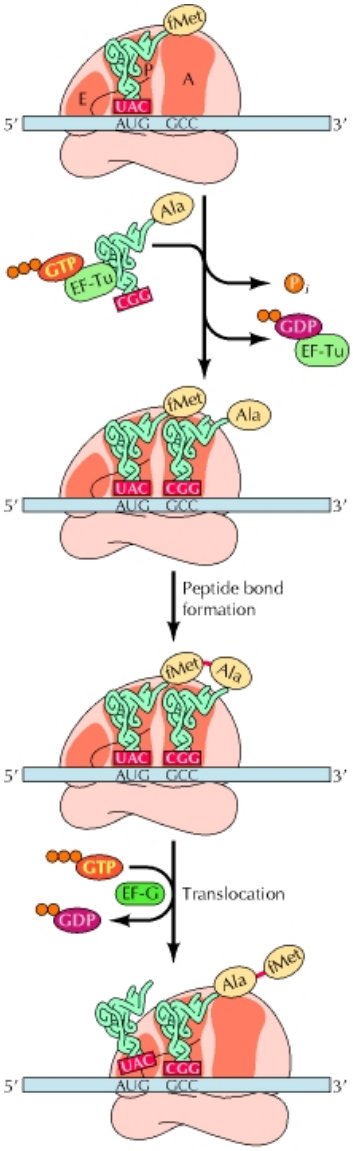


Figure 1.1. Translation Elongation. Ribosomes have three tRNA-binding sites, peptidyl (P), aminoacyl (A), and exit (E). The initiating fMet-tRNA is positioned in the P site, and the second aminoacyl tRNA (here alanyl tRNA) is brought into the A site by EF-Tu-GTP. GTP is hydrolysed, and the EF-Tu (now complexed with GDP) leaves the ribosome. A peptide bond is formed that results in the transfer of the methionine to the aminoacyl tRNA at the A site. The ribosome then moves three nucleotides along the mRNA, which translocates the peptidly (Met-Ala) tRNA to the P site and the uncharged tRNA to the E site, leaving the E site empty for the addition of another amino acid. The translocation is mediated by EF-G, coupled to GTP hydrolysis. (Figure taken from Cooper & Hausman 2008).

release factors 1 or 2) in the A-site signals translational termination. Termination occurs when the polypetptide chain is released from the ribosomes by hydrolysis of the ester linkage to the P-site tRNA. The post termination ribosome is then dissociated into ribosomal subunits (Starosta *et al*. 2014)

As described, transcription and translation are both precisely controlled in the cell. In order to understand the process of gene expression, and to allow modelling

of the system absolute quantification of mRNA and protein numbers are required. This thesis focuses on development of techniques to allow for absolute quantification of both mRNA and protein, with a focus on low copy number molecules where, as described below, the impact of stochastic fluctuation can be greater and existing techniques are less developed.

1.2 Stochastic fluctuations in living systems

Bacterial cells depend on the accurate propagation of information from gene to RNA transcript to protein to survive, and the accurate replication and propagation of their genome to reproduce. In bacterial systems, all cells in a clonal population are isogenic and carry and express almost identical chromosomes. Stochastic fluctuation, however, is ubiquitous in biological systems. This is the case because some of the components involved are present in such low numbers (DNA, for example, is regularly present at only one copy per cell) that fluctuations in the proteins responsible for gene expression such as polymerases and transcription factors can impact upon reaction rates (Pedraza *et al*. 2005). Stochastic fluctuation in bacterial systems was clearly illustrated by Elowitz *et al*. (2002) who were able to construct strains incorporating two fluorescent proteins –CFP and YFP on opposite arms of the *E. coli* chromosome, equidistant from the origin of replication (to control for the effect of variable chromosome copies). Each FP was placed under the control of the *lac* promoter. When the promoter was imperfectly repressed by the wild type *lacI* gene and the fluorescence emitted by each fluorescent protein observed by

microscopy, it is evident that individual cells express the proteins at different levels, as can be seen in Fig. 1.2B.

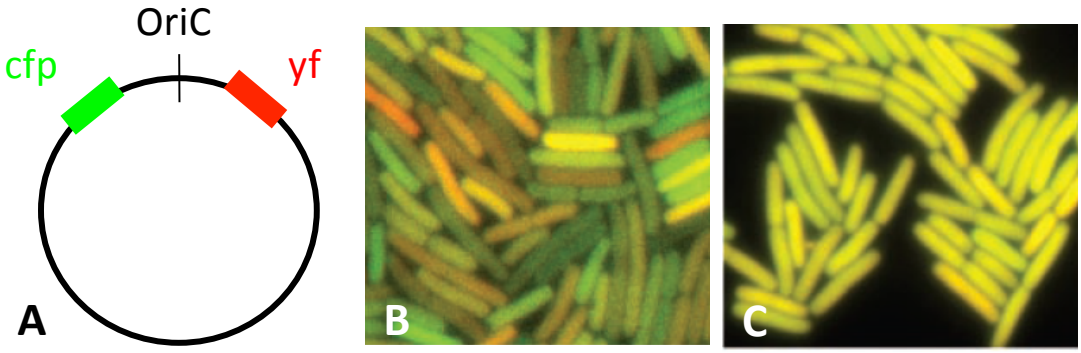


Figure 1.2. Living systems are subject to stochastic fluctuations A) Schematic representation of the *E. coli* chromosome bearing CFP (false coloured green) and YFP (false coloured red) equidistant from oriC. B) Cells from an isogenic population express CFP and YFP at different levels when the promoter controlling expression is imperfectly repressed. C) Heterogeneity is reduced, but not eliminated, when the expression of CFP and YFP is induced. (Adapted from Elowitz *et al*. 2002).

Stochastic fluctuations occur between cells of otherwise isogenic populations. While intuitively fluctuations in a system may appear like an undesirable property, this is not necessarily the case. In fact, fluctuation has been shown to be beneficial in some cases, and living systems are able to function in the presence of stochastic fluctuations (Thattai & van Oudenaarden 2001). Key functional advantages of the presence of fluctuations include probabilistic differentiation of otherwise genetically identical cells, permitting such strategies as bet hedging and division of labour. These strategies would be difficult to implement in deterministic isogenic populations (Li & Xie 2011).

An example of such beneficial stochasticity is the existence of phenotypic switches within clonal bacterial populations such as those that are believed to control the phenomenon of persistence. Persistent bacteria are those that, although genetically sensitive to antibiotics, display a growth arrest phase that allows survival of antibiotic presence. The cells are then able to grow normally once again antibiotic stress has been removed. The link between persistence and phenotypic heterogeneity that results from varied gene expression in *E. coli* was described by Balaban *et al*. (2004) who combined microfluidics and single cell microscopy to conclude that the existence of persisters can be attributed to the heterogeneity of growth rates that can be observed within bacterial populations.

A potential mechanism for such phenotypic switching was further described by Rotem *et al*. (2010). The authors describe a toxin-antitoxin module, which is implicated in the persistence of *E. coli*, and highlight the advantage conveyed by this transient phenotypic resistance to antibiotics. Toxin- antitoxin modules consist of pairs of genes that are usually found in the same operon. One gene will act as a toxin and the other gene will cancel its effect. Toxin-antitoxin modules were first discovered on plasmids but also exist on bacterial chromosomes. Rotem *et al*. worked with the hipBA toxin-antitoxin module where the HipA is the toxin and the HipB the antitoxin. Together, the proteins form a tight complex and repress their own expression. If the HipA toxin is overexpressed above a threshold then the cells enter a state of growth arrest which can be reversed by the expression of the HipB antitoxin. The authors used single cell fluorescence microscopy studies to analyse the role of the HipBA toxin- antitoxin module in the production of alternate phenotypes

in clonal populations of *E. coli*. The authors used a strain of *E.coli* in which the HipB antitoxin was expressed under the control of the native promoter and the HipA toxin was placed on a plasmid under the control of the tet promoter and fused to the fluorescent protein mCherry. Fluorescence time-lapse microscopy of this strain showed that growth arrested and rapidly growing cells were able to grow together and that within the same population HipA expression varied. The cells in which HipA was expressed above a threshold level showed a growth arrest phenotype while those that expressed HipA at a level below this threshold did not. The length of the growth arrest was determined by the size of the HipA excess within the cell. This can be seen in Fig. 1.3 below.

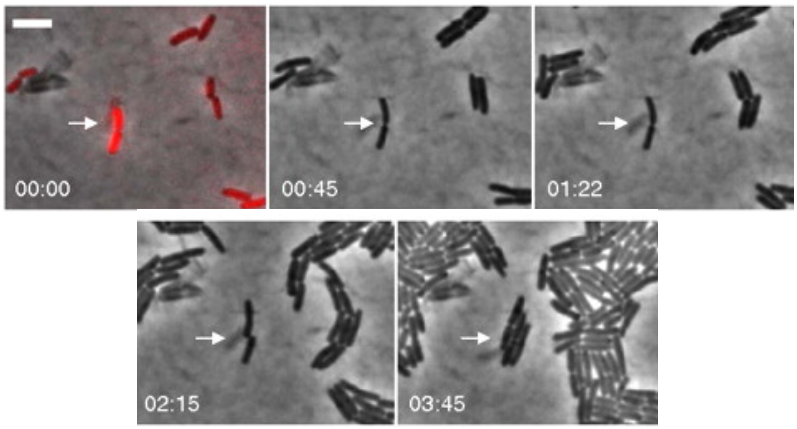


Figure 1.3. **HipA overexpression causes the cell to enter a state of growth arrest**. The first image (00:00) shows an overlay of fluorescence and phase contrast images. The fluorescence shows the relative HipA-mCherry expression across the isogenic population. The arrow highlights a cell that has an elevated level of HipA toxin and displays a growth arrested phenotype throughout the timelapse, with growth beginning over two hours after the other cells in the population. (Scale bar 4 μm). (Adapted from Rotem et al. 2010).

This example, where genetically identical cells can have drastically different phenotypes which can allow a small portion of the cellular population to survive environmental stress such as the presence of antibiotics is an excellent illustration of the fact that stochasticity, while disruptive of a highly organized system, can in fact be beneficial within living systems and can explain phenomena such as persistence.

Stochastic fluctuations can be controlled by regulation. However, such regulation can only constrain these fluctuations rather than preventing them entirely. Molecular noise is unavoidable in cells.

1.3 Stochasticity and low level expression

As discussed above, fluctuations are present in all living systems, and can have measurable impact on the phenotypes of these systems. This impact is greater when the absolute numbers of protein or mRNA involved are low. Returning to Fig 1.2, it can be seen that in an imperfectly repressed system both CFP and YFP are transcribed at different levels and microscopy reveals a noticeable heterogeneity between the cells of the population. A further figure from the same publication (Elowitz *et al*. 2002) shows that when these promoters are induced and the fluorescent proteins are overexpressed the extent of the heterogeneity is less obvious, although there is still variability (see Fig 1.2C).

In *E. coli* protein abundance varies widely from gene to gene. A study by Taniguchi *et al*. (2010) in which an YFP fusion library was made in *E. coli* by converting the C-terminal tags of an existing chromosomally affinity tagged library to YFP translational fusions. Each strain in the library had a particular gene tagged,

and 1081 strains were produced that showed no significant growth defects. The strains were analysed using a microfluidic imaging platform that allowed fluorescent imaging at the single molecule level. The study found that average protein abundance varied by five orders of magnitude from 10^{-1} molecules to 10^{4} molecules per cell. Genes essential to cellular maintenance were found to have higher abundance than other genes. 121 essential genes were present in the library and of these 108 expressed at 10 molecules or more per cell. However, around 50% of all the proteins measured were found to be present at less than 10 molecules per cell. This high prevalence of low abundance proteins, combined with the observable fact that low abundance can amplify the effect of stochastic fluctuations in systems makes a compelling argument for the study of single cells and single molecules. This is particularly true because of the proteins looked at in this study alone, 40% of the low abundance proteins remain unannotated, highlighting the current difficulty of detecting these proteins.

Within the same study RNA was examined at the single molecule level through mRNA fluorescent *in situ* hybridization (FISH), although only for those proteins that were expressed at over 100 molecules per cell. The average RNA content for each of the 137 strains examined ranged between 0.05 to 5 mRNAs per cell. The authors also reported that there was only moderate correlation between the mean mRNA number and the mean protein number between cells, citing the fact that mRNA and protein have very different life cycles within the cell. In *E. coli*, mRNAs are typically degraded within minutes, whereas most proteins have a lifetime that is longer than the lifecycle of the cell. This means that the mRNA content of a cell gives

information only about the very recent history of the cell, whereas the protein content represents a longer history of accumulated expression. It is worth noting that authors also imaged the YFP tag and mRNA FISH probes against these tags simultaneously in single cells, and here did not detect any correlation between mRNA and protein copy number, which indicates a global lack of correlation between the mRNA and protein levels of single gene products within single cells.

As can be seen, many genes are expressed at low copy number in *E. coli*, producing low (<10) numbers of proteins per cell. Together with the transient nature of mRNA molecules which have degradation times of only a few minutes, and the fact that all systems are subject to noise and that the impact of such noise is greater at low copy number it can be seen that single molecule, single cell observations are critical to further understanding gene expression in bacteria.

1.4 Measuring stochastic fluctuations

The experiments described above all aim to measure stochastic fluctuations in genetic systems. To do so, each experiment has to examine individual cells within populations, allowing an understanding of the heterogeneity that exists in the population of the protein or mRNA of interest and allowing comparisons between the distributions of different strains. To capture stochastic fluctuations in gene products from low expression genes it is necessary to go a step further and to quantify mRNA and protein at the single molecule level, as this resolution is required to measure such molecular concentrations accurately.

1.4.1 Current methods in mRNA detection

The presence of stochastic fluctuations in gene expression is unavoidable, and they can have real functional consequences for the cells and populations they occur in. It has become clear that, while informative, bulk studies in which the only quantification produced is the population mean do not shed light on fluctuations within populations and can potentially eclipse subpopulations within the whole that may be displaying different phenotypic behaviour. Below I shall briefly discuss current population level methods for mRNA detection and quantification, before moving on to outline current methods used in the detection and quantification of mRNA in single bacterial cells, highlighting techniques that show single-cell capability.

1.4.1.1 Population methods for mRNA detection and quantification

As detailed in Taniguchi *et al*. (2010) above, mRNA generally have lifetimes of only a few minutes within cells, and are frequently present at very low copy number (0.05-5 molecules per cell). In general, RNA study presents a challenge due to the susceptibility of the molecule to digestion by RNase enzymes, which are prevalent in laboratory environments as they are present in media contaminated with microbes and are secreted by humans (Harder & Schröder 2002). It is nevertheless possible to analyse RNA in bulk studies. For example this can be done through Northern blotting, where total RNA is extracted from a bulk population of cells and then separated by size through gel electrophoresis. The RNA samples are then transferred onto a nylon membrane, where they are labelled with a detectable (chemiluminescent

or radioactive) probe. The signal from the hybridized probe is then detected by X-ray film. The northern blot is advantageous in that it allows visualisation and rough estimation of RNA quantity through comparison with endogenous RNA standards (such as 5S rRNA). However, despite continual amendment and improvement to this method, which has allowed increasing accuracy and increasingly small mRNAs to be detected (Beckmann *et al*. 2010) as well as the detection of low copy number mRNAs (Kim et al. 2010) the results given remain at population level, necessarily summing the total mRNA of a population for analysis, rather than of a single cell, additionally only RNA of known sequence can be analysed.

Other bulk methods, such as RNA detection microarrays suffer from similar drawbacks in terms of inability to produce absolute quantification of specific mRNA from single cells and detection only of known mRNA. Such assays involve the binding of complementary nucleic acid sequences. Total RNA is isolated from cells of interest and (in prokaryotes) enriched for mRNA. In microarrays, columns or beads with sequences complimentary to 16s rRNA are used to remove rRNA and the remaining mRNA can then be labelled with e.g. fluorescence tags or biotin. The results are then visualized and quantified to give relative expression levels of the mRNAs being probed for within the sample taken (Bumgarner 2013). As stated, microarrays do not produce absolute numbers in quantification but rather measure relative concentration, and can do so with limited accuracy due to differences in hybridization kinetics of the different species of mRNA being detected and the impact that will have on the relative abundance of mRNAs binding. Additionally, a microarray can detect only the sequences that they array was designed for, meaning

that any mRNA species whose sequences are unknown or that are not being probed for directly will be missed. These drawbacks combined with advances in technology and corresponding reduction in cost of sequencing technology has meant that the more direct method of RNA sequencing has replaced microarrays as the most efficient means of examining gene expression in bulk populations.

Reverse Transcriptase quantitative PCR (RT-qPCR) is based on the action of the reverse transcriptase enzyme, which is able to generate cDNA sequences from RNA templates (Bustin 2000; Ginzinger 2002). For use in bulk bacterial studies, total RNA is extracted from a population of homogenized cells, before being converted into cDNA. The cDNA will then be subject to quantitative PCR. During qPCR data is collected as the PCR process occurs, combining amplification and detection. This is done through utilisation of fluorescent DNA labelling, which allows correlation of PCR product to fluorescence intensity. qPCR reactions have four main phases- the linear ground phase, the early exponential phase, the exponential phase and the plateau. During the linear ground phase PCR has just started and the fluorescence emission has not risen above background level. This is when baseline fluorescence is calculated. In the early exponential phase the PCR reaction continues and the fluorescence exceeds a threshold calculated based on the baseline. The threshold level is recorded as the cycle threshold. During the exponential phase the PCR undergoes optimal amplification – the PCR product doubles every cycle in ideal reaction conditions. The plateau stage is reached when reaction components become limited (Wong & Medrano 2005). For absolute quantification serial dilutions of standards of known concentrations are used to generate a standard curve. The

standard curve produces a linear relationship between the cycle threshold and the initial amounts of cDNA. This technique began as a bulk assay, however, combination with cell isolation technology such as microfluidics has allowed development of single RT-qPCR which will be discussed below among other single cell gene expression quantification methods.

Another bulk method that can be used for detection of mRNA (and other RNAs) is RNA-Seq. This method allows detection and quantification of the whole transcriptome. RNA-Seq involves the isolation of a population of RNA which is converted into a library of cDNA fragments before being sequenced in a high throughput manner (Wang *et al*. 2009). Both this method and that of RT-qPCR have been developed to work at the single cell level and will be discussed further in the following section.

1.4.1.2 Single cell methods for mRNA detection and quantification

1.4.1.2.1 RT-qPCR

Combining RT-qPCR with cell isolation methods allows the technique to be used in single cells. This was described by Gao *et al*. (2011), who primarily used 16s rRNA as a target due to the high expression of ribosomal RNA in the cell. Gao *et al*. describe the difficulties in making quantitative statements about RNA in bacterial cells, highlighting the fact that while mammalian cells have between $1\text{-}3x10^{-2}$ ng of total RNA, they were able to estimate from a bulk study that bacterial cells have only $3.8x10^{-5}$ng of RNA, approximately one thousandth of the quantity typical in mammalian cells. This makes it incredibly technically challenging to not only isolate

a single cell from a bulk culture, but also to recover the minute amount of RNA from the cell. In their study, Gao *et al*. initially describe a single-tube one step method for single cell RT-qPCR. As substrate they use either a single bacterial cell that has been mechanically isolated or they begin with a bulk culture that has been serially diluted to achieve a theoretical given cell number. The cells were then subjected to a combined reverse transcriptase and qPCR reaction. When aiming to detect 16s rRNA, investigators were unable to amplify RNA from single, mechanically isolated cells, and could detect RNA in serially diluted cells when the number of cells expected in the dilution was around 20, but not below. However, when RNA was isolated from a bulk culture they were able to serially dilute the isolated RNA to what they estimated to be single cell level and detect 16s RNA at that level using the single-tube method. A second, two-tube method was developed as a consequence of the single-tube limitations. The RNA isolation and cDNA synthesis was separated from the qPCR amplification and each step was optimized individually. A kit capable of isolating RNA from bacteria could be used for bulk or individual cells, whether the cells were serially diluted or mechanically isolated. 16s ribosomal RNA isolated in this manner could be converted to cDNA and amplified by qPCR through the use of a second kit. This two-tube method was able to show reproducible detection of the high copy number 16s RNA at single cell level. Furthering this, the investigators showed that they were not only able to detect the 16s RNA, but also two mRNAs – dnaK and groES. They were able to show that while 16s RNA remains fairly constant after heat shock treatment (showing only a small reduction in comparison to a no shock control), dnaK and groES both experienced an increase in

expression after heat shock. This is expected, as dnaK and groES are both involved in the heat shock response. All three target genes were detectable both in the control and heat shock conditions. An experiment was also done to determine whether the two-tube method was sensitive enough to amplify low copy number genes. Total RNA was isolated from single *E. coli* cells and was then converted to cDNA as described previously. The cDNA was then serially diluted. Again detection was being done for 16s RNA. Diluted beyond 10^{-4}, 16s RNA became difficult to detect in a reproducible manner. Even when diluted to this extent, the RNA content of 16s RNA is still much higher than that for low expression genes. This suggests that while this method of RT-qPCR is valuable for the detection and relative quantification of high copy number mRNAs within single cells, there is not yet the capacity to examine low copy number transcripts at the single cell level using this method. It is worth noting that in mammalian cells the absolute number of mRNA molecules in cells is higher, and the RT-qPCR technology is more advanced. For example White *et al*. (2011) describe a protocol for high-throughput microfluidic single-cell RT-qPCR, which allowed cell capture, cell lysis, reverse transcription and quantitative PCR in individual human cells. It is important also to note that while the method provides quantification and can be used to compare within cells and across populations and growth conditions, the output is a relative curve or a back calculated figure based on assumptions of amplification efficiency rather than an absolute number as can be achieved with microscopy techniques.

1.4.1.2.2 Single cell RNA-Seq

Single cell RNA-Seq utilizes deep sequencing technologies and single cell isolating techniques to allow the transcriptome of an individual cell to be examined and quantified. Single cell RNA-Seq of mRNA involves the isolation of individual cells. This is a non-trivial step and can be can be achieved in several different ways. Flow activated cell sorting (FACS) can be used to isolate individual cells through the use of fluorescently labelled antibodies. It is also possible to use optical tweezers which utilize laser beams to physically hold or move single cells or to isolate cells using microfluidics, a rapidly expanding array of techniques with the aim to integrate microsystems and allow culture, isolation and biochemical steps to occur within the same system (Saliba *et al*. 2014). The cells must then be lysed and the tRNA and rRNA that comprise >90% of the sample must be removed to allow study of mRNA (Saliba *et al*. 2014). This is generally done through one of two methods: 1) depletion by hybridization of the rRNA to complementary nucleotides immobilized on beads or 2) degradation through use of an exonuclease that degrades uncapped RNAs. Each of these methods inevitably cause degradation of an amount of the biologically relevant mRNA, and a complete assessment of the biasing effects of each technique has not yet been conducted (Saliba *et al*. 2014). The isolated mRNA is then converted into cDNA, which subjected to massively parallel sequencing. This provides a profile of the cell wide transcriptome that can be used to quantify transcript numbers and to compare cell-to-cell heterogeneity in populations. The strength of single cell RNA-Seq as a technique is highlighted by Wang *et al*. (2015) in a study in which they produced and validated a method for the isolation, cDNA synthesis and amplification, and next generation sequencing of single cells of the

cyanobacteria *Synechocytis* sp. PCC6803. The investigators used data gathered to compare the transcriptomes of both bulk and single cells following nitrogen starvation (24 and 72 hours). Individual live cells were picked at random from culture, and the RNA was isolated. The purified total RNA was then used as a template for amplification generate cDNA. Nine cDNA libraries were constructed for sequencing, including three single cells for 24 hours nitrogen starvation and three single cells for 72 hours nitrogen starvation, along with three bulk samples, one for zero hours nitrogen starvation and one each for 24 and 72 hours nitrogen starvation. Each was sequenced on Illumina HiSeq.

The results they gained showed that this RNA-Seq technique achieved 82-98% coverage of the Synechocytis genome from a volume of only 5-7x10^{-6} ng total RNA. The results also shed light on the utility of the technique in discerning cell-to-cell heterogeneity and comparison between single and bulk cell analysis. The data showed that there were clear changes in the transcriptional profiles of bulk cell samples between 0, 24 and 72 hours nitrogen starvation. None of the single cell profiles matched exactly the corresponding bulk cell profile, but both were similar. The transcriptomic profiles of single cells at 24 hours nitrogen starvation were separated from the profiles observed at 24 hours, a greater degree of heterogeneity was seen at 72 hours among the three cells measures. Taken together, the results support the finding that heterogeneity between cells may be one of the responses of a genetically identical bacterial populations to adverse conditions (Newman et al. 2006), . More pertinently confirmation of the RNA-Seq data shows that it is possible to isolate, amplify and sequence the mRNA content of single prokaryotic cells

through RNA-Seq. The technique is however in its infancy and remains extremely technically challenging, with each step presenting challenges as outlined.

Further to the above described techniques it is possible to examine the mRNA content of cells through use of microscopy. This approach is attractive, as it removes the need to isolate and lyse cells and instead allows direct observation of mRNA. These techniques rely on fluorescent tagging of the mRNA molecule, which can be achieved in multiple ways – Spinach, MS2 and PP7, and mRNA FISH. Each of these techniques will be described below.

1.4.1.2.3 MS2 and PP7 aptamers

The MS2 aptamer, and its derivative PP7, allow fluorescent labelling and localization of mRNA within living cells, and are now commonly used for these purposes, as well as for kinetic studies (Broude 2011). Both aptamers function in similar ways. MS2 is named after the bacteriophage MS2, and the system utilizes the expression of two components within the cell. The first component is a RNA-binding protein from the phage, MS2 coat protein, which is expressed with an intact fluorescence protein as a fusion. The second component is the target mRNA. This mRNA is tagged with a series of MS2 coat protein binding aptamers. When these two components are expressed in tandem in a cell, the MS2- FP fusion protein bind the aptamer-tagged target mRNA, which then fluoresce and can be viewed under a microscope (Broude 2011) (See Fig.1.4 for schematic). Initially this technique was developed to localize the mRNA Ash1 in yeast as there was no method for mRNA visualization available (Bertrand *et al*. 1998). Ash1 protein is involved in mating type

switching in *Saccharomyces cerevisiae*, and is now known to localize to the bud tip. To investigate the cellular localization of Ash1 a two plasmid system was constructed, the first plasmid carried the MS2 coat protein – GFP fusion and the second six MS2-binding aptamers, each with a 19-nucleotide RNA stem loop fused to the 3' UTR of Ash1 and a *lacZ* construct as a reporter mRNA (multiple aptamers were used to increase fluorescent signal coming from multiple bound MS2-GFP fusions). The MS2-GFP fusion protein was engineered to carry a nuclear localization signal which would restrict it to the nucleus unless bound to mRNA. It was shown that yeast cells expressing the MS2-GFP fusion and the Ash1/*lacZ*-aptamer contained a single, bright 'particle' localized at the bud tip. The authors confirmed that the 'particles' were indeed MS2-GFP bound to the Ash1-aptamer mRNA by performing FISH against the *lacZ* region of the mRNA reporter, and were able to not only show the position of the Ash1 mRNA in the cell, but were also able to conduct several experiments and gain insight about genes previously known to impact mRNA localization (Bertrand 1998).

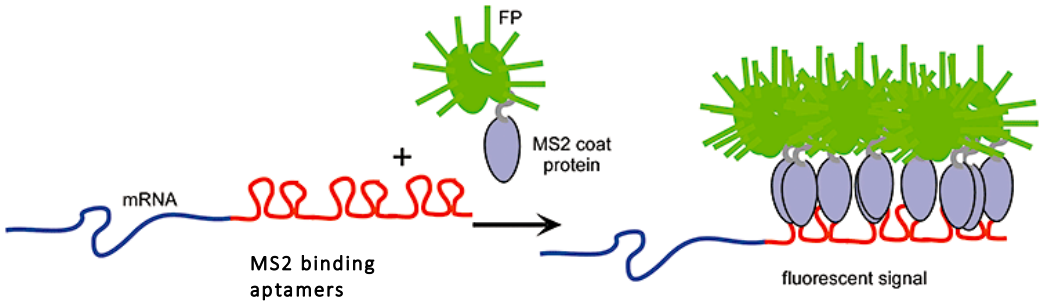


Figure 1.4. **MS2 tagging of mRNA.** The mRNA of interest is tagged with an array of MS2 binding aptamers, and MS2 is expressed as a fusion with FP. Binding of the MS2+FP to the MS2 binding aptamers allows fluorescent labelling of the mRNA. (Adapted from Broude 2011).

The PP7 aptamer system was developed as part of an effort to observe transcription initiation and elongation on an endogenous yeast gene (Larson *et al*. 2011). Here a cassette coding for 24 binding sites of the bacteriophage PP7 coat protein was inserted into the 5' untranslated region (UTR) of a target gene, specifically the POL1 promoter (which is cell cycle regulated) and the PP7-GFP fusion protein was constitutively co-expressed (see Fig 1.5 for schematic). The PP7-GFP protein was able to bind the mRNA aptamers after they were transcribed. This allowed visualization of native transcription sites, as well as assessment of the length of time the mRNA remained at the site of transcription and the amount of time for which the POL1 gene was active. The authors were able to visualize the location and intensity of the foci resulting from the PP7-GFP fusion binding to the mRNA aptamers and from them derive information about the dynamics of POL1 expression, finding that the POL1 promoter is active only during the late G1 and S phases of the cell cycle, but that initiation events while the promoter is active are stochastic and uncorrelated. The authors were also able to assess elongation and termination time during mRNA expression. This was achieved by inserting the cassette of 24 PP7 binding sites into either the 5' UTR or the 3' UTR of the 15 kb housekeeping mRNA MDN1. When the binding site cassette is inserted into the 5' UTR the dwell time and intensity of the PP7-GFP foci comes largely from the transcription time of the downstream portion of the gene, whereas when the cassette is inserted into the 3' UTR the foci appears only when the mRNA has been fully transcribed. Analysis of the intensity and duration of foci formed from 5'UTR and 3' UTR expression of the PP7 binding cassette allowed computation of the dynamics of initiation, elongation

and termination using temporal auto correction. It was found that initiation rates varied across the cell cycle, and that in G1 the mean dwell time of PP7-MDN1 transcripts is ~770 seconds and the mean dwell time for MDN1-PP7 is ~140 seconds. From these data they were able to calculate the velocity of the RNAPII on MDN1 as ~20 bases per second and the termination time as ~70 seconds. When the cells were in late s/G2 phase however PP7-MDN1 had a much shorter dwell time of ~310 seconds but a faster mean initiation rate, resulting in a lower bound estimate of the velocity of ~ 46 bases per seconds (Larson 2011). The ability to derive these numbers through use of the PP7 system highlights the utility of the technique, and its predecessor MS2, in characterizing the dynamics of gene expression in living cells. It is even possible to combine the techniques and use a two colour system to label and quantify RNA in live yeast (Hocine *et al*. 2012). Although developed initially in yeast, the protocols can also be used in a wide variety of model organisms (Tyagi 2009).

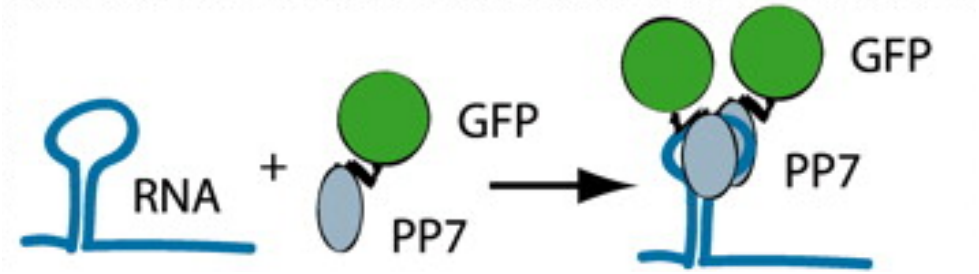


Figure 1.5. **PP7 tagging of mRNA.** The mRNA of interest is tagged with a cassette coding for 24 binding sites for the PP7 protein. PP7 is expressed as a fusion with FP. Binding of the PP7+FP to the PP7 binding sites allows fluorescent labelling of the mRNA. (Adapted from Larson 2011).

However, there are several drawbacks to the MS2/PP7 aptamer systems. It can be argued that the disruption caused by the introduction of the large coat protein binding cassettes can cause perturbation to the system, potentially by impacting the endogenous degradation of the mRNA or by impacting mRNA movement and localization (Garcia & Parker, 2015). There can also be issues with the constitutive expression of the MS2/PP7-FPs, which when unbound can generate high fluorescence background (Zhang *et al*. 2015).

1.4.1.2.4 Fluorescent RNA aptamers

A second system that utilizes RNA aptamers for RNA imaging is Spinach (Paige *et al*. 2011) (and its derivatives Spinach 2 (Strack *et al*. 2013); RNA mango (Unrau *et al*. 2014) and Broccoli (Filonov *et al*. 2014) amongst others). The Spinach system uses a short RNA aptamer (~100 nts) that is genetically fused to the mRNA of interest. The aptamer mimics GFP-like fluorescence when bound to the fluorogenic ligand 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) which is introduced externally. The ligand is structurally similar to the EGFP chromophore, and is non-toxic and membrane permeable. It fluoresces only when bound to the mRNA aptamer, reducing the problem unbound fluorescence (see Fig. 1.6 for schematic).

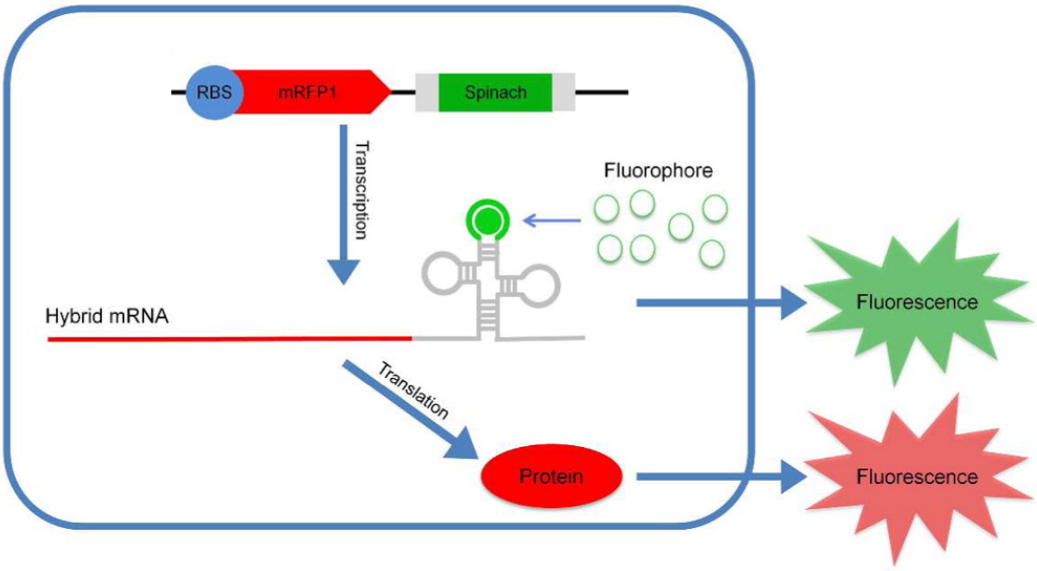


Figure 1.6. mRNA labelling with the spinach aptamer. The mRNA of interest (here that of RFP) is tagged with an aptamer. The aptamer emits GFP like fluorescence when bound to the externally introducedfluorogenic ligand DFHBI, allowing detection of the mRNA. (Taken from Pothoulakis *et al*. 2013).

Spinach has been utilised widely in studies that image high copy number non-translated RNAs (Paige *et al*. 2011; Filonov *et al*. 2014). However, when used to image lower copy number mRNAs as in Pothoulakis *et al*. (2013) the level of fluorescence produced by a single Spinach aptamer was only slightly above that of the autofluorescence in the cell. Zhang *et al*. (2015) conducted a comprehensive study into the use of the spinach aptamer and specifically a tandem array of spinach aptamers for mRNA imaging in live bacterial cells. They found that the use of multiple spinach aptamers increased the capacity for mRNA imaging (the use of 64 aptamer repeats increased fluorescence output 16 fold). While this increase in fluorescence output was greatly improved over that of a single spinach aptamer, Zhang *et al*. did not report the capacity to label at single molecule level, and noted

that increasing the number of aptamer repeats reduced the folding efficiency of the aptamer into its correct structure. They were also able to show that the mRNA they targeted (that of RFP) with the spinach arrays was not overly perturbed by the presence of the aptamer repeats. RT-qPCR quantification was done for both tagged and untagged RFP mRNA when induced by IPTG. Spinach can therefore be said to be minimally disruptive to native mRNA expression, but insufficiently bright to allow single mRNA illumination and quantification. The use of hybridisation techniques such as mRNA FISH removes the potential for expression disruption and allows single molecule detection.

1.4.1.2.5 mRNA FISH

Single molecule fluorescence *in situ* hybridization (FISH) was originally described in 1998 as a method to detect single mRNA molecules of genes of interest and determine their copy number in eukaryotic cells (Femino *et al*. 1998). Initially the method was designed to use five separate 50 nt oligonucleotide probes that were conjugated to five individual fluorophores each. Femino *et al*. were able to detect single β-actim mRNA molecules in rat kidney cells. The technique was extended to use in prokaryotes (Maamar *et al*. 2007), were a small number of probes with multiple fluorophores were used to investigate ComK, the protein that regulates competence for DNA uptake in *Bacillus subtilis*. They were able to determine that fluctuation that occurs stochastically in *comK* expression determines whether cells are competent or not, and that artificial reduction of such noise reduces the number of competent cells observed. Raj *et al*. (2008) introduced a variation on the method

whereby rather than having multiple fluorophores present on a single probe and using a small number of probes overall, they used a larger number of singly labelled probes. Raj *et al*. show that this method circumvented the issue of signal variability that can be seen with the use of a small number of probes with multiple fluorophores, as well as the false positives that arise from probe miss-binding and false negatives that occur from nonbinding events. Femino *et al*. described signal detection corresponding to the binding of only one or two probes rather than the full number. Raj *et al*. argued that this made accurate detection of the number of mRNA molecules represented by the fluorescent spots challenging. The use of multiple (12-48) singly labelled probes however allowed the detection of unimodal distributions of spot intensity, indicating a normal distribution of probes bound to a single mRNA (a multimodal distribution would be expected in the event of probes binding to multiple mRNAs (Vargas *et al*. 2005)). The Raj *et al*. method was introduced in mammalian cells, and was used in *E. coli* to quantify the copy numbers of 20 promoters by So *et al*. (2011). In 2013 a version of the technique was developed specifically for use in *E. coli* by Skinner *et al*. (2013). mRNA FISH has the advantage of being minimally disruptive to the process of gene expression, as it does not require alteration of the DNA, however to facilitate probe binding to mRNA it is necessary to fix and permeabilise the cells. The protocol outlined by Skinner *et al*. was used in this study to assess *recBD* expression in single *E. coli* cells, a schematic of the protocol can be seen in Chapter 2, Fig. 2.4.

1.4.1.2.6 mRNA detection summary

As outlined above, there are many mRNA detection techniques available for both bulk and single cell study of prokaryotic cells. Of the current single cell techniques available RT-qPCR and single cell RNA-Seq involve isolation and amplification of the cellular mRNA, which is both highly technically challenging and can introduce error. Both techniques become more challenging as mRNAs of lower copy number are being investigated, with RT-qPCR not yet able to detect low abundance mRNA. RT-qPCR requires a known mRNA sequence, while RNA-Seq allows investigation of the whole transcriptome. Microscopy techniques are not subject to issues of mRNA isolation and nucleotide amplification, however, techniques such as MS2 and PP7 as well as Spinach involve addition of sequence to the native mRNA. This is potentially disruptive to the mRNA expression and may alter ultimate mRNA number within the cell. For this reason, single molecule mRNA FISH was selected as the least disruptive technique for this study and was used to quantify the expression of the mRNA of the bacterial DNA repair protein RecBCD- *rec*B and *rec*D mRNA- in single *E. coli* cells. A summary of the RNA detection techniques outlined above can be seen in Table 1.1.

Table 1.1. Methods for RNA detection

Technique	Population or single cell experiments	Knowledge of genetic sequence required	Live cell experiments possible	Quantification type	Low copy number mRNA detection
Northern blot	Population	Yes	No	Population mean	No
Microarray	Population	Yes	No	Relative	No
RT-qPCR	Both	Yes	No	Relative	Yes
RNA-Seq	Both	No	No	Absolute	Yes
MS2/PP7	Single cell	Yes	Yes	Absolute	Yes
Fluorescent mRNA aptamers	Single cell	Yes	Yes	Absolute	No
mRNA FISH	Single cell	Yes	No	Absolute	Yes

1.4.2 Current methods in protein detection

Proteins are central to all cellular processes – they are key in cellular structure, transport of molecules, and control of growth, catalysing reactions and regulating signal transduction. In bacteria, protein lifetime in the cell is generally longer than that of mRNA, although half-life varies from protein to protein, and protein is generally diluted from cells by cellular division rather than degradation (Hintsche & Klumpp 2013). Protein number per cell varies widely from protein to protein (five orders of magnitude from 10^{-1} to 10^{4} molecules per cell), and one study in *E. coli* (Taniguchi *et al*. 2010) indicated that about half of the proteins investigated (1018 in total) were present at less than 10 molecules per cell. This highlights the need to be able to investigate protein at a very low copy numbers, as an abundance of proteins

are likely to be expressed in this range. However, as with mRNA protein quantification has traditionally been done in bulk conditions.

1.4.2.1 Population methods for protein detection and quantification

As with mRNA, initial quantification of protein was largely done in bulk studies. Western blots, have been widely used to separate and identify proteins, and can be applied in a semi quantitative manner. As with Northern blots described above, Western blots involve isolation of total protein from a bulk cell population, followed by size separation by gel electrophoresis. The proteins are then transferred to a solid support and the membrane is then incubated with primary and secondary antibodies that are specific to the protein of interest. A relative comparison of protein levels can be made, however there is likely to be variation between the loading of each lane in the original gel electrophoresis as well as differences in the rate of transfer to the membrane between different lanes. Additionally detection signal may not be linear across the detection range of samples, meaning western blots are generally semi quantitative (Mahmood & Yang 2012). Similar to Northern blots, detection of very low levels of protein can be challenging with Western blots.

Another bulk method for the detection of protein is the use of immunoassays (such as ELIZA and EIA) where protein is detected by the production of a measurable signal on binding to an antibody, whether that be radioactive isotopes producing radiation, light emission, or colour change induced by an enzyme as is described in Lequin (2005). It is also possible to use mass spectrometry to quantify protein. Mass spectrometry can be used to determine the molecular mass of proteins

(Mass spectrometers measure mass/charge), and the molecular masses of their constituent parts following fragmentation, as we as for protein quantification. This has proved a powerful tool in terms of protein characterization (McLafferty 2008). Mass spectrometry relies on converting protein into an ionized, gaseous form and analysing them in an electric or magnetic field. While mass spectrometry is not naturally quantitative due to differences in detection capacity and ionization efficiency between peptides in the same sample (although it is possible to compare the peak intensity of the same peptide across samples), it is also possible to assess relative quantities of protein in samples through mass spectrometry. This can be done by incorporating isotopically labelled amino acids in vivo using a technique called stable isotope labelling with amino acids in cell culture (SILAC) or by post-labelling of peptides using isotope labelled molecules that can be covalently bonded proteins using a technique called isobaric tags for relative and absolute quantification (iTRAQ). These techniques allow labelling of peptides from different samples with heavy or light isotopes, mixing of the samples and taking the ratio of heavy to light isotopes to indicate the relative abundance of the proteins in each sample. Quantitative, condition dependant *E. coli* proteome analysis has been conducted using protein extraction and sample fractionation and quantitative mass spectrometry (Schmidt *et al*. 2015). The authors were able to quantify protein in 22 different experimental growth conditions, and were able to determine protein abundance for 55% (>2300) of *E. coli* proteins. This was done through efficient protein extraction and mass spectrometry. Cells were lysed and protein extracted and proteolyzed. Peptide mixtures were then either further fractionated or directly analysed in

biological triplicates using liquid chromatography mass spectrometry and quantified using label free quantification. Then, cellular concentrations of 41 proteins involved in the glycolysis pathway were determined through stable isotope dilution and selected reaction monitoring. This meant that heavy labelled reference peptides were synthesized for each protein, which could be spiked into each sample. This allowed determination of absolute quantities for the corresponding proteins by selected reaction monitoring. Flow cytometry was then used to determine the number of cells taken for liquid chromatography mass spectrometry, allowing the protein quantities determined to be converted into protein copies per cell.

Ribosome profiling and polysome profiling are two distinct methods that can be used to quantify protein expression from bulk samples. These techniques both rely on measuring the process of translation, rather than identifying proteins directly. Polysomes consist of mRNA molecules on which multiple ribosomes are bound. As these structures have more mass than ribosomal subunits or single ribosomes, they migrate more quickly through sucrose gradients (Noll 2008). In this manner, it is possible to resolve polysomes that have different numbers of ribosomes (and therefore different levels of translation) from cellular lysate and analyse the fractions produced using techniques such as northern blotting and RT-qPCR to give a snapshot of the translational profile of the initial cellular population (Qin & Fredrick 2013)

Ribosome profiling relies on the fact that the position of a translating ribosome can be determined by the fact that a ribosome protects a discreet footprint on its template mRNA (~30 nts) (Steitz 1969). The process is conducted by isolation of mRNA from a sample, followed by digestion of the RNA that is not protected by

ribosomes and isolation of the ribosome-RNA complexes. The ribosomes are then stripped from the mRNA which is reverse transcribed to cDNA for strand specific amplification for sequencing. The sequences obtained can then be mapped to the genomic sequence to determine a translational profile (Ingolia *et al*. 2009). Ribosome profiling is highly adaptable to different organisms and has been used widely across cell types (Ingolia *et al*. 2011; Stern-Ginossar *et al*. 2012; Bazzini *et al*. 2012). Ribosome profiling has several advantages, including the high sensitivity of the technique and the precise quantification of translation in samples. Ribosome profiling is able to detect all but the rarest translation events. Finally ribosome profiling provides a snap shot of the process of translation at the moment it was interrupted. This is valuable as it allows tracking of changes in the process of translation as they occur, as was done by Andreev *et al*. (Andreev *et al*. 2015) in a neural cell line. They used time resolved ribosome profiling to determine that oxygen and glucose deprivation of these cells impacts the translation profile within 20 minutes, with the translation of ~3000 genes being impacted in the first hour of deprivation. Due to the precise nature of the ribosome profiling data they were able to determine that oxygen and glucose deprecation alters translation rates as well as impacting the stringency of start codon recognition. Ribosome profiling has also been used in *E. coli* (Li *et al*. 2014), in a genome wide study that examined the absolute synthesis rate for the cellular proteins produced by 3041 genes. However, ribosome profiling does also have limitations, and these include the potential to introduce technical artefacts, for example by impacting the translation activity while trying to arrest it prior to experimentation. It is also necessary with ribosome

profiling to infer the rate of protein synthesis. This is done by considering the average density of ribosomes along the stretch of mRNA of interest. This can be somewhat inaccurate as it does rely on the assumption that all ribosomes that load onto mRNA complete translation as well as assuming that the translation rate for all mRNAs is similar. A further limitation is that ribosome profiling requires a large amount of input material, and cannot yet be applied to single cells. It is likely that this limitation will be circumvented in the future (Brar & Weissman 2015).

While the techniques described above are able to give an indication as to the presence or absence of a protein, or to suggest relative abundance within a population, they do not provide absolute quantification within single cells. More recently techniques that do provide quantitative data in single cells have been developed. The challenges involved in single cell protein analysis are similar to those in mRNA detection, specifically given that the absolute number of a given protein in cells can be very small, and isolating and individual cell can be extremely technically challenging. I will now discus various methods that are commonly used to identify and quantify protein number in individual cells.

1.4.2.2 Single cell methods for protein detection and quantification

1.4.2.2.1 Flow cytometry and microfluidic flow cytometry

Flow cytometry was originally developed in 1969 (Hulett *et al*. 1969) by the Herzenberg lab as a means for automated separation of mammalian cells as a function of intracellular fluorescence. The system involves engineering cells to be differently fluorescent and then sorted in liquid flow (Bonner *et al*. 1972). Initially

the technique was used to discern between and sort cells with one or two different fluorescent markers, however as availability of markers increased and technology advanced this number has increased into the teens, such as was done in Perfetto *et al*. (2004). It is now possible to take measurements of multiple proteins in the same cell, allowing relative measurement of protein quantity. The technique has also been combined with microfluidics, and integrated devices have been made that are able to combine sample handling, flow cytometry and cell sorting as was done in Srivastava (2009) to investigate phosphoprofiling of macrophage response to *E. coli* lipopolysaccharide stimulation.

1.4.2.2.2 Separation based methods

Separation based methods for protein analysis have potential, given the advantages of separating protein isolate into individual proteins before analysis. High resolution separation allows for chromatography or electrophoresis techniques to give an unbiased measure of the entire proteome. However, current separation techniques such as HPLC or slab electrophoresis lack the capacity to process very small amount protein and so are unsuited to single cell analysis. However more recent techniques such as capillary electrophoresis are able to capture these very small amounts of proteins. Huang *et al*. (2007) describe a method that allow for manipulation, capture and lysis of a single cell followed by chemical separation and analysis of the lysate. To do this they used single molecule fluorescence detection, capillaries and microfluidic channels. For investigation of eukaryotic cells and cyanobacteria they produced a PDMS chip that has three sections- cell manipulation,

electrophoretic separation and single molecule counting (where proteins were not naturally fluorescent, fluorescently labelled antibodies were added to tag target proteins). The single molecule counting was done by monitoring the number of fluorescence bursts generated when the molecules flowed through a small detection volume. With this method Huang *et al*. were able to examine the response of the unicellular cyanobacterium *Synechoccus* to the depletion of nitrogen–containing nutrients in their growth culture. They found that the order of degradation of proteins involved in collecting light energy for photosynthesis differed from that observed in bulk culture when examined at the single molecule level. This displays the utility of single molecule measurements in determining the true distribution of proteins in populations.

1.4.2.2.3 Genetic and chemical probes

Protein analysis in single cells occurs frequently through the use of fluorescent tagging. This can be done in multiple ways and can be utilized at population or single cell level, and can be scaled down to single molecule level. Labelling proteins with fluorescent tags allows visualization, localization and can allow quantification. The most widely used fluorescent tags are auto fluorescent proteins, such as GFP and its many derivatives. Fluorescent proteins are generally around 25 kDa in size, and their entire structure is important in producing fluorescence (Cranfill *et al*. 2016). The structures consist of 11 β-strands that surround an α-helix in the centre, where the capacity for fluorescence arises from a few amino acids (Ormo *et al*. 1996). When the protein is being synthesized a reaction occurs among fluorescence producing

amino acid residues. The fluorescence depends on rigid maintenance of the structure and the chemical environment within the β-strands (Follenius-Wund *et al*. 2003) (See Fig 1.7 for illustration of structure).

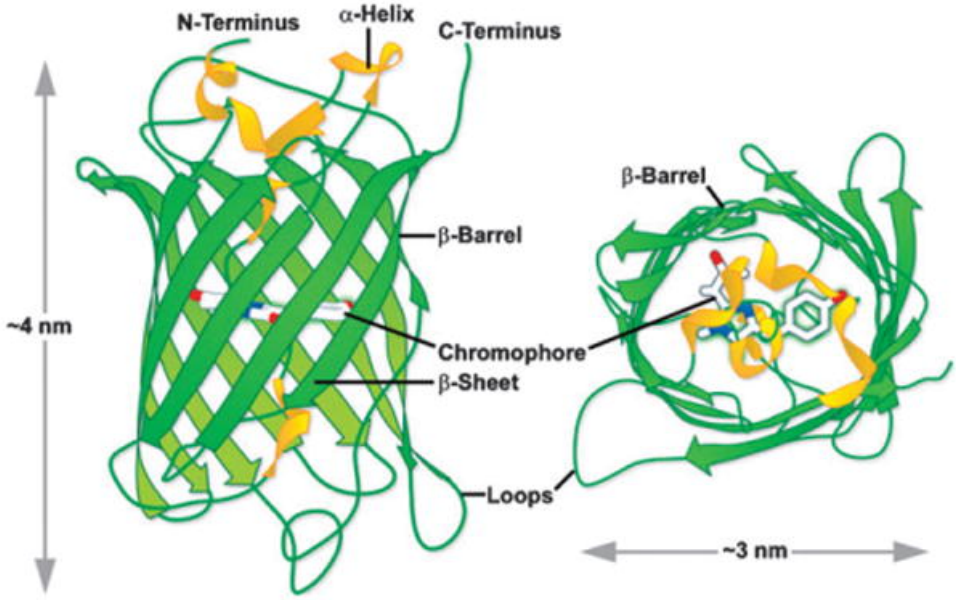


Figure 1.7. **GFP β-barrel structure.** (Taken from Day & Davidson 2009)

The identification of GFP from the jellyfish *Aequorea* the 1960s (Shimomura *et al*. 1962) marked the beginning of a revival in the field of cell biology, but it would be a further 30 years before the sequence of GFP was cloned in full (Prasher *et al*. 1992) and expressed for the first time out-with its native jellyfish as a marker for gene expression (Chalfie *et al*. 1994). The potential utility of the protein in expression, quantification and localization studies was noted immediately, as was the potential to artificially enhance the fluorescence properties of the native protein by altering its structure. Initially this was done through optimisation of the absorption profile to a single peak, as well as decreasing the maturation time and increasing the photostability of the GFP protein (Heim *et al*. 1995). Expansion of the colour palate

also occurred, and GFP derivatives in blue (Heim & Tsien 1996), cyan (Heim *et al*. 1994), and yellow (Ormo *et al*. 1996) were all developed and optimised. Additionally, a fluorescent protein that would emit in red was sought, both for increased capacity in multi-colour imaging and because cellular auto fluorescence is decreased in this region of the colour spectra (Shcherbo *et al*. 2009). DsRed was among the first fluorescent proteins to be cloned from the Anthozoa and characterised (Matz *et al*. 1999), and as with GFP mutagenesis was employed to improve the utility of the protein for imaging (increasing the maturation rate, increasing the intrinsic brightness, and reducing the tendency of the protein to aggregate). This gave rise to DsRed2 and further DsRed derivatives (Bevis & Glick 2002; Strack *et al*. 2008). In addition, directive evolution was used to produce the monomeric RFP mRFP1, as well as a variety of derivatives known as the mFruit FPs (Shaner *et al*. 2004; Shaner *et al*. 2008). For further details concerning the history of fluorescent proteins and details of the current array of proteins available see Day and Davidson (2009). Fluorescent proteins are now used to measure and visualise protein ubiquitously across all domains of life. They are used to confirm gene expression (Zhao *et al*. 1998; Yeh *et al*. 1995; Jach *et al*. 2001), to conduct localization studies (Feilmeier *et al*. 2000; Lequin *et al*. 2003; Fernandez-Abalos *et al*. 1998) and are used to assess protein-protein interactions through techniques such as fluorescence resonance energy transfer. Additionally, fluorescent proteins are used in extensively in quantitative work. For example, the use of microfluidics assisted cell sorting (MACS) to mechanically slow down proteins diffusing in the cytoplasm has allowed for quantification of the low copy number stress response proteins RpoS and SprE

when fused to fluorescent proteins (Okumus *et al*. 2016). Despite this wide usage and utility fluorescent proteins do have disadvantages. Principally, their photo-physical properties are not as good as those of organic dyes for use in fluorescence imaging. They are limited by chemistry in terms of their colour, brightness and photo stability, which are all key elements in the success of fluorescence microscopy, particularly at the single molecule level. Additionally, fluorescent proteins are often denatured by fixation and consequently lose their fluorescence (Segala *et al*. 2015). However, other fusion tags are available. Specifically, self-labelling protein tags (Crivat & Taraska 2012). Such tags are generally fused to the protein of interest (like FPs) but are not innately fluorescent. They become fluorescent when exposed to a fluorescent ligand.

The first of these self-labelling tags developed was the SNAP tag (Keppler *et al*. 2003; Keppler *et al*. 2004), where labelling with a small molecule occurs through the use of O^6-alkylguanine-DNA alkyltransferase (AGT). AGT irreversibly transfers the alkyl group from its substrate, O^6-alkylguanine or O^6-benzylguanine to one of its cysteine residues, covalently linking substrate and enzyme. To allow labelling of AGT, organic fluorescent substrates were made and added to cells with genetic fusions of AGT that specifically label the proteins. A variant of this system, the CLIP tag. The CLIP also utilises the AGT protein, mutated to allow use of a different substrate, O^2- benzylcytosine. SNAP and CLIP tags can be used in both eukaryotes and prokaryotes (Liss *et al*. 2016) and the enhanced photo-physical properties of the dyes that are ligated to the enzymes substrates have allowed the enhancement of super-resolution microscopy techniques that allow for imaging at resolutions below

that of the diffraction limit. A further self-labelling enzyme is the HaloTag. The HaloTag was developed by Los *et al*. (2008) as a multifunctional enzyme tag that would allow binding of protein to fluorescent dyes, affinity handles or even solid surfaces. The HaloTag itself if a modified haloalkane dehalogenase with a mutated histidine in its active site. The enzyme was modified such that when exposed to a substrate it binds it covalently. For microscopy, this allows the binding of ligand bearing an organic fluorescent tag (see Fig.1.8 for schematic of Halo, SNAP and CLIP tags and their respective fluorescent ligands). The HaloTag has been widely used in fluorescence microscopy for protein detection and visualisation, in both eukaryotes and prokaryotes (Huybrechts *et al*. 2009; Gallo *et al*. 2011; Daniels *et al*. 2012; Ke *et al*. 2016a; Barlag *et al*. 2016) and has been used at the single molecule level in yeast (Reck-Peterson *et al*. 2006). In this work, the HaloTag labelling method was developed for use at the single molecule level in *E. coli*. Additionally, the protocol was combined with that of mRNA FISH to provide a method for simultaneous labelling of mRNA and protein inside individual cells. An existing method for this is described below.

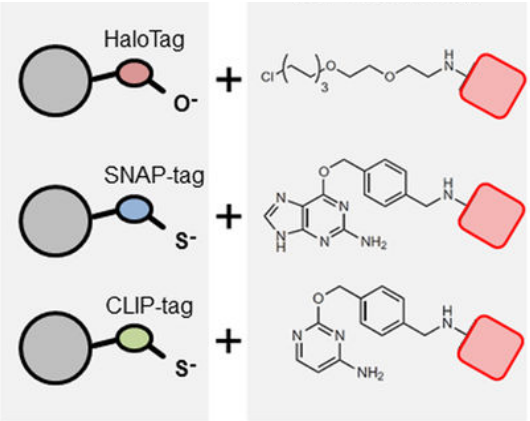


Figure 1.8. HaloTag, SNAP and CLIP tags expressed with protein bind to fluorogenic ligands. Each of these self-labelling enzymes is fused to a protein of interest and expressed. They bind to externally introduced ligands that are fused to fluorophores that allow fluorescence imaging and protein detection.

1.4.2.2.4 Simultaneous detection of mRNA and protein in single cells

Methods that are able to simultaneously detect protein and mRNA within individual cells are currently lacking, despite the evident utility such techniques would provide in deciphering gene expression dynamics. Currently such investigation has been done by Taniguchi *et al* (2010) who were able to detect both YFP and its mRNA through mRNA FISH for 137 proteins that were expressed at over 100 copies per cell. Additionally, a specific protocol was developed utilising mRNA FISH and antibody staining in *Drosophila melanogaster*. Here, the mRNA of the gene *hunchback* was labelled as was its regulating transcription factor Bicoid at different stages of the embryo development (Xu *et al*. 2015). This protocol was recently developed, and it was found that smFISH followed by antibody staining was able to preserve the signal produced from both the *hunchback* mRNA and the Bicoid protein labelling, allowing for single molecule detection of each. Xu *et al*. were able to detect and quantify *hunchback* mRNA, and found numbers in agreement with those previously published. They were also able to show that the antibody signal in the Bicoid channel was proportional to Bicod concentration by measuring the immunofluorescence and autofluorescence signal produced in a strain where the Biocoid protein was fused to GFP. The levels of the Bicod GFP fusion were also in agreement with previous studies. The quantification of both the *hunchback* mRNA and the Bicoid protein allowed for analysis of the transcriptional response, displaying the utility of such dual detection. The system described by Xu *et al*. was also used in *E. coli* by Sepulveda *et al*. (2016). As stated above this work will describe a method for detection of both protein and mRNA in individual *E. coli* cells, which will allow

for greater insight into gene expression at the level of both transcription and translation.

1.4.2.2.5 Protein detection summary

As outlined above, there are many protein detection techniques available for both bulk and single cell study of prokaryotic cells. Of the current single cell techniques available, all rely on fluorescence tagging. In the case of flow cytometry fluorescence is required to different populations of cells, and/or to sort them. In methods based on separation of proteins fluorescence is used once again to indicate the presence or absence of a protein within a population. By far the most common means for protein detection in individual cells is through tagging with either genetic or chemical probes. Genetic fusion of GFP is the classic method for protein detection, localisation and identification. However, despite the constant production of auto fluorescent proteins with enhanced photo-physical properties, chemical fluorophores are known to be brighter and more photostable. This has led to a proliferation in the use of self-labelling enzymes such as the SNAP, CLIP tags, and HaloTags. This work makes novel use of the HaloTag for detection of single protein molecules in *E. coli* with standard epifluorescence. The proteins examined are the three subunits of the bacterial DNA repair protein RecBCD. Additionally an alternative method for the detection of both mRNA and protein in single cells is described through use of the mRNA FISH and HaloTag protocols.

Table 1.2. Methods for protein detection.

Technique	Population or single cell experiments	Knowledge of genetic sequence required	Live cell experiments possible	Quantification type	Low copy number protein detection
Western blot	Population	Yes	No	Population mean	No
Immunoassay	Population	Yes	No	Relative	No
Mass Spectrometry	Population	Yes	No	Relative	No
Ribosome profiling	Population	No	No	Absolute	Yes
Flow cytometry/ microfluidic flow cytometry	Single cell	Yes	Yes	Relative	Yes
Fluorescent proteins	Single cell	Yes	Yes	Absolute	Yes
Self-labelling enzymes	Single cell	Yes	No	Absolute	Yes

1.5 RecBCD copy number

The RecBCD enzyme was chosen for use in this work because the protein is currently reported to be present at low copy number within *E. coli* cells, but reported numbers vary and accurate, single molecule study in functional protein has been lacking, as has investigation of mRNA quantity. Historically RecBCD has been reported to be present at about 10 molecules per cell, a figure that is often quoted (Taylor & Smith 1980; Taylor & Smith 1999; Dillingham & Kowalczykowski 2008; Smith 2012) and the origin cited as Eichler & Lehman (1977), a bulk study concerning the role of ATP in phosphodiester bond hydrolysis as catalysed by the RecBC nuclease wherein the *E. coli* cells were lysed and the protein complex was isolated through extensive treatment before any estimate of protein per cell could be inferred. More recently the study described above (Taniguchi *et al*. 2010)

investigated the numbers of RecB and RecD through use of YFP tagging, finding 0.61 and 4.76 molecules per cell respectively, and do not report findings for RecC. A further study, conducted by Li *et al*. (2014) and also mentioned above investigated RecBCD expression through use of ribosome profiling, and reported approximately 100 molecules per cell for each subunit. These results are conflicting in two ways. Firstly, Li *et al*. observe equal expression of each subunit, while Taniguchi *et al*. report much higher expression of RecD than RecB, and secondly the absolute numbers concluded by Li *et al*. are far greater than those seen by Taniguchi *et al*. This work aims to accurately quantify both the protein and mRNA content of RecBCD in isogenic *E. coli* cells. Here I will discuss the expression, structure and function of the RecBCD enzyme.

1.5.1 RecBCD expression

RecBCD is a heterotrimeric enzyme, and while spatially close on the *E. coli* chromosome, the genes that express each of the enzymes subunits are not expressed as a single operon and were not discovered simultaneously. The enzyme was first known as Exonuclease V (Emmerson 1968). The *recC* and *recB* genes were first described by Willetts and Mount (1969), who established that the *recC* and *recB* genes on the chromosome lay between *thyA* (a gene which encodes the protein thymidylate synthease, which is involved in the production of dTMP, a required precursor for DNA biosynthesis (Belfort *et al*. 1983)) and *argA* (a gene which encodes the protein N-acetylglutamate synthase, which functions in the arginine biosynthesis pathway (Shi *et al*. 2015)). Willetts and Mount show the genes to be in

the order *thyA-recC-recB-argA*. This mapping was further refined by Dykstra *et al*. (1984), who showed that the structural gene for *E. coli* protease III (*ptrA*) maps between the *recC* and *recB* genes, and was modified again in 1986 by Amundsen *et al*., who identified the rec*D* gene as the third subunit of the Exonuclease V enzyme, and suggested the alternative name of RecBCD for the complete enzyme complex. Amundsen *et al*. located the *recD* gene as being present between *recB* and *argA* on the *E. coli* chromosome, making the final gene order *thyA-recC-ptrA-recB-recD-argA*. A schematic of this structure can be seen in see Fig 1.9. It is worth noting that the presence of *ptrA* between *recC* and *recB* is not ubiquitous among bacteria, but rather appears only in *E. coli* and closely related species such as *Salmonella* (Cromie 2009).

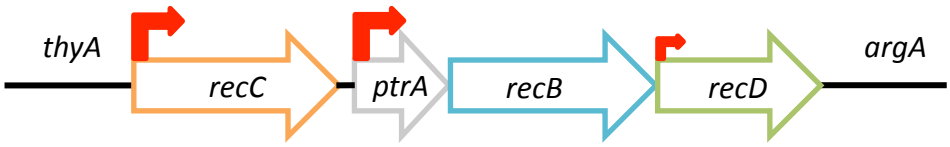


Figure 1.9. **Chromosomal structure of *recC*, *recB* and *recD***. RecBCD is expressed from two distinct operons. *recC* is expressed under the control of its own promoter (shown here with red arrows). While *recB* and *recD* form an operon (with *ptrA*), expressed primarily from a promoter upstream of *ptrA* but with a minor internal promoter for *recD* only.

As can be seen in Fig 1.8, the *recBCD* genes are not expressed as a single operon. *recB* and *recD* are expressed as an operon under the control of the same promoter, with a minor internal promoter for *recD* alone (Amundsen *et al*. 1986) and *recC* is expressed under the control of its own promoter (Finch *et al*. 1986).

In terms of regulation of expression very little is known concerning the *recBCD* genes. Unlike many DNA repair proteins they do not form a part of the bacterial SOS response, the global DNA damage response in *E. coli*. However, it is worth noting that close to the *ptrA* promoter there is an aberrant version of a LexA binding motif (LexA is the transcriptional repressor that regulates the SOS response), although this motif is not bound by LexA in vitro, and binding in vivo does not lead to transcriptional repression of the target gene (Wade *et al*. 2005).

1.5.2 RecBCD structure and function

The three RecBCD subunits come together in complex to form a highly efficient protein machine which has both helicase and nuclease activity, that is able to unwind DNA in vitro at a rate of about 1000 bp per second (Roman *et al*. 1989). A single RecBCD molecule can unwind over 40 000 base pairs (Bianco *et al*. 2001). In the cell, RecBCD binds to DNA double strand ends and degrades DNA until it reaches a Chi site (Chi is a cis acting DNA octomer that reads 5'- GCTGGTGG-3', and has an important role in mediating the function of RecBCD enzyme as described below). The specific qualities of each subunit are described below, and each contribute to the overall function of the enzyme complex.

Purified RecB protein is an ssDNA dependent ATPase, as well as being a DNA helicase with 3' -> 5' activity and having a domain which interacts with the RecA protein (Spies & Kowalczykowski 2005). The precise role of the RecC subunit is unclear, although the existence of RecC with mutants with reduced or absent Chi recognition strongly suggests a role in Chi recognition for the RecC subunit (Handa

et al. 2012). The RecD subunit is a DNA dependent ATPase, as well as being a DNA helicase with 5' ->3' activity. As the helicase activity of RecB is 3'-5' and the helicase activity of RecD is 5'-3', the RecBCD holoenzyme is able to translocate in the same direction along the DNA, with each of the helicase subunits operating on a different polarity.

RecBCD is capable of rapidly unwinding and digesting DNA. However, within the cell these highly destructive capabilities are employed in the repair of DNA double strand breaks. When a DSB occurs RecBCD binds to the available dsDNA end, before translocating along the DNA unwinding and degrading the duplex. This continues until RecBCD it reaches a Chi site. Chi sites are strand specific, and will only be recognized when encountered by RecBCD travelling in the correct orientation (i.e. when RecBCD encounters the 3' end of the Chi sequence first when translocating). Chi recognition is a regulating step in the function of RecBCD. Once recognized, the RecD helicase is disengaged, while the 3' ended strand continues to be unwound by the RecB helicase. The exact molecular mechanism of RecBCD is unclear, with different conditions producing different results in in vitro study. In particular, the frequency with which the nuclease cuts increases with increasing concentration of free Mg^{2+} ions, while the translocation rate of the complex increases with increasing ATP:Mg^{2+} concentration (Dillingham & Kowalczykowski, 2008) . For example, when ATP concentration is less than that of magnesium, the 3' ended strand (that is unwound by RecB) is rapidly digested before Chi recognition, while the 5' ended strand (that is unwound by RecD) is cleaved intermittently. Following Chi recognition, the 3' ended strand is no longer digested, but the nuclease continues

to act on the ‘5 ended strand (Anderson & Kowalczykowski 1998). However, when the ATP concentration is greater than that of magnesium a different pattern of cleavage is observed, and nuclease action is seen only once, ~5 nts upstream from the 3’ end of the Chi sequence (Taylor *et al*. 1985).

In either case, the RecB helicase is believed to translocate more slowly along the DNA than its RecD counterpart, which causes the formation of an ssDNA loop in front of the translocating RecB molecule (see Fig 1.10 for schematic of RecBCD function). Following Chi recognition, RecA is loaded onto this ssDNA loop by RecD. This promotes RecA nucleoprotein filament formation and the subsequent homology search and strand invasion which allows DNA repair by homologous recombination to occur (Anderson 1997). RecBCD is the first enzyme in the pathway to the repair of DNA double stand breaks, and initiates the homologous recombination repair pathway. As such, and considering the understood low copy number of RecBCD and the lack of DNA repair specific regulation, there is potential phenotypic impact of stochasticity in expression of RecB, RecC and RecD. These features make RecBCD an excellent candidate for single cell, single molecule study.

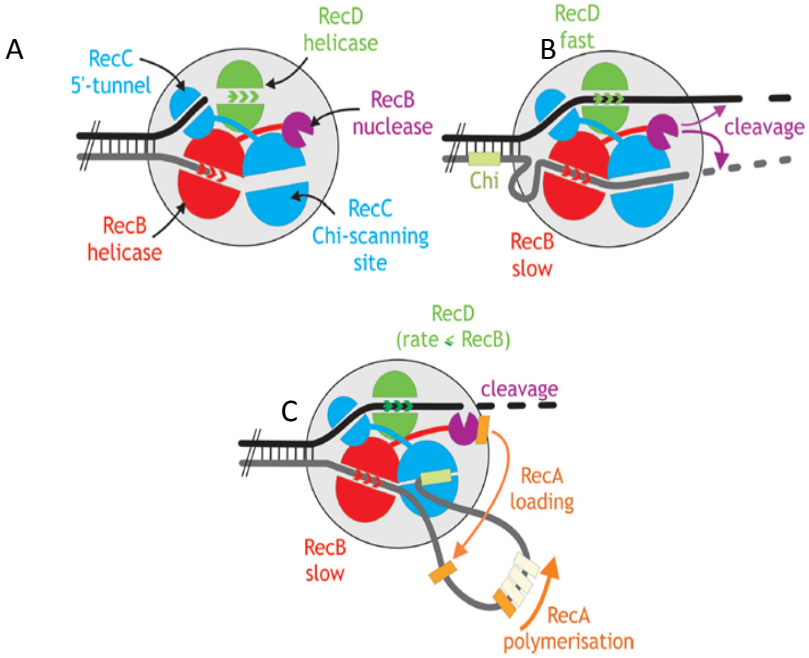


Figure 1.10. **Model of RecBCD enzyme mechanism.** A) RecBCD loads on to a double strand DNA end. B) RecBCD translocates along the DNA duplex, processing each strand at a different rate, producing a single strand DNA loop in front of the slower moving RecB protein. C) Chi recognition causes loading of RecA onto the single strand DNA loop, creating a RecA nucleoprotein filament formation. (Adapted from Dillingham & Kowalczykowski 2008).

1.6 Scope of this thesis

As described above, bacterial gene expression is a stochastic process in which there is always fluctuation. The impact of such fluctuation is enhanced when the molecules involved (DNA, mRNA and protein) are present at low levels. In order to better understand the process of gene expression and the impact that stochastic fluctuations can have, we must be able to detect and quantify mRNA and protein even at very low levels. Current techniques for low copy number protein and mRNA detection are

outlined above with their various advantages and disadvantages. As shown, fluorescence microscopy provides a way to examine protein and mRNA expression at very low levels without causing excessive disruption to the cellular system being investigated. In this work, mRNA FISH is used to detect low copy number *recB* and *recD* mRNA in *E. coli*, and the HaloTag-TMR labelling technique is developed and characterized for quantitative use in single bacterial cells. RecB, RecC and RecD protein are quantified. All measurements were taken within isogenic populations with the aim of understanding and evaluating the distributions of each molecule to gain insight into the process of RecBCD expression at the levels of both transcription and translation. Additionally a method combining the mRNA FISH and HaloTag protocols is detailed which allowed simultaneous detection of mRNA and protein within single cells. This thesis introduces and thoroughly characterises a new method for quantitative detection of low copy number proteins within single bacterial cells (HaloTag-TMR labelling), as well as presenting quantitative data for mRNA and protein copy number of the bacterial DNA repair protein RecBCD. Finally this thesis introduces proof of principle for a method combining HaloTag and FISH labelling that allows simultaneous mRNA and protein detection and is a substantial improvement over existing methods.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Stock solutions

20% (w/v) Glycerol

20% of glycerol dissolved in dH_2O and autoclaved at 115 ^oC for 15 minutes.

20% (w/v) Arabinose

20% of arabinose dissolved in dH_2O and autoclaved at 115 ^oC for 15 minutes.

20% (w/v) Glucose

20% of glucose dissolved in dH_2O and autoclaved at 115 ^oC for 15 minutes.

20% (w/v) Sucrose

20% of sucrose dissolved in dH_2O and autoclaved at 115 ^oC for 15 minutes.

Chloramphenicol

50 mg mL^{-1} of chloramphenicol (Cm) were dissolved in 100% ethanol and stored at -20 ^oC. Cm was used at a final concentration of 50 $\mu g\ ml^{-1.}$

Kanamycin

50 mg mL^{-1} of kanamycin (km) was dissolved in sterile Milli-Q water and stored at -20 ^oC. Km was used at a concentration of 50 $\mu g\ ml^{-1}$.

Ampicillin

50 mg ml^{-1} of ampicillin (amp) was dissolved in sterile Milli-Q water and stored at -20 ^{o}C. Amp was used at a concentration of 50 μg ml^{-1}.

Nalidixic Acid

20 mg ml^{-1} of nalidixic acid (nal) was dissolved in sterile Milli-Q water and stored at -20 ^{o}C. Nal was used at a concentration of 20 μg ml^{-1}.

1 M $MgS0_4$

Made up to 1 M in sterile Milli-Q water. Sterilised using a 0.2μM syringe filter. Stored at room temperature. Used at 2 mM.

0.1 M $CaCl_2$

Made up to 0.1 M in sterile Milli-Q water. Sterilised using a 0.2μM syringe filter. Stored at room temperature Used at 0.1 mM.

SafeView nucleic acid stain

Purchased from NBS Biologicals (#NBS-SV1), stored at 4^{o}C. 5 μl added to 100 ml agarose for gel electrophoresis.

16% formaldehyde

Purchased from Thermo Scientific (# 28906).

Formamide

Purchased from Ambion (#AM9342).

2.1.2 Culture media

All solutions made on site were prepared to the required volume in distilled water and autoclaved.

L-Broth	**1 Litre**
bacto-tryptone (Difco),	10g
yeast extract (Difco)	5g
NaCl	10g

pH adjusted to 7.5 with NaOH.

LB Agar

15 g Bacto-was added to 1 L of L-broth prior to autoclaving.

4 x M9 salts	**1 Litre**
88.5 mM KH_2PO_4	12g
197 mM Na_2HPO_4	28g
34 mM NaCl	2g
75 mM NH_4Cl	4g

Phage Buffer	**1 Litre**
22 mM KH_2PO_4	3g
49 mM Na_2HPO_4	7g

85 mM NaCl	5g
1 mM $MgSO_4$	0.1g
1 mM $CaCl_2$	0.1g
1% (w/v) gelatine	10g

SOC outgrowth media

Purchased from New England Biolabs (NEB, #B9035).

2.1.3 Imaging media

10 % LB imaging media	**100 ml**
4 X M9 salts	25 ml
1M $MgSO_4$	200 µl
0.1 M $CaCl_2$	100 µl
L-Broth	10 ml
20% glucose	1 ml

Made up to 100 ml with sterile water.

Amino acid imaging media	**100 ml**
1 X M9 salts	25 ml
2mM $MgSO_4$	200 µl
0.1 mM $CaCl_2$	100 µl

0.2% glucose	1 ml
MEM Essential Medium	2 ml
MEM Non- Essential Amino Acids	1 ml

Made up to 100 ml with sterile water.

2.1.4 Buffers

10 x PBS

Purchased from Fisher Scientific (#BP399-500). Used at 10X and 1X. Diluted in Milli-Q water for 1X.

20 x SSC

Purchased from Ambion (#AM9770). Used at 20X and 1X. Diluted in DEPC-treated water for 1X.

50 X Tris-acetate (TAE)	**1 Litre**
2 M Tris-base	242 g
0.95M Glacial acetic acid	57.1 ml
0.05 M EDTA	14.6 g

TE Buffer

Purchased from Ambion (#AM9849)

100% DMSO

Purchased from Sigma Aldrich (#472301)

HaloTag fixation solution

16% formaldehyde (Thermo Scientific #10751395)	156 µl
10x PBS	100 µl

Made up to 1 ml with Milli-Q water

FISH wash solution

Formamide	353 µl
20x SSC	100 µl

Made up to 1 ml with Milli-Q water

FISH hybridisation solution

Dextran Sulfate Sodium Salt (Sigma-Aldrich #D8906-10G)	353 µl
Formamide	100 µl
E. coli tRNA (Sigma-Aldrich #R1753)	10 mg
20x SSC	1 ml
200 mM Ribonucleoside Vanadyl Complex (NEB #S1402S)	100 µl

Made up to 10 ml with DEPC-treated water

2.1.5 Fluorophores

5 nmol *recB* mRNA FISH probes

Stellaris DNA FISH probes with TAMRA Dye purchased from Biosearch Technologies (#SMF-1001-5). 48 probes of 20 nucleotides, with a minimum of at least 2 nucleotides between binding sites were designed against the 3543 base pair *recB* gene by algorithm provided at:

https://www.biosearchtech.com/support/education/stellaris-rna-fish

Probes were diluted to 100 μM and 25 μM by resuspension in RNAse-free TE buffer.

Table 2.1. DNA oligonucleotide probes for *recB* mRNA

Probe Name	**Sequence (5'-3')**
recB _1	CTG TAA GGG CAA GCG CAA AG
recB _2	GCA ATC GTA AAG GTT TTG CC
recB _3	TCG TGG ATA TTG CTA CGG AT
recB _4	GTT CGG CTA ACA ACA ACC AC
recB_ 5	AAA GGC ATT CAG GTT GAG CA
recB_ 6	AAT CAG CTG CTG CTC AAA CA
recB _7	AAA GAC GAC CTG GGC TAT TT
recB _8	CGC CTT GCA GAT AAC GAT TA
recB _9	CTG CTG TTT TAC CGT ATC AA
recB _10	ACCAGAAGATTCGATCAGCG
recB _11	CGG TTA AAC TTG CGT CGA TC
recB_12	ATC TTG TCG ATC CAT TTA GC
recB _13	CCG GCA ACT GAT AAC TGT TT

recB_14	CTT CGT GCG ATC TTC TAA GA
recB_15	TTG ATC GAT CGC CTC AAA CA
recB_16	CAG ATC GCG GAT CGA CAA TG
recB_17	AGC CGA CTT AAC ATG TCA TC
recB_18	CGC CAA TTA GCA ACA ATG CG
recB_19	TAC GCG CCT TCA TAT AAG TG
recB_20	TGT CTA AAG TGT AGT GGG CG
recB_21	AAG CTT ATT CAC GCT GTT CA
recB_22	TCG CGA AAC ATG AAC GCG TC
recB _23	AAA CGG AAG GGA TTT CCA GC
recB _24	CCT GCA ACA ACC AAA GCA TT
recB _25	CAG ATT TGC CGA TAA CCA TC
recB _26	GTT TTC AGC AAT GTT ACG CG
recB _27	TTG TAG CAG TTC GCT GAT AT
recB _28	TGG ATC GTG ACA ATC TGC AC
recB _29	TGG ACG CGG AAA TTG GTG AT
recB _30	ATC GTG ATA AAA CGC CTG CT
recB _31	TTA AGA TCC AGA ACT GCC TC
recB _32	AAG CAA ACG CAG ATC TTC CG
recB _33	AAT GCC AAA CCG AAC GTG TC
recB _34	AAC GCT TCA ATA CAG GTG CG

recB _35	TTG GTT ATC ACC AGT TTG TG
recB _36	TCA GCT CTG CTG TAG AAA CA
recB _37	AAA CCA GAG TAG CTG GTG AC
recB _38	TGA TGT GGT GTT AAC GTC GG
recB _39	TCA ACC GGC TGG GTA AAA TC
recB _40	TTA TTG CGG GCG GAA AGT TG
recB _41	GAT AAA ACT CCA TCT CCA CC
recB _42	AAC GTA TCA AGC TGA CTG GC
recB _43	GCC AGG GTA TAA AGC TGA TA
recB _44	CAA TGC GAT GGC GCA GAT AA
recB _45	TGG TGC TCA TAG TCG TAA TC
recB _46	ACA GAT AAA TAA CGC CGC CA
recB _47	GAT GTT CTT TAT CAA CG CCA
recB _48	CAT ACC GGC AAA CAT CTC AT

5 nmol *recD* mRNA FISH probes

Stellaris DNA FISH probes with TAMRA Dye purchased from Biosearch Technologies (#SMF-1001-5). 48 probes of 20 nucleotides, with a minimum of at least 2 nucleotides between binding sites were designed against the 1827 base pair *recD* gene by algorithm provided at:

https://www.biosearchtech.com/support/education/stellaris-rna-fish

Probes were diluted to 100 μM and 25 μM by resuspension in RNAse-free TE buffer.

Table 2.2. DNA oligonucleotide probes for *recD* mRNA

Probe Name	Sequence (5'-3')
recD_1	GTC AGG GCA AAT TGC ACA TC
recD_2	ATC ATG ACT TAA CAG TGC CG
recD _3	CAA ACA AAC GTG TCC CTC TC
recD _4	CGT TAT TTT CCA GTC GTG AA
recD _5	GAT TTC ACT GAC ACA GGT CG
recD _6	ATT CTT CCC AAT TTT GTA GC
recD _7	AAA GAC GAT CGC CAC AGA GG
recD _8	TTA CAC CAC ATG CGA TTC AA
recD _9	CTT CGT TGA AAA AGC GTG CC
recD _10	AGC TTC ATC AAC CTC AAT GG
recD _11	GGA AAA AGT TTG TCC AGG GT
recD _12	CGC AAC TTT TTG CCA GTT AA
recD _13	TGA ATT AAC GCT GCC AGC AA
recD _14	TGC CGA GAG ATT CGG TTA AG
recD _15	CAG CGG TAA CTG TCG CAA AG
recD _16	CGG AAT GCG TTT CTT TTG TT
recD _17	CAA TCG GTG CAA AGT GCT GG
recD _18	GAT GAC GTA AAC GCT GGC TA
recD _19	ATC AAG ATG CAG CGG GTT AC

recD _20	GAT CGA TCA TTG ACG CTT CA
recD _21	TCG ATC AGT CTC GAC ATC AT
recD _22	GAT CGC CGA GAA AGA TCA CT
recD _23	TGG CAT AAG CGC AGA TAT CG
recD _24	GAG ACT GTC GCG CAA AGA TG
recD _25	CGA TAG CTT TTT TGC AGC AG
recD _26	ACT GAC CAA TGC CAG AAT CG
recD _27	TTT ATC ACC ACG GTT GAT CG
recD _28	CCT GCT GAA AAA CGG TTT TC
recD _29	GCC GTT TTT CGA TAT CAG TA
recD _30	AAT ATA ATC TTC GCC GCT CT
recD _31	CAA GAG CTT CCT CAA GCA TC
recD _32	CAG CAG ATC CAG ATA ACG TC
recD _33	CCT GAA TGA TTA AAT CCG GC
recD _34	CGC ACA AAA GCT GGT ACT CA
recD _35	CTC AAT TCG CTC ATT CAG TC
recD _36	TGC GCT TCT GTT GCA TAA AC
recD _37	AAC GAG AGT GCG GAT GAC GA
recD _38	GTC ATT ACG GGC AAT CAT CA
recD _39	CAT TAA ACA ACC CAA GCG CG
recD _40	ATC CAG CGC AAT ACC GAT AT
recD _41	ACT CGG TTG CAC AGA CTT AA
recD _42	TAG TTT CGT GCT CTG GCA AG

recD _43	TGC GAT TTA TGT ACC GTC AT
recD _44	AAA ATC AAC GCC GCA TGG TC
recD _45	ATA AAC CAG CTC TCG CGT TA
recD _46	GCA CTT AAT ATG CGC TCA TC
recD _47	CTC AGT ACG AGT GGC GAT TG
recD _48	CGT GAA CTA AAC AAC GCC GC

5 nmol HaloTag mRNA FISH probes

Stellaris DNA FISH probes with Fluorescein Dye (#SMF-1025-5). 26 probes of 20 nucleotides, with a minimum of at least 2 nucleotides between binding sites were designed against 891 base pair *HaloTag* gene by algorithm provided at:

https://www.biosearchtech.com/support/education/stellaris-rna-fish

Probes were diluted to 100 μM and 25 μM by resuspension in RNAse- free TE buffer.

Table 2.3. DNA oligonucleotide probes for *HaloTag* mRNA

Probe Name	**Sequence (5'-3')**
HaloTag _1	TCG AAT GGA AAG CCA GTA CC
HaloTag _2	CAG GAC TTC CAC ATA ATG GG
HaloTag _3	AAC ATC GAC GTA GTG CAT GC
HaloTag _4	TAC CGT GCA GGA ACA GCA CA
HaloTag _5	CAC ACG TAG GAG GAG GTC GG

HaloTag _6	AAC ATG CGG GAT GAT GTT GC
HaloTag _7	CTG GAG CAA TGC AGC GAT GG
HaloTag _8	GAT TTG CCC ATA CCG ATC AG
HaloTag _9	AAT AAC CCA GGT CTG GTT TG
HaloTag _10	AAG CGG ACG TGG TCG TCG AA
HaloTag _11	AGG GCT TCG ATG AAG GCA TC
HaloTag _12	TCG TGA ATG ACC AGG ACG AC
HaloTag _13	AGT GGA AAC CCA GAG CGG AG
HaloTag _14	AAT ACC TTT GAC GCG CTC TG
HaloTag _15	GGG CGG ATG AAC TCC ATA AA
HaloTag _16	AAG GTC TCG CGG GCA AAT TC
HaloTag _17	TGA TCG ATG ATC AGC TTG CG
HaloTag _18	GCG TAC CCT CGA TAA AAA CG
HaloTag _19	TAA TGG TCC ATC TCG ACT TC
HaloTag _20	CAA CAG GAT TCA GGA ACG GC
HaloTag _21	ATT GGC AGC TCG TTT GGG AA
HaloTag _22	CAG TCC ATG TAT TCT TCG AC
HaloTag _23	AGA ACA GCA GCT TCG GGA CA
HaloTag _24	TGC AGT TAG GCA GGC TTT TG
HaloTag _25	TTC TTG CAG CAG ATT CAG AC
HaloTag_26	CCG GAA ATC TCC AGA GTA GA

5 mM HaloTag TMR Ligand

Purchased from Promega #G8252. Diluted to 5 µM working stock in 100 % DMSO.

2.2 Methods

2.2.1 Bacterial methods

Bacterial stocks stored at -80^{o}C

500 µl of an overnight culture was mixed with 1 ml of 20% (w/v) glycerol and stored at -80°C.

Overnight cultures

Overnight cultures were prepared by inoculation of relevant media from either a single isolated colony on an agar plate or a glycerol. The relevant antibiotics were added and cultures were placed at the required shaking speed (250 rpm) and temperature (30^{o}C or 37 ^{o}C).

Transformation of *E. coli* by $CaCl_2$ treatment followed by a heat shock

NEB Turbo Competent *E. coli* (High efficiency) cells were purchased (#C2984H) and chemical transformations were conducted following the instructions of the manufacturer.

Production of electrocompetent *E. coli* cells for transformation by electroporation

In order to transform cells by electroporation, a stock of electrocompetent cells of the appropriate strain must be available. These stocks were generated by culturing the appropriate strain at the appropriate temperature overnight. An over-day culture was made with a 1 in 100 dilution of the overnight culture, grown to an OD_{600} of 1 before being placed on ice for one hour. The culture was then centrifuged and washed once with cold water before being concentrated in H_2O and washed once more in H_2O. The culture was then diluted with 10% glycerol and dispensed into microcentrifuge tubes before being flash frozen in liquid nitrogen and stored at -80°C until required.

Transformation *E. coli* by electroporation

A micro centrifuge of competent cells as described above was thawed on ice and 2µl of the desired plasmid miniprep was added to the defrosted cells. The cells were then transferred to a pre-cooled cuvette for electroporation. Immediately after this, pre-warmed SOC was added to the cuvette and the total volume transferred to a microcentrifuge tube. The tube was placed on a shaker for one hour at the appropriate temperature. The mixture was then plated on appropriate plates and incubated at the relevant temperature overnight.

Serial dilution of *E. coli* cultures

Overnight cultures were diluted to OD_{600} 0.3 which was considered the 10^0 sample. This sample was then sequentially diluted to 10^{-6}.

Plasmid mediated gene replacement (PMGR)

Plasmid mediated gene replacement is an in-out cloning method first described by Link *et al*. (1997) that uses a multicopy plasmid to carry a genetic modification and

integrate it into the chromosome. The plasmid, pTOF24, has a temperature sensitive replication initiator protein (repA101TS), chloramphenicol (CmR) and kanamycin (KnR) resistance genes which act as positive selection markers, and a sacB gene which encodes for a levansucrase and acts as a negative selector. The KnR gene is flanked by PstI and SalI which allows the resistance to be replaced by the desired cloning region of interest. pTOF24 derivatives that had been modified to contain the desired mutation (through Gibson Assembly, see Section 2.2.2 DNA cloning techniques) were streaked on Cm plates and grown overnight at 42°C. The replication initiator of pTOF24 will not initiate at this temperature, allowing selection of cells that have integrated the plasmid (with the CmR gene) into their chromosome. The largest colonies from this step were then re-streaked on Cm plates and once again grown at 42°C. This step was done to purify the integrants. Single colonies from these plates were then picked and inoculated in LB at 30°C overnight with no selection. This allows the selection of cells that have excised the plasmid from the chromosome. 100 µl of a 10^{-5} dilution was then plated in an LB agar with 5% sucrose and grown at 30°C overnight, allowing negative selection as cells that have the pTOF24 plasmid on their chromosome are sensitive to sucrose. Colonies from the 5% sucrose plates were then patched onto Cm, sucrose and LB plates and grown at 30°C overnight. Cells that have excised the plasmid from their chromosome grew on sucrose and LB but not on Cm (see Fig. 2.1). The genotype of colonies from these plates was then checked with PCR, as detailed in Section 2.2.2 (DNA cloning techniques).

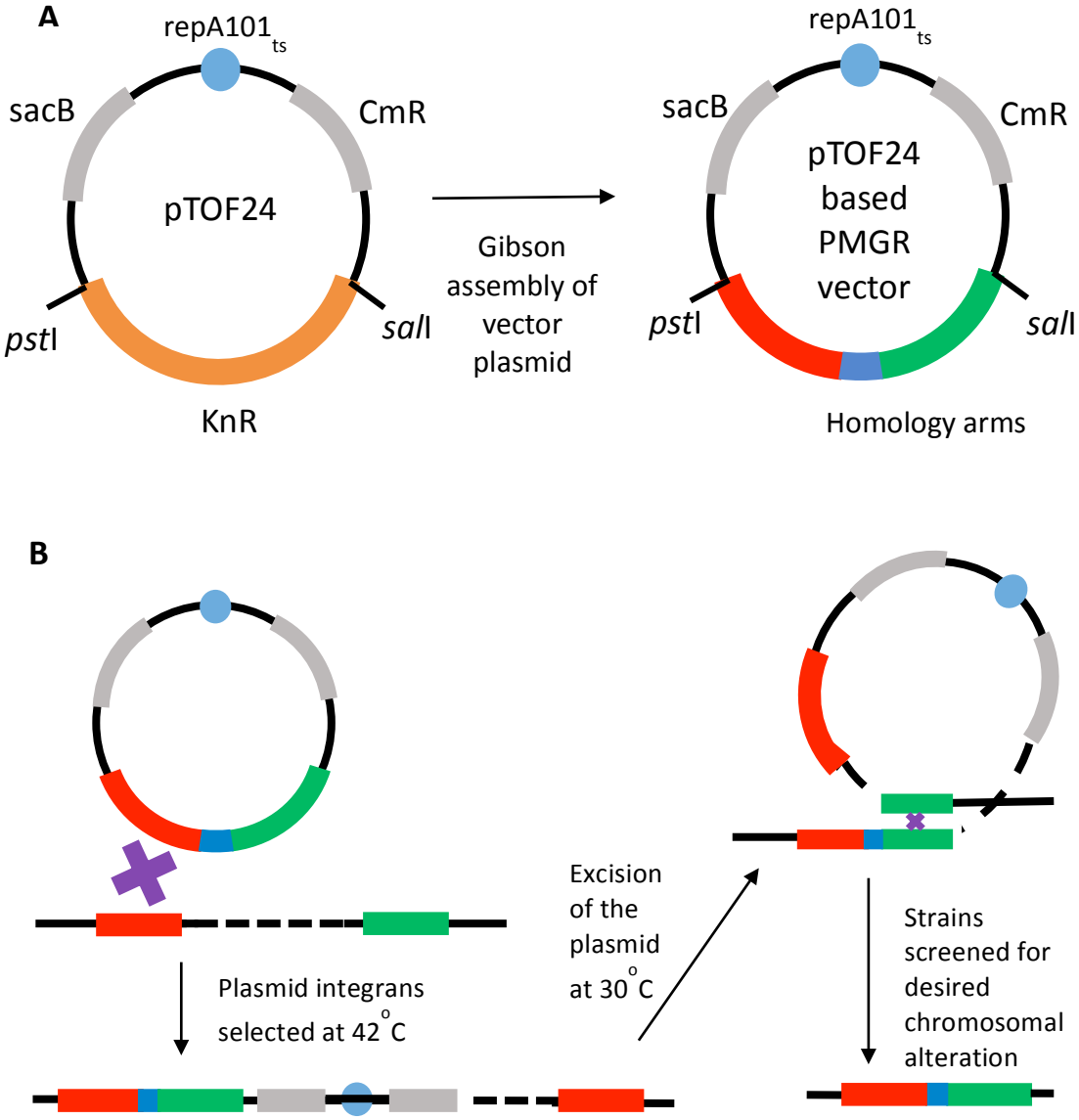


Figure 2.1. Plasmid mediated gene replacement. A) PTOF24 derivatives for use in PMGR are constructed through PCR and Gibson assembly using the Pst1 and Sal1 restriction sites of the pTOF24 plasmid. repA101ts codes for a temperature-sensitive initiator protein that initiates at 30°C but not 42°C. CmR and KnR code for chloramphenicol and kanamycin resistance respectively. SacB encodes a levan sucrose, which converts sucrose into a product that is toxic for *E. coli* and so is used as a negative selector. B) Chromosomal modification using PTOF24 vectors. When cells transformed with the plasmid are grown at 42°C, the plasmid can only be replicated if it integrates into the chromosome. PTOF24 vectors are designed to have two regions of homology to the chromosome. This allows chromosomal integration at 42°C , which occurs through RecA mediated homologous recombination between one of the regions of homology on the plasmid and the corresponding chromosomal region. This results in the integration of the entire plasmid onto the chromosome. When growing such an integren at 30 °C, the plasmid is able to excise from the chromosome, which also occurs through RecA mediated homologous recombination. When integration occurs at the first region of homology (red) and excision occurs at the second (green), the WT region of the chromosome is replaced with the modified DNA region from the plasmid.

2.2.2 DNA cloning techniques

Genomic DNA extraction for PCR

The Promega Wizard Genomic DNA purification Kit (# A1120) was used following the instructions of the manufacturer. The DNA was stored at -20°C.

Plasmid DNA preparation for PCR

The QIAGEN QIAprep® Spin Miniprep Kit (# 27106) was used to purify plasmid DNA following the instructions of the manufacturer.

Polymerase chain reaction for confirmation of strains

To check the genetic content of newly generated strains, PCR using Promega GoTaq® polymerase (# 9PIM300) were prepared: 10 µl GoTaq® reaction buffer; 1 µl 10 mM dNTP mix (Thermo Scientific #R0192); 1 µM upstream primer; 1 µM downstream primer; 0.25 µl GoTaq® polymerase and ;1 µl template DNA. This gave a final concentration of 50 µl.

Polymerase Chain Reaction for Gibson Assembly

This technique was used to amplify a DNA sequence which could then be seamlessly ligated into a plasmid vector. Primers of ~60 bps were designed. The last ~20 bps of the upstream primer (when considered 5'-3') and the last ~20 bps of the downstream primer (when considered 5'-3') are used to amplify the desired gene/ DNA fragment. The remaining ~40 bps of each primer are designed to be homologous to the vector the DNA will be cloned into (see Fig. 2.2). This fragment is then used in Gibson Assembly to create a clean fragment insertion. All amplification was done with NEB

Q5® High Fidelity DNA Polymerase (#M0491S) to ensure accurate amplification of the initial template: 10 µl 5x Q5® reaction buffer; 1 µl 10 nM NEB dNTPS; 1 µM upstream primer; 1 µM downstream primer; 1 µl template DNA; and 0.5 µl Q5® polymerase. This gave a final volume of 50 µl.

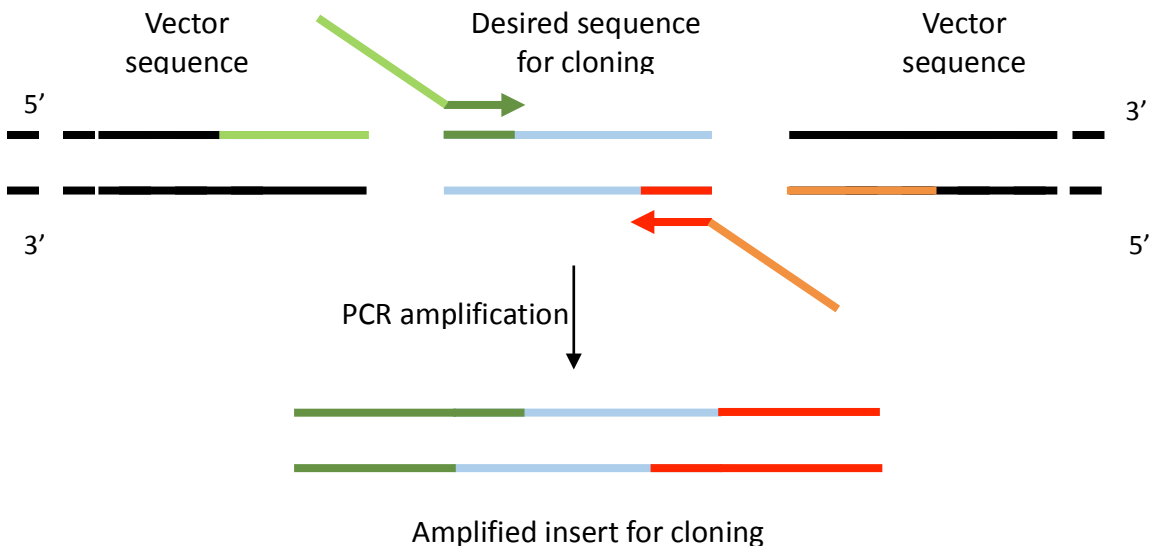


Figure 2.2. Polymerase Chain Reaction for Gibson Assembly. To amplify fragments for Gibson assembly primers are designed that have ~20 bp homology to the sequence being amplified (5'-3 shown in dark green, 3'-5' shown in dark red), and ~40 bp homology to the vector sequence that the insert will be ligated into(5'-3 shown in light green, 3'-5' shown in orange). After PCR amplification the fragment produced contains the desired sequence for cloning and the homologous ends for use in Gibson Assembly.

Gibson Assembly of DNA fragments

This technique (also known as isothermal assembly) allows fusion of double strand DNA (dsDNA) fragments with sufficiently homologous ends, and was introduced by Gibson *et al*. (2009). Fragments of interest were first PCR amplified (as described

above) and gel purified. The cut vector was also gel purified. An isothermal assembly reaction buffer was prepared as described below:

Gibson Assembly reaction buffer	**6 ml**
1M Tris-HCl pH 7.5	3 ml
1M $MgCl_2$ (Amresco #E525-100ML)	300 µL
10 mM dNTP mix (Fermentas #10319879)	600 µL
1M DTT (Fermentas #10699530)	300 µL
Polyethylene Glycol 8000 (AESAR #43443.36)	1.5 g
Nicotinamide adenine dinucleotide (Applichem Lifescience, A1124.0005)	20 mg
Made up to 6 ml with double distilled H_2O	

This mix was aliquoted into 320 µL volumes and all but one stored at -20°C (stable for up to one year). For use, the following were added to the remaining buffer aliquot:

Isothermal assembly reaction enzymes

T5 Exonuclease (NEB #M0363S)	1.2 µL
Phusion polymerase (Finnzymes # 10024537)	20 µL
Taq Ligase (NEB #M0208L)	160 µL
Reaction buffer	320 µL
Double distilled H_2O	700 µL

This gave a 1201.2 µL volume which was then aliquoted into 15 µL volumes. These aliquots could be stored at -20°C for up to one year. To perform Gibson Assembly,

the DNA fragments to be ligated (up to a total 5 µL volume combined) and the above described 15 µL aliquots were combined and the mix placed at 50°C for 30 minutes, which is sufficient for the completion of the reaction (see Fig. 2.3 for details).

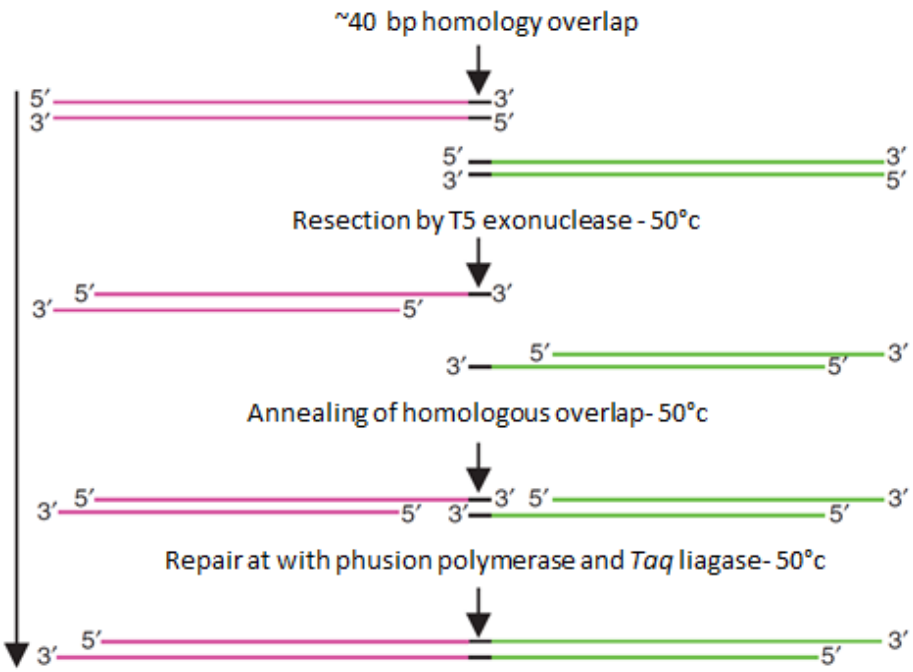


Figure 2.3. Gibson assembly reaction. Gibson Assembly allows the joining of two adjacent DNA molecules with sufficiently homologous ends. The T5 exonuclease removes 5′ nucleotides from each DNA molecule allowing the complementary single stranded DNA overhangs to anneal. DNA polymerase fills the gaps and DNA ligase seals the nicks to give a fully formed double stranded DNA molecule. The T5 exonuclease is heat-liable and so is inactivated during the incubation period. (Adapted from Gibson *et al*. 2009).

Restriction digestion of PCR purified DNA

Restriction enzymes and buffers were obtained from NEB and digestions were carried out following the instructions of the manufacturer.

Sequencing of DNA

Each sample was submitted for Sanger sequencing at Edinburgh Genomics in a 6 µL (3.2 pmole/µL of the appropriate primer and the required concentration of purified PCR product).

Length of purified PCR product (bp)	Required DNA Quantity (ng)
100-200	1-3
200-500	3-10
500-1000	5-20
1000-2000	10-40
> 2000	40-100

PCR product purification for cloning (QIAGEN kit)

The QIAGEN QIAquick PCR purification kit (Cat. No. 28104) was used to purify DNA fragments before cloning, following the instructions of the manufacturer.

Gel electrophoresis for detection of PCR products or plasmid DNA

DNA fragments from PCR or plasmid digestions were separated on a 1% (w/v) agarose gel. If the DNA fragment was being recovered for further use, Seakem GTG Agarose (Lonza, #50074) was used to produce the gel and the QIAquick PCR Extraction Kit was used to extract the DNA following the instructions of the manufacturer (#28704). If the DNA fragment was being visualised to determine length only, UltraPure Agarose (ThermoFisher, #16500500) was used to produce the gel. The appropriate amount of agarose was dissolved in 1 X TAE and SafeView

(NBS Biologicals #NBS-SV1) was added to allow visualisation of the DNA under UV light. Gels were run at 80-120 V for up to 2 hours and DNA was visualised using a UV box (BioRad). The fragment size was determined using DNA ladders (HyperLadder 1kb, Bioline, # BIO-33053 and Supercoiled DNA Ladder, #N0472S). When required, DNA was quantified with a Nanodrop (ND-1000v3.5).

2.2.3 Phenotypic testing

T4Gene2 test for normal RecBCD activity

This phenotypic test uses enterobacteria phage T4. This phage infects *E. coli* and undergoes a lytic lifecycle, causing the death of infected cells and the destruction of the membranes. T4 phage has a double-stranded DNA genome (Miller *et al*, 2003) and the T4Gene2 mutant lacks gene2 and its protein. In wild type T4, gene2 protein protects the DNA double strand ends from RecBCD exonuclease degradation on entering an *E. coli* cell. In the absence of gene2, the phage DNA is not protected and is degraded by RecBCD on entering *E. coli* cells (Portakal, 2008). A plaque assay can test for RecBCD exonuclease activity. A strain with RecBCD exonuclease functionality will degrade the phage DNA and prevent cell death and membrane destruction, meaning few plaques will form when the strain is grown on a plate with T4. However, a strain that lacks RecBCD exonuclease activity will fail to degrade the phage DNA, allowing the T4 to complete its lytic lifecycle and cause cell death and membrane destruction resulting in many plaques when the strain is grown on a plate with T4. The strains to be tested and appropriate controls were grown overnight

in 10 ml LB. The following day, 3 x 300 µl of each strain were incubated for 10 minutes with 100 µl of 3 serial dilutions of T4gene 2 phage (dilutions were made in phage buffer). This 400 µl volume was then mixed with 3 ml of soft agar (1.2 ml melted LB agar and 1.8 ml LB and incubated at 50^{o}C) before being poured onto a LB agar plate. The soft agar was allowed to solidify and the plates were incubated at 37^{o}C overnight. The following day, the plaques for each strain at each dilution of phage were counted.

Nalidixic acid test for RecBCD-dependant DNA repair activity

Nalidixic acid functions by inhibiting DNA gyrase and Topoisomerase IV. This inhibition induces DNA double-strand breaks, which must then be repaired by the action of RecBCD (Newmark *et al*. 2005). Strains that are incapable of such repair will form fewer colonies when plated in serial dilution on Nal^{+} plates compared to the same strains plated on LB only plates. Strains to be tested and the appropriate controls were grown overnight in 10 ml LB. The following day, these strains were diluted 1 in 250 and grown to OD 0.2-0.3. The strains were then normalised to OD 0.2 before being spotted in serial dilution onto 2 µg/ml Nal and LB only plates followed by incubation at 37^{o}C overnight. The number of colony forming units could then be compared with and without nalidixic acid.

2.2.4 Protein and mRNA detection techniques

2.2.4.1 Protein detection by western blot

All western blot experiments presented in this work were performed by Lorna McLaren.

Cell Culture and Protein Lysate

A 10ml overnight culture of each strain was grown in a shaking 37°C incubator. The following morning, a day culture was prepared: 200ul of the overnight culture was added to 55ml of LB and grown at 37°C incubator until an OD_{600} of between 0.2 and 0.3 was reached. 50ml of the culture was spun in a refrigerated centrifuge at 4°C at 4,000rpm for 8 minutes. The cells were kept on ice throughout the following procedure: the supernatant was removed and the pellet was resuspended in 1ml RIPA buffer plus sodium orthovanadate, PMSF and protease cocktail inhibitor (Santa Cruz, #sc-24948), according to manufacturer's instructions. 2ul of benzonase (Sigma Aldrich, #9025-65-4) was added to the lysate and incubated at 37°C for 30 minutes. 300ul of the cell lysate was added to 100ul of 4 x loading buffer (Life Technologies, #NP0007), dispensed into smaller volumes and stored at -20°C until required for protein gels. The remaining 700ul of lysate was stored at -80°C until required.

SDS-PAGE and Western Blot

Nupage precast gels (Life Technologies, C-NP0321-X) were used with the Nupage gel system in accordance with manufacturer's instructions. 12.5ul of cell lysate was loaded per well and 5ul of the prestained Fermentas PageRuler protein marker

(Thermo Scientific, 11854544). A total of 21ug of Halo standard protein (Promega, #G4491) was digested with ProTEV Plus protease (Promega, V6101) at 30°C for 30 minutes to cleave the Halo protein from the GST protein. From this digest, a stock dilution of 100pg/ul of total standard protein was prepared. 500 pg of the digested standard Halo protein was directly added to the cell lysate of BW. After loading, the gel was run at 200V for 70 minutes in MOPS running buffer (Nupage, NP0001), removed from the cast and the proteins transferred onto L-PVDF membrane (GE Healthcare Amersham, # 10600023) for 1 hour at 250mA using Nupage transfer buffer (Life Technologies, NP00061) and the Mini Trans blot electrophoretic transfer cell (Biorad, #1703930), according to manufacturer's instructions. After transfer, the membrane was rinsed in PBS and blocked for 1 hour in 4% ECL Prime blocking agent (GE Healthcare Amersham, #RPN418) in 1% Tween 20 (VWR, #663684B) in PBS (PBST) at room temperature in an orbital shaker. The blot was incubated in the Halo antibody (mouse monoclonal anti-Halo antibody (Promega, #G9211) diluted 1:1,000 in 4% prime blocking agent PBST at 4°C overnight in a moist chamber and the following day was washed with PBST for 90 minutes at room temperature on an orbital shaker with multiple changes of PBST. The membrane was incubated in the secondary antibody, donkey anti-mouse horse radish peroxidase-labelled antibody (Abcam, #ab6820), diluted 1:30,000 in 4% Prime blocking agent in PBST for 1 hour at room temperature on an orbital shaker. The membrane was washed for 90 minutes on an orbital shaker at room temperature with multiple changes of wash to reduce background. The Halo proteins were detected using ECL Prime substrate (GE Healthcare Amersham, #RPN2232) for 5 minutes, according to

manufacturer's instructions. Hyperfilm ECL photographic film (GE Healthcare Amersham, # 28906836) was used according to manufacturer's instructions and the film was developed in a Konica Minolta (#SRX-101A) developer.

2.2.4.2 Quantitative labelling with HaloTag-TMR

This work aimed to establish labelling protein with the HaloTag as an accurate method for precise measurement of protein number in *E. coli*. Overnight cultures of the strains of interest were set up in 10 ml 10% LB imaging media (used throughout work with the HaloTag). The following day, a dilution of 1 in 250 or 1 in 500 was made in imaging media and the strains were grown to OD_{600} 0.2-0.3 at 37 ^{o}C. A volume of cells equivalent to 1 ml at OD_{600} 0.2 was then pelleted by centrifugation (10 minutes, 4000 RPM, 4 ^{o}C) and resuspended in 1 ml fresh media. To this, 10 µl of 500µM HaloTag TMR ligand was added to give a final TMR concentration of 5µM (or 10 µl 100% DMSO to no TMR controls). All samples were incubated shaking at 37^{o}C in the dark for one hour to allow labelling of the HaloTag with the TMR ligand. The cells were then pelleted by centrifugation (3 minutes, 8000 RPM, 4 ^{o}C) and a suction pump was used to remove the supernatant, increasing speed and efficiency of TMR removal. The cells were then washed as quickly as possible five times, each with 1 ml of fresh media. The microcentrifuge tube was changed with each wash. Following this the cells were incubated in fixation solution for one hour to allow complete fixation. The cells were pelleted and then washed with 1 ml 1X PBS twice more before being mounted on an agarose pad for imaging. This protocol allowed visualization of individual TMR molecules bound to RecBHalo. A schematic of the

protocol can be seen in Figure 2.4. All steps involving formaldehyde were conducted in the fume hood and waste was disposed of following institute guidelines.

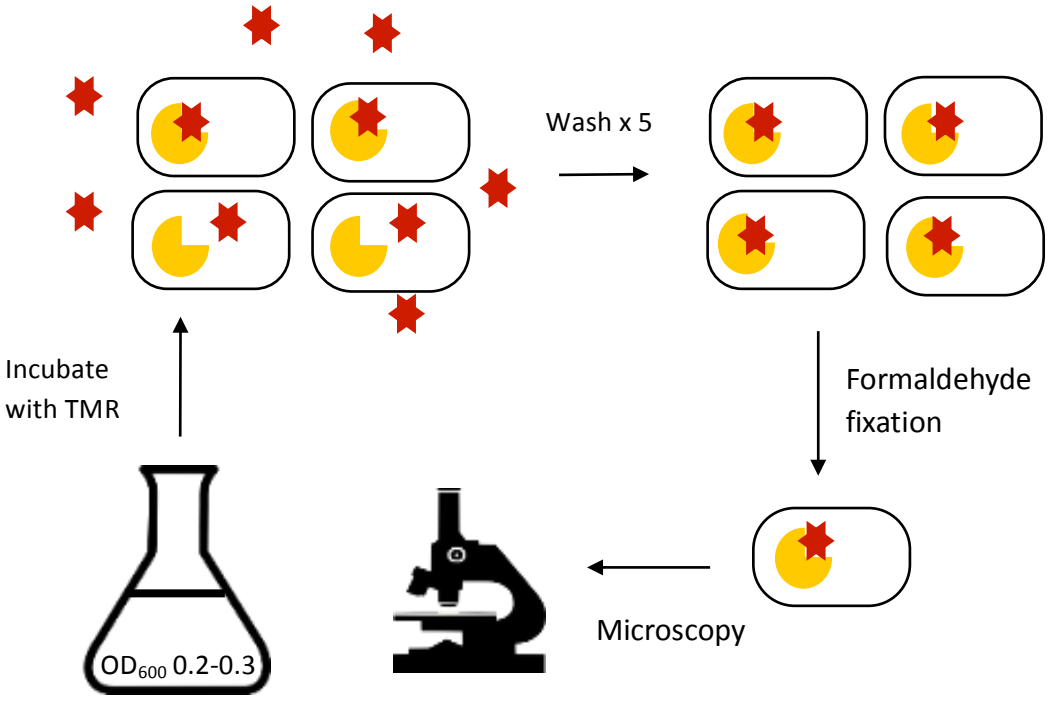


Figure 2.4. Labelling of protein with the HaloTag. Cells were first grown to OD_{600} 0.2-0.3 before being labelled with TMR for one hour, washed extensively, fixed with formaldehyde and imaged using an epifluorescence microscope.

2.2.4.3 Labelling of mRNA with single molecule fluorescence *in situ* hybridisation (smFISH)

This protocol was modified from Skinner *et al*. (2013). A schematic of the process can be seen in figure 2.5. Overnight cultures of the desired strains were incubated shaking at 37°C in amino acid imaging media (used throughout). 1 in 500 to 1 in 250

of this overnight culture was incubated in imaging media overday and grown to an OD_{600} of 0.2. The equivalent of 5 ml at OD_{600} 0.2 of the overday culture was then pelleted by centrifugation (3500 RPM, 5 minutes, 4 ^{o}C). The pellet was resuspended in ice-cold 1 X PBS and transferred into a microcentrifuge tube. The cells were then fixed in 3.2% formaldehyde solution for 30 minutes, centrifuged and washed twice with 1 X PBS. The cells were then resuspended in DEPC-treated water and incubated overnight in 70% ethanol to allow permeabilisation. In low light conditions, the cells were centrifuged and then resuspended in 40% formamide wash solution and incubated for 5 minutes before being spun down once more and then treated with 40% formamide hybridisation solution. This resulted in a final concentration of 1µM of the FISH probes (as seen in Table 2.1, Table 2.2 and Table 2.3). This mix was incubated at 30°C in the dark overnight to allow the probes to hybridise to their target DNA. The following day, 1 ml of 40% wash solution was added and the mix was centrifuged before resuspension and incubation in 1 ml 40% formamide wash solution for one hour. This was repeated three times in total. Finally, the cells were resuspended as needed in 1 X PBS and mounted onto agarose pads for imaging. All steps involving formaldehyde and formamide were conducted in the fume hood and waste was disposed of following institute guidelines. Probes were designed against MG1655 *recB* and *recD* transcripts and *ΔrecB* and *ΔrecD* strains were used as negative controls, see Table 2.7 for strain details. Probes were also designed against the HaloTag gene for use in the combined HaloFISH protocol.

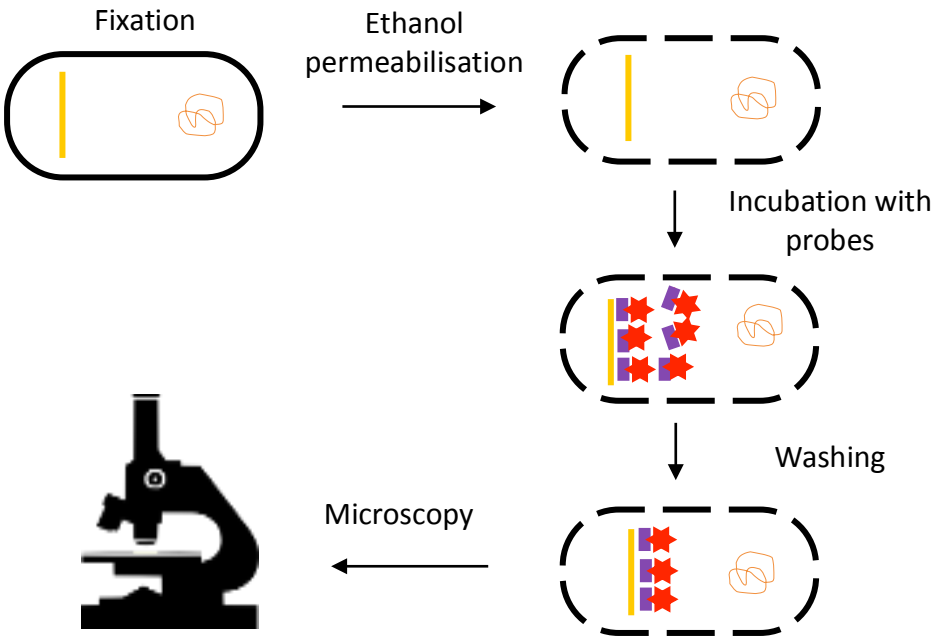


Figure 2.5. Labelling of mRNA with mRNA FISH. Cells were fixed in formaldehyde, permeablised in ethanol, incubated with mRNA FISH probes, washed and then imaged under an epifluorescence microscope.

2.2.4.4 Fluorescent *in situ* hybridisation combined with HaloTag labelling

Overnight cultures of the strains of interest were set up in 10 ml 10% LB imaging media (used throughout). The following day, a dilution of 1 in 500 was prepared using imaging media and the strains were grown to OD_{600} 0.2-0.3 at 37°C. A volume of cells equivalent to 25 ml at OD_{600} 0.2 was then pelleted by centrifugation (10 minutes, 4000 RPM, 4°C) and resuspended in 1 ml fresh media. To this, 20 µl of 500 µM HaloTag TMR ligand was added (or 20 µl of 100% DMSO to the no TMR

controls). All samples were incubated shaking at 37^{o}C in the dark for one hour to allow labelling of the HaloTag with the TMR ligand. The cells were then pelleted by centrifugation (3 minutes, 8000 RPM, 4^{o}C) and a suction pump was used to remove the supernatant, increasing speed and efficiency of TMR removal. The cells were then washed as quickly as possible five times, each with 1 ml of fresh media. The microcentrifuge tube was changed with each wash. After the final wash, the pellet was resuspended in ice-cold 1 X PBS and transferred into a microcentrifuge tube as in the mRNA FISH protocol. The cells were then fixed in 3.2% formaldehyde solution for one hour , centrifuged and washed twice with 1 X PBS. The cells were then resuspended in DEPC-treated water and incubated for seven days in 70% ethanol to allow permeabilisation. In low light conditions, the cells were centrifuged and then resuspended in 40% formamide wash solution and incubated for 5 minutes to allow acclimatisation. The cells were centrifuged and then treated with 40% formamide hybridisation solution to give a final concentration of 1 μM FISH probes. This mix was incubated at 30°C in the dark overnight. 1 ml 40% fomamide wash solution was added and the mix was centrifuged before resuspension and incubation in 1ml of the 40% formamide wash solution for one hour. This was repeated three times in total. Finally the cells were resuspended as needed in 1 X PBS and mounted onto agarose pads for imaging. Fig. 2.6 shows a schematic of the protocol. All steps involving formaldehyde and formamide were conducted in the fume hood and waste was disposed of following institute guidelines.

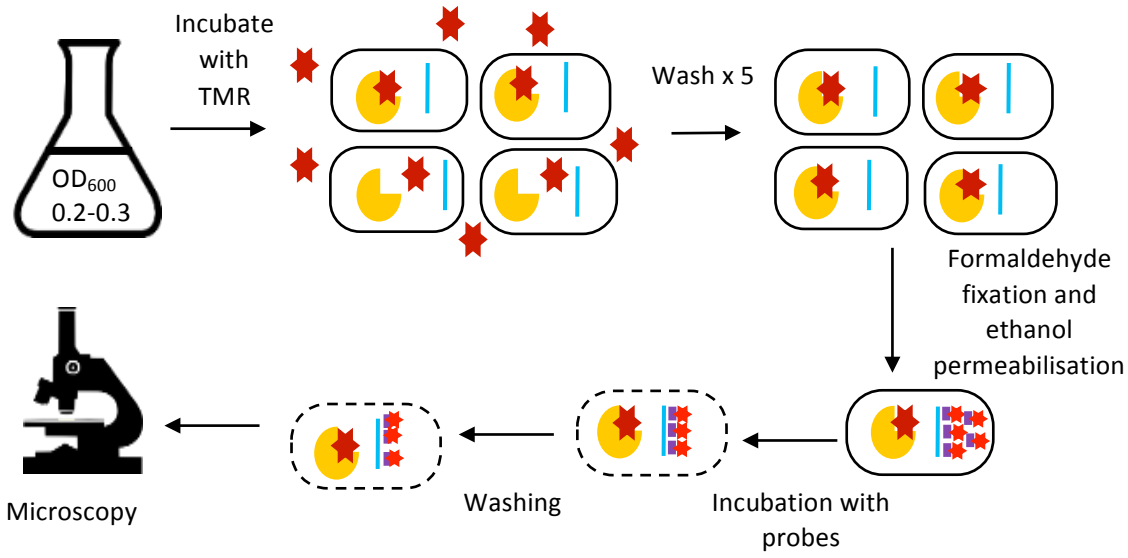


Figure 2.6 Labelling of protein and mRNA with HaloFISH. Cells were grown to OD_{600} 0.2-0.3, labelled with TMR, washed extensively and fixed. The cells were then permeabilised in ethanol and incubated with the FISH probes. The cells were washed and imaged using an epifluorescence microscope.

2.2.4.5 Overexpression of mRNA or protein with arabinose

High copy number mRNA/ protein was produced by deliberate overexpression under the control of the arabinose inducible *ara*BAD promoter. These experiments were conducted using the bacterial strain BW27783. The strains were transformed with either the pBADHalo plasmid for overexpression of the HaloTag or the pBAD*recD* plasmid for overexpression of *recD* mRNA for mRNA FISH (see Table 2.8 for details of plasmids). BW27783 was used because in MG1655 genes placed under the *ara*BAD promoter are expressed in an all-or-none manner, meaning that it is the percentage of induced cells in the population that will increase with increasing

induction, rather than the degree of induction in each individual cell. The BW27783 strain (Khlebnikov, 2001) was designed to allow arabinose induction to vary the amount of expression in individual cells. This is achieved by placing the *ara*E gene (which encodes a low-affinity high-capacity transporter) under the control of a constitutive promoter. In wild type cells, this gene is under the control of an arabinose dependent promoter, meaning that only cells that happen to take in arabinose will then produce more transporters, producing the bistable population described above. However, when the promoter is constitutive all cells in the population will take up arabinose and respond to the concentration of arabinose in the culture medium. The HaloTag labelling protocol was conducted as described in 2.2.4.1 and mRNA FISH as described in 2.2.4.2. Concentrations between 10^{-5}% and 1% arabinose were used for induction, and were added to the overday growth media. 20% glycerol was used in place of 20% glucose in the relevant growth media, as glucose inhibits *ara*BAD expression (Guzman *et al*. 1995). In addition to this, 20 µL of 500 µM HaloTag TMR ligand was added when the HaloTag was overexpressed to ensure an excess of TMR ligand was available for binding to the HaloTag.

2.2.5 Microscopy

2.2.5.1 Preparation of 2% agarose pad for imaging

All imaging performed using 2% agarose pads as they were found to increase image quality. The pads are placed on microscope slides using two Gene Frame (Thermo Scientific, #AB-0577) adhesive mounts. 350 µl of melted 2% agarose is pipetted into

the frames and a second microscope slide was used to compress the melted agar and ensure even setting. The pads were left for 10 minutes to set and then allowed a further 10 minutes to dry after the upper slide has been removed and before the sample was mounted.

2.2.5.2 Imaging conditions

All cells imaged in this work were fixed as described above and mounted on No. 1.5 coverslips with agarose pads. The cells were resuspended for image acquisition in 1X PBS. All images were collected on a Nikon Ti inverted microscope equipped with a 100X Plan Apo NA 1.45 objective lens and the Perfect Focus system was used for continuous maintenance of focus. The filters used were purchased from Chroma. Z-stacks were acquired in fluorescence channels unless otherwise stated (6 images over 1μm). Metamorph was used for image acquisition.

Single molecule HaloTag imaging

To image single TMR molecules, the excitation filter #ET545/30x was used with the emission filter #ET620/60m. The dichroic mirror used was #T570LP. All single molecule HaloTag-TMR imaging was done with an exposure time of two seconds and with the EM gain set to 300. For photobleaching analysis, Z-stacks were not acquired.

High copy number HaloTag imaging

To image high numbers of TMR molecules, the excitation filter #ET545/30x was used with the emission filter #ET620/60m. The dichroic mirror used was #T570LP. All high copy number TMR imaging was done with an exposure time of two seconds and with the EM gain set to 4.

mRNA FISH imaging

To image TAMRA probes for mRNA FISH, at high or low mRNA expression, the excitation filter #ET545/30x was used with the emission filter #ET620/60m. The dichroic mirror used was #T570LP. All mRNA FISH imaging was done with an exposure time of two seconds and with the EM gain set to 4.

HaloFISH imaging

For HaloFISH imaging, two channels were used. The first channel imaged protein bound to TMR, with the excitation filter #ET545/30x and the emission filter #ET620/60m and the dichroic mirror #T570LP. For HaloFISH, the exposure time used in this channel was one second and the EM gain was set to 100. The second channel imaged mRNA to which probes carrying the dye fluorescein were bound. To image these probes, the excitation filter #ET480/40x, the emission filter #ET535/50m and the dichroic mirror #T510LPXR were used. For HaloFISH the exposure time used in this channel was one second and the EM gain was set to 100.

2.2.5.3 Image analysis

Single molecule HaloTag analysis

The spot finding analysis used in this work for single molecule HaloTag was adapted for use in the lab by Alessia Lepore, and based on previous work (see below). The spot detection method can be summarised as a two-step procedure: finding the cells (segmentation) and counting the spots in each cell (counting). To find the cells in the image, we used the natural autoflourescence observed in the first image of the Z-stacks acquired (a maximum projection image- an image summing all of the Z-stacks- was done to amplify the signal). We then detect the edges of each cell using a segmentation script developed in MATLAB. For each detected cell, the characteristics were recorded (area, perimeter, cell length and width). To count the number of diffraction-limited spots, we compute the maximum projection image from the acquired z-stacks images (without the first z-stack frame that contained the auto-fluoresce signal). This allows a better detection of the single spots over the background. We proceed to find the spots in each cell by cropping an area of 30x35 pixel around the centroid of the cell and filtering by a band pass filter to remove high-frequency noise and low-frequency features. Each local maximum within a size of 6x6 pixel with intensity above a local threshold was counted as a spot. The software analysis uses band pass filter and peak finder from previously developed and published software (available at: http://physics.georgetown.edu/matlab/, as used in Okumus *et al*. 2016).

Arabinose induction HaloTag Analysis

Arabinose induction images were analysed using a script produced by Sebastian Jaramillo-Riveri. Fluorescence images were segmented by a custom-made algorithm implemented in MATLAB. The algorithm was designed to detect pixel outliers in the

distribution of intensity values. Parameters that gave good segmentation were arbitrarily chosen and used to segment the whole dataset. The result was manually curated to remove false positives. The natural logarithm of the image was used to reduce cell-to-cell variability and the image was background subtracted. A High Pass filter, then a Gaussian filter, were used to sharpen and reduce the noise. From here, operations based on the distribution of pixel intensity were used to distinguish "cell" from "non-cell" pixels. The cell density in the images was low, meaning most of the pixels were not in cells. This allowed the determination of a distribution of 'non-cell' intensities. With this "non-cell" distribution, the images were thresholded using an intensity value that tolerated an arbitrary frequency of "non-cell" pixels (typically 0.1%). Finally, segmentation results were manually curated to remove false positives or cells incorrectly segmented.

Single molecule mRNA FISH Analysis

This analysis was done manually in ImageJ. For each image, the Z-stack image with the best focus was selected. Strains of interest were compared with negative controls and a minimum threshold was selected for true signal. Each image was examined individually and the number of cells noted, along with the number of cells containing foci. For each focus, the brightest 3x3 square of pixels was selected and the mean of the pixels in the square was noted. Within each cell that had a pixel, three further 3x3 pixel squares were selected and the mean noted. These numbers would then be averaged and subtracted from the mean of the 9 pixel square surrounding the foci to give a method for background correcting each focus. The background corrected mean pixel intensity for each foci was then binned in 200 au bins and a histogram of

focus intensity produced. From this, an estimate of the fluorescence intensity of a single mRNA could be estimated, allowing the number of mRNA per cell to be calculated.

Arabinose induction FISH analysis

This analysis was done manually in Image J. For each image Z-stacks acquired in the fluorescence channel were maximally projected and overlaid with the corresponding brightfield image. The outline of each cell was then manually drawn on the image in the brightfield channel and the mean intensity per pixel per cell extracted from the fluorescence channel and recorded. For each image, three measurements of approximately the same size as a single cell were taken outwith the cells, and mean background was calculated and subtracted from the mean cellular intensities acquired to give a measurement of background corrected mean cell intensity for each cell.

2.2.6 DNA primers, DNA sequences and gBlocks, bacterial strains and plasmids

2.2.6.1 DNA primers

All primers used in this work are listed in Table 2.5.

Table 2.4. DNA Primers

Name	Sequence (5' to 3')	Purpose
oht1	TAT GTC TAT TGC TGG TCT CGG TAC CCG ACC TGC ACT TTC CGC CCG CAA TAA ACA GG	Generate 576 base pair fragment from

oht2	GTC GCA TCC GGC AAT TAC GTT TTG CAA TTT CAT TAC GCC T	MG1655, used to make pHT1 by Gibson Assembly
oht3	AGG CGT AAT GAA ATT GCA AAA CGT AAT TGC CGG ATG CGA C	Generate 570 base pair fragment from MG1655, used to make pHT1 by Gibson Assembly
oht4	GCG CCG CTA CAG GGC GCG TCC CAT TCG CCA CCG GTC GAC GCC GAG CAG ATT CGT CG	
oht5	GGT GCT ACG CCT GAA TAA GTG ATA A	Generate 935 base pair fragment that ensures the above fragments have been inserted into the pTOF plasmid by Gibson Assembly to make pHT1
oht6	ATT TAG AGC TTG ACG GGG AAA GC	
oht9	CGG CGT AAG CCT GAG TCA	Generate a 2356 base pair fragment if *recD* is present and a 817 base pair fragment if *recD* has been removed
oht10	CGG TAC GCA GAT TGT GAT GG	
oht28	TTG GGC TAG AAA TAA TTT TGT TTA AGA GAT CTG GATCCA TAT GAC CCT GGA GGA GG	Generate a 1924 base pair fragment from MG1655 used to make pHT3 by Gibson Assembly
oht27	AGA ATC AGT GAT GGT GAT GGT GAT GCT CGA GGC GGC CTT ATT CCC GTG AAC TAA ACA	
oht29	GCC GCT TAT GTC TAT TGC TGG TCT CGG TAC CCG ACC TGC AAG GCG GTA ATT ACC CAG ATG	Generate 1434 base pair

oht30	AAT TTC ACT ACC ATC TCC AGG TGC TCC AGA ACC GGA AAT CTC CAG AGT AG	fragment from MG1655 used to make pHT3 by Gibson Assembly
oht33	CTT GAA ATC AGC GGC TCT GGA GCA CCT GGA GCC GCC TTT CCC CGC CCG CTG ACC GTT GAA	Generate a 525 base pair fragment used to make pHT4 by Gibson Assembly
oht34	ATG CGC CGC TAC AGG GCG CGT CCC ATT CGC CAC CGG TCG AGA TTA ATA TCG CGC AGC AAC	
oht35	GTC GTC GGT TCA GGG CAG	Generate a 2890 base pair fragment that ensures the above fragments have been inserted into the pTOF plasmid by Gibson Assembly to make pHT4
oht36	CTG CGC GTA ACC ACC ACA	
oht41	ACG CTC GGC GAA GAA GCA TCG A	Generate fragments for sequencing HT43
oht42	AAC CCA GGT CTG GTT TGT CGG A	
oht43	TCC GAC AAA CCA GAC CTG GGT T	
oht44	GAA CCG GAA ATC TCC AGA GTA G	
oht45	CTA CTC TGG AGA TTT CCG GTT C	
oht46	CTC CCC AGC AAT GGG AAG TTC G	
oht47	CGA ACT TCC CAT TGC TGG GGA G	
oht48	CTA TTT CAC GCG GCA GCG	

Oht 82	GCC GCT TAT GTC TAT TGC TGG TCT CGG TAC CCG ACC TGC AAG GCG GTA ATT ACC CAG ATG	Generate a 1056bp fragment from pBH35 for use in Gibson Assembly of pHT8
Oht83	CTT CCT TGG CAG CCG CCT CTT TCG CTG CGG CTT CAG CCA GAC CGG AAA TCT CCA GAG TAG	
Oht86	CGA GAT CGC GCG CTG GCT GTC TAC TCT GGA GAT TTC CGG TTC TGG AGC ACC TGG AGC C	Generate a 469bp fragment from pBH35 for use in Gibson Assembly of pHT8
Oht87	ATG CGC CGC TAC AGG GCG CGT CCC ATT CGC CAC CGG TCG AGA TTA ATA TCG CGC AGC AAC	
oSF1	TTT CTC CAT ACC CGT TTT TTT GGG CTA GCG AAT TCG AGC TAA AGA GGA GAA AGG ATC CAT	Generate a 1007 bp fragment from pBH35 used to make pSF1 by Gibson Assembly
oSF2	CTT CTC TCA TCC GCC AAA ACA GCC AAG CTT GCA TGC CTT AAC CGG AAA TCT CCA GAG TAG	
oSF3	CCA TAA GAT TAG CGG ATC CTA CC	Generate a 1126 base pair fragment that ensures the above fragment has been inserted into the pBAD33 plasmid by Gibson Assembly to make pHT1
oSF4	CCG CCA GGC AAA TTC TGT TTT A	
oSF5	GGC AAA TTC TGG CCA TTC GTC C	Generate fragments for sequencing of pSF1
oSF6	GGA CGA ATG GCC AGA ATT TGC C	

2.2.6.2. DNA sequences and gBlocks

The DNA sequence for the HaloTag (Los *et al*. 2008) and a codon-optimised version designed for this work are listed in Table 2.5, along with the two linker sequences used between tandem HaloTags. Table 2.6 contains gBlock sequences used in the construction of strains carrying two HaloTag genes. These strains and their uses are discussed in Chapter 3.

Table 2.5. HaloTag, codon optimized HaloTag and large and small linkers

Fragment	Sequence
Original HaloTag	GGA TCC GAA ATC GGT ACT GGC TTT CCA TTC GAC CCC CAT TAT GTG GAA GTC CTG GGC GAG CGC ATG CAC TAC GTC GAT GTT GGT CCG CGC GAT GGC ACC CCT GTG CTG TTC CTG CAC GGT AAC CCG ACC TCC TCC TAC GTG TGG CGC AAC ATC ATC CCG CAT GTT GCA CCG ACC CAT CGC TGC ATT GCT CCA GAC CTG ATC GGT ATG GGC AAA TCC GAC AAA CCA GAC CTG GGT TAT TTC TTC GAC GAC CAC GTC CGC TTC ATG GAT GCC TTC ATC GAA GCC CTG GGT CTG GAA GAG GTC GTC CTG GTC ATT CAC GAC TGG GGC TCC GCT CTG GGT TTC CAC TGG GCC AAG CGC AAT CCA GAG CGC GTC AAA GGT ATT GCA TTT ATG GAG TTC ATC CGC CCT ATC CCG ACC TGG GAC GAA TGG CCA GAA TTT GCC CGC GAG ACC TTC CAG GCC TTC CGC ACC ACC GAC GTC GGC CGC AAG CTG ATC ATC GAT CAG AAC GTT TTT ATC GAG GGT ACG CTG CCG ATG GGT GTC GTC CGC CCG CTG ACT GAA GTC GAG ATG GAC CAT TAC CGC GAG CCG TTC CTG AAT CCT GTT GAC CGC GAG CCA CTG TGG CGC TTC CCA AAC GAG CTG CCA ATC GCC GGT GAG CCA GCG AAC ATC GTC GCG CTG GTC GAA GAA TAC ATG GAC TGG CTG CAC CAG TCC CCT GTC CCG AAG CTG CTG TTC TGG GGC ACC CCA GGC GTT CTG ATC CCA CCG GCC GAA GCC GCT CGC CTG GCC AAA AGC CTG CCT AAC TGC AAG GCT GTG GAC ATC GGC CCG GGT CTG AAT CTG CTG CAA GAA GAC AAC CCG GAC CTG ATC GGC AGC GAG ATC GCG CGC TGG CTG TCT ACT CTG GAG ATT TCC GGT
Codon optimized HaloTag	GGT AGT GAA ATT GGG ACA GGC TTT CCT TTT GAT CCT CAT TAC GTC GAA GTG CTG GGC GAG CGC ATG CAT TAC GTC GAC GTA GGA CCA CGC GAT GGG ACC CCA GTG CTG TTC CTG CAC GGC AAT CCG ACG AGT TCA TAT GTT TGG CGG AAC ATT ATC CCT CAT GTG GCA CCT ACG CAT CGT TGC ATT GCA CCC GAC CTG ATC GGA ATG GGT AAA AGC GAT AAA CCG GAT TTG GGT TATT TTT TCG ATG ATC ACG TGC GGT TCA TGG ACG CGT TTA TCG AGG CGC TGG GGC TGG AAG AAG TAG TGC TGG TTA TTC ACG ACT GGG GTA GCG CGC TGG GTT TTC ACT GGG CCA AAC GTA ACC

	CGG AAC GCG TTA AAG GCA TTG CAT TTA TGG AAT TTA TCC GTC CGA TCC CGA CGT GGG ATG AGT GGC CGG AGT TTG CTC GGG AAA CCT TTC AGG CGT TCC GTA CCA CTG ATG TCG GTC GCA AGC TGA TCA TTG ATC AGA ATG TAT TCA TTG AGG GCA CCC TGC CAA TGG GGG TTG TGC GGC CGC TGA CCG AAG TGG AAA TGG ACC ATT ACC GTG AGC CAT TTC TGA ATC CAG TGG ATC GCG AAC CTC TGT GGC GCT TTC CGA ACG AAC TTC CCA TTG CTG GGG AGC CGG CAA ATA TCG TTG CGC TGG TGG AGG AGT ACA TGG ATT GGC TTC ACC AGA GCC CGG TAC CAA AGC TGC TGT TTT GGG GCA CGC CAG GTG TCC TTA TCC CGC CGG CCG AGG CTG CGC GTC TG GCG AAA AGT CTC CCG AAC TGC AAA GCT GTT GAC ATC GGC CCT GGC CTT AAT CTC CTG CAG GAA GAC AAC CCA GAT CTG ATC GGC AGC GAG ATT GCC CGT TGG CTC AGC ACC CTT GAA ATC AGC GGC
Small linker (pHT4)	TCT GGA GCA CCT GGA GAT
Large linker (pHT8)	CTG GCT GAA GCC GCA GCG AAA GAG GCG GCT GCC AAG GAA GCT GCC GCT AAG GAG GCT GCT GCA AAA GAA GCT GCG GCG AAA GCT GCC GCA

Table 2.6. gBlocks

Name	**Sequence (5' to 3')**	**Purpose**
coHalo1	TCT GGA GCA CCT GGA GAT GGT AGT GAA ATT GGG ACA GGC TTT CCT TTT GAT CCT CAT TAC GTC GAA GTG CTG GGC GAG CGC ATG CAT TAC GTC GAC GTA GGA CCA CGC GAT GGG ACC CCA GTG CTG TTC CTG CAC GGC AAT CCG ACG AGT TCA TAT GTT TGG CGG AAC ATT ATC CCT CAT GTG GCA CCT ACG CAT CGT TGC ATT GCA CCC GAC CTG ATC GGA ATG GGT AAA AGC GAT AAA CCG GAT TTG GGT TAT TTT TTC GAT GAT CAC GTG CGG TTC ATG GAC GCG TTT ATC GAG GCG CTG GGG CTG GAA GAA GTA GTG CTG GTT ATT CAC GAC TGG GGT AGC GCG CTG GGT TTT CAC TGG GCC AAA CGT AAC CCG GAA CGC GTT AAA GGC ATT GCA TTT ATG GAA TTT ATC CGT CCG ATC CCG ACG TGG GAT GAG TGG CCG GAG TTT GCT CGG GAA ACC TTT CAG GCG TTC CGT ACC ACT GAT GTC GGT CGC AAG CTG ATC ATT GAT CAG AAT GTA TTC ATT GAG GGC ACC CTG CCA ATG GGG GTT GTG CGG CCG CTG ACC GAA GTG GAA ATG GAC CAT TAC CGT GAG CCA TTT CTG AAT CCA GTG	891 bp HaloTag gene codon optimised for expression in *E. coli* plus 18 bp linker on each end, used as Gibson Assembly fragment in construction of pHT4

	GAT CGC GAA CCT CTG TGG CGC TTT CCG AAC GAA CTT CCC ATT GCT GGG GAG CCG GCA AAT ATC GTT GCG CTG GTG GAG GAG TAC ATG GAT TGG CTT CAC CAG AGC CCG GTA CCA AAG CTG CTG TTT TGG GGC ACG CCA GGT GTC CTT ATC CCG CCG GCC GAG GCT GCG CGT CTG GCG AAA AGT CTC CCG AAC TGC AAA GCT GTT GAC ATC GGC CCT GGC CTT AAT CTC CTG CAG GAA GAC AAC CCA GAT CTG ATC GGC AGC GAG ATT GCC CGT TGG CTC AGC ACC CTT GAA ATC AGC GGC TCT GGA GCA CCT GGA	
coHalo2	CGA GAT CGC GCG CTG GCT GTC TAC TCT GGA GAT TTC CGG TCT GGC TGA AGC CGC AGC GAA AGA GGC GGC TGC CAA GGA AGC TGC CGC TAA GGA GGC TGC TGC AAA AGA AGC TGC GGC GAA AGC TGC CGC AGA TGG ATC CGA AAT CGG TAC TGG CTT TCC ATT CGA CCC CCA TTA TGT GGA AGT CCT GGG CGA GCG CAT GCA CTA CGT CGA TGT TGG TCC GCG CGA TGG CAC CCC TGT GCT GTT CCT GCA CGG TAA CCC GAC CTC CTC CTA CGT GTG GCG CAA CAT CAT CCC GCA TGT TGC ACC GAC CCA TCG CTG CAT TGC TCC AGA CCT GAT CGG TAT GGG CAA ATC CGA CAA ACC AGA CCT GGG TTA TTT CTT CGA CGA CCA CGT CCG CTT CAT GGA TGC CTT CAT CGA AGC CCT GGG TCT GGA AGA GGT CGT CCT GGT CAT TCA CGA CTG GGG CTC CGC TCT GGG TTT CCA CTG GGC CAA GCG CAA TCC AGA GCG CGT CAA AGG TAT TGC ATT TAT GGA GTT CAT CCG CCC TAT CCC GAC CTG GGA CGA	552 bp sequence containing 40 bp homology to the original HaloTag sequence, the 90 bp linker detailed above and one half of the codon optimised HaloTag gene, used as Gibson Assembly fragment in construction of pHT8
coHalo3	CAT TTA TGG AGT TCA TCC GCC CTA TCC CGA CCT GGG ACG AAT GGC CAG AAT TTG CCC GCG AGA CCT TCC AGG CCT TCC GCA CCA CCG ACG TCG GCC GCA AGC TGA TCA TCG ATC AGA ACG TTT TTA TCG AGG GTA CGC TGC CGA TGG GTG TCG TCC GCC CGC TGA CTG AAG TCG AGA TGG ACC ATT ACC GCG AGC CGT TCC TGA ATC CTG TTG ACC GCG AGC CAC TGT GGC GCT TCC CAA ACG AGC TGC CAA TCG CCG GTG AGC CAG CGA ACA TCG TCG CGC TGG TCG AAG AAT ACA TGG ACT GGC TGC ACC AGT CCC CTG TCC CGA AGC TGC TGT TCT GGG GCA CCC CAG GCG TTC TGA TCC CAC CGG CCG AAG CCG CTC GCC TGG CCA AAA GCC TGC CTA ACT GCA AGG CTG TGG ACA TCG GCC CGG GTC TGA ATC TGC TGC AAG AAG ACA ACC CGG ACC TGA TCG GCA GCG AGA TCG CGC GCT GGC TGT CTA CTC TGG AGA TTT CCG GTT CTG	552 bp sequence containing the second half of the codon optimised HaloTag gene, along with 40 bp of *recB* homology, used as Gibson Assembly fragment in construction of pHT8

	GAG CAC CTG GAG CCG CCT TTC CCC GCC CGC TGA CCG	

2.2.6.3 *E. coli* strains

All *E. coli* strains used in this work are listed in Table 2.7.

Table 2.7. *E. coli* strains

Strain ID	Strain Name	Genotype of Interest	Background	Source
HT35	BW	*Δ(araFGH) Φ(ΔaraEp PCP8−araE)*	BW27783	(Khlebinkov, 2001)
HT28	MG	*F- lambda- rph-1*	MG1655	Meriem El Karoui
HT29	RecBHalo	*recB:halo*	MG1655	Meriem El Karoui
HT43	RecBHcoH1	*recB:halo:cohalo*	MG1655	This work
HT50	RecBHcoH2	*recB:halo:cohalo*	MG1655	This Work
HT51	RecCHalo	*recC:halo*	MG1655	Meriem El Karoui
HT52	RecDHalo	*recD:halo*	MG1655	Meriem El Karoui
HT10	Δ*rec*DMG	*ΔrecD*	MG1655	This work
DL5757	Δ*rec*B	*ΔrecB*	MG1655	From David Leach
DL4576	Δ*rec*C	*ΔrecC*	BW27784	From David Leach

2.2.6.4 Plasmids

All plasmids used in this work are listed in Table 2.8

Table 2.8. Plasmids

Plasmid ID	Plasmid name	Purpose	Source
	pTOF24	pTOF24 plasmid which allows PMGR	(Merlin *et al*. 2002)
	pBAD33	Plasmid containing arabinose operon	(Guzman *et al*. 1995)
	pUCBB-pBAD-eGFP	Plasmid containing arabinose operon and eGFP	(Vick *et al*. 2011)
pBH35	pTOFRecBHalo	pTOF24 plasmid containing *recB* homology ends and original HaloTag sequence for integration by PMGR into MG1655	Meriem El Karoui
pHT4	pTOFRecBHalo1	pTOF24 plasmid containing *recB* homology ends, original HaloTag, small linker and codon optimised HaloTag sequence for integration by PMGR into MG1655 to give RecBHcoH1	This work
pHT8	pTOFRecBHalo2	pTOF plasmid containing *recB* homology ends, original HaloTag, large linker and codon optimised HaloTag sequence for integration by PMGR into MG1655 to give RecBHcoH2	This work
pHT1	pTOF*recD*	pTOF24 plasmid that allows removal of *recD* gene from MG1655 through PMGR	This work
pSF1	pBAD*halo*	pBAD33 with HaloTag	Suraya Fawcett for this work
pHT3	pBAD*recD*	pUCBB-pBAD-recD	This work

Chapter 3

Developing Quantitative single cell HaloTag-TMR labelling in *E. coli* and application of the method to each subunit of the bacterial DNA repair protein RecBCD

3.1 Introduction

The HaloTag enzyme published by Los *et al.* (2008) is a modified haloalkane dehalogenase, which was desinged to covalently bind synthetic ligands comprised of a chloroalkane linker attached to one of a range of useful molecules- affinity handles, solid surfaces or organic fluorescent dyes. This chapter focuses on the use of the HaloTag enzyme as one half of a amino acid detection system in conjuntion with the flourescent HaloTag specific TMR ligand. An ever expanding array of fluorescent ligands are available for use with the HaloTag (Grimm *et al*. 2015). TMR was selected for use in this study for severl resons. Firstly, this fluorophore emits in red, a part of the colour spectra that is not compromised by cellular autoflourescence. Secondly, the TMR ligand was known to be permeable to the bacterial membrane (Ke *et al*. 2016) and finally, TMR is not toxic to the E. coli cells, which grow normally in its presence.

In this Chapter, I will describe the development of the HaloTag labelling protocol to allow visualisation and quantification of protein in *E. coli* using epifluorescence microscopy (protocol described in Chapter 2 Section 2.2.4.2).

I will first outline the characterization of the protocol, progressing from labelling highly expressed HaloTag to labelling individual RecBHalo fusion proteins the RecBCD subunits were selected because, as described in Chapter 1, they are known to be expressed at low copy number. I will show that the method is capable of single molecule labelling, and that the labelling efficiency is equal to that of fluorescent proteins such as GFP. I will then show the quantitative data gathered using the HaloTag-TMR labelling protocol for the protein subunits RecB, RecC and RecD and compare the results of this quantification with others existing in the literature.

3.2 Characterization of the method

3.2.1 Specific detection of HaloTag-TMR

To ensure the specificity of the TMR ligand for the HaloTag, a plasmid was constructed with the HaloTag gene expressed under the control of the arabinose promoter. The resulting plasmid, pBAD*halo*, was transformed into *E. coli* BW27783 (this strain allows for homogenous induction of the arabinose promoter as explained in Chapter 2 Section 2.2.4.5). TMR labelling was compared in a strain carrying the induced plasmid and in a wild type strain that did not carry the plasmid. On induction with arabinose (10^{-4}%) followed by HaloTag labelling and imaging with fluorescence microscopy, specific detection of the HaloTag is evident (Fig. 3.1). Bright, diffuse signal can be seen in the pBAD*halo*+ strain, while the strain without

the plasmid has only a small amount of nonspecific signal. When quantified, 99.6% of 889 pBAD*halo*+ cells were found to show specific signal when TMR and the HaloTag were present, while in pBAD*halo*- cells 100% of 611 cells did not show diffuse signal above background.

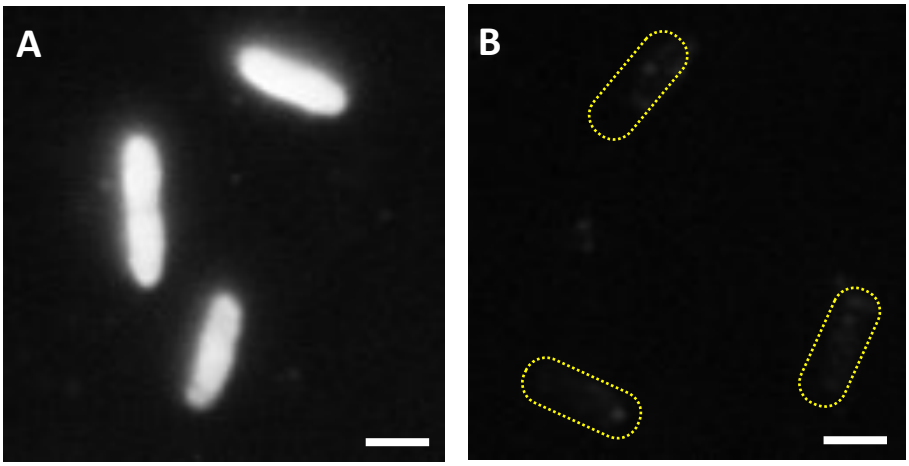


Figure 3.1. Specific detection of TMR bound to induced HaloTag protein. A) BW27783 + pBAD*halo* + TMR + 10^{-4}% arabinose. Bright, diffuse signal in cells resulting from TMR binding to the available HaloTag protein. B) BW27783 + TMR + 10^{-4}% arabinose. No HaloTag is expressed, limited non-specific signal/ cellular autofluorescence is seen. (Scale bar = 1µm.)

3.2.2 Detection of a protein across a wide range of expression levels

To test whether HaloTag-TMR labelling allows relative quantification of protein across a wide range of expression levels, the pBAD*halo* plasmid described above was used with a wide range of arabinose induction levels (10^{-5}% - 1%). The experiment was repeated twice per induction level, using >100 cells per condition.

The results obtained showed that the HaloTag-TMR protocol labels well across a wide range of induction level. Increased mean fluorescent signal per cell is detected corresponding with increased arabinose induction, with a slight decrease at 1% arabinose induction (see Fig. 3.2).

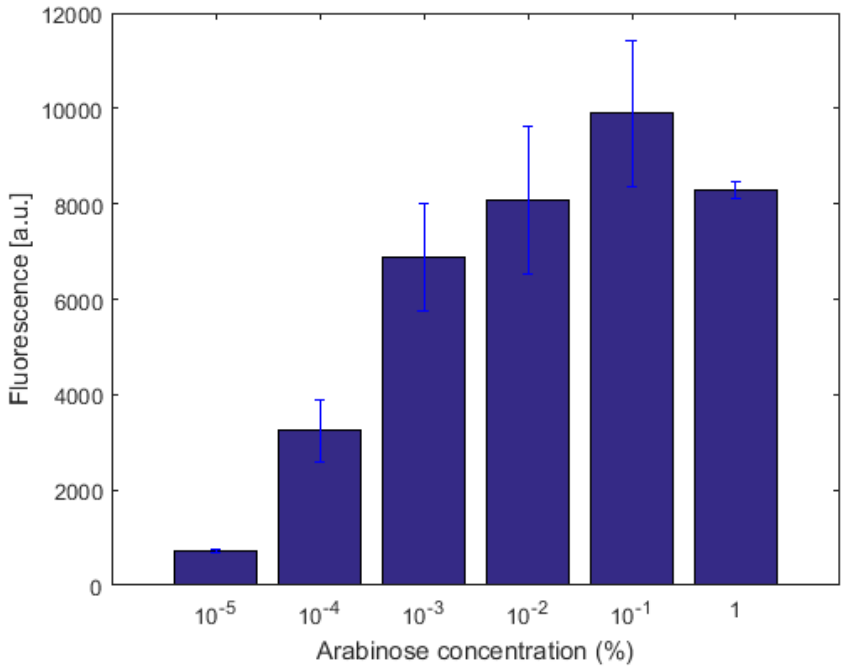


Figure 3.2. Mean fluorescence detection per cell following induction at varying concentrations of arabinose induction. Between 10^{-5}% and 10^{-1}% arabinose induction, mean fluorescence detection per cell increases, followed by a slight decrease at 1% induction. Error bars indicate the SD from the mean of two data sets per condition.

These data compare favourably with those detailed by Khlebnikov *et al*. (2001), who developed the BW27783 strain used in this work. In their publication they produced four strains that allow homogenous induction of expression using arabinose inducible promoters (BW27783, BW27784, BW27786, BW27378) and expressed *gfpuv* (a

GFP derivative) on a plasmid under the control of an arabinose inducible promoter (pCSAK50). They quantified the induced fluorescence to give a culture average (fluorescence/OD_{600}) as can be seen in Fig. 3.3.

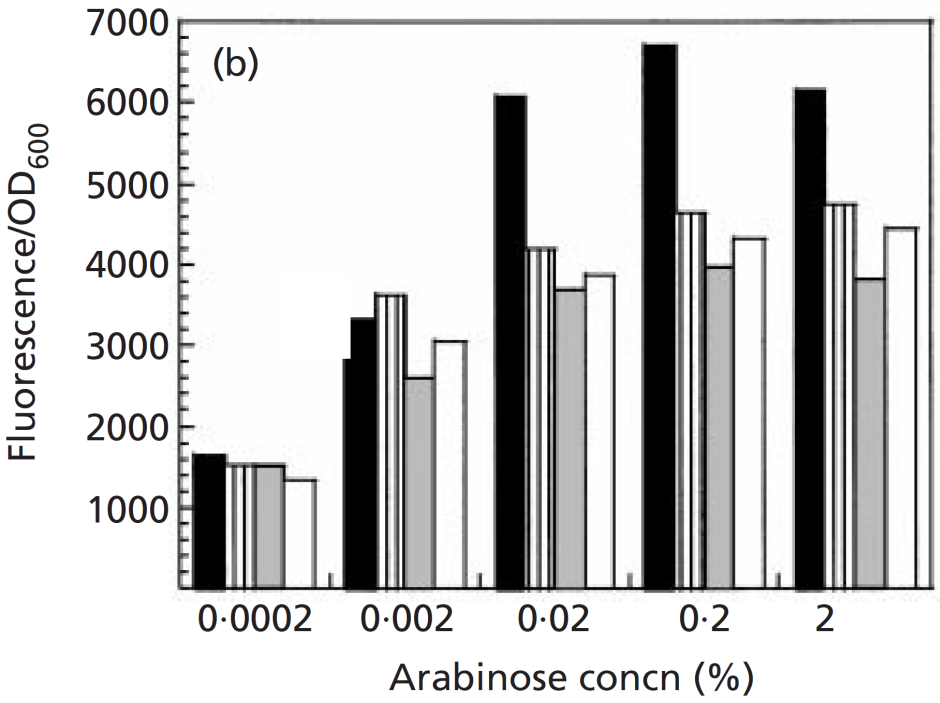


Figure 3.3. Culture-averaged fluorescence (fluorescence/OD_{600}) of *E. coli* cultures containing promoters of different strengths for *ara*E. The measured fluorescence is that of *gfpuv*, expressed in each strain under the control of an arabinose inducible promoter on plasmid pCSAK50. Four different strains are represented - BW27783 (as used in this work, black), BW27784 (vertical line), BW27786 (grey) and BW27378 (white). Arabinose induction is done with concentrations of arabinose between 2^{-4}% and 2%. Between 2^{-4}% and 2^{-1}% arabinose induction, culture average fluorescence increases, followed by a slight decrease at 2% induction. Adapted from Khlebnikov *et al*. (2001).

Both this work, and that done by Khlebnikov *et al.* display an increase in fluorescence detection in the BW27783 strain as arabinose induction is increased, as well as a slight decrease at the highest level of induction used (1% and 2%

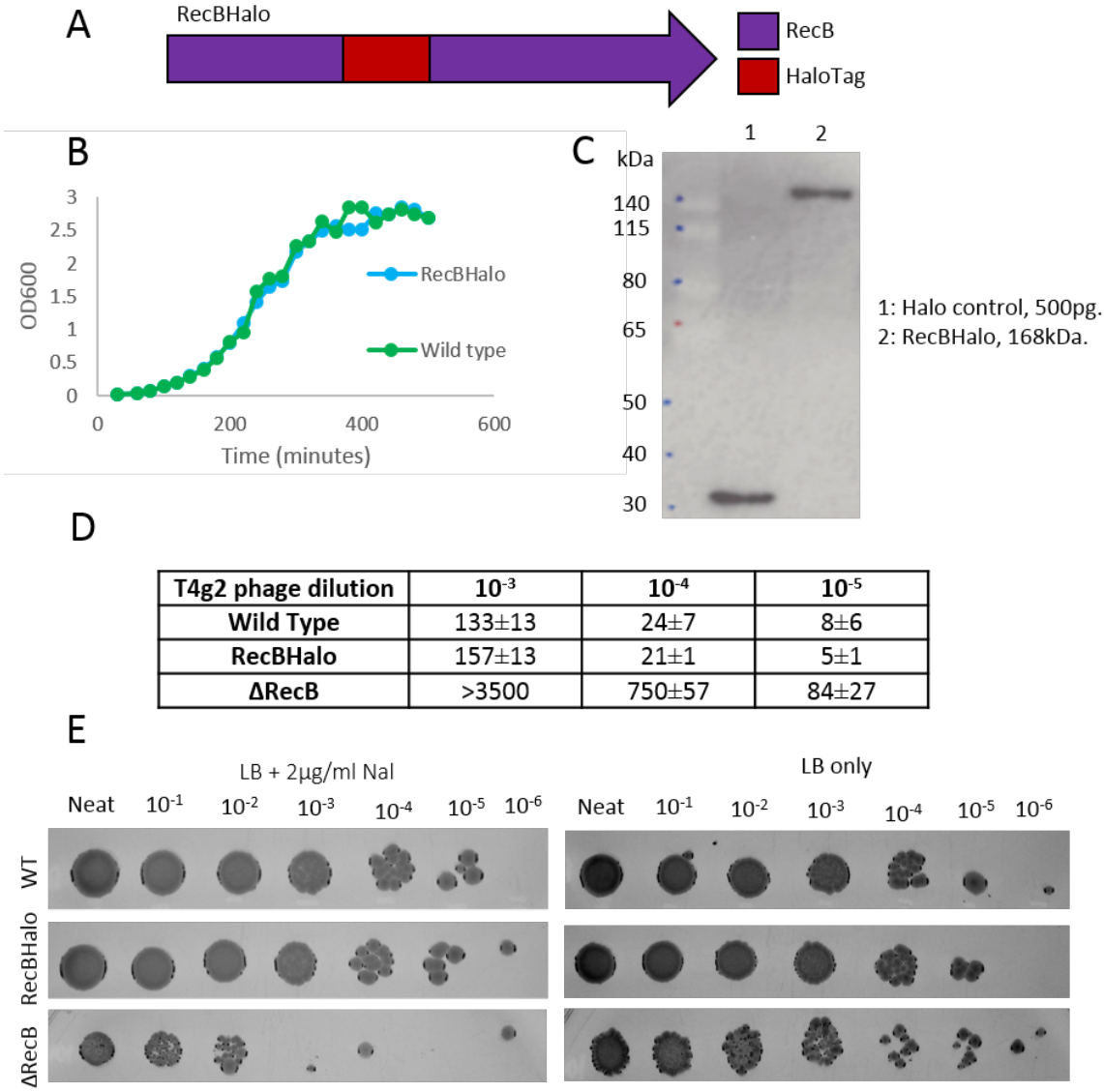
respectively). This decrease is likely to be due to the extremely high level of protein production having a detrimental impact on cellular metabolism, as is well documented (Dong *et al*. 1995; Chou 2007; Scott *et al*. 2010). The comparable data indicates that the HaloTag-TMR protocol allows for semi-quantitative labelling across a wide range of detection in a manner similar to FPs such as *gfpuv*. The next step in the development of this protocol for quantitative labelling was to attempt to label at low levels of expression, specifically to try and detect single molecules using the HaloTag-TMR protocol

3.2.3 Detection of low copy number protein at single molecule level

Having established specific detection of the HaloTag with HaloTag-TMR and confirmed that detection over a wide range of protein expression levels was possible, confirmation of detection at low copy number, specifically single molecule detection, was the next step.

For this, it was decided to use the HaloTag in conjunction with a known low copy number protein, RecB (see Chapter 1). A strain containing a fusion of RecB and the HaloTag (RecBHalo, see Chapter 2 Table 2.7 for strain details and Fig. 3.4a. for schematic) had previously been constructed in other work by Meriem El Karoui. In this strain the 891 bp HaloTag sequence was inserted into the chromosomal sequence of MG1655 at Ser47 of the native RecB gene. The HaloTag was inserted into the centre of the RecB sequence rather than being fused to either the C- or N-terminal as the mature structure of RecB and its incorporation into the RecBCD heterotrimer buries the terminal ends within the structure and would prevent TMR

binding and interferes with RecBCD complex formation. The central fusion however allows RecBCD



T4g2 phage dilution	10^{-3}	10^{-4}	10^{-5}
Wild Type	133±13	24±7	8±6
RecBHalo	157±13	21±1	5±1
ΔRecB	>3500	750±57	84±27

Figure 3.4. RecBHalo strain function and viability. A) Schematic image of the HaloTag sequence insertion into the RecB gene sequence. B) Representative growth curve for RecBHalo and wild type strains, showing no difference in growth between the two. C) Western blot showing purified HaloTag protein in lane 1 and the 168 kDa RecBHalo protein in lane 2 (HaloTag= 34 kDa, RecB= 134 kDa)(Western blots performed by Lorna McLaren). D) Plaque formation following T4g2 phenotypic test for RecB function. WT and RecBHalo strains show formation of very few plaques in comparison to the ΔRecB strain at each phage dilution. E) Serial dilution of WT, RecBHalo and ΔRecB strains on LB+ 2µg/ml nalidixic acid and LB only plates. WT and RecBHalo cells show comparable viability on nalidixic acid while ΔRecB shows

assembly and ensures that the HaloTag is available for TMR binding (see Fig 3.4a). To ensure that the construct was functional, producing RecBHalo and not impacting the viability of the cells multiple tests were carried out. Growth curve data shows that the presence of the RecBHalo construct has no significant impact on growth rate compared to wild type *E. coli* (see Fig. 3.4b.) with doubling times of 44 and 43 minutes respectively in 10% LB imaging media. Fig. 3.4c displays a western blot that confirms expression of the HaloTag in the RecBHalo strain (Western blot was carried out by Lorna McLaren).

A T4Gene2 phenotypic test was performed to assess the exonuclease function of the RecBHalo strain in comparison to that of wild type *E. coli*. As described in Chapter 2, T4Gene2 is a mutant strain of the phage T4 which is degraded on exposure to functional RecBCD and therefore unable to complete its lytic lifestyle and produce plaques. As can be seen in Fig. 3.4b. in two repeats of the test wild type and RecBHalo strains displayed similar, low numbers of plaque formation at each dilution (10^{-3}, 10^{-4}, 10^{-5}) while a ΔRecB strain consistently showed high levels of plaque formation across each dilution. Additionally, a nalidixic acid assay was performed. As described in Chapter Two nalidixic acid induces double strand breaks. Such breaks can be repaired by cells that carry functional RecBCD but not by those that do not. Wild type, RecBHalo and ΔRecB strains were serially diluted to 10^{-6} and plated on both LB and 2 μg/ml nalidixic acid. As can be seen in Fig. 3.4c, when exposed to nalidixic acid colony formation in the ΔRecB strain is much reduced after the 10^{-3} dilution, while in RecBHalo and wild type it continues to the 10^{-6} and 10^{-5}

dilutions. This indicates that the presence of the HaloTag on the RecB gene is not impeding its function in double strand break repair or the viability of the strain. When plated on LB all strains display colony formation in dilutions to 10^{-5}.

As was seen in the high protein expression strain, pBAD*halo*+, labelling of RecBHalo with TMR allows specific detection of fluorescence in the presence of the HaloTag, and no labelling is seen when the HaloTag is absent in the wild type strain (see Fig. 3.5).

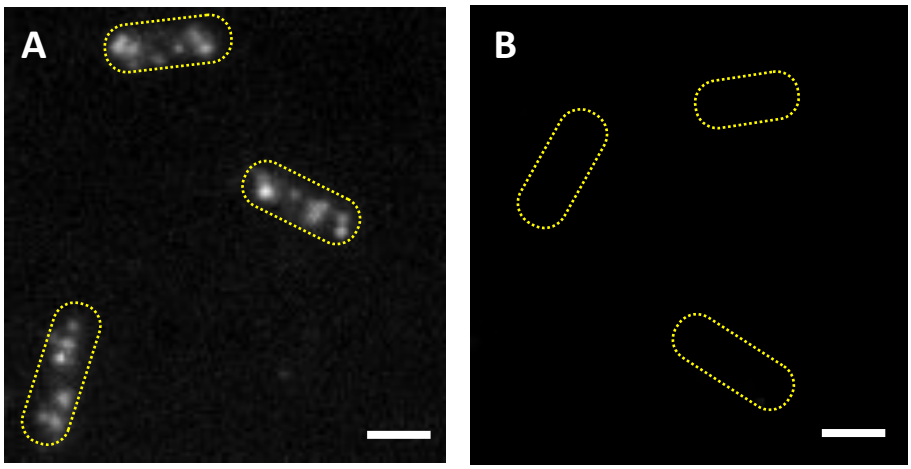


Figure 3.5. Specific detection of TMR bound to low copy number RecBHalo. A) RecBHalo + TMR. Individual diffraction limited foci can be seen following RecBHalo labelling with TMR. B) Wild type + TMR. No HaloTag is expressed and limited non-specific signal/ cellular autofluorescence can be seen. (Scale bar =

However, unlike the high protein expression conditions, the signal observed when imaging a labelled low copy number protein with fluorescence microscopy is not diffuse throughout the cell. When detecting a low number of proteins in cells spatial separation of the proteins being detected allows the visualisation of discrete diffraction limited foci. Such foci are observed because the diffraction limit of light

imposed by the microscope prevents resolution of objects smaller than ~250 nm. This means that direct resolution of proteins within cells is not possible with standard epifluorescence microscopy, as they tend to have absolute sizes that are far below this threshold (GFP for example has a length of 4.7nm (Hink *et al*. 2000)). However, light emitted from florescent proteins or organic dyes can be detected and are visualised in the form of diffraction limited spots. In the case of RecBHalo labelled with TMR, these foci represent a labelled RecBHalo molecule bound to a single TMR. RecBHalo molecules labelled with TMR can be seen in Fig. 3.5, along with a wild type control that does not contain the HaloTag insertion in the RecB gene and therefore does not show specific labelling when exposed to TMR.

It is possible that a single diffraction limited spot represents more than one TMR molecule, as the molecules could be closer together than 250nm within the cell. Confirmation that the diffraction limited foci observed in Fig. 3.5 do indeed represent single molecules of RecBHalo labelled with individual TMR fluorophores can be achieved through the use of bleaching curves. Single fluorophores display a single stepwise intensity drop when photobleached. This means that when the fluorescence intensity of the focus is graphed over time the moment at which a fluorophore bleaches (stops emitting light) can be seen through a corresponding drop in fluorescence intensity. Should two fluorophores be present, two such drops in intensity would be observed, and so on for increasing fluorophore number. Therefore it is possible to quantify the number of fluorophores present in a focus by quantifying the number of intensity drops observed over time. A representative example of this can be seen in Fig. 3.6. This figure shows a distinct bleaching pattern for each of

three individual foci within a single cell. Individual fluorophores can have variable lifetimes, but each can be seen to bleach in a stepwise manner during the 400 second exposure.

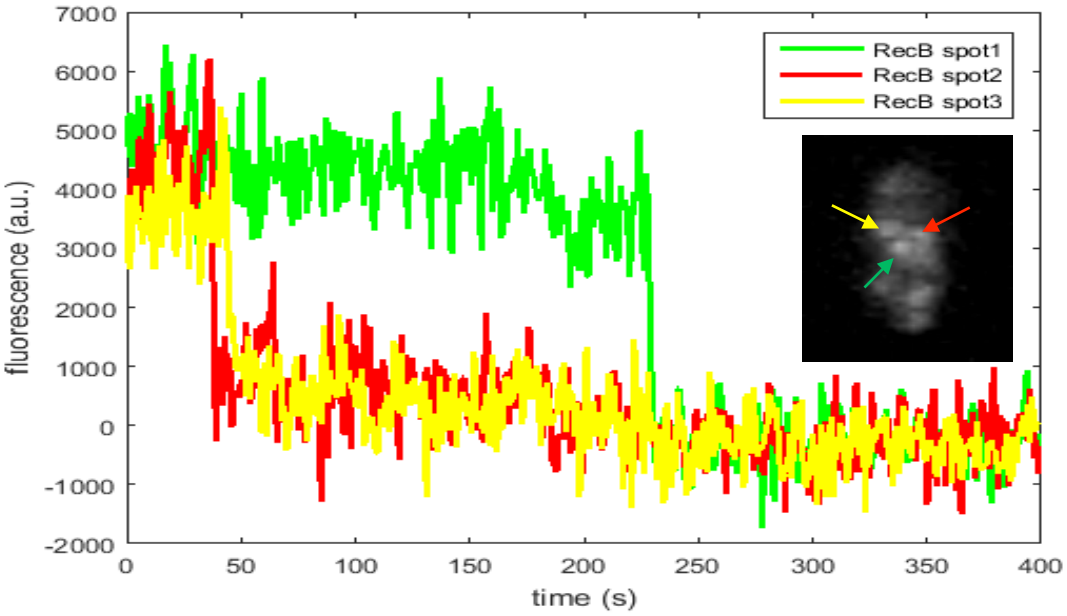


Figure 3.6. The HaloTag-TMR protocol allows visualization of single TMR molecules which can then be quantified. Each trace corresponds to a single TMR focus. Each foci emits light for between 30-250 seconds before bleaching causing the fluorescence signal to drop to background level.

3.2.4 Assessment of HaloTag-TMR labelling efficiency

After determining that the HaloTag-TMR protocol was able to label low copy number RecBHalo at single molecule level the labelling efficiency of the HaloTag-TMR protocol was assessed. Such assessment is challenging as it requires determination of how many RecBHalo molecules are not being labelled inside cells.

To assess the labelling efficiency it was decided to add an additional HaloTag to the original RecBHalo strain. This would produce RecB proteins that could be

labelled by two TMR molecules. Theoretically, quantification of the foci present in the single vs double HaloTag strain would give an indication of whether the labelling was efficient or inefficient. The diffraction limit of light and the close proximity of the TMR molecules once bound to the same protein complex mean that only one focus would be distinguishable per RecBHalo/HaloHalo whether it was bound to one or two TMR. Therefore, if the labelling conditions were highly efficient then on average we would see very slightly more foci in the double HaloTag strain than in the single HaloTag strain (if there were 10 RecB molecules available and labelling was occurring at 80% we would see 8 molecules in the single HaloTag strain and 8-10 in the double HaloTag strain), but if the labelling was inefficient we would anticipate seeing a much higher average and wider distribution in the double HaloTag strain than in the single HaloTag strain (if there were 10 RecB molecules available and labelling was occurring at 40% we would see 4 molecules in the single HaloTag strain and 4-8 in the double HaloTag strain) (See Fig. 3.7c). These expected results are based on three assumptions, the first of which is that the TMR available is sufficient to bind to all available HaloTag proteins, the second is that the binding of TMR to the HaloTag is independent of the number of HaloTags present, and the third is that the RecB will be expressed at the same level independent of having one or two HaloTags fused to it.

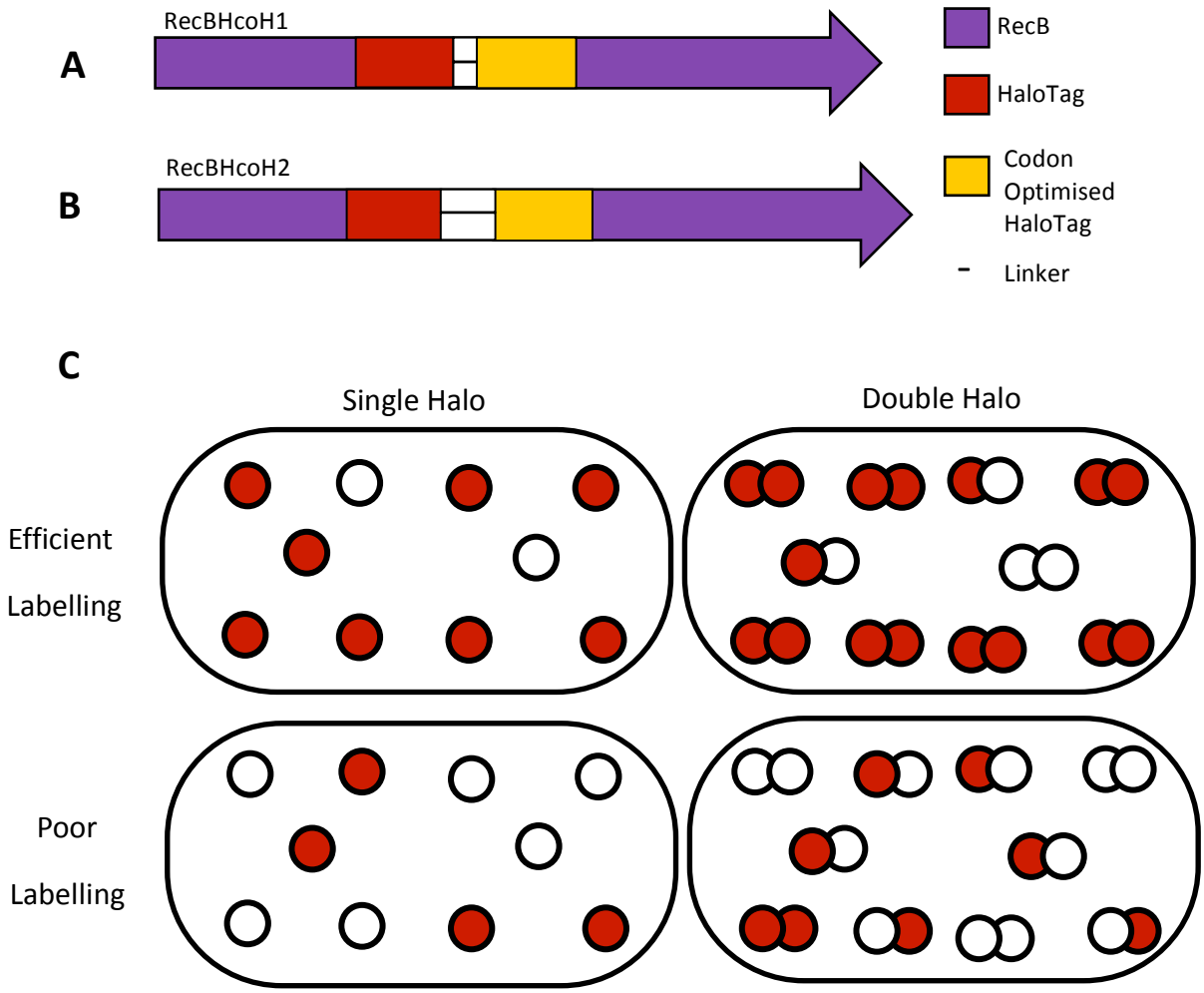


Figure 3.7. RecB double HaloTag constructs and experimental design. A) RecBHcoH1- Initial RecB double HaloTag construct, containing a HaloTag gene, a codon optimised HaloTag gene, and a small (six amino acids) linker between the two. B) RecBHcoH2- Second RecB double HaloTag construct, containing a HaloTag gene, a codon optimised HaloTag gene, and a large (31 amino acids) linker between the two. C) Experimental design - The double HaloTag strain has twice as many HaloTags available to bind TMR. Therefore, if labelling is efficient then we would expect to see fractionally more TMR molecules per cell in the double HaloTag strain, as only one of the HaloTags has to be labelled for the protein to be visible. However if labelling is poor we should be able to see far more TMR molecules per cell in the double HaloTag strain than in the single HaloTag strain, as it is unlikely that the available TMR will bind only to HaloTag pairs on single proteins.

The second HaloTag that was added to the RecBHalo sequence was a version that had been codon-optimised for expression in *E. coli* (see Chapter 2 Table 2.5 for details). The codon-optimisation was done to allow for simpler strain assembly by PMGR (as described in Chapter 2) which relies on regions of sequence homology to insert or delete genes from the chromosome. Initially a RecB double HaloTag strain was built with only a short linker sequence (6 amino acids) between the two HaloTag sequences (RecBHcoH1, see Fig. 3.7A for schematic and Chapter 2 Table 2.5 for details). The results obtained with RecBHalo and RecBHcoH1 were unexpected. We saw neither of the predicted distributions outlined above, but rather observed that RecBHcoH1 was consistently producing very slightly fewer foci than the single HaloTag strain, with RecBHcoH1 giving means of 5.3± 1.5and 4.9± 1.0 foci per cell and RecBHalo giving means of 5.7± 2.4 and 5.0± 1.0 foci per cell (see Fig. 3.8 for distributions).

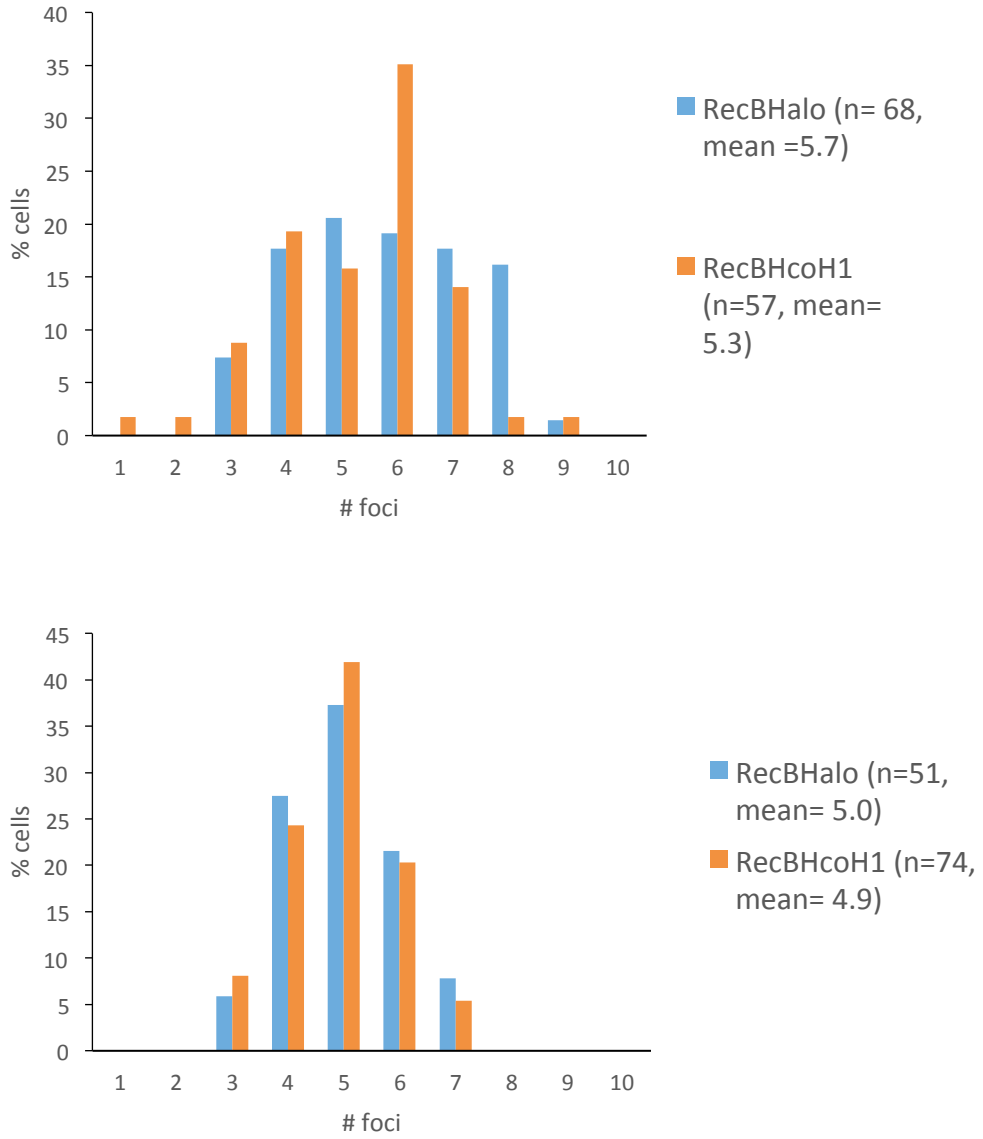


Figure 3.8. Distributions of foci in RecBHalo/RecBHcoH1 (5µM TMR). The distributions of foci in RecBHalo and RecBHcoH1 are similar within repeats, however, in each repeat the double HaloTag strain consistently has slightly fewer foci on average.

These unexpected results indicated that there was an issue with one of the assumptions stated above. It seemed possible that the addition of the second HaloTag was preventing the free binding of TMR molecules, or that both TMR molecules

were binding but that their close proximity was causing fluorophore quenching. Consequently, it was decided to introduce a longer, stiffer linker between the two copies of the HaloTag sequence. This was to try and introduce greater spatial separation between the binding sites of the two HaloTag proteins. A second strain was constructed with a longer 31 amino acid linker between the two HaloTag sequences (RecBHcoH2, see Fig. 3.7B for schematic and Chapter 2 Table 2.5 for sequence details). For both RecBHcoH1 and RecBHcoH2, T4Gene2 and nalidixic acid assays were performed and showed strain functionality and viability to be equal to that of wild type, and growth curves showed no growth defects compared to wild type (see Appendix Fig. A1.).

However, results obtained with the RecBHcoH2 strain were very similar to those obtained with RecBHcoH1. RecBHcoH2 also consistently produced very slightly fewer foci than the single RecBHalo strain. In three repeats the RecBHalo strain gave means of 5.0± 1.26, 4.9± 1.29 and 4.9± 1.17 foci per cell and RecBHcoH2 gave means of 4.5± 1.28, 4.8± 1.21 and 4.5± 1.5 foci per cell (see Fig. 3.9 for distributions). This indicated that the length of the linker and the efficiency of the TMR binding to the HaloTag may not be altered by the presence of the second HaloTag.

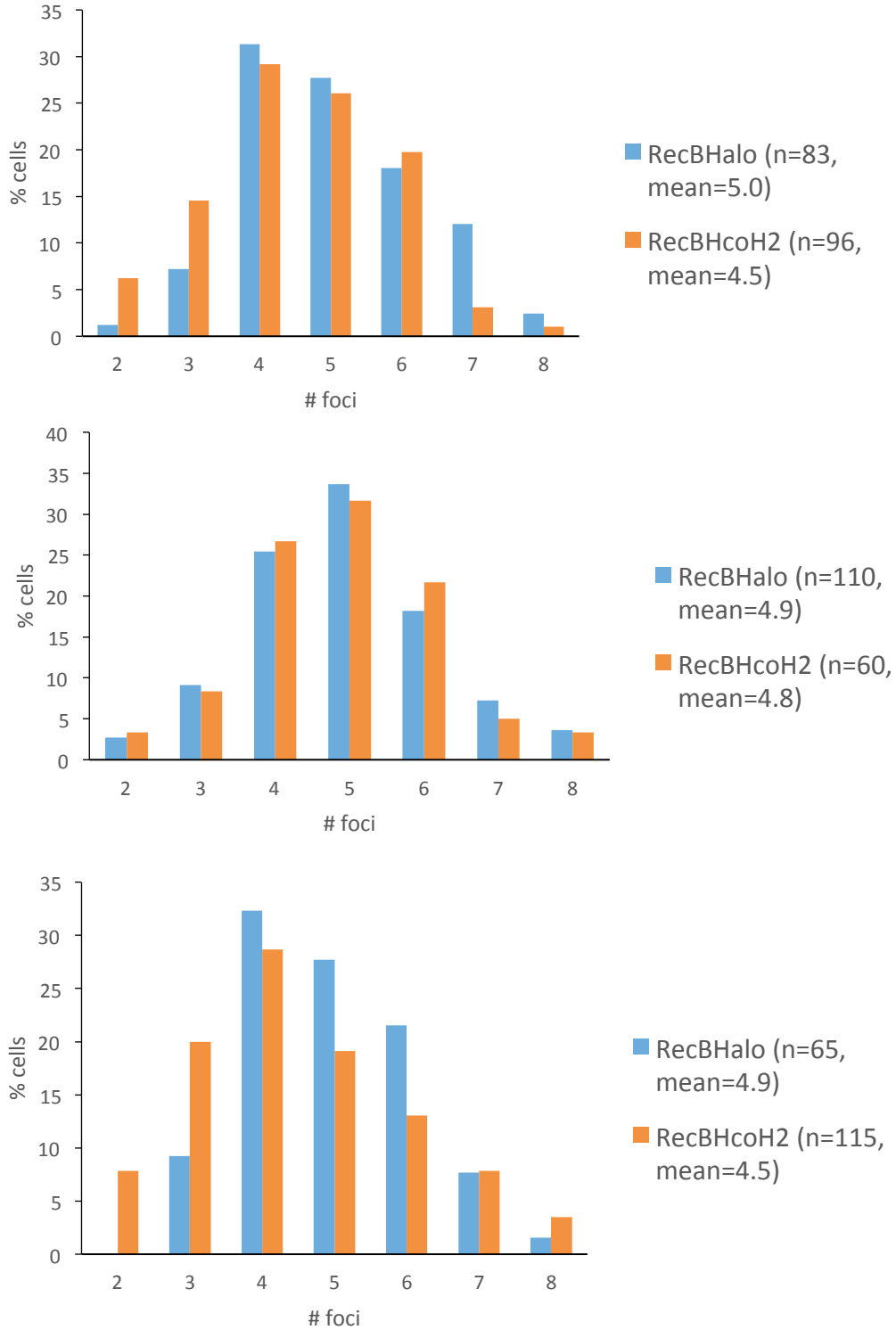


Figure 3.9. Distributions of foci in RecBHalo/RecBHcoH2 (5µM TMR). The distributions of foci in RecBHalo and RecBHcoH2 are similar within repeats, however, once again each repeat the double HaloTag strain consistently has slightly fewer foci on average.

The single HaloTag/ double HaloTag experiments described above indicate that the labelling efficiency is not low, as we do not see much increase in mean number in the double HaloTag strain than in the single HaloTag strain. However, it is clear that one of the underlying assumptions of the experiment is incorrect.

The first assumption, that the TMR available is sufficient to bind to all available HaloTag proteins has been shown to be true through experiments with using variable TMR concentration. Experiments were done with final concentrations of 0.05µM and 0.5µM TMR and the results contrasted with the averaged data from the above described 5µM TMR repeats (5µm TMR is the standard final concentration as described in Chapter 2, experiments conduced with 10µM TMR confirmed this, showing equivalent RecBHalo detection). This was repeated several times in both the single and double HaloTag strains and the average distribution observed in each strain for each concentration can be seen in Fig. 3.10 (>170 cells in each condition).

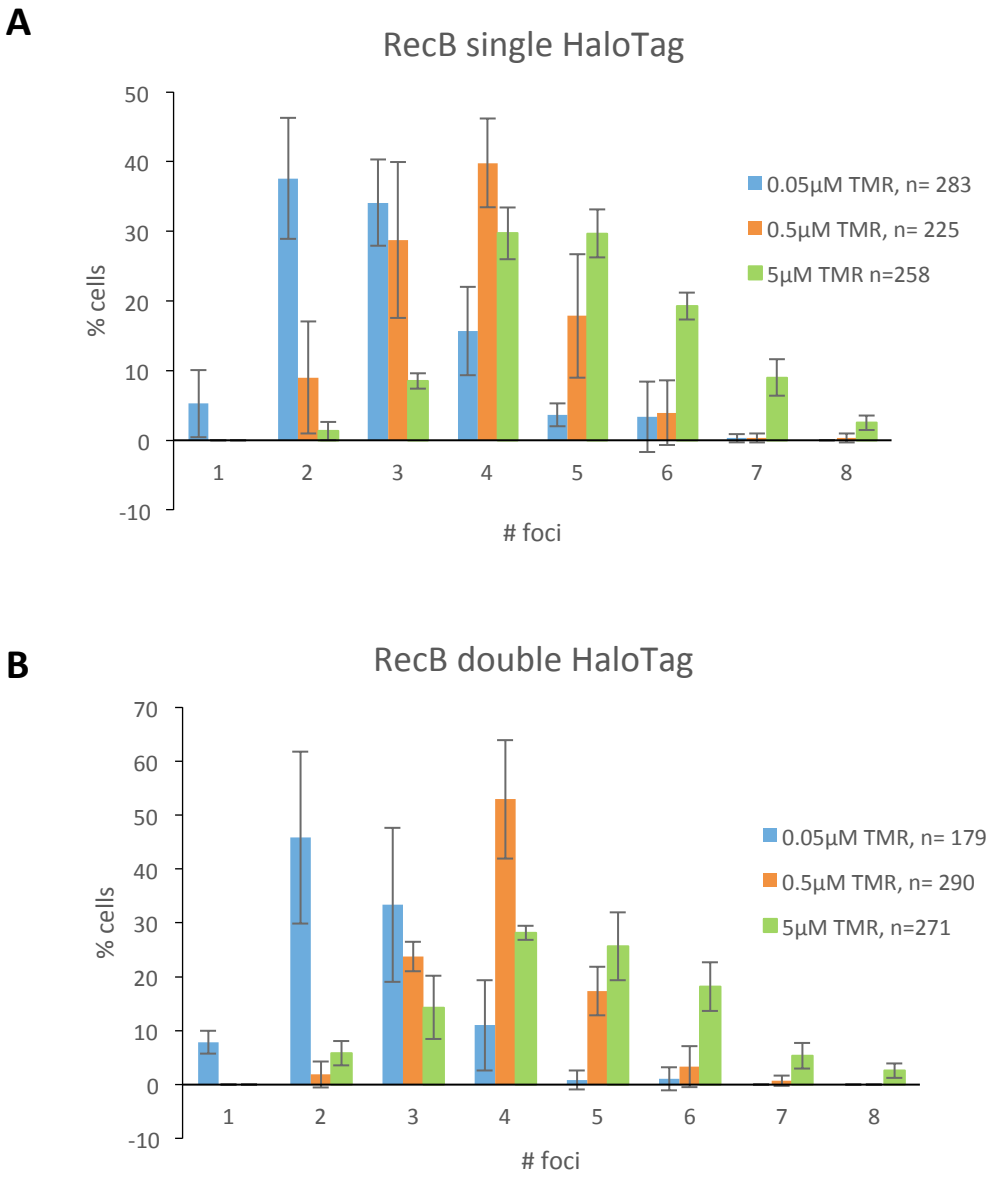


Figure 3.10. Distributions of foci detected with 0.05µM, 0.5µM and 5µM TMR in single and double HaloTag strains. A) Detection of RecBHalo with different TMR concentrations in the single HaloTag strain. 0.05µM TMR gives far fewer foci than 0.5µM and 5µM TMR, which detect similar distributions of RecBHalo. B) Detection of RecBHalo with different TMR concentrations in the double HaloTag strain. 0.05µM TMR gives far fewer foci than 0.5µM and 5µM TMR, which detect similar distributions of RecBHCOH2. Error bars indicate standard deviation of cell number across 3 repeats for each labelling method.

While 0.05μM appears to under label the available HaloTag TMR, giving a mean number of foci of 2.9 ± 1.1, in the single HaloTag strain and 2.6 ± 1.0 in the double HaloTag strain, 0.5μM TMR provides very similar distributions to 5μM TMR in each strain, with means of 3.9 ± 1.0 and 4.0 ± 0.9 for single and double HaloTag respectively. 5μM TMR gives a mean of 4.9 ± 1.3 for single HaloTag and 4.6 ± 1.4 for double HaloTag. The 0.5μM and 5μM TMR means fall within one standard deviation of each other for both strains. This, combined with the capacity to detect overexpressed HaloTag strongly indicates that the TMR is not limiting in this experiment.

This finding indicates that it is likely to be one of the remaining assumptions - that the binding of TMR to the HaloTag is independent of the number of HaloTags present, or that the RecB is expressed at the same level independent of the number of HaloTags bound to it- that is incorrect. It seems likely that the expression of the whole RecBHcoH1/2 complex that is altered by the addition of the second HaloTag, suggesting that there may be marginally less RecBHcoH1/2 available to label than there is RecBHalo.

It is worth noting that other attempts were made to quantify the fluorescence being emitted from the foci in both strains to see if it was possible to assign threshold values to the presence of one or two TMR fluorophores. This would have allowed an understanding of the frequency of labelling with two TMR molecules. However, while the TMR fluorophore is very bright and very photostable there is sufficient variability in fluorescence intensity produced by each individual fluorophore that made this kind of quantification impossible (some fluorophores were observed to

emit twice as much fluorescence as others). Additionally, quantification of expression through western blot was assessed. However, the antibodies used were against the HaloTag, and the presence of two tags on one strain prevented quantitative comparison of RecBHalo and RecBHcoH1/2 protein yield. Taken together, it is clear that while this experiment indicates that the labelling efficiency of the protocol is not poor, it is not sufficient to confidently state that the labelling efficiency is sufficient for quantitative measurement.

An alternative approach to determining labelling efficiency was to compare the data gathered in this study to that found when labelling RecB with the fluorescent protein GFP (work carried out by Meriem El Karoui). As discussed in Chapter 1 fluorescent proteins are widely used in cellular biology, and unlike the HaloTag protein GFP is intrinsically fluorescent and does not depend on the binding of a secondary ligand to allow protein detection as HaloTag-TMR does. GFP is reported allow detection of 80% of the proteins it is fused too, with misfolding events and maturation time accounting for the loss of 20% efficiency (Okumus *et al*. 2016). Fig. 3.11 displays the mean distribution found when labelling RecB with a single HaloTag (RecBHalo) and compares the distribution to that measured when labelling RecB with GFP. (The RecBGFP data was not generated as part of this work).

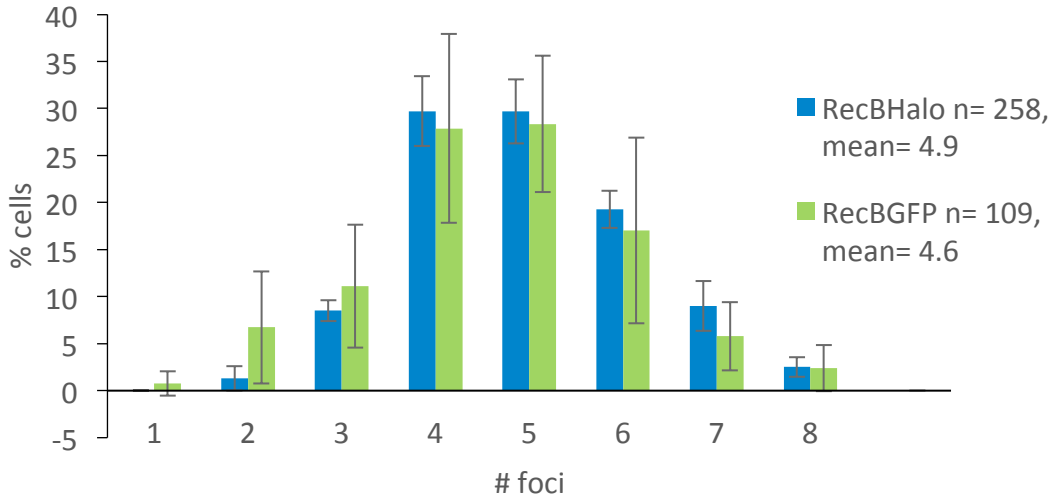


Figure 3.11. Distributions of RecB foci when labelled with HaloTag-TMR (5µM TMR) and GFP. The distributions observed with the HaloTag-TMR labelling system and the fluorescent protein GFP are very similar. Error bars indicate standard deviation of cell number across 3 repeats for each labelling method.

The comparison of RecBHalo and RecBGFP data clearly shows that the labelling of RecBHalo with the HaloTag-TMR protocol is highly efficient, as both the distributions and the mean number of foci per cell observed with RecBHalo (4.9± 0.05) and RecBGFP (4.6± 0.39) are very similar.

3.2.5 Assessment of HaloTag-TMR labelling reproducibility

The RecBHalo data shown in Fig. 3.8 and Fig. 3.9 also highlight the day to day reproducibility of the HaloTag-TMR method. The HaloTag-TMR method for labelling RecB consistently shows a distribution of foci between 2 and 9 per cell across 3 experimental repeats. The summed data and error bars displayed in Fig. 3.11 show the standard deviation of the normalised data for each number of foci per day.

This high reproducibility, along with the comparability of the data with that generated though the RecBGFP fusion provides confidence in the quantitative detection capabilities of the method. Taken overall, the introduction and characterisation of this technique as presented above allows the technique to be used to investigate low copy number proteins and make quantitative statements regarding their copy number in *E. coli*.

3.3 Quantitative detection of RecB, RecC and RecD

3.3.1 Quantification of RecBHalo

RecB was found to be present in all cells, with between 2 and 8 copies of the RecBHalo fusion protein detected in all cells for which the length was measured as <3.5μm. This threshold was introduced to ensure that only cells with comparable gene copy numbers were considered, as chromosomal copy number is known to vary throughout the *E. coli* cell cycle (Bipatnath *et al*. 1998). Fig. 3.12 shows the distribution of RecBHalo molecules observed when all cells are considered in comparison to the distribution seen when the threshold is applied. As can be seen in the figure the all cell data has a distribution that is skewed towards higher levels of expression, indicative of higher gene copy number. For this reason all quantification shown will be representative of only the <3.5 μm cells. For RecBHalo this data can be seen with error bars in Fig. 3.11.

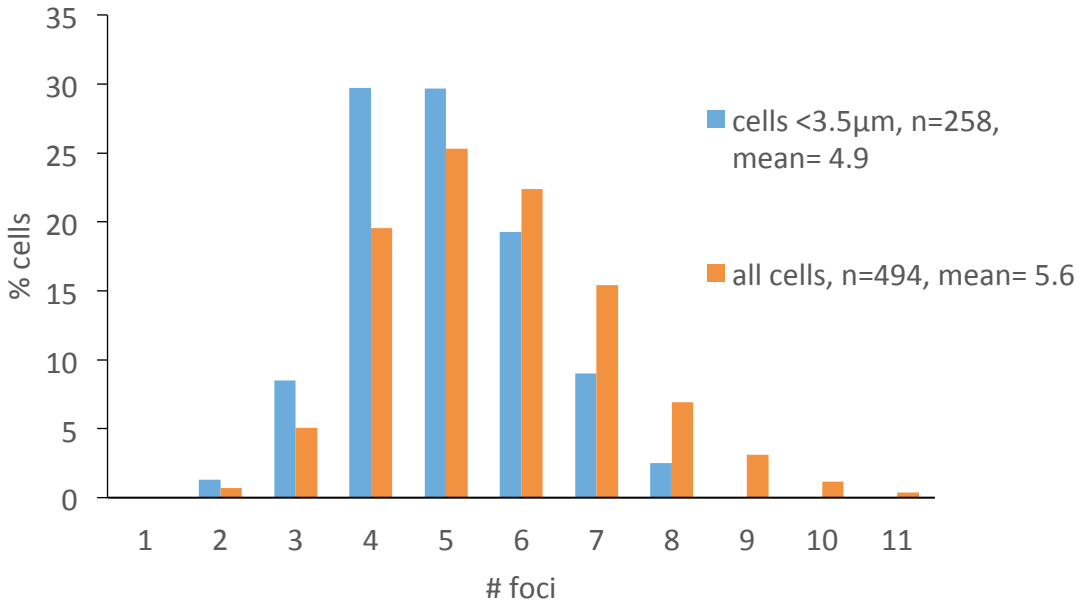


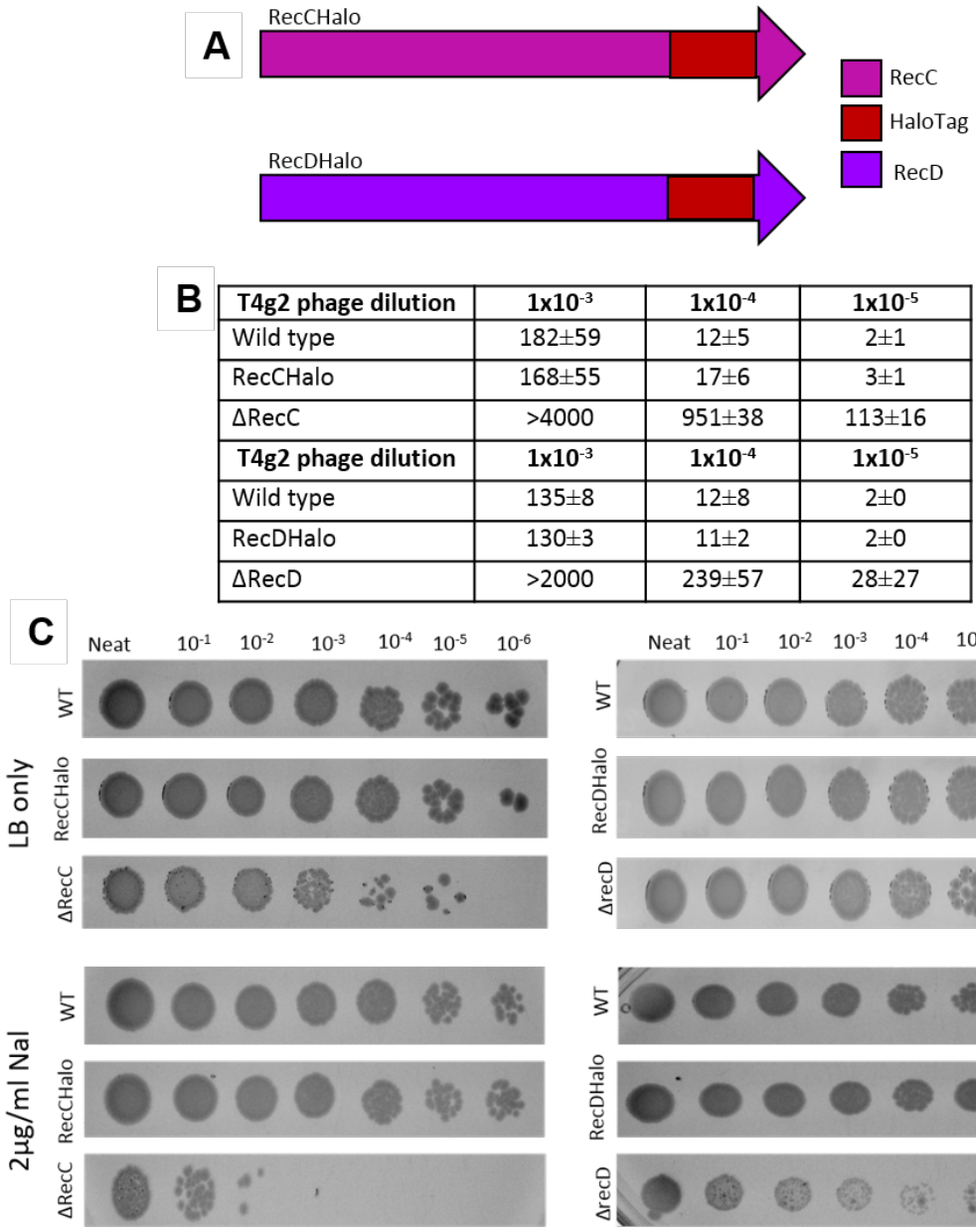
Figure 3.12. Distributions of RecB foci when observing cells <3.5µm and cells of all lengths. When cells of all lengths are considered the distribution is skewed towards higher numbers of foci compared to the distribution given by cells <3.5µm, likely due to increased gene copy number in larger cells.

The mean number of molecules detected per cell was 4.9. These data conflict with that found by Taniguchi *et al* (2010), who reported a mean of 0.6 RecB molecules per cell in their study using C-terminal YFP tags (details in Chapter 1) . However, it is possible that the Taniguchi *et al*. study is under labelling RecB as a result of the C-terminal fusion. As described above the C-terminus of RecB is buried within the structure of the RecBCD enzyme. It is possible that the YFP is unable to fold correctly in this position and so is unable to fluoresce, or that the C-terminal fusion is preventing the assembly of the RecBCD enzyme, causing reduced viability. The case for under labelling of RecB in particular by Taniguchi *et al*. is strengthened by their reported mean of 4.8 molecules per cell for RecD, a protein that is reported to be expressed as part of an operon with RecB and therefore translated from the

same mRNA as described in Chapter 1. This number was in agreement with the quantification of RecD in this study, which is described below along with RecC quantification.

3.3.2 Phenotypic and viability testing of RecDHalo and RecCHalo

Following quantification of RecBHalo, RecDHalo and RecCHalo were analysed. A schematic of the constructs used (made in other work by Meriem El Karoui) can be seen in Fig. 3.13A. As part of this work, each strain was phenotypically tested using the T4Gene2 test described above. Both RecDHalo and RecCHalo were shown to be comparable to wild type, forming very few plaques at each dilution compared to the ΔrecD and ΔrecC controls (see Fig. 3.13B). Additionally, nalidixic acid viability assays were conducted for RecDHalo and RecCHalo. The strains were serially diluted to 10-^{6} and plated on both LB and 2 μg/ml nalidixic acid. As can be seen in Fig. 3.13C, when exposed to nalidixic acid colony formation in the ΔRecC strain is much reduced after the 10^{-2} dilution, while in RecCHalo and wild type it continues to the 10^{-6} dilutions. Equally the ΔRecC strain sees a sharp fall in colony formation after the 10^{-2} dilution, while the RecDHalo and wild type strains show colony formation to 10^{-6}. This indicates that the presence of the HaloTag on the RecC and RecD genes is not impeding their function or the viability of the strain. When plated on LB all strains display colony formation in dilutions to 10^{-6}. A western blot was performed that confirmed HaloTag expression in both strains and growth curves showed that the strains had no growth defects in comparison to wild type (see Appendix Fig. A2.).



T4g2 phage dilution	$1x10^{-3}$	$1x10^{-4}$	$1x10^{-5}$
Wild type	182±59	12±5	2±1
RecCHalo	168±55	17±6	3±1
ΔRecC	>4000	951±38	113±16
T4g2 phage dilution	**$1x10^{-3}$**	**$1x10^{-4}$**	**$1x10^{-5}$**
Wild type	135±8	12±8	2±0
RecDHalo	130±3	11±2	2±0
ΔRecD	>2000	239±57	28±27

Figure 3.13. **Phenotypic and viability assays for RecCHalo and RecBHalo.** A) Schematic representation of the RecCHalo and RecDHalo chromosomal insertions used in this work (constructed by Meriem El Karoui).B) Number of plaques formed following T4g2 phenotypic test for normal RecC and RecD function in RecBHalo and RecDHalo. WT, RecCHalo and RecDHalo strains show formation of very few plaques in comparison to the ΔRecC and ΔRecD strain at each phage dilution. C) Nalidixic acid assay for RecCHalo and RecDHalo. Serial dilution of wild type, RecCHalo, RecDHalo ΔRecC and ΔRecD strains on 2µg/ml nalidixic acid and LB only plates. WT and RecBHalo cells show comparable viability on nalidixic acid while ΔRecC and ΔRecD show reduced viability on nalidixic acid.

3.3.3 Quantification of RecDHalo

For RecDHalo the number of molecules observed per cell varied between 2 and 8. The mean number of foci observed in each data set were 4.7± 1.18, 4.5± 1.11 and 4.3± 1.13. The distributions for each repeat of the experiment, as well as the combined data can be seen in Fig. 3. 14. The mean number of molecules observed for the whole data set was 4.5± 0.17 foci per cell, which is in excellent agreement with the mean published by Taniguchi *et al*. (2010)(4.8). Both the mean and the distribution corresponds well to those measured for RecB in this study as would be expected for genes expressed together as an operon.

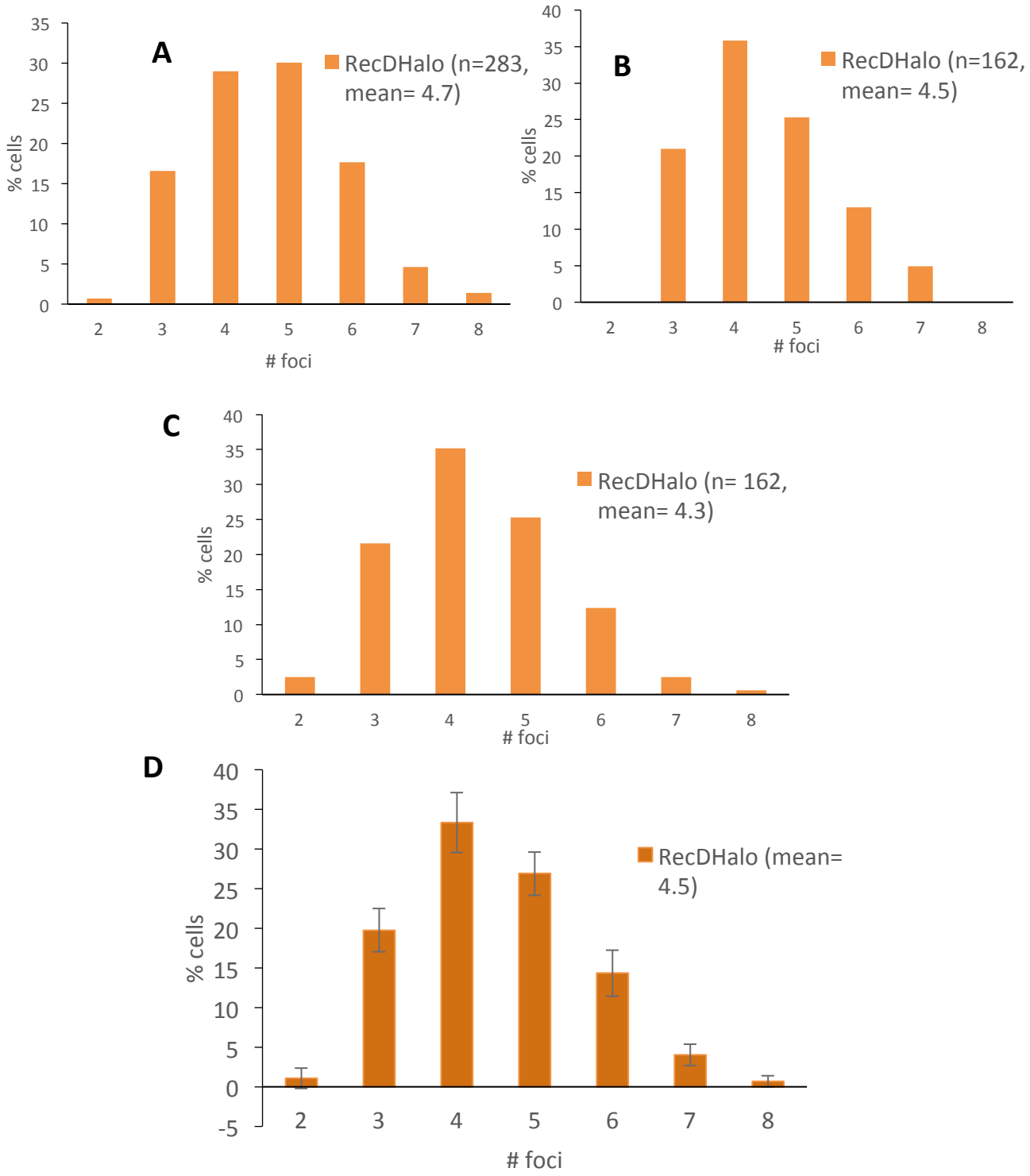


Figure 3.14. Distributions of RecD foci when labelled with HaloTag-TMR (5µM TMR). RecD labelling with HaloTag-TMR. A-C) Distributions of individual repeats HaloTag labelling. Cells are shown to contain between 2 and 8 RecCHalo foci. D) Averaged distribution data (n= 584). Error bars indicate one standard deviation.

3.3.4 Quantification of RecCHalo

The same quantification was done for RecCHalo, which is not expressed in conjunction with RecB and RecD but rather has its own promoter (see Chapter 1 for details). However, the protein numbers observed were very similar. The number of foci found per cell varied between 2 and 8, and the mean number of foci detected in each of the three repeats were 4.7± 1.0, 4.7± 1.12 and 4.6± 1.19. The mean of the whole dataset was 4.7± 0.05. The distribution for each repeat and the averaged distribution can be seen in Fig.3.15.

Taniguchi *et al*. did not report a protein number for RecC, however the ribosome profiling study conducted by Li *et al*. (2014) and outlined in Chapter 1 reported copy numbers of ~100 molecules per cell for each of the RecB, RecC and RecD subunits. While the absolute number reported by Li *et al*. conflicts with the data presented here, the ratio of the subunits observed is the same. It should be noted that ribosome profiling does not directly measure protein number, but rather examines the ribosome density on an mRNA to calculate the absolute protein synthesis rate, which allows the estimate of the absolute protein copy number. This method of calculation both assumes that the proteins are stable once produced, and introduces potential for error in the final estimation of protein copy number. These issues are not present when visualising protein directly through use of quantitative HaloTag-TMR labelling.

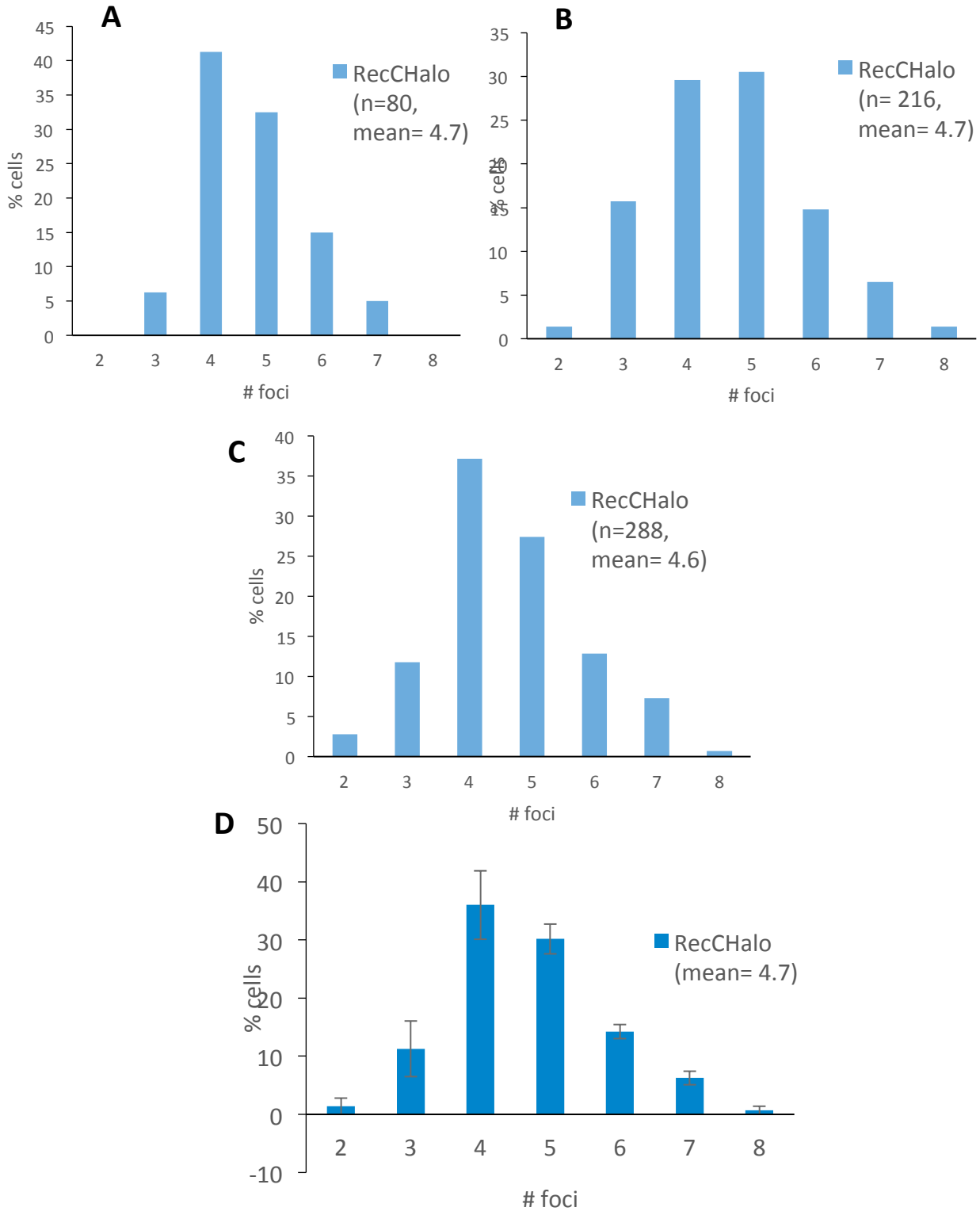


Figure 3.15. Distributions of RecC foci when labelled with HaloTag-TMR. RecC labelling with HaloTag-TMR. A-C) Distributions of individual repeats HaloTag labelling. Cells are shown to contain between 2 and 8 RecCHalo foci. D) Averaged distribution data (n= 584). Error bars indicate one standard deviation.

3.4 Conclusions and future work

3.4.1 Conclusions

In conclusion, the HaloTag-TMR labelling method is able to allow quantitative detection and labelling of single molecules in E. coli, as well as detecting protein over a wide range of expression. The technique has been used to quantify all three subunits of the bacterial DNA repair protein RecBCD, providing insight into the distribution of the molecule within *E. coli* populations. Original estimates of 10 molecules per cell (Dillingham & Kowalczykowski 2008; Smith 2012; Taylor & Smith 1999) have been shown to be in the correct order of magnitude, however the range has been narrowed to between 2 and 8 copies of each subunit per cell, with a mean between 4 and 5 molecules for each subunit. These results confirm the findings of Taniguchi *et al*. for RecD, who reported a mean of 4.8 molecules per cell. The results also introduce copy numbers of 4.9 and 4.7 for RecB and RecC respectively. Taniguchi *et al*. reported a mean of 0.6 molecules per cell for RecB, however, as they did not perform phenotypic or viability tests as shown above, it seems likely that this is due to either misfolding of the reporter YFP or reduction of cell viability resulting in under labelling of RecB. It is of note that RecBD and RecC are observed in cells at similar frequency, as they are not expressed as an operon. However, Li *et al*. observed such proportional synthesis of protein subunits for many multi-subunit proteins, both those expressed from single operons and those expressed from multiple operons. This indicates that, while their estimate of the absolute protein copy number for RecB, RecC and RecD may be incorrect, the ratio 1:1:1 ratio

observed through ribosome profiling is likely to be genuine (see Chapter 5 for further discussion).

3.4.1 Future work

The work presented here introduces the Quantitative HaloTag-TMR labelling method, and refines current estimates for the copy number of the RecBCD enzyme in *E. coli* cells with single molecule accuracy. Future work using this method could benefit from use of recently introduced derivatives of TMR and other dyes, such as the Janelia Fluor series introduced in Grimm *et al*. (2015). These dyes have been structurally modified to improve brightness and photostability while maintaining their spectral properties and cell permeability. This increased array of dyes is very useful in single molecule study, and could be combined with the HaloTag-TMT labelling protocol to allow for live cell single particle tracking, or for detection of multiple proteins simultaneously.

Chapter 4

Establishing single molecule mRNA FISH for quantitative mRNA detection, applying the method to *recD* and *recB* mRNA and introducing combined HaloFISH

4.1 Introduction

In this study, the smFISH method described by Skinner *et al*. (2013) and discussed in Chapter 1 was taken from the literature and optimized to allow for labelling and quantification of *recD* and *recB* mRNA in single *E. coli* cells. smFISH is highly technically challenging and is not routinely used in bacterial study, meaning it was necessary to characterize the method thoroughly in the lab before moving on to quantitative mRNA analysis. As discussed in Chapter 1, mRNA are frequently present at only a few copies per cell and survive for only a few minutes before degradation. As such, mRNA is likely to be susceptible to stochastic fluctuations in gene expression. The smFISH method is sensitive enough to detect mRNA at very low copy number, without perturbing the mRNA expression. This makes the technique ideal for use in the case of RecBCD mRNA. As highlighted in Chapter 1, there are currently very few techniques that are able to simultaneously quantify protein and mRNA in single cells, despite the clear utility of such techniques in the investigation of gene expression. Xu *et al*. (2015) developed a protocol specifically to allow for simultaneous mRNA and protein detection, where a combination of

smFISH and immunofluorescence was used to detect and quantify the mRNA of zygotic drosophila gene *hunchback* and the transcription factor protein *bicoid* that regulates it (see Chapter 1 for further details). Taniguchi *et al*. (2010) were able to perform mRNA FISH on 137 library strains containing chromosomal fusions of YFP to the c-terminal ends of highly expressed proteins (>100 molecules per cell) and detect both the YFP indicating protein binding and the red fluorescence emitted from the smFISH probes (see Chapter 1 for further details). The ability to monitor both mRNA and protein would allow assessment of the expression of individual genes at the level of both transcription and translation or even to investigate the dynamics between protein transcription factors and the mRNA they regulate. This Chapter will detail results gained using a new protocol designed through the fusion of both the HaloTag-TMR and smFISH techniques. This was done as a proof of principle experiment and shows that the two protocols can be combined efficiently to allow for simultaneous detection of protein and mRNA within a single cell using fluorescence microscopy. This is particularly useful, as the protocols for HaloTag-TMR labeling and mRNA FISH are very similar and can be combined more easily than mRRNA FISH and immunofluorescence as described above. The technique has not yet been optimized for use at the single molecule level but preliminary results suggest that it is able to discriminate between different levels of expression. When fully optimized the HaloFISH technique will allow quantification of both mRNA and protein copy number within individual cells.

4.2 Establishing mRNA detection with smFISH

4.2.1 mRNA FISH is able to detect *recD* mRNA specifically

In this work, single molecule mRNA FISH was performed using multiple DNA oligonucleotide probes labelled individually with fluorophores and complementary to target mRNA (details of all probe sets can be seen in Chapter 2, Tables 2.1-2.3). *recD* mRNA was initially probed against to allow establishment of the technique. This subunit was chosen to allow the simple generation of a negative control strain. *E. coli* null mutants for RecB and RecC are known to be sensitive to DNA damaging agents and be recombination deficient (Willetts & Mount 1969; Willetts *et al*. 1969). However, null mutants for RecD remain highly viable and are proficient in DNA repair and recombination (Chaudhury & Smith 1984; Amundsen *et al*. 1986). The recombination deficiency in RecB and RecC mutants results poor viability and cultures containing many dead cells, meaning RecD mutant cultures are easier to manipulate and provide more reliable controls. In this work a *ΔrecD* strain was constructed through use of Gibson assembly and PMGR as described in Chapter 2. The strain produced was a derivative of MG1655 (Δ*recD*MG). T4Gene2 and nalidixic acid assays were performed as described in previous chapters to confirm the absence of the *recD* gene and its protein. The results of these tests can be seen in Fig. 4.1. In the T4Gene2 phage assay Δ*recD*MG displays a lack of wild type exonuclease functionality, allowing the formation of far more plaques in response to the phage than the wild type control (see Fig. 4.1A). Similarly, the nalidixic acid assay displays a drop in viability when exposed to nalidixic acid in the Δ*recD* strain as compared to

the wild type, while growth on LB is unaffected by the loss of the *recD* gene and its product.

A

T4G2 phage dilution	10^{-3}	10^{-4}	10^{-5}
Wild type	135±8	12±8	2±0
Δ*recD*MG	>4000	493±36	55±13

B

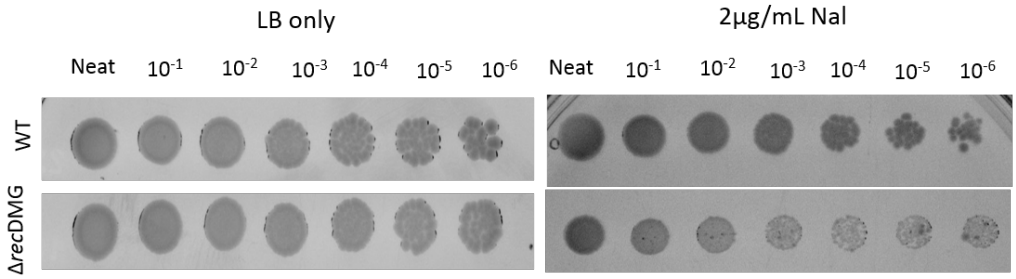


Figure 4.1. Phenotypic and viability assays for Δ*rec*DMG. A) Number of plaques formed following T4g2 phenotypic test for normal RecD function in WT and Δ*rec*DMG. The WT strain shows formation of very few plaques in comparison to the Δ*rec*DMG at each phage dilution. B) Nalidixic acid assay for WT and Δ*rec*DMG. Serial dilution of WT and Δ*rec*DMG strains on 2µg/ml nalidixic acid and LB only plates. WT cells show comparable viability on nalidixic acid while Δ*rec*DMG shows reduced viability on nalidixic acid.

After production of a viable negative control strain for wild type *E. coli* (Δ*recD*MG), investigation of probe binding to *recD* mRNA was conducted through the use of several tests. The first was done to test the specificity of the probes, the second to assess the permeability of the cells to the *recD* probes and a third to investigate the dynamic range of the method. Additionally the possibility that the oligonucleotide smFISH probes were binding to the DNA rather than the mRNA was ruled out. To address the first point, a wild type strain and the Δ*recD*MG strain were

each exposed to smFISH as described in Chapter 2. As can be seen in Fig. 4.2, a bright focus representing binding of the *recD* probes to the *recD* mRNA is evident in the wild type strain, while in the *ΔrecD*MG strain no focus is detected.

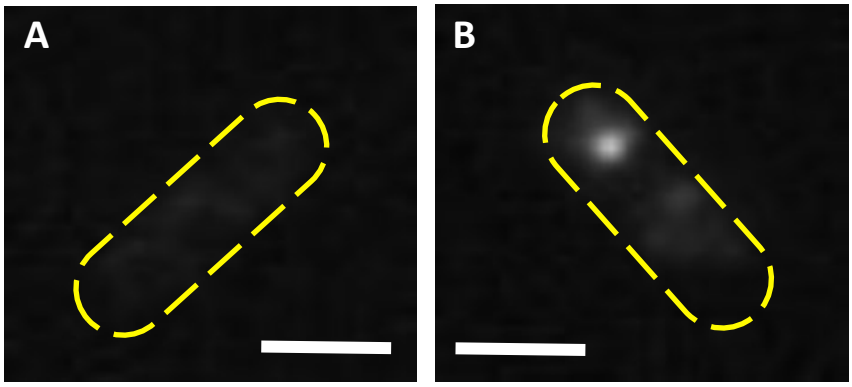


Figure 4.2. mRNA FISH detects *rec*D mRNA specifically in single cells. Both (A) Δ*rec*DMG and (B) wild type strains were exposed to the smFISH protocol, and signal specific to the binding of multiple single-fluorophore labelled oligonucleotide probes can be seen only in the wild type strain that is able to express *rec*D mRNA.

The images in Fig.4.2 however represent only one cell. The histogram in Fig. 4.3 displays the percentage of cells per population that contained foci. In the Δ*recD*MG population (2791 cells), 0.4% of the cells contain at least one focus, however the occurrence of a low number of false positives is expected due to nonspecific binding of probes, and 99.6% contain no foci. In the wild type population (3303 cells) however, 19.6% of cells contained at least one focus, with 81.4% containing no foci. As stated above and discussed in the introduction, the presence of foci in only 19.6%

of the cells is not unexpected given that mRNA are frequently present in low copy number.

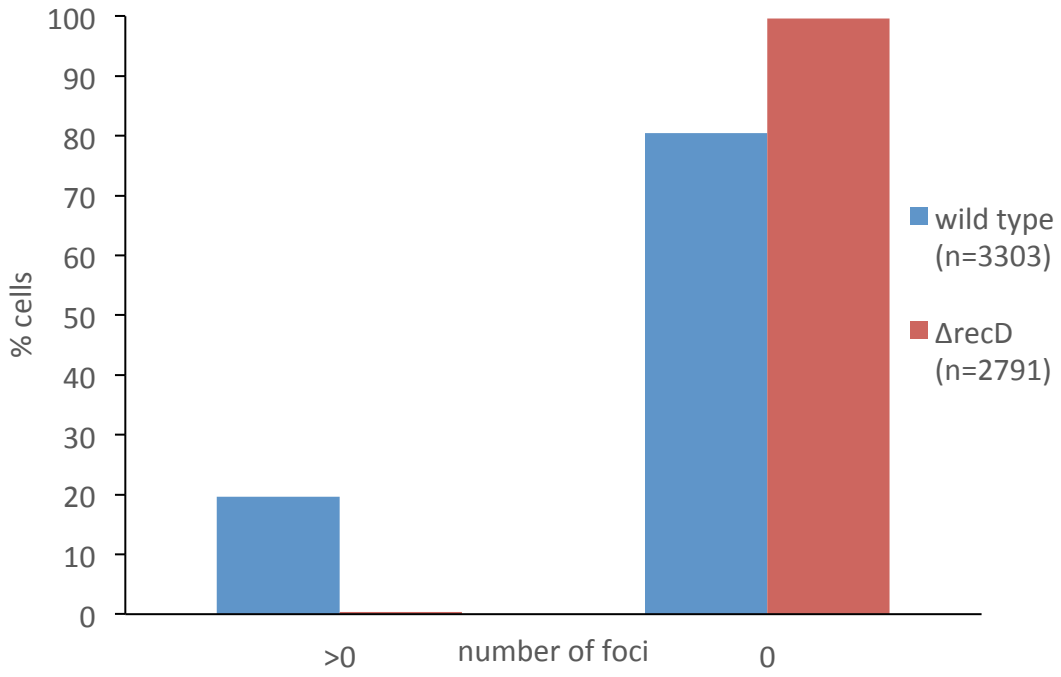


Figure 4.3. mRNA FISH detects *rec*D mRNA specifically across the population. Histogram showing whole-population results for *rec*D detection in wild type and Δ*rec*D using smFISH. A small number (0.4%) of false positives are detected in the Δ*rec*D strain through spurious probe binding, while 19.6% of wild type cells were found to contain *rec*D mRNA. This result is consistent with expectations for mRNA of low expression genes.

However, having established that only ~20% of wild type cells display labelled foci following exposure to smFISH, the possibility that the *recD* probes were not entering cells with 100% efficiency was considered. To assess this possibility of a permeability issue it was decided to overexpress the *recD* gene under the control of an inducible promoter to ensure its expression and availability for probe binding and

detection. Here the BW strain was used, and arabinose induction of the araBAD operon as explained in previous chapters was used to control the expression of the *recD* gene (see Chapter 2 Table 2.8 for details of pBAD*recD* plasmid).

On induction with arabinose (10^{-5}%) followed by exposure to smFISH, examination of 529 cells over two experimental repeats showed that 88% of the cells contained signal above background, while 12% of the cells displayed no detection of overexpressed *recD* mRNA as can be seen in Fig. 4.4. This was found to be the case across a range of induction levels.

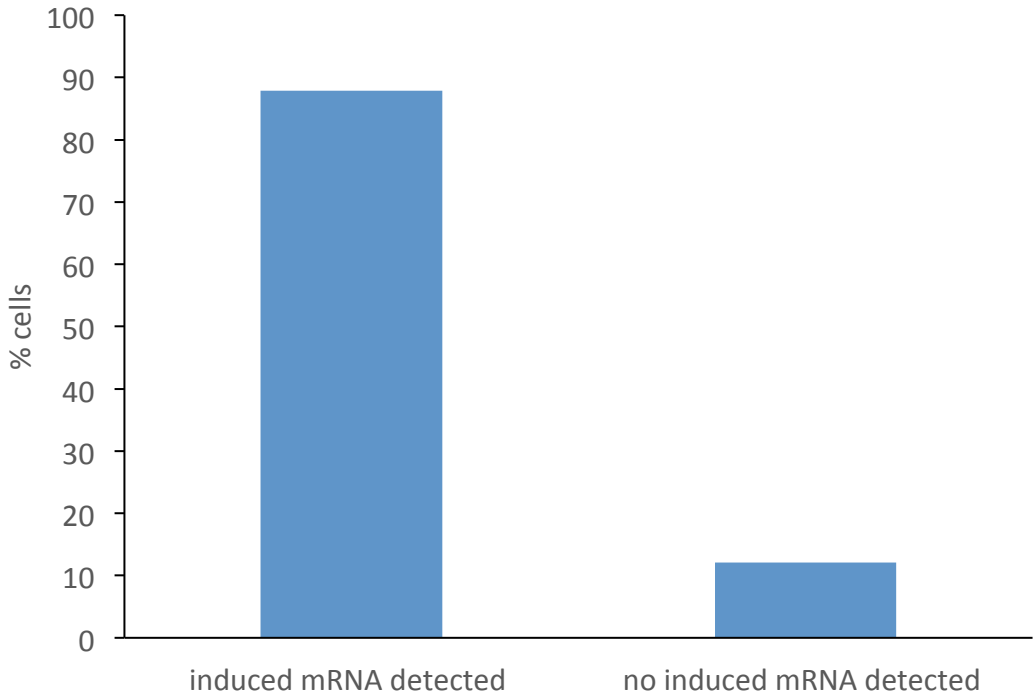


Figure 4.4 *rec*D overexpression is detected in 87% of cells. The oligonucleotide probes were able to successfully enter and bind to *rec*D mRNA in 88% of cells induced with 10^{-5}% arabinose. However there is a 12% population of cells in which overexpression of *rec*D is not detected.

The data shown in Fig. 4.4 confirms that 87% of the cells are permeable to the *recD* probes, and that the probes bind to available mRNA. However, while possible, it is not necessarily the case that the 12% of cells that do not show arabinose detection are impermeable to the probes. It is possible that these cells do not contain overexpressed *recD* mRNA. This could occur if the pBAD*recD* plasmid was unstable within the cell. pBAD*recD* is a high copy number plasmid, and when RecD production was induced with 1% arabinose culture growth was seriously impeded. This would provide a strong selective pressure on the cells to lose the plasmid, preventing RecD overexpression and *recD* detection. It is possible that plasmid loss was occurring within a percentage of cells at induction levels lower than 1% arabinose. It is also possible that at 10^{-5}% arabinose, there is simply not enough arabinose present in the media to fully induce all promoters, meaning that some cells remain uninduced.

The dynamic range of the method was tested through use of different levels of arabinose induction. Arabinose induction was performed in Δ*recD*MG at 10^{-6}%, 2.5x10^{-6}%, 5x 10^{-6}%, 7.5x10^{-6}%, 10^{-5}%, and 10^{-4}% arabinose, and the results obtained strongly indicated that the smFISH protocol was effective at detecting mRNA not only at single molecule level, but across a wide range of expression levels as can be seen in Fig. 4.5. This figure displays data from >230 cells per condition between 10^{-6} and 10^{-5} and >25 cells for 10^{-4} (the low number of cells analyzed at 10^{-4}% arabinose is due to the necessity of using the same imaging conditions across all concentrations to allow for comparison, at 10^{-4}% arabinose induction the fluorescent signal from the smFISH probes was beginning to saturate

the camera). At 10^{-6} and $2.5x10^{-6}$ very little induction is observed due to the minute amount of arabinose present, however at $5x10^{-6}$% arabinose and above a linear increase in the detection of mRNA by the oligonucleotide probes consistent with the linear increase in arabinose induction is observed, as can be seen in Fig.4.5.

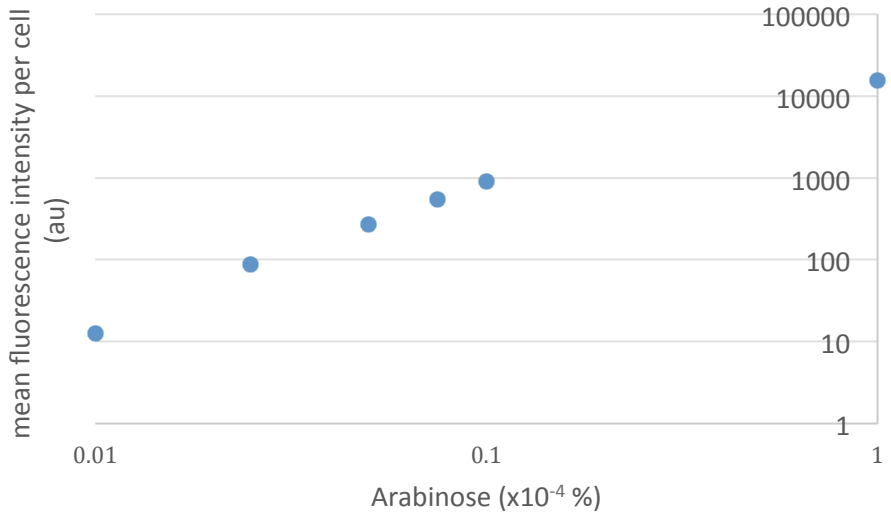


Figure 4.5. mRNA FISH can detect changes in *rec*D mRNA concentration across a wide range of induction levels. As induction of *rec*D expression from pBAD*rec*D is increased with increasing arabinose, resultant detection of *rec*D mRNA by smFISH increases correspondingly across a wide range of induction levels.

Finally, as the probes used to detect *recD* mRNA are DNA based and complementary to the antisense strand of the *recD* gene, it is possible that the smFISH probes could bind to the DNA and produce foci. The conditions generally used for DNA FISH however are dissimilar to those used for RNA FISH, as the harsh conditions required to denature the chromosome for DNA FISH can damage

single stranded RNA, with concentrations of 70% or more formamide being required to facilitate probe hybridization rather than the 40% formamide recommended in smFISH (Barakat & Gribnau 2014). This suggests that in our conditions DNA is not readily labelled. Moreover, given that *recD* is a low expression gene the system has an evident internal control. As can be seen in Fig. 4.3, 81.4% of wild type cells (which all carry the *recD* gene) display no foci specific to the binding of the *recD* smFISH probe set, strongly indicating that the *recD* DNA is not being detected. Thus it can be conclude that the probes are indeed binding to and allowing detection of *recD* mRNA.

4.3 Quantitative detection of *recD* and *recB* mRNA

4.3.1 Quantification of *recD* transcript number

Once confident that the technique was labelling *recD* mRNA in at least 87% of cells quantitative analysis could take place. In 3303 cells a total of 716 foci were identified, giving a mean of 0.22 foci per cell. 647 of the 3303 cells (19.6%) were found to contain one or more foci. Specifically, 583 (17.6%) contain only one focus and 64 (1.9%) contain more than one foci- 59 contain 2 foci, 4 contain 3 foci and one contained 4 foci. This information gives us an indication of the distribution of *recD* mRNA across the population of cells at a fixed point in time, however, further information can be obtained by plotting the intensity data of each of the foci. This data is valuable because it is possible that the individual foci described above represent more than one mRNA that is being bound by more than one set of 48 labelled probes. As was described in Chapter 3 the diffraction limit of light means

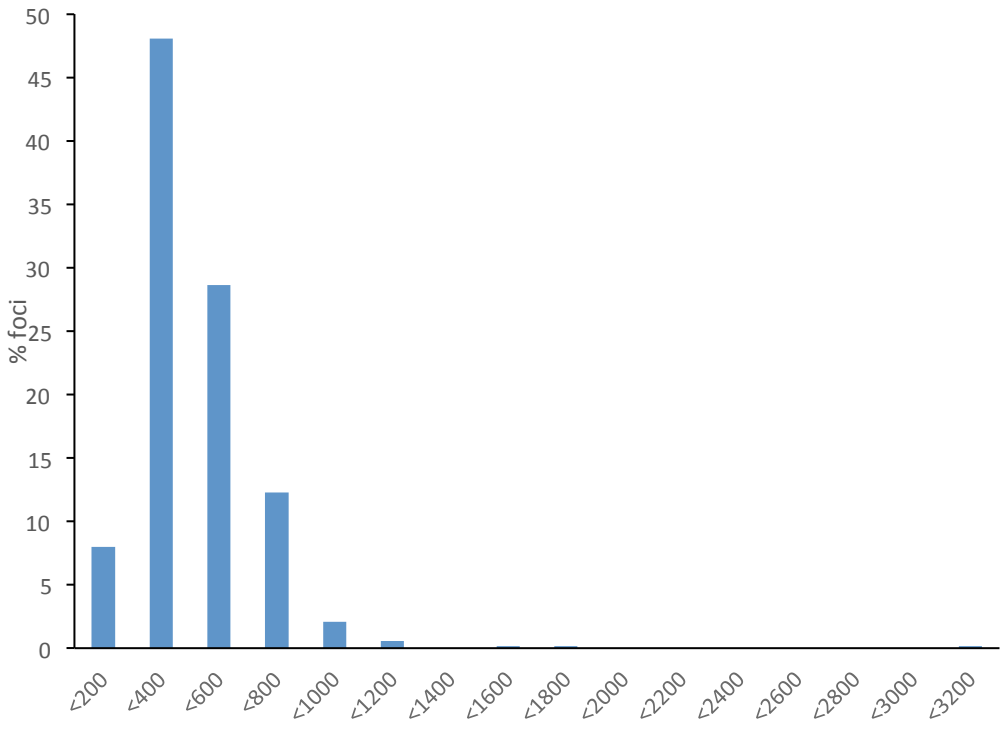
that under a standard epifluorescence microscope a single focus represents a diffraction limited spot, the details of which cannot be resolved below 250 nm (Thompson *et al*. 2002). It is therefore possible that more than one mRNA is present in the 250 nm space, but that the individual foci cannot be resolved. Additionally, as each probe set consists of 48 probes, and as it is unlikely that 100% binding efficiency will be achieved on each mRNA, a range of foci intensities will correspond to a single number of mRNAs detected. Plotting a histogram of the foci intensity however provides means to assign a fluorescence value to a single mRNA, allowing interpretation of the number of mRNA present in any given diffraction limited spot.

Skinner *et al*. (2013) plot intensity histograms that allow definition of an intensity value for single-mRNA, and then use this number to convert the total spot intensity in each cell into a number of target mRNA. The authors argue that in a low mRNA expression sample, where individual foci are spatially separated, the most frequent spot intensity observed will correspond to the presence of a single mRNA. In samples that have a mean of three mRNA per cell or fewer, Skinner *et al*. found that a well-defined peak was present representing a single mRNA. To estimate the single-mRNA intensity, Skinner *et al*. fit a Gaussian around the peak representing the most frequent mRNA intensity, and the mean of this Gaussian was taken as the intensity value corresponding to a single mRNA. The total cellular intensities were then divided by this number and the value rounded to the closest non-zero integer to give an estimate of mRNA per foci. In this work a slightly different approach was taken, as the data obtained did not fit a Gaussian distribution (this is likely due to

truncation of data by the threshold applied to discount false positives). Here, the focus intensity data were binned (as was done in Skinner *et al*.) and a histogram produced as can be seen in Fig. 4.6A (bins of 200 au were selected as they reflected the expected distribution of the FISH data, giving multiple peaks corresponding to multiple mRNA copies). The mean intensity of the predominant species was taken as an estimate of the brightness of a single mRNA (for *recD*, 296 au). As was done in Skinner *et al*., The total fluorescence intensity per cell was summed (this accounted for cells where there is more than one focus) and divided by this single mRNA estimate, and mRNA number per cell was assigned according to the nearest non-zero integer. The assigned number of mRNA, the number (and percentage) of cells carrying this assigned number of transcripts and the total number of mRNA represented in these cells, as well as the total number of mRNA molecules in the data set can be seen in Fig. 4.6B.

For *recD*, calculation of the number of mRNA per cell alters the mean from 0.22 foci per cell to 0.31 mRNA per cell as while 716 foci were detected, these foci represent 1026 mRNA. Of the 19.6% of cells in which mRNA was detected 11.6% contained only one mRNA, 5.9% contained two mRNA and 1.4% contained three mRNA. The remainder contained between four and 11 mRNA (see Fig. 4.6B for full distribution).

A



B

Number of mRNA	Number (Percentage) of cells	Number of individual mRNA
0	2656 (80.4)	0
1	383 (11.6)	383
2	194 (5.9)	388
3	47 (1.4)	141
4	11 (0.3)	44
5	8 (0.2)	40
6	2 (0.1)	12
7	1 (0.0)	7
11	1 (0.0)	11
Total	**3303 (100)**	**1026**
	Mean mRNA per cell	**0.31**

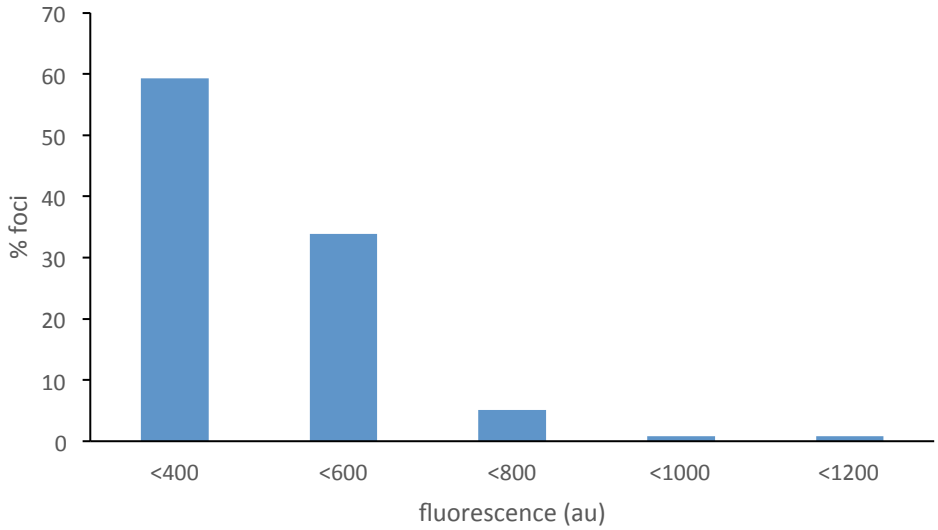
Figure 4.6. *rec*D quantification with smFISH. A) Focus intensity histogram. Data were divided into 200 au bins and the most frequent species identified. The mean value within this bin (296 au) was assigned as an estimate for the fluorescent intensity of a single mRNA, and total intensity values per cell were divided by this number and rounded to the nearest non-zero integer, which was assigned as the number of mRNA per cell. B) Table displaying assigned mRNA number, the number (and percentage) of cells found to carry each number of mRNA and the total mRNA represented within these cells. The total number of cells carrying mRNA, the total number of mRNA detected and the mean mRNA

4.3.2 Quantification of *recB* transcript number

Having established the smFISH protocol in the lab, I next quantified the *recB* mRNA content of wild type *E. coli*. There were 118 foci imaged in 667 cells, giving a mean of 0.18 foci per cell. 109 cells had at least one focus (16.3%). Specifically, 101 cells were found to have one focus exactly (15.1%) and 8 cells were found to have more than one focus (1.2%) with 7 cells containing 2 foci and only 1 cell containing 3 foci. The analysis outlined above, whereby a histogram of foci intensity was plotted and an estimate of the fluorescence intensity of a single mRNA was made, was applied to the analysis of *recB* mRNA also. The histogram produced can be seen in Fig.4.7A, and the estimated single mRNA intensity calculated was 333 au. The total fluorescence intensity per cell was divided by this number, and a number of mRNA per cell assigned (as can be seen in Fig.4.7B). As with *recD*, the mean number of mRNA per cell is slightly higher than the mean number of foci per cell with 0.21 mRNA per cell being observed. Of the 16.3% of cells in which mRNA was detected 12.4 % contained only one mRNA, 3.0% contained two mRNA, 0.6% contained three mRNA and 0.3% contained four mRNA. The remainder contained between four and 11 mRNA (see Fig. 4.7B for full distribution).

Quantification of *rec*C transcript number was also attempted, however technical difficulties meant that reproducible results were not obtained, and so the data are not presented here.

A



B

Number of mRNA	Number (Percentage) of cells	Total mRNA represented by foci
0	558 (83.7)	0
1	83 (12.4)	83
2	20 (3.0)	40
3	4 (0.6)	12
4	2 (0.3)	8
Total	**667 (109)**	**143**
	mean mRNA per cell	**0.21**

Figure 4.7. *recB* quantification with smFISH. A) Focus intensity histogram. Data were divided into 200 au bins and the most frequent species identified. The mean value within this bin (333) was assigned as an estimate for the fluorescent intensity of a single mRNA, and total intensity values per cell were divided by this number and rounded to the nearest non-zero integer, which was assigned as the number of mRNA per cell. B) Table displaying assigned mRNA number, the number (and percentage) of cells found to carry each number of mRNA and the total mRNA represented within these cells. The total number of cells carrying mRNA, the total number of mRNA detected and the mean mRNA per cell are also shown.

4.3.3 Comparison of *recD* and *recB* transcript number

As can be seen in Fig 4.6B and Fig 4.7B, the final numbers of mRNA per cell, and the distributions of mRNA number per cell, are very similar for *recD* and *recB*. *recD* is observed at a frequency of 0.31 mRNA per cell and *recB* is observed at 0.21 per cell. (These numbers increase only very slightly to 0.34 and 0.24 if we account for a potential 12% of the cells being impermeable to both sets of probes, as discussed section 4.2.1). The similarity of these means, and the similarity of the mRNA copy number distributions observed in each population (displayed in Fig 4.8), are as expected given that *recBD* form an operon and are expressed as a single polycistronic mRNA (as described in Chapter 1), meaning that in each case the probes are likely to be labeling different parts of the same transcript. It is possible that the slightly elevated mean and the wider distribution observed for *recD* is an artifact of the larger sample size used (3303 cells for *recD* and 667 cells for *recB*). However, as described in Chapter 1 there is also an internal promoter within the *recBD* operon that controls expression of *recD* only within the cell, meaning it is possible that the levels of *recD* mRNA in the cell are slightly elevated above that of *recB* mRNA.

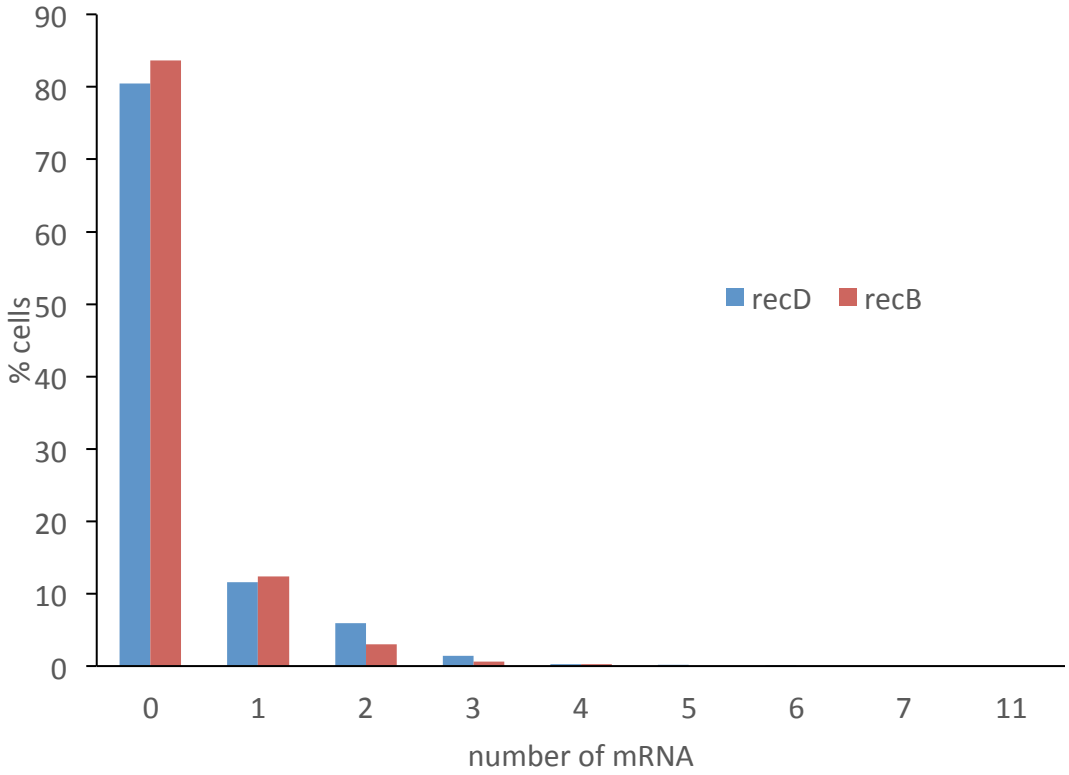


Figure 4.8. Distributions of *rec*D and *recB* mRNA. *recD* and *recB* mRNA are present in cells at very low levels. For each gene, ~80% of cells examined were found to contain no mRNA. The mRNA present in the remaining 20% of cells was distributed similarly for *rec*D and *rec*B, with the largest proportion of cells containing only one mRNA, and the proportion of cells decreasing for two, three and four mRNA. When probing for *rec*D a very small number of cells were found to contain 5-11 mRNA.

4.4 Combining HaloTag-TMR and smFISH to label protein and mRNA simultaneously within cells

4.4.1 Protocol combination

Quantitative HaloTag-TMR labelling of protein and quantitative smFISH of mRNA are both methods that allow precise assessment of gene expression products. The combination of these two techniques would allow quantification of both mRNA and protein within the same cell, with the potential to push the method to the single molecule level for both mRNA and protein. Simultaneous detection of both mRNA and protein within single cells allows understanding of the correlation between mRNA and the protein that it codes for, as well as for analysis of mRNA and the protein transcription factors that control their production. As such, simultaneous mRNA and protein detection give greater insight into the dynamics of gene expression than the detection of either individually. The two methods are highly compatible, in contrast to the protocol combining smFISH and immunofluorescence described in Xu *et al*. (2015) where combination of the protocols is challenging due to conflicting reaction conditions. The HaloTag-TMR protocol produces fixed cells that already carry labelled protein while the smFISH protocol begins with the fixation of cells before permeabilisation and further labelling (See Chapter 2 for details). To produce proof of principle for HaloFISH, oligonucleotide probes were designed against the HaloTag gene itself (See Chapter 2 Table 2.3 for details). This was done for three reasons. Firstly, the use of probes against HaloTag mRNA allowed direct visualization of both the HaloTag protein and its corresponding mRNA. Secondly, wild type *E. coli* could be used as a negative control as it lacks both the HaloTag gene and the mRNA and protein it produces. Finally, it was essential that spectrally separable probes were used to detect both mRNA and protein, to allow visualization of and discrimination between both molecules in the

same cell. For this reason the HaloTag mRNA probes were designed to be tagged with the fluorophore fluorescein, which emits light in green. The TMR fluorophore that is used in the HaloTag-TMR protein labelling technique emits light in red. A schematic of the combined HaloFISH protocol can be found in Chapter 2.

4.4.2 Proof of principle for the HaloFISH protocol

To test whether the smFISH and HaloTag protocols could be combined and allow for detection of both mRNA and protein, production of the HaloTag protein was overexpressed under the control of an inducible promoter. This was done using the pBAD*halo* plasmid and BW27783 strain described in Chapter 3 and induction was done with 10^{-3}% and 10^{-4}% arabinose. The protocols were combined successfully, although it was discovered that the fluorescein bound smFISH probes did not enter the cells as efficiently as the TAMRA bound *recD* and *recB* probes, and the cells needed to be permeabilised in ethanol for six nights rather than one. However, detection of both protein bound to TMR and mRNA bound to fluorescent oligonucleotide probes was possible at high levels of expression. This can be seen in Fig.4.9.

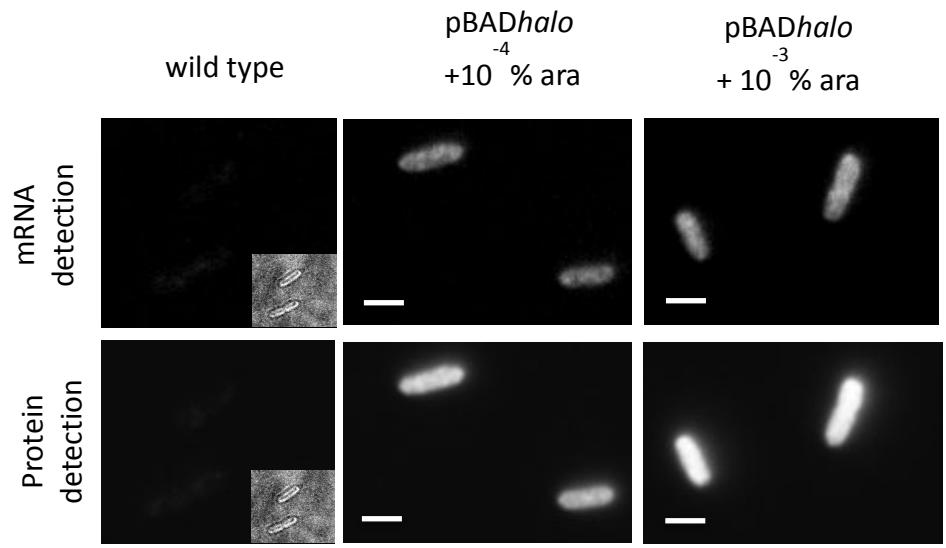


Figure 4.9 Simultaneous labeling of highly expressed mRNA and protein is possible with HaloFISH. In the pBAD*halo* strains induced with arabinose, signal is detected from both the mRNA binding probes and the HaloTag-TMR. No signal is seen in the absence of the HaloTag gene. Scale bars represent 1 μm.

This figure displays only representative cells, however independent quantification of the mean fluorescence intensity of the cells in both the protein (red) and mRNA (green) detection channels was done using automated analysis as described in Chapter 2. As can be seen in Fig 4.10, there is an increase in mean fluorescence intensity of the cells in both channels corresponding to an increase in protein and mRNA expression when there is increased arabinose induction (≥55 cells per condition were analysed).

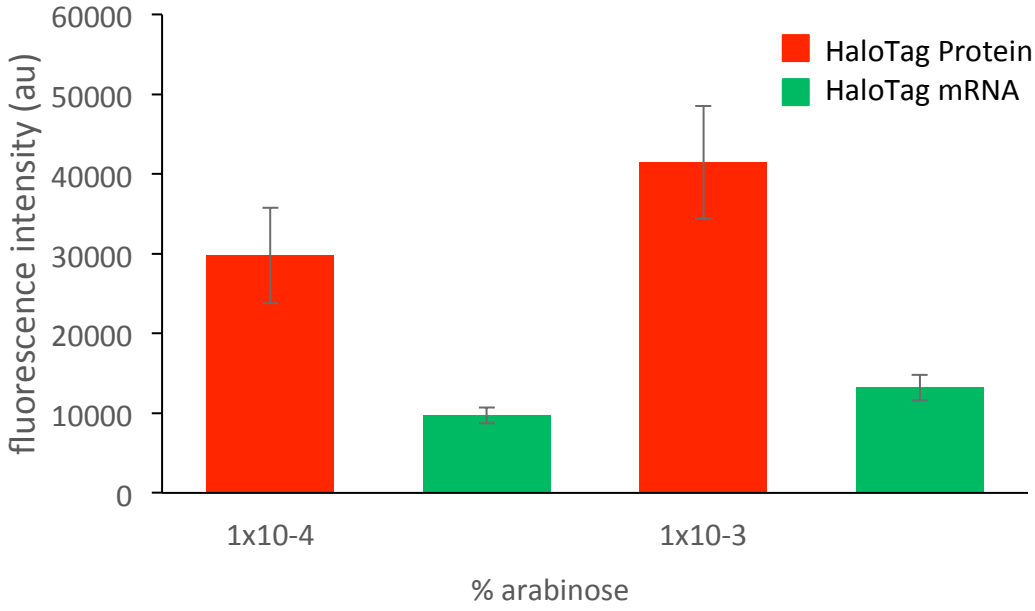


Figure 4.10. Signal is detected on induction of mRNA and protein expression with arabinose, and the amount of signal increases as the level of induction increases. Histogram displaying detection of HaloTag protein (red) and HaloTag mRNA (green) when the expression of the gene is induced with varying quantities of arabinose. This data suggests that the fluorescence detected varies with the amount of protein/ mRNA available for binding by probes.

This initial experimentation provided proof of principle for the technique. However, as it stands this technique is limited to the detection of high copy number protein and mRNA. Attempts were made to detect the HaloTag protein and mRNA using the RecBHalo strain described in Chapter 3, where the HaloTag is fused to the native RecB gene and therefore produced in low copy number, but no mRNA was detected (see Section 4.2.5 below). Further development and characterization would be required to push the method to single molecule level and make it truly quantitative. However, as quantitative results have been obtained for each technique

individually, and as the techniques have been proven to be compatible it seems highly likely that such single molecule simultaneous detection will be possible.

4.5 Conclusions and further work

4.5.1 smFISH

The smFISH protocol was successfully taken from the literature and implemented in the lab, and shown to be capable of specifically detecting mRNA expressed at both high and low copy number within cells. Detection at the level of single mRNA molecules was established, and a method for discerning the number of mRNA molecules represented by a single diffraction limited focus was implemented. The technique was used to quantify the mRNA of the *recB* and *recD* genes, which are known to be expressed as an operon. The quantification achieved showed good agreement in average number of mRNA per cell for each gene (0.31 for *recD* and 0.21 for *recB*), as well as the distributions of mRNA per cell. It is possible that the internal promoter within the *recBD* operon is responsible for the slight elevation in *recD* level above that of *recB*. The distributions of *recB* and *recD* as shown in Figure 4.8 correlate well with those seen in other work investigating low copy number mRNA using mRNA FISH. Further work for smFISH would include increasing the sample size for the *recB* quantification, and quantification of the mRNA of the *rec*C gene. This gene encodes the RecC subunit of the RecBCD holoenzyme but is known to be expressed under the control of a different promoter to the *recB* and *recD* genes.

4.5.2 HaloFISH

Proof of principle was produced for detection of both mRNA and protein in the combined HaloFISH protocol. In addition to this it was established that the technique can detect changes in mRNA and protein concentration, as it was able to detect increased signal for both mRNA and protein when the induction of expression was increased. This is critical as it shows that the conditions used in each step of the protocol are not causing signal degradation from previous steps. Further work to optimize the technique and allow detection of single molecules of both mRNA and protein is required. As described above simultaneous low copy number detection of the HaloTag mRNA and protein was attempted, however no mRNA was visualized. There are several potential reasons for this – firstly the probe set used to bind HaloTag mRNA contained only 26 probes, rather than the 48 used for *recB* and *recD* detection. This was a consequence of the shorter length of the HaloTag gene (891 bp) compared to the RecB (3543 bp) and RecD (1827 bp) and the constraints of the algorithm used to generate the probes (see Chapter 2 for details). This problem could be overcome through use of one of the strains that contains not one but two HaloTag gene sequences described in Chapter 3 and a complementary probe set. The longer mRNA produced in this strain would allow for a full set of 48 probes to be bound to the mRNA, increasing the fluorescence output expected per mRNA.

Additionally, as described in Chapter 1, cellular autoflourescence is generally observed in green. However, to combine the HaloTag and smFISH protocols it was necessary to select probes that were spectrally separable from the red TMR used in the HaloTag procedure, which resulted in the use of fluorescein for mRNA detection.

As well as introducing the issue of increased autofluorescent background, the probes that were labelled with fluorescein were found to be substantially less able to permeate the cells than those labelled with TAMRA as was done for *recD* and *recB* detection, with cells needing to be permeabilised in ethanol for up to 6 days to allow signal detection of fluorescein in contrast to the overnight permeabilisation carried out for the TAMRA probes. As described in Chapter 3, however, there is an ever increasing repertoire of fluorophores that are compatible with the HaloTag enzyme. Specifically, recent development of the Janelia Fluor series to include a HaloTag marker dye that can be excited with far-red light (JF_{646}, Grimm *et al*. 2017) could prove very useful in circumventing both the autofluorescence and permeability issues as described above, allowing imaging of both protein and mRNA with bright, spectrally separable and photostable fluorophores.

Taken overall, smFISH is a highly precise method for the detection of low copy number protein in fixed *E. coli* cells, and one that does not require perturbation to the expression of the mRNA to allow detection. The technique is however highly technically challenging, requires prior knowledge of the mRNA sequence to be examined and extremely low copy number mRNAs such as those produced by the *recBD* operon require very large sample sizes of cells to be examined to allow for meaningful analysis. The combination of the technique with that of HaloTag protein detection shows promise as a means of simultaneous, quantitative protein and mRNA detection.

Chapter 5

Discussion

5.1 HaloTag-TMR labelling of RecB, RecC and RecD protein

5.1.1 Quantitative HaloTag-TMR labelling of low copy number protein

As outlined in Chapter 3, this thesis introduces and characterizes a method that allows quantitative detection of low copy number protein at the single molecule level in *E. coli*. The quantitative HaloTag-TMR protocol was used to label three subunits of the bacterial DNA repair enzyme RecBCD. As discussed in the introduction, protein identification, detection and localization is routinely done through use of either fluorescent proteins such as GFP (or one of its many derivatives) or through the use of chemical fluorophores. For single molecule protein detection, use of inherently fluorescent proteins is challenging. This is because they are limited by their structural biology in their colour palette, which is restrictive of how many molecules can be labelled simultaneously, have a tendency to denature when in cells subjected to fixation, and have limited brightness and photosability. These features are critically important in single molecule microscopy, where the absolute quantity of fluorescence emitted per fluorophore is very low, and any enhancement to the fluorescence output can be highly significant in terms of increasing signal. The HaloTag is a modified enzyme that is capable of binding a chloroalkane linker, in this case fused to bright, photostable, organic fluorophore TMR which allows one-to-one labelling of a protein of interest that has been genetically fused to the HaloTag.

The TMR ligand used in the protocol is known to be able to permeate the *E. coli* membrane and label proteins in both the periplasm and cytoplasm of the cells (Ke *et al*. 2016). However, HaloTag-TMR had not yet been used quantitatively at a single molecule level. In Chapter 3 section 3.2 the method is thoroughly characterised and its advantages highlighted, specific detection is seen at low levels of protein expression (individual RecBHalo molecules) and high levels of protein expression (overexpressed HaloTag protein), showing the dynamic range of the technique, which includes the ability to detect individual molecules as well as to detect increased protein numbers over 5 orders of magnitude of induction. Additionally, the labelling efficiency of the technique was proved to be equal to that of GFP (~80%). The generation of this reliable, reproducible single molecule labelling protocol for protein in fixed *E. coli* cells provides a novel technique that complements those described in the introduction, allowing detailed analysis of the protein content of cells at single molecule level using a bright, photostable fluorophore.

As with all techniques, quantitative HaloTag-TMR labelling has disadvantages that counterbalance the advantages outlined above. One such drawback is the fact that the HaloTag must be inserted into the gene sequence for the protein of interest. This increases the time and labour involved in the technique as constructs must be build and strains tested for functionality and viability as was described in Chapter 3, section 3.2.4. Production of functional protein of interest that carries the HaloTag may also be non-trivial, as placing the insert in the C- or N-terminal may not allow for labelling or may impede function (the RecBHalo used in this work has the HaloTag integrated into the coding sequence, see Fig. 3.4 for schematic).

Additionally the need to insert the HaloTag sequence into the gene sequence coding for the protein of interest means that the gene sequence for the protein must be known, unlike in genome wide techniques such as ribosome profiling. Combination of the current technique with microfluidic technology however could increase this without extensive optimisation. Use of the HaloTag and the fluorescent TMR ligand is necessarily more expensive than use of inherently fluorescent proteins, as the ligand must be purchased (#G8252). A further disadvantage of the use of the HaloTag with TMR in comparison to the use of fluorescent proteins is that protein labelling occurs during incubation with TMR, and the TMR must be removed through thorough washing to allow imaging. If cells are not fixed shortly after incubation in TMR and washing (as is done in the quantitative HaloTag-TMR protocol) they will continue to grow and synthesise protein. This protein will remain unlabelled and cannot be imaged or quantified. While this feature creates the opportunity for pulse-chase experimentation, it also means that the utility of the HaloTag is limited in live cells, where inherently fluorescent proteins can be imaged through microfluidic assisted cell screening (Okumus *et al*. 2016).

Taken overall the Quantitative HaloTag-TMR labelling protocol presented in this work has both advantages and limmitations, however the utility of the technique in the detection and quantification of single protein molecules in fixed cells using standard epifluorescence microscopy is evident. The technique has allowed for detailed quantification of the copy numbers of RecB, RecC and RecD and is well suited to detection and quantification of such low copy number proteins.

5.1.2 Quantification of RecB, RecC and RecD protein

In addition to introduction of the quantitative HaloTag-TMR labelling protein, quantification was completed for each of the subunits of the RecBCD enzyme complex. RecB was found to have a mean of 4.9 molecules per cell, RecC was found to have a mean of 4.7 molecules per cell and RecD was found to have a mean of 4.5 molecules per cell. In all subunits the range of molecules detected per cell was consistently between 2 and 8. Prior to this work, several different copy numbers have been reported in the literature for RecBCD. Most recently, Taniguchi *et al*. (2010) reported 0.6 molecules per cell for RecB and 4.8 molecules per cell for RecD using c-terminal YFP fusions and Li *et al*. (2014) reported ~100 molecules per cell for each subunit through use of ribosome profiling and back-calculation of protein synthesis rate.

Good agreement is seen between this study and that of Taniguchi *et al*. when looking at the RecD subunit, where the mean observed by Taniguchi *et al*. is 4.8 and the mean observed in this study is 4.5. However, the numbers differ for the RecB subunit, where this study observed a mean of 4.9 molecules per cell and Taniguchi *et al*. saw only 0.6. As described in Chapter 3, the HaloTag fused to the RecB protein in this study was designed specifically to ensure that the RecBCD complex would be functional, and to allow access of the TMR to the HaloTag enzyme. Functionality and viability tests were carried out and the RecBHalo strain was shown perform as well as wild type cells. No such tests were conducted by Taniguchi *et al*. who performed only growth curves to ensure no significant growth defects were present in the strains they examined. This, combined with agreement between the means

observed for the RecD subunit suggest that the RecBYFP fusion protein examined by Taniguchi *et al*. may have lacked functionality, leading to reduced viability and fewer live cells which would reduce the mean number of RecB molecules observed. It is also possible that the placement of the YFP on the c-terminus of RecB caused increased YFP misfolding, which would also reduce the mean number of RecB molecules observed.

The ribosome profiling study conducted by Li *et al*. reports a synthesis rate of around 100 molecules of RecB, RecC and RecD. This study, however, does not look directly at protein copy number within cells. Rather, they used the ribosome density on the body of a gene to deduce the absolute synthesis rate (in molecules produced per generation), with corrections applied to account for the elevated density of ribosomes seen at the start of open reading frames and at internal Shine-Dalgarno sequences and normalisation of the average ribosome density per protein by the total number of proteins synthesized throughout doubling time of the cells, and then estimated the absolute protein copy number assuming the proteins are stable within the cell. It is possible that the RecB, RecC and RecD proteins are unstable within the cell, and that degradation of the protein is responsible for the low copy number observed. It is also possible that the extensive back calculation required for this method of protein copy number calculation has artificially increased the numbers of RecB, RecC and RecD protein being observed. Indeed this possibility seems likely. Another low copy number protein, the transcriptional repressor LacI, has been measured elsewhere as occurring at a copy number of ≤ 20 copies per cell (Gilbert, Walter and Müller-Hill 1966; Johan Elf, Gene-Wei Li 2007).Li *et al*., however,

report a copy number of >250 molecules per cell. This indicates that the calculation used to estimate protein copy number may be inaccurate, at least for low copy number proteins. As was shown in Chapter 3, the quantitative HaloTag-TMR method is both specific at single molecule level and sensitive to alterations in expression, meaning that if ~100 molecules per cell were present it is likely they would be detected by direct visualisation as well as through ribosome profiling.

There is some consistency between the Li *et al*. study and this work however, in that each detect study finds that the RecB, RecC and RecD subunits are present in cells in ratio very close to 1:1:1. This similarity is not necessarily expected due to the expression of RecC on one operon and RecB and RecD on another. In the study by Li *et al*. they highlight the huge energetic cost to the cell of protein synthesis, and discuss the fact that this high energetic cost (estimated at ~50% of energy consumption in rapidly growing bacterial cells) makes protein synthesis a key regulatory step in determining cellular function. They ask whether cells maintain tight control over the production of individual components for protein complexes such as RecBCD, and introduce the concept of proportional synthesis, whereby subunits of multiprotein complexes are synthesized in ratios corresponding to their final subunit stoichiometry at the level of translation. Looking specifically at protein complexes with known stoichiometry, they identified 64 different complexes (composed of a total of 212 different subunits) and found that 92% of them displayed proportional synthesis. The study saw that such proportional synthesis was seen in both membrane and cytosolic protein complexes, and that it was seen for protein complexes that were composed of subunits encoded on a single polycistronic mRNA

such as the ATP operon as well as those that, like RecBCD, have multiple distinct transcripts. In their study, Li *et al*. highlight the lack of current understanding of translation initiation rate determination, saying that current models for the strength of ribosome binding sites do not account for the proportional synthesis they observe. It is possible that proportional synthesis is regulated through RNA secondary structure or translational autoregulation, which may contribute to the precise, proportional pattern of translation they observe for protein complexes.

The finding of this ribosome profiling study, combined with the similar number of RecB, RecC and RecD molecules observed per cell in this study, indicate that RecBCD is being expressed in a proportional manner, suggesting a form of translational regulation such as those described above is present for the complex that is not yet understood.

5.2 smFISH and HaloFISH labelling

5.2.1 smFISH labelling of mRNA

As outlined in Chapter 4, this thesis presents quantitative data for *rec*B and *recD* mRNA in *E. coli* cells through use of the smFISH protocol published by Skinner *et al* (2014). As detailed in the introduction smFISH allows detection and visualisation of mRNA in fixed cells without perturbation of mRNA synthesis through addition of extra sequence that is relied upon in other microscopy techniques, as well as avoiding the complications and potential artefacts that come with mRNA isolation and nucleotide amplification required for molecular biology approaches to mRNA

quantification. smFISH was optimised with care for use in *recB* and *recD* quantification. Specific detection of *recD* was shown at single molecule level as well as when the mRNA was overexpressed from a plasmid, although ~10% of the population of cells were impermeable to the smFISH probes (this was accounted for in analysis). The dynamic range of the protocol was also displayed, with proportional detection occurring as *recD* expression was varied over several orders of magnitude (10^{-6}-10^{-4}% arabinose induction of pBAD promoter) as well as detection of individual mRNA at low copy number expression. These results are comparable to the range observed by So *et al*. (2011) who investigated the expression of *lacZ* using mRNA FISH under the control of the P_{lac} promoter and were able to detect mRNA between 0.001mM and 1mM IPTG induction. These features make smFISH highly useful for investigation of low copy number proteins. However, the technique also has limitations. One such limitation is that the need to permeabilise the cells prior to hybridisation of the probes means that mRNA FISH is performed only in fixed cells, unlike systems such as MS2 and PP7 which can be used in live cells. Further to this and similar to the HaloTag system described above, to allow mRNA FISH the mRNA sequence of interest must be known prior to investigation in contrast to systems such as RNA Seq. Additionally optimal conditions vary depending on the combination of probe and fluorophore used for each mRNA. The technique is technically challenging, and time consuming for investigation of a large number of different mRNAs.

Taken overall mRNA FISH allows detection and quantification of very low copy number mRNA in single cells and has allowed an understanding of the copy number

of *recB* and *recD* mRNA, however it does require careful optimisation and is both technically challenging and time consuming.

5.2.2 Quantification of *recB* and *recD* mRNA

Quantification of *recB* and *recD* mRNA in this work has shown that cells carry an average 0.21 and 0.31 mRNAs respectively. These results are in reasonable agreement with those gained by T. Hwa (personal communication) through RNA Seq, which suggest approximately 0.1 mRNA per cell for each mRNA. *recD* and *recB* would be expected to produce similar mRNA numbers as seen here as the *recBD* genes form a single polycistronic operon (as discussed in Chapter 1). Therefore the probes are likely to be labelling different parts of the same mRNA (there does appear to be an internal, weaker promotor that allows expression of *recD* only, see Chapter 1 for details).

mRNA are known to be short lived in bacterial cells, and very few mRNA per cell are generally observed. Taniguchi *et al*. reported mean copy numbers of between 0.05-5 molecules per cell for the genes they examined, which compares with the data observed in this study. Additionally, the results observed here for *recB* and *recD* expression are comparable to those observed by So *et al*. (2011) when investigating *lac*Z expression under the control of the P_{lac} with no or very low induction. When observing mean mRNA numbers of 0.1 (no induction) and 0.7 (0.01 mM IPTG) per cell they also observed distributions of mRNA per cell that showed 80-90% of cells containing no mRNA and the remainder containing between one and 20 mRNAs with the majority containing between one and four mRNA copies. A Poisson

distribution is observed for the final mRNA number per cell in both So *et al*. and this work, as is expected for transcription at very low expression levels. Here, further work could be done to quantify the mRNA numbers observed for the *rec*C gene, which would allow a complete comparison of the *recBCD* mRNA levels.

5.3 HaloFISH detection of mRNA and protein

In addition to the quantification of RecBCD protein and mRNA, a novel method for simultaneous protein and mRNA detection was introduced in this work – HaloFISH. This method combined the HaloTag-TMR protein labelling protocol with the smFISH mRNA detection protocol to give a proof of principle for a means of identifying both gene expression products within single cells. Currently techniques that assess both mRNA and protein within single cells are lacking. Combining immunofluorescence and smFISH (Xu *et al*. 2015) allows detection of single molecules of both mRNA and protein, however the protocol is cumbersome. Taniguchi *et al.* (2010) were able to simultaneously quantify YFP and smFISH signal, but were able to detect mRNA only where protein copy number exceeds 100 molecules per cell. Further techniques are required in this area, as simultaneous quantification of mRNA and protein in cells allows greater insight into the process of gene expression as it allows direct observation of relative quantities of each molecule within a single cell at a given time. This can be done for a single mRNA and the protein that is translated from it, or as was done in Xu *et al*. a transcription factor and the mRNA that is expressed under its control can be examined. The protocol outlined in this work is currently able to detect mRNA and protein that have been

overexpressed in cells and quantification has shown increased detection of the HaloTag protein and mRNA when overexpression has been increased (10^{-4}-10^{-3}% arabinose induction of pBAD promoter). Given the successful detection of single molecules of protein and mRNA in *E. coli* with the individual techniques it is likely that further optimisation and use of novel dyes for HaloTag labelling such as those in the Janelia Fluor (Grimm *et al*. 2015; Grimm *et al*. 2017) series as described in Chapter 3 and 4 would allow the technique to function at single cell level.

5.4 Conclusion

In conclusion, the work presented in this thesis provides a novel method for the detection and quantification of low copy number protein in individual bacterial cells, Quantitative HaloTag-TMR labelling. This technique complements those described in Chapter 1. Use of the Quantitative HaloTag-TMR labelling technique has allowed thorough assessment of the copy number of RecB (4.9) RecC (4.7) and RecD (4.5) proteins in cells and the data gathered is in line with both the general expectation of low copy number and with the finding that these subunits are present in a 1:1:1 ratio within cells (Taniguchi *et al*.). Taken together this work shows both that the Quantitative HaloTag-TMR labelling technique is a good option when quantifying low copy number protein, and that while RecBCD is present in low copy number, the copy number of each subunit does not fluctuate widely within cells, varying between 2 and 8 copies for cells of small size (<3.5μm). This, combined with the equal ratios of protein produced from two different operons hints towards some form of regulation of RecBCD expression that is currently not understood. This work also

describes the use of mRNA FISH to quantify very low copy number mRNA. This provides direct measurement of mRNA copy number per cell for *recD* (0.31) and *recB* (0.21) for the first time. Finally this thesis introduces proof of principle for a novel technique for simultaneous mRNA and protein detection in individual cells, HaloFISH. The proof of principle opens the door to pushing the technique to single molecule level (as has been done for each technique individually). Simultaneous quantification of mRNA and protein is incredibly useful as it allows deeper examination of gene expression as it allows comparison of protein and mRNA production within single cells.

References

Amundsen, S.K., Taylor, A.F., Chaudhury, A.M., Smith, G.R., 1986. recD: the gene for an essential third subunit of exonuclease V. *Proceedings of the National Academy of Sciences of the United States of America*, 83(15), 5558–5562.

Anderson, D.G., 1997. The Translocating RecBCD Enzyme Stimulates Recombination by Directing RecA Protein onto ssDNA in a Chi-Regulated Manner. *Cell*, 90, 77–86.

Anderson, D.G. & Kowalczykowski, S.C., 1998. SSB protein controls RecBCD enzyme nuclease activity during unwinding: a new role for looped intermediates. *Journal of molecular biology*, 282(2), 275–85.

Andreev, D.E., O'Connor, P., Zhdanov, A.V., Dmitriev, R.I., Shatsky, I.N., Papkovsky, D.B., Baranov, P.V., 2015. Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biology*, 16(1), 90

Balaban, N.Q., 2004. Bacterial Persistence as a Phenotypic Switch. *Science*, 305(5690), 1622–1625.

Barakat, T.S. & Gribnau, J., 2014. Combined DNA-RNA fluorescent in situ hybridization (FISH) to study X chromosome inactivation in differentiated female mouse embryonic stem cells. *Journal of visualized experiments : JoVE*, (88), e51628.

Bazzini, A.A., Lee, M.T. & Giraldez, A.J., 2012. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science*, 336(6078), 233–237.

Beckmann, B.M. Grunweller, A., Weber, M.H.W, Hartmann, R.K., 2010. Northern blot detection of endogenous small RNAs (approximately14 nt) in bacterial total RNA extracts. *Nucleic acids research*, 38(14), e147.

Belfort, M., Maley, G., Pedersen-Lane, J., Maley, F., 1983. Primary structure of the Escherichia coli thyA gene and its thymidylate synthase product. *Proceedings of the National Academy of Sciences of the United States of America*, 80(16), 4914–8.

Bertrand, E. Chartrand, P., Scaefer, M., Shenoy, S.M., Singer, R.H., Long, R.M., 1998. Localization of ASH1 mRNA particles in living yeast. *Molecular cell*, 2(4), 437–45.

Bevis, B.J. & Glick, B.S., 2002. Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nature Biotechnology*, 20(1), 83–7.

Bianco, P.R., brewer, L.R., Corzett, M., Balhorn, R., Yeh, Y., Kowalczykowski, S.C., Baskin, J., 2001. Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature*, 409(6818), 374–8.

Bipatnath, M., Dennis, P.P. & Bremer, H., 1998. Initiation and Velocity of Chromosome Replication in Escherichia coli B / r and K-12. *Journal of bacteriology*, 180(2), 265–273.

Blattner, F.R. et al., 1997. The Complete Genome Sequence of Escherichia coli K-12. *Science*, 277, 1453-1462.

Bonner, W.A., Hulett, H.R., Sweet, R.G., Herzenberg, L.A., 1972. Fluorescence activated cell sorting. *Review of Scientific Instruments*, 43(3), 404–409.

Brar, G.A. & Weissman, J.S., 2015. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology*, 16(11), 651–664.

Broude, N.E., 2011. Analysis of RNA localization and metabolism in single live bacterial cells: Achievements and challenges. *Molecular Microbiology*, 80(5), 1137–1147.

Bumgarner, R., 2013. Overview of dna microarrays: Types, applications, and their future. *Current Protocols in Molecular Biology*, (SUPPL.101), 1–11.

Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, 25(2), 169–193.

Chalfie, M., Ty, Y., Euskirchen, G., Ward, W.W., Prasher, D.C., 1994. Green fluorescent protein as a marker for gene expression. *Science (New York, N.Y.)*, 263(1988), 802–805.

Chaudhury, A.M. & Smith, G.R., 1984. A new class of Escherichia coli recBC mutants: implications for the role of RecBC enzyme in homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 81(24), 7850–7854.

Chou, C.P., 2007. Engineering cell physiology to enhance recombinant protein production in Escherichia coli. *Applied Microbiology and Biotechnology*, 76(3), 521–532.

Cooper

Cooper, G.M. & Hausman, R.E., 2008. *The Cell A Molecular Approach.*

Cranfill, P.J. et al., 2016. Quantitative assessment of fluorescent proteins. *Nature Methods*, 13(7), 557–562.

Crick, F., 1970. Central dogma of molecular biology. *Nature*, 227(5258), pp.561–3.

Crivat, G. & Taraska, J.W., 2012. Imaging proteins inside cells with fluorescent tags. *Trends in Biotechnology*, 30(1), 8–16.

Cromie, G.A., 2009. Phylogenetic ubiquity and shuffling of the bacterial RecBCD and AddAB recombination complexes. *Journal of Bacteriology*, 191(16),5076–5084.

Day, R.N. & Davidson, M.W., 2009. The fluorescent protein palette: tools for cellular imaging. *Chemical Society reviews*, 38(10), 2887–2921.

Dillingham, M.S. & Kowalczykowski, S.C., 2008. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiology and Molecular Biology Reviews*, 72(4), 642–671.

Dong, H. et al., 1995. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, 177(6), 1497–1504.

Dykstra, C.C., Prasher, D. & Kushner, S.R., 1984. Physical and biochemical analysis of the cloned recB and recC genes of Escherichia coli K-12. *Journal of Bacteriology*, 157(1), 21–7.

Eichler, D.C. & Lehman, I.R., 1977. On the role of ATP in phosphodiester bond hydrolysis catalyzed by the recBC deoxyribonuclease of Escherichia coli. *Journal of Biological Chemistry*, 252(2), 499–503.

Elowitz, M.B. et al., 2002. Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, 297(5584), 1183–6.

Emmerson, P.T., 1968. Recombination Deficient Mutants of Escherichia Coli K12 That Map Between thyA and argA. *Genetics*, 60, 19–30.

Feilmeier, B.J. et al., 2000. Green Fluorescent Protein Functions as a Reporter for Protein Localization in Escherichia coli Green Fluorescent Protein Functions as a Reporter for Protein Localization in Escherichia coli †. *Journal of Bacteriology*, 182(14), 4068–4076.

Femino, A.M. et al., 1998. Visualization of single RNA transcripts in situ. *Science (New York, N.Y.)*, 280(5363), 585–590.

Fernandez-Abalos, J.M. et al., 1998. Plant-adapted green fluorescent protein is a versatile vital reporter for gene expression, protein localization and mitosis in the filamentous fungus, Aspergillus nidulans. *Molecular microbiology*, 27(1), 121–130.

Filonov, G.S. et al., 2014. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society*, 136(46), 16299–16308.

Finch, P.W. et al., 1986. Complete nucleotide sequence of the Escherichia coli recC gene and of the thyA-recC intergenic region. *Nucleic Acids Research*, 14(11), 4437–4451.

Follenius-Wund, A. et al., 2003. Fluorescent Derivatives of the GFP Chromophore Give a New Insight into the GFP Fluorescence Process. *Biophysical Journal*, 85(3), 1839–1850.

Gao, W., Zhang, W. & Meldrum, D.R., 2011. RT-qPCR based quantitative analysis of gene expression in single bacterial cells. *Journal of Microbiological Methods*, 85(3), 221–227.

Garcia, J.F. & Parker, R., 2015. MS2 coat protein bound to yeast mRNAs block 5' to 3' degredation and trap mRNA decay products: implications for the localisation of mRNAs by MS2-MCP system. *RNA*, 21, 1393–1395.

Gibson, D.G. et al., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* , 6(5), 12–16.

Gilbert, Walter and Müller-Hill, B., 1966. The Isolation of the Lac repressor. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 12(1), 41–43.

Ginzinger, D.G., 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Experimental Hematology*, 30(6), 503–512.

Gowrishankar, J. & Harinarayanan, R., 2004. Why is transcription coupled to translation in bacteria? *Molecular microbiology*, 54(3), 598–603.

Grimm, J.B. et al., 2017. A general method to fine-tune fluorophores for live-cell and

in vivo imaging. *bioRxiv*, 127613.

Grimm, J.B. et al., 2015. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods*, 12(3), pp.244–250.

Guzman, L.-M. et al., 1995. Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter. *Journal of Bacteriology*, 177(14), 4121–4130.

Handa, N. et al., 2012. Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, χ, by RecBCD enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23), 8901–6.

Harder, J. & Schröder, J.M., 2002. RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *Journal of Biological Chemistry*, 277(48), 46779–46784.

Heim, R., Cubitt, A.B. & Tsien, R.Y., 1995. Improved green fluorescence. *Nature*, 373, 663–664.

Heim, R., Prashert, D.C. & Tsien, R.Y., 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 12501–12504.

Heim, R. & Tsien, R.Y., 1996. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology*, 6(2), 178–182.

Hink, M.A. et al., 2000. Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. *Journal of Biological Chemistry*, 275(23), 17556–17560.

Hintsche, M. & Klumpp, S., 2013. Dilution and the theoretical description of growth-rate dependent gene expression. *Journal of biological engineering*, 7(1), 22.

Hocine, S. et al., 2012. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nature Methods*, 10(2), 119–121.

Huang, B., 2007. Counting Low – Copy Number Proteins in a Single Cell. *Science* , 81,81–85.

Hulett, A.H.R. et al., 1969. Cell Sorting : Automated Separation of Mammalian Cells as a Function of Intracellular Fluorescence Cell Sorting : Automated Separation of Mammalian Cells as a Function of Intracellular Fluorescence. *Science*, 166(3906), 747–749.

Hurwitz, J., 2005. The discovery of RNA polymerase. *Journal of Biological Chemistry*, 280(52), 42477–42485.

Ingolia, N.T. et al., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, 324(5924), 218–223.

Ingolia, N.T., Lareau, L.F. & Weissman, J.S., 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), 789–802.

Jach, G. et al., 2001. Use of red fluorescent protein from Discosoma sp. (dsRED) as a reporter for plant gene expression. *Plant Journal*, 28(4), 483–491.

Johan Elf, Gene-Wei Li, X.S.X., 2007. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science*, 316(May), 1191–1195.

Ke, N. et al., 2016a. Visualization of periplasmic and cytoplasmic proteins with a self-labeling protein tag. *Journal of Bacteriology*, 198(7), 1035–1043.

Ke, N. et al., 2016b. Visualization of periplasmic and cytoplasmic proteins with a self-labeling protein tag. *Journal of Bacteriology*, 198(7), 1035–1043.

Keppler, A. et al., 2003. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature biotechnology*, 21(1), 86–9.

Keppler, A. et al., 2004. Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods*, 32(4),437–444.

Khlebnikov, A. et al., 2001. Homogeneous expression of the P BAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology*, 3241–3247.

Kohara, Y., Akiyama, K. & Isono, K., 1987. The physical map of the whole E. coli chromosome: Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, 50(3), 495–508.

Larson, D.R. et al., 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*, 332(6028), 475–478.

Lequin, R.M., 2005. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clinical Chemistry*, 51(12), 2415–2418.

Lequin, S., Chassagne, D. & Bellat, J.-P., 2003. Global analysis of protein localization in budding yeast . *Nature*, 425(6959), 686–691.

Li, G.W. et al., 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3), 624–635.

Li, G.W. & Xie, X.S., 2011. Central dogma at the single-molecule level in living cells. *Nature*, 475(7356), 308–315.

Link, A.J. & Phillips, D., 1997. Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type Escherichia coli : Application to Open Reading Frame Characterization. *Journal of Bacteriology* , 179(20), 6228–6237.

Liss, V. et al., 2016. Self-labelling enzymes as universal tags for fluorescence microscopy, super-resolution microscopy and electron microscopy. *Scientific Reports*, 5(1), 17740.

Los, G. V et al., 2008. HaloTag: A Novel Protein Labelling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology*, 3(6), 373–382.

Maamar, H., Raj, A. & Dubnau, D., 2007. Noise in gene expression determines cell fate in Bacillus subtilis. *Science (New York, N.Y.)*, 317(5837), 526–9.

Mahmood, T. & Yang, P.C., 2012. Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, 4(9), 429–434.

Matz, M. V et al., 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology*, 969–973.

McLafferty, F.W., 2008. Mass spectrometry across the sciences. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47), 18088–18089.

Merlin, C., McAteer, S. & Masters, M., 2002. Tools for Characterization of Escherichia coli Genes of Unknown Function. *Journal of Bacteriology*, 184(16), 4573–4581.

Murakami, K.S., 2015. Structural biology of bacterial RNA polymerase. *Biomolecules*, 5(2), 848–864.

Newmark, K.G. et al., 2005. Genetic analysis of the requirements for SOS induction by nalidixic acid in Escherichia coli. *Gene*, 356(1–2), 69–76.

Noll, H., 2008. The discovery of Polyribosomes. *BioEssays*, 30(11-12):1220-34

Okumus, B. et al., 2016. Mechanical slowing-down of cytoplasmic diffusion allows in vivo counting of proteins in individual cells. *Nature Communications*, 7, 11641.

Orm , M. et al., 1996. Crystal Structure of the Aequorea victoria Green Fluorescent Protein. *Science*, 273(5280), 1392–1395.

Paige, J.S., Wu, K. & Jaffrey, S.R., 2011. RNA mimics of green fluorescent protein. *Science*, 333(6042), 642–646.

Pedraza, J.M., van Oudenaarden, A. & Science, B., 2005. Noise propagation in gene networks. *Science*, 307(5717), 1965–1969.

Perfetto, S.P., Chattopadhyay, P.K. & Roederer, M., 2004. Innovation: Seventeen-colour flow cytometry: unravelling the immune system. *Nature Reviews Immunology*, 4(8), 648–655.

Pothoulakis, G. et al., 2013. The Spinach RNA aptamer as a characterisation tool for synthetic biology. *ACS Synthetic Biology*, 3(3), 130830130257008.

Prasher, D.C. et al., 1992. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111(2), 229–233.

Qin, D. & Fredrick, K., 2013. Analysis of polysomes from bacteria, *Methods Enzymol,* 530:159-72.

Raj, A. et al., 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 5(10), 877–879.

Reck-Peterson, S.L. et al., 2006. Single-Molecule Analysis of Dynein Processivity and Stepping Behavior. *Cell*, 126(2), 335–348.

Roman, L.J., Eggleston, a K. & Kowalczykowski, S.C., 1992. Processivity of the DNA helicase activity of Escherichia coli recBCD enzyme. *The Journal of biological chemistry*, 267(6), 4207–14.

Rotem, E. et al., 2010. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proceedings of the National Academy of Sciences*, 107(28), 12541–12546.

Saliba, A.E. et al., 2014. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Research*, 42(14), 8845–8860.

Schmeing, T.M. & Ramakrishnan, V., 2009. What recent ribosome structures have revealed about the mechanism of translation. *Nature*, 461(7268), 1234–1242.

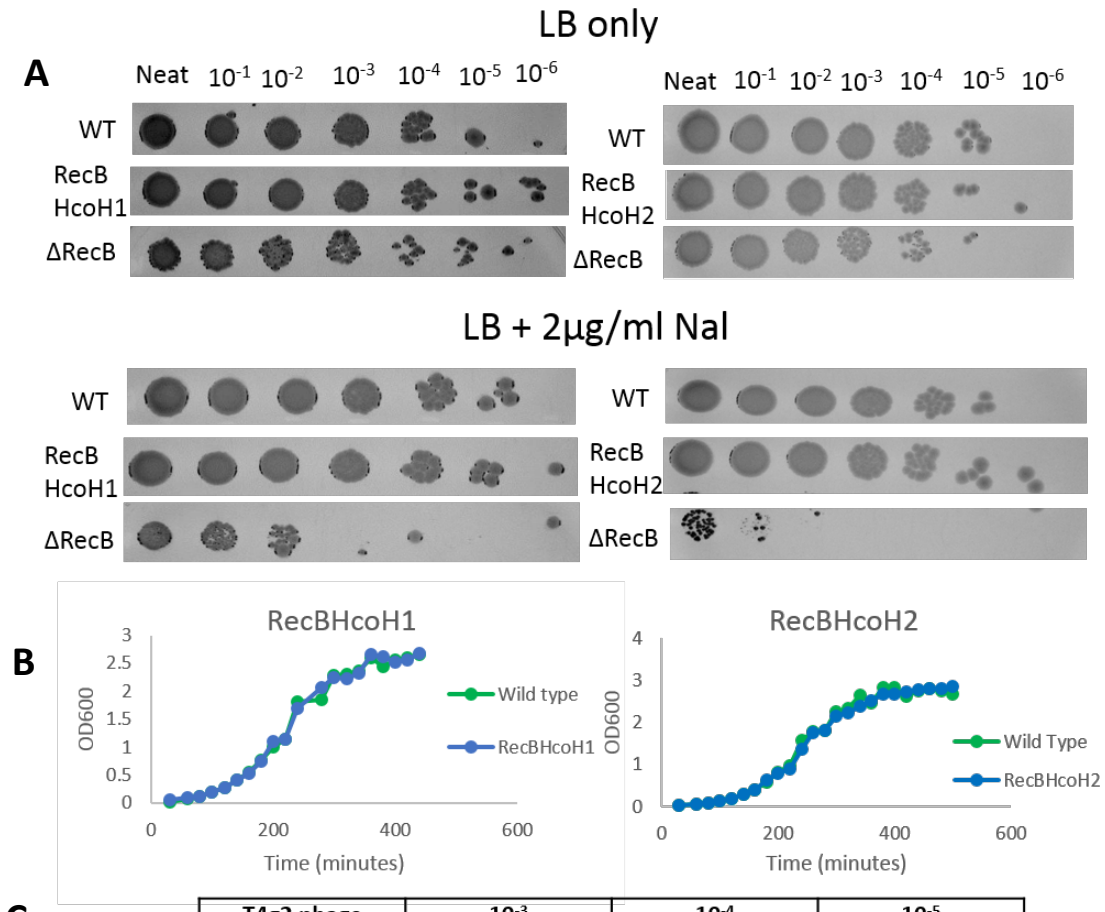
Schmidt, A. et al., 2015. The quantitative and condition-dependent Escherichia coli

proteome. *Nature biotechnology*, 34(1), 104–110.
Scott, M. et al., 2010. Interdependence of Cell Growth Origins and Consequences. *Science*, 330, 1099–1102.
Segala, M.G.P. et al., 2015. Fixation-resistant photoactivatable fluorescent proteins for correlative light and electron microscopy. *Nature Methods*, 12(3), 215–218.
Shaner, N.C. et al., 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology*, 22(12), 1567–1572.
Shaner, N.C. et al., 2008. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nature Methods*, 5(6), 545–551.
Shcherbo, D. et al., 2009. Far-red fluorescent tags for protein imaging in living tissues. *Biochemical Journal*, 418(3), 567–574.
Shi, D., Allewell, N.M. & Tuchman, M., 2015. The N-acetylglutamate synthase family: Structures, function and mechanisms. *International Journal of Molecular Sciences*, 16(6), 13004–13022.
Shimomura, O., Johnson, F.H. & Saiga, Y., 1962. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *Journal of Cellular and Comparative Physiology*, 59(3), 223–239.
Skinner, S.O. et al., 2013. Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nature protocols*, 8(6), 1100–13.
Smith, G.R., 2012. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiology and molecular biology reviews : MMBR*, 76(2),217–28.
So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L.A., et al., 2011. General properties of transcriptional time series in Escherichia coli. *Nature genetics*, 43(6), 554–60.
So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L. a, et al., 2011. General properties of transcriptional time series in Escherichia coli. *Nature genetics*, 43(6), 554–60.
Spies, M. et al., 2003. A molecular throttle: The recombination hotspot Chi controls DNA translocation by the RecBCD helicase. *Cell*, 114(5), 647–654.
Spies, M. & Kowalczykowski, S.C., 2005. Homologous Recombination by the RecBCD and RecF Pathways. In *The Bacterial Chromosome*. American Society of Microbiology, 389–403.
Srivastava, N. et al., 2009. Fully integrated microfluidic platform enabling automated phosphoprofiling of macrophage response. *Analytical Chemistry*, 81(9), 3261–3269.
Starosta, A.L. et al., 2014. The bacterial translation stress response. *FEMS microbiology reviews*, 38(6), 1172–1201.
Steitz, J.A., 1969. Nucleotide sequences of the ribosomal binding sites of bacteriophage R17 RNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 34, 621–630.
Stern-Ginossar, N. et al., 2012. Decoding Human Cytomegalovirus. *Science*,

338(6110), 1088–1093.
Strack, R.L. et al., 2008. A noncytotoxic DsRed variant for whole-cell labeling. *Nature Methods*, 5(11), 955–957.
Strack, R.L., Disney, M.D. & Jaffrey, S.R., 2013. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA. *Nature Methods*, 10(12), 1219–1224.
Taniguchi, Y. et al., 2010. Quantifying E. coli proteome and transcriptome with single- molecule sensitivity in single cells. *Science*, 329(5991), 533–538.
Taylor, A.F. et al., 1985. RecBC enzyme nicking at chi sites during DNA unwinding: Location and orientation-dependence of the cutting. *Cell*, 41(1), 153–163.
Taylor, A.F. & Smith, G.R., 1999. Regulation of homologous recombination: Chi inactivates RecBCD enzyme by disassembly of the three subunits. *Genes and Development*, 13(7), 890–900.
Taylor, a & Smith, G.R., 1980. Unwinding and rewinding of DNA by the RecBC enzyme. *Cell*, 22(2 Pt 2), 447–457.
Thattai, M. & van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), 8614–9.
Thompson, R.E., Larson, D.R. & Webb, W.W., 2002. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophysical Journal*, 82(5), 2775–2783.
Tyagi, S., 2009. Imaging intracellular RNA distribution and dynamics in living cells. *Nature Methods*, 6(5), 331–338.
Unrau, P.J. et al., 2014. RNA Mango aptamer-fluorophore : a bright , high affinity , complex for RNA labeling and tracking. *ACS Chemical Biology*, 9, 2412-2420
Vargas, D.Y. et al., 2005. Mechanism of mRNA transport in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 102(47), 17008–17013.
Vick, J.E. et al., 2011. Optimized compatible set of BioBrick??? vectors for metabolic pathway engineering. *Applied Microbiology and Biotechnology*, 92(6), 1275–1286.
Wade, J.T. et al., 2005. Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. *Genes and Development* , 2619–2630.
Wang, J. et al., 2015. RNA-seq based transcriptomic analysis of single bacterial cells. *Integr. Biol.*, 7(11), 1466–1476.
Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), 57–63.
White, A.K. et al., 2011. High-throughput microfluidic single-cell RT-qPCR. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), 13999–4004.
Willetts, N.S., Clark, A.J. & Low, B., 1969. Genetic location of certain mutations conferring recombination deficiency in Escherichia coli. *Journal of Bacteriology*,

97(1), 244–249.
Willetts, N.S. & Mount, D.W., 1969. Genetic Analysis of Recombination-Deficient Mutants of Exchericia coli K-12 Carrying rec Mutations Cotransducible with thyA. *Journal of Bacteriology*, 100(2), 923–934.
Wong, M.L. & Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. *BioTechniques*, 39(1), 75–85.
Xu, H. et al., 2015. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nature Methods*, 12(8), 739-744.
Yeh, E., Gustafson, K. & Boulianne, G.L., 1995. Green fluorescent protein as a vital marker and reporter of gene expression in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 92(15), 7036–40.
Zhang, J. et al., 2015. Tandem Spinach Array for mRNA Imaging in Living Bacterial Cells. *Scientific Reports*, 5, 17295.
Zhao, H. et al., 1998. Use of Green Fluorescent Protein To Assess Urease Gene Expression by Uropathogenic Proteus mirabilis during Experimental Ascending Urinary Tract Infection Use of Green Fluorescent Protein To Assess Urease Gene Expression by Uropathogenic Proteus mirabilis. *Infection and Immunity*, 66(1), 330–5.

Appendix



C

T4g2 phage dilution	10^{-3}	10^{-4}	10^{-5}
Wild type	133±13	24±7	8±6
RecBHcoH1	130±30	10±7	3±2
RecBHcoH2	137±54	18±6	3±0
ΔRecB	>3500	750±57	84±27

Figure A.1. Phenotypic tests for double HaloTag strains. A) Nalidixic acid assay for RecBHcoH1 and RecBHcoH2. Serial dilution of wild type, RecBHcoH1 and ΔRecB, and wild type, RecBHcoH2 and ΔRecB on 2µg/ml nalidixic acid and LB only plates. WT and RecBHalo cells show comparable viability on nalidixic acid while ΔRecB cells show reduced viability on nalidixic acid. B) Representative growth curves for RecBHcoH1 and RecBHcoH2 with wild type, no growth differences are apparent.C) Plaque formation following T4g2 phenotypic test for RecB function. WT and RecBHcoH1 and RecBHcoH2 strains show formation of very few plaques in comparison to the ΔRecB strain at each phage dilution.

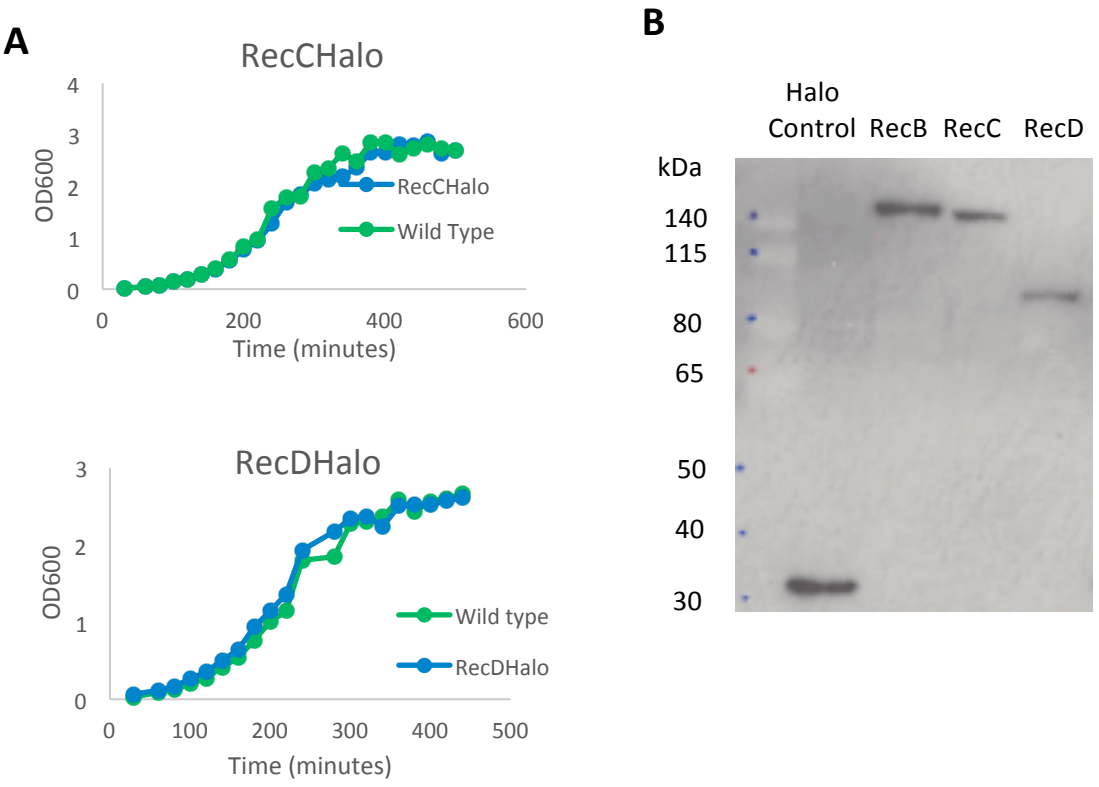


Figure A2. Western blot and growth curves for RecCHalo and RecDHalo. A) Representative growth curves for RecCHalo and RecDHalowith wild type, no growth differences are apparent. B) Western blot showing purified HaloTag protein in lane 1, the 168 kDa RecBHalo protein in lane 2, the 161kDa RecCHalo protein in lane 3 and the 101KDa protein in lane 4 (HaloTag= 34 kDa, RecB= 134 kDa)(Western blot performed by Lorna McLaren)