

**Newcastle**  
University

HETEROGENEITY IN *BACILLUS*  
*SUBTILIS*:  
GROWTH PHASE DEPENDENT  
ACTIVITY AND NOISE

Tom Ewen

Thesis submitted for the degree of  
Doctor of Philosophy

Institute for Cell and Molecular Biosciences  
Newcastle University

March 2017

## Abstract

Stochastic noise is a naturally occurring phenomena in all chemical reactions. The processes involved in gene expression are subject to the stochastic, random nature of chemical interactions. Single cell measurements of fluorescent activity have been used to deconstruct the sources of gene expression noise. These systems rely on the use of inducible gene expression from negatively regulated promoters. Gene expression noise is quantified by the fluorescent activity of the reporter. Single time-point assays have defined the contribution of transcription and translation as sources of gene expression noise.

This thesis investigated gene expression noise in *Bacillus subtilis*. It extends the single time-point assays and places noise within the context of the bacterial growth curve. Two main findings were concluded from the data. Firstly, there is growth phase dependent fluorescent activity, due to the accumulation of a stable fluorescent protein. Secondly, high noise levels in gene expression are a transcription dependent feature of cells in stationary phase.

Investigating the synthetic gene circuits responsible for these phenotypes highlight the importance of fully characterising the system. Differences in transcription, translation and fluorescent activity were observed in response to the architecture of the gene circuits.

## **Dedication**

This thesis is dedicated to my wife, Karen Ewen.

Of course.

## **Acknowledgements**

I would like to thank my supervisors Prof. Leendert Hamoen, Dr. Phillip Aldridge and Prof. Darren Wilkinson for their continued help and support. I was fortunate to experience three contrasting approaches to student supervision. Prof. Colin Harwood and Prof. Nikolay Zenkin challenged my approach, direction and methodology in our assessment meetings. They were instrumental in maintaining the focus required to produce this work.

I would also like to thank past and present members of the laboratory groups I have been a part of. It has been a pleasure to have Henrik, Pamela, Chris Sauer, Chris Birchall, Martin, Simon, Rita, Declan and Lauren to chat with and discuss all things scientific.

Contributions to this work are included in the main body of this thesis. In particular I would like to acknowledge Alex Henderson for the strains provided and Prof. Darren Wilkinson for the dynamic model.

This PhD was funded for three years by the Biotechnology and Biological Sciences Research Council (BBSRC). While writing up my results I have been fortunate to have additional financial support from the Institute of Cell and Molecular Bioscience at Newcastle and a number of family members: Nuala, Jono & Eileen, Nick & Linda, Aidan & Sal. Completing this work would have been extremely difficult without this support.

# Table of Contents

<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1. Heterogeneity in <i>Bacillus subtilis</i> .....	1
1.1.1. Regulation of motility .....	2
1.1.2. Heterogeneity in Biofilms.....	6
1.1.3. Genetic competence .....	9
1.1.4. Sporulation.....	11
1.2. Stochasticity in gene expression is a source of heterogeneity .....	13
1.2.1. Noise in gene expression .....	13
1.2.2. Measuring intrinsic and extrinsic noise .....	15
1.2.3. Bursting patterns of transcription and translation .....	16
1.2.4. Noise in transcription and translation .....	18
1.2.5. Noise in gene networks .....	19
1.3. Synthetic Biology tools and applications .....	21
1.3.1. Synthetic gene circuits .....	22
1.3.2. Biosensors .....	25
1.3.3. Fluorescent reporters.....	27
1.4. Aims .....	29
<b>Chapter 2. Materials and Methods.....</b>	<b>30</b>
2.1. Bacterial Growth Conditions.....	30
2.1.1. General growth conditions .....	30
2.1.2. Standard assay conditions .....	30
2.1.3. Preparation of chemically competent cells .....	30
2.1.4. Transformation by heat shock.....	31
2.1.5. Transformation into competent <i>B. subtilis</i> .....	31
2.1.6. Antibiotic concentrations .....	31

2.2.	Molecular Cloning.....	32
2.2.1.	Preparation of plasmid DNA .....	32
2.2.2.	Preparation of chromosomal DNA .....	32
2.2.3.	Polymerase chain reaction .....	32
2.2.4.	DNA agarose gel electrophoresis.....	32
2.2.5.	Gel extraction.....	33
2.2.6.	Digestion .....	33
2.2.7.	Ligation of DNA .....	33
2.2.8.	Gibson Assembly .....	33
2.2.9.	Screening.....	34
2.3.	Northern Blotting .....	34
2.3.1.	Generating single stranded RNA probes.....	34
2.3.2.	Standard RNA purification .....	34
2.3.3.	RNA stability purification.....	34
2.3.4.	RNA Agarose gel electrophoresis.....	35
2.3.5.	Blotting .....	35
2.3.6.	Hybridisation and visualisation.....	36
2.3.7.	mRNA quantification.....	36
2.4.	Protein Immunoblotting .....	37
2.4.1.	Preparation of whole cell lysate.....	37
2.4.2.	Protein gel electrophoresis.....	37
2.4.3.	Coomassie stain .....	37
2.4.4.	Membrane transfer .....	37
2.4.5.	Protein detection .....	38
2.5.	Fluorescence Microscopy.....	38
2.5.1.	Cell Profiler analysis of microscope images.....	38
2.5.2.	MicrobeTracker analysis.....	40

2.5.3. Quality control .....	41
2.6. Table of strains .....	42
2.7. Table of plasmids .....	44
2.8. Table of oligonucleotides .....	50
<b>Chapter 3. Quality control .....</b>	<b>55</b>
3.1. Introduction .....	55
3.2. Results and discussion.....	56
3.3. Conclusions .....	61
<b>Chapter 4. Growth phase dependent heterogeneity in the expression of a single gene 62</b>	
4.1. Introduction .....	62
4.2. Results and discussion.....	66
4.2.1. Growth phase dependent induction sfGFP* .....	66
4.2.2. Background fluorescence in <i>Bacillus subtilis</i> .....	71
4.2.3. Inducer concentration affects fluorescence levels and noise .....	75
4.2.4. Does protein stability affect fluorescence levels and noise in sfGFP* reporter strains? 77	
4.3. Conclusions .....	80
<b>Chapter 5. LacI regulation .....</b>	<b>82</b>
5.1. Introduction .....	82
5.2. Results and discussion.....	86
5.2.1. Transcript lengths determined by Northern Blotting .....	86
5.2.2. sfGFP* activity is dependent on the location of the repressor.....	87
5.2.3. Noise in gene expression is not determined by lacI location.....	90
5.2.4. Repressor location alters the efficiency of response in the system.....	91
5.2.5. $\Delta$ lacI used to investigate the growth phase dependent phenotype in sfGFP* ....	94
5.3. Conclusions .....	96
<b>Chapter 6. Transcription and translation effects in the reporter system .....</b>	<b>98</b>

6.1.	Introduction .....	98
6.1.1.	Transcription .....	98
6.1.2.	Translation .....	99
6.2.	Results .....	100
6.2.1.	Is transcription from $P_{\text{hyperspank}}$ responsible for the observed phenotypes?.....	100
6.2.2.	Translation of the five amino acid N-terminal linker, MEFLQ.....	103
6.2.3.	Codon optimisation of sfGFP .....	104
6.2.4.	Alternative reporters .....	107
6.3.	Conclusions .....	114
<b>Chapter 7. Dual gene reporter constructs.....</b>		<b>116</b>
7.1.	Introduction .....	116
7.2.	Results and discussion.....	118
7.2.1.	Does expression from a monocistronic or bicistronic operon alter the activity of sfGFP* or mCherry2?.....	118
7.2.2.	Does expression from a monocistronic or bicistronic operon alter the noise produced with sfGFP* or mCherry2 reporters?.....	120
7.3.	Conclusions .....	122
<b>Chapter 8. Correlating fluorescent activity with RNA and protein levels.....</b>		<b>123</b>
8.1.	Introduction .....	123
8.2.	Results and discussion.....	123
8.2.1.	Do transcription levels match fluorescent activity?.....	123
8.2.2.	Does the reporter gene alter the transcript stability? .....	126
8.2.3.	Do protein levels match the activity profile?.....	130
8.2.4.	Conditions resulting in mCherry2 heterogeneity .....	134
8.2.5.	Modelling sfGFP activity.....	136
8.3.	Conclusions .....	138
<b>Chapter 9. Discussion.....</b>		<b>140</b>
9.1.	Summary of findings .....	140



9.2.	Implications .....	141
9.3.	Future work .....	143
9.4.	Concluding remarks .....	143
<b>Chapter 10.</b>	<b>Appendix .....</b>	<b>146</b>
10.1.	Growth Media, Buffers and Solutions .....	146
10.1.1.	Lysogeny broth (LB) .....	146
10.1.2.	LB agar .....	146
10.1.3.	Spizizen minimal media (SMM) .....	146
10.1.4.	Minimal competence media (MM) .....	146
10.1.5.	Starvation media.....	147
10.1.6.	RNA loading dye.....	147
10.1.7.	10 X MOPS running buffer .....	147
10.1.8.	20 X Saline-sodium citrate buffer (SSC) .....	147
10.1.9.	Denaturation solution .....	148
10.1.10.	Neutralisation solution .....	148
10.1.11.	Malic acid .....	148
10.1.12.	Pre-hybridisation solution .....	148
10.1.13.	Hybridisation solution .....	149
10.1.14.	Blocking reagent .....	149
10.1.15.	Blocking buffer .....	149
10.1.16.	Wash solution 1 .....	149
10.1.17.	Wash solution 2.....	149
10.1.18.	Tween wash buffer .....	150
10.1.19.	Alkaline phosphatase buffer.....	150
10.1.20.	Phosphate buffered solution (PBS) .....	150
10.1.21.	Phosphate Milk Tween (PMT).....	150

10.1.22.	Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) buffer	151
10.1.23.	Tricine gel buffer.....	151
10.1.24.	10 X Anode running buffer .....	151
10.1.25.	10 X Cathode Running Buffer .....	151
10.1.26.	Transfer buffer.....	152
10.1.27.	DNA agarose gel (1%) .....	152
10.1.28.	RNA agarose gel (1.2%) .....	152
10.1.29.	Stacking acrylamide gel (3.96%) .....	152
10.1.30.	Separating acrylamide gel (12%) .....	153
10.2.	Supplementary figures .....	154
10.3.	Supplementary tables.....	159
<b>Chapter 11.</b>	<b>References .....</b>	<b>174</b>

## List of Figures

Figure 1. Heterogeneity in the formation and dispersal of a biofilm.....	2
Figure 2. Hysteresis in bistable systems .....	3
Figure 3. Motility in <i>B. subtilis</i> is controlled by a positive feedback loop producing a bistable phenotype.....	5
Figure 4. Heterogeneous signalling responses in biofilm formation.....	7
Figure 5. Regulation of biofilm gene expression.....	8
Figure 6. Two fluorescent reporter genes are used to quantify intrinsic noise .....	16
Figure 7. Fluorescent intensity of labelled mRNA corresponds to the number of mRNA molecules .....	17
Figure 8. Noise in gene expression results from a combination of transcription and translation.....	19
Figure 9. Propagation of noise can be investigated using a synthetic gene network.....	20
Figure 10. Noise propagation in gene networks .....	21
Figure 11. Logic gates are the simplest functional unit of digital circuits.....	23
Figure 12. Layered AND gates respond to four input signals with a binary on/off response	24
Figure 13. Synthetic logic gates designed to compute the edge between light and dark.....	25
Figure 14. Biosensor communication between cell lines.....	26
Figure 15. Fluorescent cells illuminate the surrounding agarose .....	56
Figure 16. Limitations of the image analysis software Cell Profile .....	58
Figure 17. R script used to identify and omit outlying microscope data from the final analysis .....	59
Figure 18. Outlying data is omitted using a quality control pipeline.....	60
Figure 19. Schematic drawings of the constructs used to quantify noise in gene expression	63
Figure 20. Phenotypic noise isolated from varied transcriptional efficiency and varied translational efficiency.....	65
Figure 21. Phase contrast and fluorescent microscope images of the sfGFP* strain induced with 1 mM IPTG in a standard assay.....	67
Figure 22. Growth phase effects fluorescence and heterogeneity in the expression of a single gene.....	70
Figure 23. Background fluorescence and noise in the $\Delta sfGFP$ control strain .....	72
Figure 24. Fluorescence and noise in the sfGFP* reporter strain .....	74

Figure 25. Transcription rate alters the fluorescence and noise in the <i>sfGFP*</i> reporter strain	76
Figure 26. Protein stability alters <i>sfGFP*</i> activity and noise.....	79
Figure 27. Tetrameric LacI binding to operator sites in $P_{\text{hyperspank}}$ regulates transcription .....	83
Figure 28. Schematic drawing of the constructs used to investigate repressor location effects	85
Figure 29. Co-transcription of reporter/repressor is a feature of the <i>lacI</i> strain .....	88
Figure 30. Regulator location affects <i>sfGFP*</i> activity.....	89
Figure 31. Regulator location makes little difference to the noise in gene expression .....	91
Figure 32. <i>lacI</i> location alters the response to varied inducer concentrations .....	93
Figure 33. <i>lacI</i> location and growth phase alter the system sensitivity to inducer concentrations .....	95
Figure 34. Growth phase dependent <i>sfGFP*</i> activity is not caused by LacI regulation.....	96
Figure 35. Growth phase dependent <i>sfGFP*</i> activity is a feature of the $P_{\text{hyperspank}}$ and $P_{\text{xyI}}$ promoters while noise is restricted to $P_{\text{hyperspank}}$ .....	101
Figure 36. The N-terminal translational linker present on <i>sfGFP*</i> is not responsible for the fluorescent profile in $P_{\text{hyperspank}}\text{-sfGFP}$ strains .....	104
Figure 37. Predicted translation speed in <i>sfGFP</i> variants.....	106
Figure 38. Codon optimisation of <i>sfGFP</i> increases the fluorescent output of the gene without altering the growth phase dependent activity.....	107
Figure 39. Growth phase dependent fluorescence does not occur with an mCherry2 reporter .....	110
Figure 40. Growth phase dependent differences in <i>sfGFP*</i> and mCherry2 fluorescence ....	111
Figure 41. An artefact generates high noise with an mCherry2 reporter when compared to <i>sfGFP*</i> .....	113
Figure 42. Monocistronic and bicistronic reporter constructs .....	117
Figure 43. Gene order in bicistronic operons alters the fluorescent output of the system....	119
Figure 44. Gene order in bicistronic operons generates little difference to the noise profiles .....	121
Figure 45. Transcript variation in $P_{\text{hyperspank}}$ strains .....	125
Figure 46. mRNA degradation in the single gene <i>sfGFP*</i> and <i>mCherry2</i> strains.....	128
Figure 47. mRNA degradation in the bicistronic <i>sfGFP*-mCherry2</i> and <i>mCherry2-sfGFP*</i> reporter strains .....	129

Figure 48. Protein levels and fluorescent activity are monitored concurrently with growth in the single reporter gene <i>sfGFP*</i> and <i>mCherry2</i> constructs .....	131
Figure 49. Comparison of protein levels and fluorescent activity in the bicistronic <i>sfGFP*</i> - <i>mCherry2</i> and <i>mCherry2-sfGFP*</i> reporter constructs.....	133
Figure 50. Microscope images of the <i>sfGFP*</i> and <i>mCherry2</i> strains after 24 hours growth	135
Figure 51. Fluorescence in the <i>sfGFP*</i> and <i>mCherry2</i> strains after 24 hours growth.....	136
Figure 52. Deterministic dynamic model of growth, gene expression and dilution.....	137
Figure 53. Predicted secondary structures in <i>sfGFP*</i> and <i>mCherry</i> transcripts .....	145
Figure 54. Preliminary data establishing IPTG induction parameters for the expression of <i>sfGFP*</i> .....	154
Figure 55. Transcript variation in the single gene <i>sfGFP*</i> strain .....	155
Figure 56. Transcript variation in the single gene <i>mCherry2</i> strain .....	156
Figure 57. Transcript variation in the bicistronic <i>sfGFP*-mCherry2</i> strain.....	157
Figure 58. Transcript variation in the bicistronic <i>mCherry2-sfGFP*</i> strain.....	158

## List of Tables

Table 1. Antibiotic concentrations used with <i>B. subtilis</i> and <i>E. coli</i> .	31
Table 2. Strains used in this thesis	43
Table 3. Plasmids used in this thesis	49
Table 4. Oligonucleotides used in this thesis	54
Table 5. Parameters used in MicrobeTracker image analysis	159
Table 6. Relative sfGFP* fluorescence in the auto-regulating lacI strain, induced with varied IPTG concentrations	160
Table 7. Relative sfGFP* fluorescence in the negatively regulated lacI-cis strain, induced with varied IPTG concentrations	160
Table 8. Relative sfGFP* fluorescence in the negatively regulated lacI-trans strain, induced with varied IPTG concentrations	160
Table 9. Cell numbers corresponding to the data presented in Figure 22, page 70.	160
Table 10. Cell numbers corresponding to the data presented in Figure 23, page 72.	161
Table 11. Cell numbers corresponding to the data presented in Figure 24, page 74.	161
Table 12. Cell numbers corresponding to the data presented in Figure 25, page 76.	161
Table 13. Cell numbers corresponding to the data presented in Figure 26, page 79.	162
Table 14. Cell numbers corresponding to the data presented in Figure 18 <b>Error! Reference source not found.</b> , page 60.	162
Table 15. Cell numbers corresponding to the data presented in Figure 30, page 89 and Figure 31, page 91.	162
Table 16. Cell numbers corresponding to the data presented in Figure 32, page 93 and Figure 33, page 95.	163
Table 17. Cell numbers corresponding to the data presented in Figure 34, page 96.	163
Table 18. Cell numbers corresponding to the data presented in Figure 35, page 101.	164
Table 19. Cell numbers corresponding to the data presented in Figure 36, page 104.	164
Table 20. Cell numbers corresponding to the data presented in Figure 38, page 107.	164
Table 21. Cell numbers corresponding to the data presented in Figure 39, page 110 and Figure 41, page 113.	165
Table 22. Cell numbers corresponding to the data presented in Figure 43, page 119 and Figure 44, page 121.	165
Table 23. Cell numbers corresponding to the data presented in Figure 48, page 131.	165

Table 24. Cell numbers corresponding to the data presented in Figure 49, page 133.....	166
Table 25. Cell numbers corresponding to the data presented in Figure 55, page 155.....	166
Table 26. Cell numbers corresponding to the data presented in Figure 56, page 156.....	166
Table 27. Cell numbers corresponding to the data presented in Figure 57, page 157.....	166
Table 28. Cell numbers corresponding to the data presented in Figure 58, page 158.....	167
Table 29. Comparison of fluorescence in LacI regulated strains with full (1 mM) IPTG induction .....	167
Table 30. Comparison of fluorescence in LacI regulated strains without IPTG induction ..	168
Table 31. Welch two-sample t-test comparing sfGFP* activity in the negative auto- regulating lacI strain under varied induction conditions.....	169
Table 32. Welch two-sample t-test comparing sfGFP* activity in the negatively regulated lacI-cis strain under varied induction conditions .....	170
Table 33. Welch two-sample t-test comparing sfGFP* activity in the negatively regulated lacI-trans strain under varied induction conditions.....	171
Table 34. Fluorescent output of the sfGFP* reporter is reduced by the N-terminal translational linker coding for MEFLQ .....	171
Table 35. Fluorescent output of the sfGFP reporter is increased by codon optimisation.....	172
Table 36. Paired t-test comparing the relative mCherry mRNA levels during different growth phases.....	172
Table 37. mRNA half-lives of strains in exponential growth and in stationary phase .....	173

## Abbreviations

<b>APS</b>	Ammonium persulfate
<b>BA</b>	Benzyl alcohol anhydrous
<b>bp</b>	Base pairs
<b>CAPS</b>	N-cyclohexyl-3-aminopropanesulfonic acid
<b>CCD</b>	Charge-coupled device (digital camera)
<b>CFP</b>	Cyan fluorescent protein
<b>CSV</b>	Comma separated value file
<b>DAPI</b>	4',6-Diamidino-2-phenylindole dihydrochloride
<b>dH<sub>2</sub>O</b>	Deionised H <sub>2</sub> O
<b>DMSO</b>	Dimethyl sulfoxide
<b>DNA</b>	Deoxyribonucleic Acid
<b>dNTPs</b>	Deoxyribonucleotide mix of bases A, T, G, C
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>GFP</b>	Green fluorescent protein
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>LB</b>	Luria Bertani media
<b>MOPS</b>	4-Morpholinepropanesulfonic acid
<b>mRNA</b>	Messenger ribonucleic acid
<b>OD</b>	Optical Density at 600 nm
<b>ORF</b>	Open reading frame
<b>PCR</b>	Polymerase chain reaction
<b>ROI</b>	Region of interest
<b>RNA</b>	Ribonucleic Acid
<b>rRNA</b>	Ribosomal ribonucleic acid
<b>sfGFP</b>	Super-folding green fluorescent protein
<b>TAE</b>	Tris Acetate-EDTA buffer
<b>TEMED</b>	Tetramethylethylenediamine
<b>tRNA</b>	Transfer ribonucleic acid
<b>QC</b>	Quality control
<b>qPCR</b>	Quantitative polymerase chain reaction
<b>wt</b>	Wild-type



**YFP**

Yellow fluorescent protein

# Chapter 1. Introduction

## 1.1. Heterogeneity in *Bacillus subtilis*

*B. subtilis* is a model organism in which to study the physiology of bacteria (Kunst *et al.*, 1997). This Gram-positive bacterium is a non-pathogenic, soil dwelling species exhibiting distinct morphological features (Bais *et al.*, 2004; Kearns *et al.*, 2005; Schultz *et al.*, 2009; McKenney *et al.*, 2013; Belas, 2014). An isogenic population of *B. subtilis* can contain distinct subpopulations, defined by differences in morphology, physiology or a combination of the two. Heterogeneous phenotypes of *B. subtilis* include motility, production of biofilm, genetic competence, cannibalism and sporulation (Errington, 2003; Gonzalez-Pastor *et al.*, 2003; Smits *et al.*, 2005; Mordini *et al.*, 2013; Winkelman *et al.*, 2013). During exponential growth a distinct group of cells within the population can become motile. Motile cells synthesise multiple flagellar, whip-like filaments rotating to propel *B. subtilis* into new environmental niches (Ito *et al.*, 2005; Moriya *et al.*, 2006). The motile phenotype can be lost in response to its environment where sessile cells are then able to colonise an area from within a biofilm (Figure 1)(Ababneh and Herman, 2015). There is versatility in the physiological behaviour of cells within a biofilm. Only a subpopulation of cells produce the matrix forming the biofilm (Lopez *et al.*, 2009a). Over time, changing conditions trigger a minority of cells to become genetically competent (Mirouze *et al.*, 2012). Competence enables this subpopulation to actively transport exogenous DNA from the surrounding area into the cell (Dubnau, 1991). DNA is then integrated into the chromosome of the competent cell. To manage increasingly unfavourable conditions, *B. subtilis* can initiate differentiation into an endospore (hereafter referred to a spore). Sporulation is a bimodal process resulting in cells encased within a protective exterior (McKenney *et al.*, 2013; Siebring *et al.*, 2014). The protection of a spore enables the bacterium to remain metabolically dormant until more favourable conditions are sensed (Chen *et al.*, 2014). Investment of the resources required for sporulation can be postponed through cannibalism from within the bacterial community (Hofler *et al.*, 2016).

Importantly, these distinct phenotypes occur within a genetically identical community of bacteria. The physiology underpinning cell-fate decisions in *B. subtilis* is also distinct. Heterogeneity is engineered through temporal responses to stimuli, paracrine signalling and stochastic responses from within bistable systems (Lopez *et al.*, 2009b; Norman *et al.*, 2013; Gamba *et al.*, 2015). The mechanistic diversity of *B. subtilis* responses to “identical”

environmental conditions are a pre-requisite for its ability to respond in a heterogeneous manner. Consideration of these mechanisms are therefore important within the context of heterogeneity.

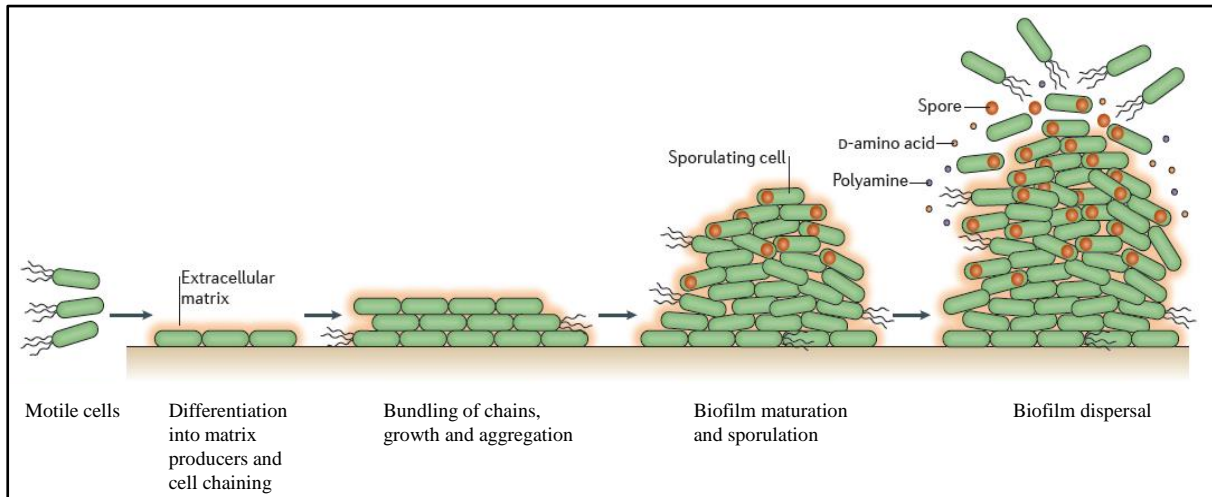


Figure 1. Heterogeneity in the formation and dispersal of a biofilm. The isogenic population of *B. subtilis* has phenotypically distinct subpopulations within the community. Adapted from Vlamakis *et al.*, 2013.

### 1.1.1. Regulation of motility

Motility in *B. subtilis* requires assembly of multiple flagella. The flagellum uses proton motive force to drive rotation of an articulated hook attached to a long, whip-like filament (Shioi *et al.*, 1980). Rotation of multiple flagella propel the bacterium through its environment. This differentiation strategy is only adopted by a subpopulation of the community (Nishihara and Freese, 1975). This form of diversification allows non motile cells the protection of a growing colony, while motile cells are able to sample new environments by swimming through a liquid or swarming across solids surfaces. A switch is required to produce the mixed community where some cells are motile and others are not (Norman *et al.*, 2013). For motile cells to benefit from a new environmental niche they must be able to stay there. It is therefore necessary that the motile phenotype can be switched off. Regulation of motility in *B. subtilis* is an example of epigenetic bistability. Bistable systems can exist in two steady states, “ON” or “OFF” (Ferrell, 2002). It is not possible to maintain the intermediate state between “ON” and “OFF” for extended periods of time. In addition to this the system must exhibit some level of hysteresis (Figure 2). This is defined by a difference in energy input required to transition between states, dependent on the starting state of the system. The stimulus/response curve from the “OFF” to “ON” position is

different from the stimulus/response curve from “ON” to “OFF”. The phenotype is epigenetic as it is passed on to progeny cells without altering the chromosomal DNA of the bacterium. The motile phenotype in *B. subtilis* requires activation of a positive feedback loop (Rao *et al.*, 2004). In this context positive feedback of a master regulator protein leads to an increase in the expression of its own gene (Mordini *et al.*, 2013). Levels of the protein will naturally fluctuate due to the stochastic nature of gene expression (Wilkinson, 2009). The overall

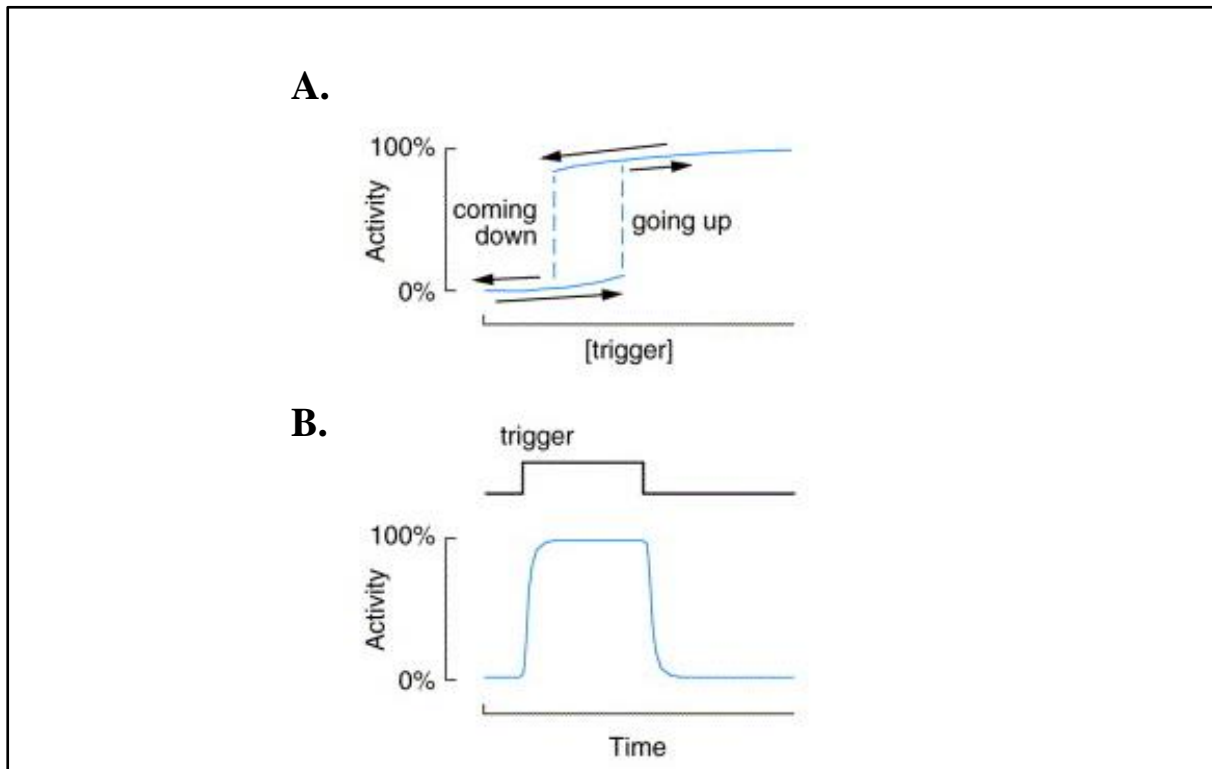


Figure 2. Hysteresis in bistable systems. **A**, a difference in stimulus/response curves is the defining feature of hysteresis. Moving from the “OFF” to “ON” state requires a different energy input from the “ON” to “OFF” state. **B**, it is not possible for a bistable system to exist for extended time in intermediate state between “OFF” and “ON”. Adapted from Ferrell (2002).

levels remain low until a threshold is reached. On reaching the threshold the master regulator will continue to upregulate its own expression until maximum output is reached. In this way positive feedback prevents maintenance of an intermediate state. The system is either “ON” or “OFF”.

SwrA is the putative master regular of flagellar biosynthesis in *B. subtilis* (Patrick and Kearns, 2009; Patrick and Kearns, 2012; Mordini *et al.*, 2013). The positive feedback loop occurs between  $\sigma^D$  and SwrA (Figure 3). SwrA interacts with DegU~P, co-operatively

binding to the  $\sigma^A$  dependent promoter of the *fla/che* operon (Mordini *et al.*, 2013). The *fla/che* operon is comprised of 32 genes, required for the synthesis of the hook and basal body of the flagellum (Mukherjee and Kearns, 2014). The penultimate gene in this operon is *sigD* coding for  $\sigma^D$ , the alternative motility sigma factor.  $\sigma^D$  completes the positive feedback loop by its action on the  $\sigma^D$  dependent promoter of the *swrA* operon (Estacio *et al.*, 1998). In the simplest sense the bistable switch is turned “ON” by SwrA-DegU~P and “OFF” by SlrA-SinR (Figure 3 A). The SlrA-SinR heterodimer inhibits expression of the entire operon, reducing levels of  $\sigma^D$  in the system. In addition to this, the complex acts downstream of *sigD* repressing *hag*, the gene coding for flagellin (Cozy *et al.*, 2012). DegU~P is both positive and negative in its regulation. High levels of DegU~P repress the  $\sigma^A$  promoter of the *fla/che* operon (Murray *et al.*, 2009). This mechanism allows DegU~P to fine tune the conditions where the switch is turned “ON”. Switching the system “OFF” requires downregulation of the *fla/che* operon.

Motility in *B. subtilis* demonstrates the use of stochasticity in generating distinct phenotypes. Fluctuating levels of the regulatory components are, in a subpopulation, able to engage a positive feedback loop driving differentiation. This form of bistability is also a heritable trait. Progeny cells have been shown to inherit the motility phenotype of the mother cell (Cozy and Kearns, 2010). This phenotype was observable through mutations to the autolysins responsible for cell separation. The mutations prevented mother/progeny cell separation, producing a “chain locked” phenotype. Motile and non-motile cells remained physically held together in growing chains of bacteria. Motile cells were held in position within the chain and selectively labelled with GFP. *gfp* was placed behind  $P_{hag}$  to produce green cells where flagella were fully synthesised. The motility phenotype in the growing chain of cells demonstrated the epigenetic nature of the bistable switch. In one chain there was a series of green motile cells, followed by a series of dark, non-motile cells, followed by a series of green motile cells. The bistable switch in this growing chain has been fixed in the “ON” position for multiple generations. In one progeny cell the switch has then moved to the “OFF” position and been held there for multiple generations, resulting in the dark phenotype. Finally the switch is changed to the “ON” position again and expression via  $P_{hag}$  produces green cells. The epigenetic motility phenotype shows hysteresis in the bistable switch. The positive feedback loop is more likely to remain fixed in one state, rather than continually switch between states. This is due to the difference in energy required to move the switch

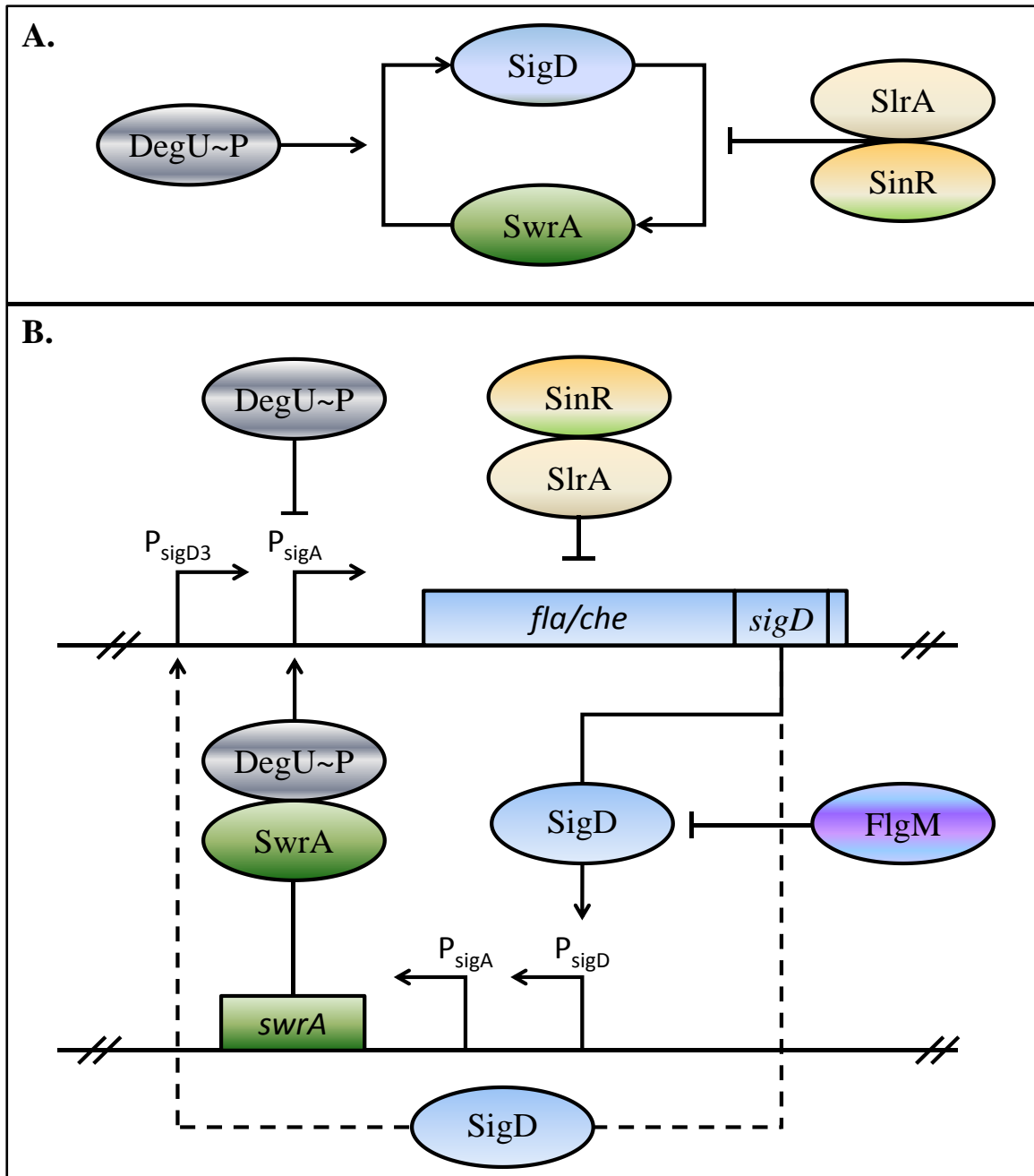



Figure 3. Motility in *B. subtilis* is controlled by a positive feedback loop producing a bistable phenotype. Genes/operons are depicted by coloured rectangles and proteins by coloured ovals. Promoters are indicated with bent arrows. The activity of proteins on the promoters are depicted by arrows for activation and T-bars for repression. Weaker activation is indicated by dashed arrows. **A**, simple schematic depicting the positive feedback between SwrA and sigma factor D. **B**, Expression of the flagellar gene operon *fla/che* is activated by the DegU~P-SwrA complex on  $P_{sigA}$  (Mordini *et al.*, 2013). *swrA* expression occurs predominantly through  $\sigma^D$  action on  $P_{sigD}$  with little activation through  $P_{sigA}$  (Calvio *et al.*, 2008). The inhibition of  $\sigma^D$  by FlgM is removed on completion of the hook/basal body of the flagellum (Caramori *et al.*, 1996). The action of  $\sigma^D$  on the weaker  $P_{sigD3}$  is not sufficient to initiate a state switch. In the absence of SwrA,  $\sigma^D$  is able to maintain flagellar gene expression in the “ON” state (Cozy *et al.*, 2012).

from “OFF” to “ON”. Switching “ON” requires different chemical interactions than switching the system “OFF”.

### **1.1.2. Heterogeneity in Biofilms**

*B. subtilis* is able to produce biofilms, supporting the colonisation of the bacterial community (Bais *et al.*, 2004). Biofilms protect cells within it against physical interference and biological or chemical assaults (Morikawa, 2006). They are formed at the interface between different mediums: liquid-air, liquid-solid and in the laboratory, agar-air (Branda *et al.*, 2001; Wilking *et al.*, 2011; Hollenbeck *et al.*, 2016). Formation requires expression of genes leading to production of an exopolymeric matrix. This extracellular matrix comprises peptides, lipopolysaccharides, lipids, glycolipids and nucleic acid (Marvasi *et al.*, 2010). The biofilm originates where motile cells contact with a surface and differentiate into non-motile, matrix producing cells (Figure 1). Cleavage of peptidoglycan between parent and progeny halts, resulting in chain growth and aggregation within the matrix. The biofilm matures with dynamic cell differentiation into: motile, non-motile, matrix producing, competent and sporulating phenotypes. Dispersal of the matrix occurs through the action of excreted D-amino acids (Kolodkin-Gal *et al.*, 2010). 

The formation of a biofilm requires both quorum sensing and paracrine signalling. Pheromones are exported from the cell, generating a unidirectional response and heterogeneous phenotypes (Figure 4). All cells within the growing community export ComX. However, only a subpopulation respond to this signal and produce surfactin (Lopez *et al.*, 2009a). Extracellular ComX is bound by the sensor kinase ComP. In turn ComP phosphorylates ComA, the master regulator of surfactin biosynthesis. The role of surfactin is the defining characteristic of paracrine signalling. Those cells exporting surfactin do not respond to it themselves. Cells responding to surfactin produce the matrix and do not produce surfactin. The proposed mechanism for this lack of surfactin production is an interference of matrix with ComX. The downstream activation of ComP is not possible, stabilising a heterogeneous response to ComX (Lopez *et al.*, 2009b). In cells responding to surfactin Spo0A~P increases. The sensor kinase KinC phosphorylates Spo0A (Spo0A~P) at reduced intracellular potassium concentrations. The phosphorylation state of Spo0A determines the expression profile of over 100 genes, with intermediate levels corresponding to matrix production (Vlamakis *et al.*, 2013). Both KinC and KinB respond to intracellular potassium concentration. Increased potassium levels result in sliding motility. The balance

between potassium levels, KinB and KinC regulates the temporal activation of sliding motility/biofilm formation via the phosphorylation state of SPO0A and SPO0F (Grau *et al.*, 2015; Narula *et al.*, 2015). Spo0A~P activates expression of *sinI*. SinI interacts with the master regulator of biofilm formation SinR (Chai *et al.*, 2009). The SinI-SinR interaction

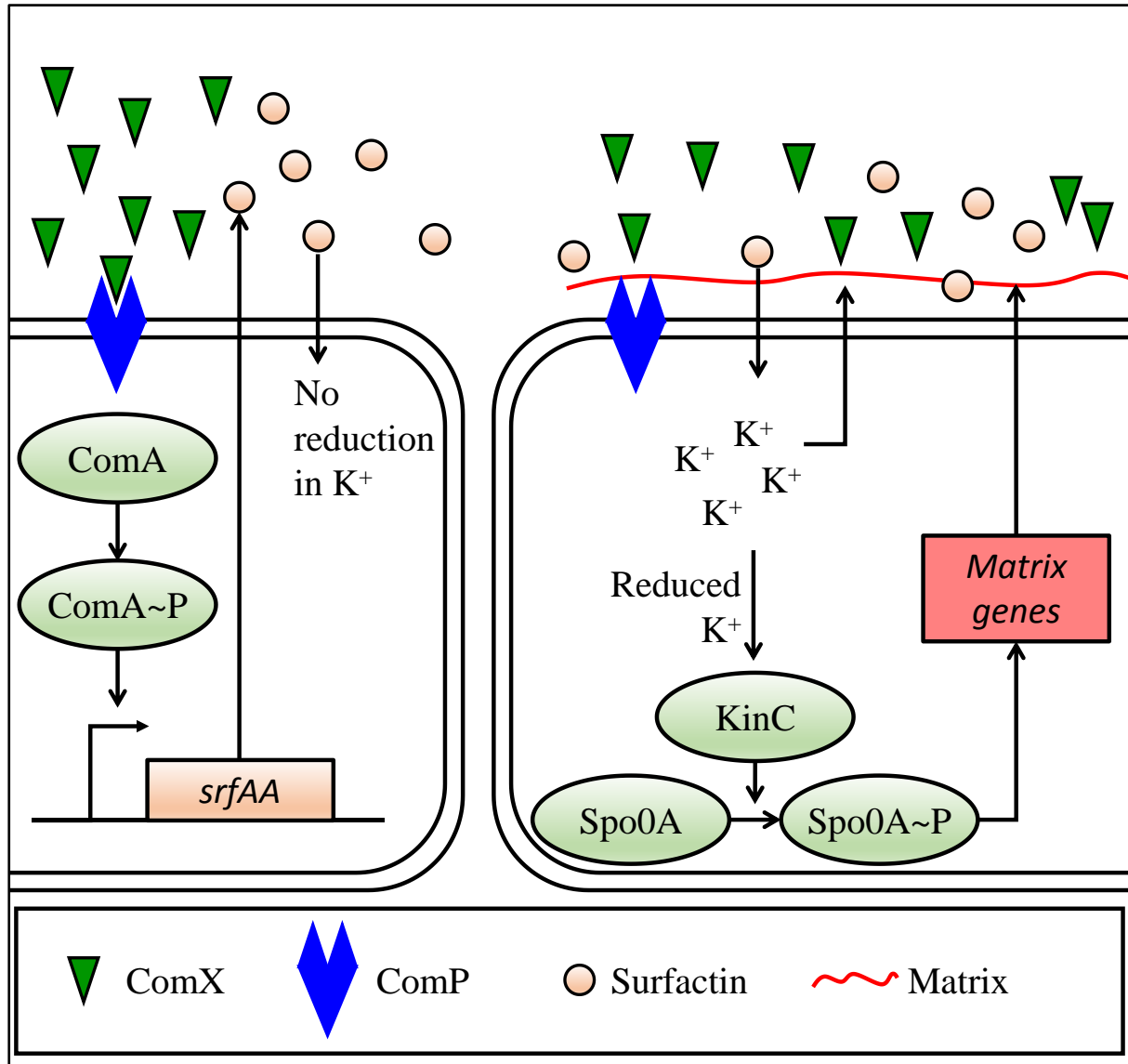


Figure 4. Heterogeneous signalling responses in biofilm formation. *B. subtilis* cells exhibit different responses to the extracellular signalling molecules ComX and Surfactin (Lopez *et al.*, 2009b). In one subpopulation ComX is able to bind to the kinase ComP, triggering surfactin production. Cells producing surfactin do not themselves respond to it. In the second subpopulation Surfactin leads to potassium leakage and matrix production via the kinase KinC phosphorylation of Spo0A. Matrix producing cells do not respond to extracellular ComX.

prevents the SinR negative regulation of biofilm genes (Figure 5). In the absence of SinI, SinR represses expression of the two operons responsible for matrix components, *epsA-O* and



*tap-sipW-tasA* and also represses transcription of *slrR*. SlrR forms part of a negative feedback loop functioning to switch matrix gene production on and off. The negative feedback loop has two positive switches mediated by SinI-SinR and SlrA-SinR protein-protein interactions. In addition to the upregulation of matrix genes these interactions allow expression of SlrR leading to the SlrR-SlrA protein-protein interaction. The SlrR-SlrA complex reduces the repressive action of SlrA on the SinR, the master regulator (Chai *et al.*, 2009). Whilst *sinI* is positively regulated by Spo0A~P, *slrA* is negatively regulated by the action of YwcC. The triggers responsible for altering YwcC repression of SlrA are not fully understood. YwcC is known to repress transcription of *swrA* and its own gene expression (Chai *et al.*, 2012).

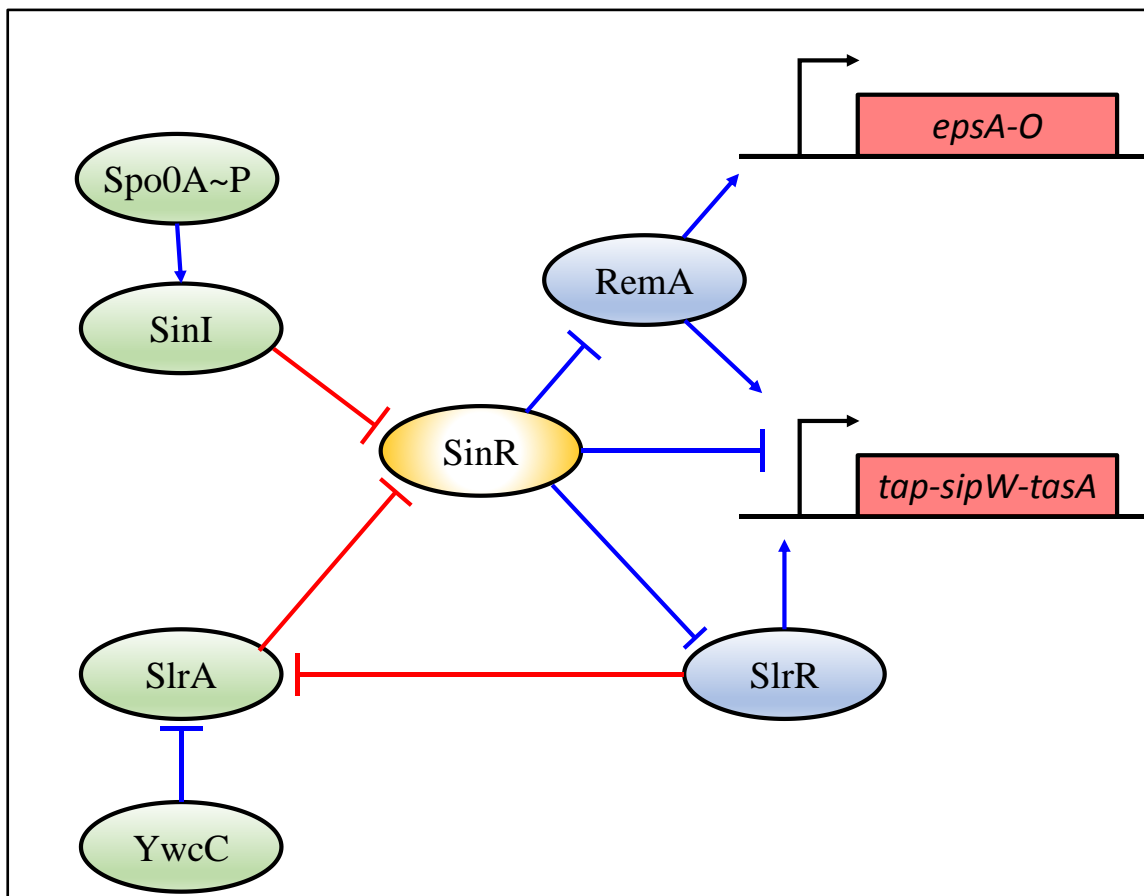


Figure 5. Regulation of biofilm gene expression. SinR is the master regulator of matrix genes in the two operons *epsA-O* and *tap-sipW-tasA*. SinR functions as an anti-activator through inhibition of the DNA binding protein RemA (Winkelman *et al.*, 2013). The Spo0A~P pathway primarily activates matrix production. Activation through YwcC is triggered through an unknown signal. Arrows indicate activation, T-bars denote repression. Blue arrows/T-bars, activation and repression occur at the level of transcription. Red T-bars, repression is through protein-protein interactions. Adapted from Chai *et al.*, 2009.

Biofilm formation requires cells in the community to be non motile. The SinR-SlrA heteromer prevents SinR repression of matrix production. In addition to this the SinR-SlrA heteromer downregulates motility through repression of  $\sigma^D$  (Chai *et al.*, 2010). Downregulation of flagellar genes will prevent synthesis of new flagella. To reduce motility in matrix producing cells a molecular clutch is engaged. On derepression of the *epsA-O* operon the bi-functional glycosyltransferase EpsE interacts with the flagellar motor switch protein, FliG to reduce motility. The effects of EpsE-FliG interaction was demonstrated by placing *epsE* under the control of an IPTG inducible promoter (Blair *et al.*, 2008). Cells were tethered by a single flagellum and rotation around the flagellum timed in EpsE +/- cells. In cells expressing *epsE* rotation was reduced but not abolished compared to cells without EpsE. This is consistent with a model where EpsE acts as a molecular clutch. The transfer of power into rotation is removed without physically immobilising the flagellum. This regulatory system acts as a two component switch to turn motility “OFF” in matrix producing cells. Motility gene expression is reduced and existing flagella rotation is disengaged in the matrix “ON” population.

### **1.1.3. Genetic competence**

Genetic competence is a finely tuned bistable phenotype. It exists within a system of global regulation functioning to generate multiple, distinct phenotypes. To become competent, cells in the community must generate a heterogeneous response to the same environmental cues. Competence provides *B. subtilis* with the ability to integrate exogenous DNA into the chromosome. This bistable phenotype is a growth phase dependent response undertaken by a sub-population of the isogenic community (Dubnau, 1991; Smits *et al.*, 2005). During this process DNA replication, cell elongation and division is arrested until exogenous DNA has been successfully integrated. This phenotype ultimately leads to cell death where no successful integration takes place (Haijema *et al.*, 2001). Successful DNA integration may provide no selective advantage, or may be detrimental to the cell. This differentiation strategy has significant risk, both in terms of the input of resources and the likelihood of an advantageous outcome. To maintain fitness of the community there must be tight control of this risky differentiation strategy.

As with biofilm development, cells destined for competence respond to the extracellular signalling molecule ComX (Magnuson *et al.*, 1994). As the density of the population increases, so too does the ComX signal. This causes phosphorylation of ComA via

membrane bound ComP, upregulating expression of the small molecule ComS (Dsouza *et al.*, 1993). ComS competitively binds MecA in the MecA-ClpC-ClpP complex which degrades both ComS and the master regulator of competence ComK (Vansinderen and Venema, 1994; Ogura *et al.*, 1999). Increase in population density increases the levels of ComS thus reducing the degradation of ComK (Turgay *et al.*, 1998). ComK is both the master regulator of competence and a positive regulator of its own expression (Smits *et al.*, 2005). ComK is the only transcriptional regulator required to engage the positive feedback loop. Above threshold levels ComK will continue to upregulate its own expression. This fixes the bistable switch in the competence “ON” position (Smits *et al.*, 2005).

Maintaining the competence “OFF” state requires tight control of *comK* expression. The MecA-ClpC-ClpP complex continually removes ComK from the system (Hamoen *et al.*, 2003b). In addition to this the transcriptional regulators AbrB, CodY and Rok function to down-regulate expression of *comK*, maintaining low ComK levels in exponential phase. (Schultz *et al.*, 2009) AbrB is a transition state regulator repressing *comK* expression during exponential growth (Hamoen *et al.*, 2003a). CodY represses stationary phase genes where nutrients are in abundance, with derepression occurring in nutrient limited conditions (Ratnayake-Lecamwasam *et al.*, 2001). Rok represses the transcription of *comK* directly (Hoa *et al.*, 2002). Rok is repressed by SinR and AbrB (Hoa *et al.*, 2002). These growth phase dependent interactions maintain low basal levels of ComK in exponential growth and allow a transient increase during stationary phase. The temporal expression dynamics in stationary phase increase ComK levels closer to the threshold where noise in gene expression can switch “ON” the auto-stimulatory loop (Suel *et al.*, 2006). Kre (previously known as YkyB) has recently been found to reduce the half-life of *comK* mRNA, increasing noise in the system (Gamba *et al.*, 2015; Kampf and Stuelke, 2015). Increased noise in ComK levels is required for a sub-population of cells to differentiate into a genetically competent state. Reducing noise in the expression of ComK reduces competence in the population (Suel *et al.*, 2006). Stochastic noise driving differentiation must be temporally regulated. The probability of a cell switching to a competent state will be increased with time. The longer ComK levels remain close to the threshold the greater the chance of the threshold being past. Varying Spo0A~P levels determine the time window in which competence can occur. Spo0A~P levels increases as the community approach stationary phase. This increase first represses expression of *rok*, leading to an increase in *comK* expression. As the levels of

Spo0A~P continue to increase repression of *comK* then occurs through Spo0A~P binding at two operator sites in  $P_{comK}$  (Mirouze *et al.*, 2012).

The phosphorylation state of a second global regulator also functions to limit the time spent where cells can become competent. The DNA binding protein DegU positively regulates ComK (Hamoen *et al.*, 2000). Phosphorylated DegU (DegU~P) negatively regulates expression of the *srfA* operon, reducing the ComS levels required for ComK expression (Hamoen *et al.*, 2000; Murray *et al.*, 2009). In the dephosphorylated state DegU binds to DNA in the *comK* promoter region facilitating binding of ComK dimers. ComK dimers bind to A/T rich sequences either side of the DegU binding region known as K-boxes (Hamoen *et al.*, 2002). This increases ComK levels within the cell which negatively regulates the expression of *rok*, coding for the repression of competence protein (Hoa *et al.*, 2002). This further de-represses expression of *comK* increasing the chance of a switch to competence. Once ComK levels have reached sufficiently high levels DegU binding is no longer required to facilitate ComK binding to its own promoter. ComK levels are increased and maintained at high levels where expression of competence genes can take place (Smits *et al.*, 2005).

Regulation of the competent phenotype in *B. subtilis* is a complex, growth phase dependent process. Differentiation into the competent state can be summarised as a simple bistable switch. Hysteresis maintains the switch in the “ON” position through the activity of ComK on its own promoter. The switch is held in the “OFF” position by the constant removal of ComK via MecA-ClpC-ClpP mediated degradation. Regulation of this system increases the probability of a state switch occurring, only in the early stages of stationary phase. ComK levels are delicately balanced allowing stochastic fluctuations to drive the differentiation strategy. The phosphorylation state of global regulators time limits switching between competent states increasing the hysteresis in the system. This well characterised phenotype is an elegant example of bacteria utilising inherent gene expression noise to their selective advantage.

#### **1.1.4. Sporulation**

Sporulation is a terminal differentiation strategy well studied in *B. subtilis*. The formation of an endospore (hereafter referred to as a spore) allows the cell to survive in a metabolically dormant state for extended periods of time. The longevity of *Bacillus* spores remains unknown with some authors reporting viability in the millions of years (Cano and Borucki, 1995; Vreeland *et al.*, 2000). The mature *B. subtilis* spore has a protective exterior comprised

of four distinct layers: the inner forespore membrane, cortex, outer forespore membrane and the coat (McKenney *et al.*, 2013). It secures the chromosome in a partially dehydrated core which is resistant to mechanical shearing, ultraviolet radiation, extreme heat and chemical attack from enzymes, acids and alkalis (Nicholson *et al.*, 2000). The sporulation phenotype is stimulated by nutrient limited conditions and high cell density (Errington, 2003; Fujita and Losick, 2005). The master regulator of sporulation is phosphorylated Spo0A (Spo0A~P) (Burbulys *et al.*, 1991). Decreasing nutrient availability and increasing cell density increase the levels of Spo0A~P within the cells. At intermediate levels of Spo0A~P, differentiation into a genetically competent state is favoured. At higher levels of Spo0A~P heterogeneous gene expression drives differentiation of a sub-population into spores (Fujita and Losick, 2005; Kuchina *et al.*, 2011).

The commitment to sporulation is a response undertaken by a subpopulation of the bacterial community. It results from heterogeneous signalling in the Spo0A~P phosphorelay (de Jong *et al.*, 2010b). In a subpopulation of the colony this increases the concentration of Spo0A~P above the threshold level required to initiate sporulation. The commitment to sporulate is known to correlate with increasing *spo0A* expression levels (Narula *et al.*, 2012). Unlike competence there is a gradual, rather than bimodal increase in the concentration of the master regulator. Spo0A activity depends on its phosphorylation by protein kinases KinA-E (Jiang *et al.*, 2000). The system has a graded response to KinA phosphorylation in the relay (Kuchina *et al.*, 2011). Membrane localised kinases KinA and KinB respond to signals, autophosphorylating then transferring the phosphate to Spo0F. The tertiary complex YmcA-YlbF-YaaT increase the rate Spo0A phosphorylation (Carabetta *et al.*, 2013). The action of YmcA-YlbF-YaaT on the phosphorelay is not fully understood, however Carabetta *et al.* found YmcA interacts with both Spo0B and Spo0F. Spo0F is phosphorylated in the relay, in turn transferring the phosphate to Spo0B where it can then phosphorylate Spo0A forming Spo0A~P (de Jong *et al.*, 2010b). KinC is also membrane bound and phosphorylates Spo0A directly. Increasing Spo0A~P levels de-represses *spo0H* via inhibition of the transition state regulator AbrB. *spo0H* encodes SigH which upregulates *kinA*, *spo0F* and *spo0A* causing a positive feedback loop (Fujita *et al.*, 2005). The phosphorylation state of Spo0A is regulated through inhibition of kinase activity and the action of phosphatases. KipI and Sda inhibit the kinase activity of KinA and KinB. Rap proteins RapA, RapB, RapE and RapH dephosphorylate Spo0F~P and Spo0E dephosphorylates Spo0A~P directly (Higgins and Dworkin, 2012). Overexpression of KinA or KinB leads to sporulation regardless of nutrient

availability (Fujita *et al.*, 2005). Tracking the promoter activity of genes within the phosphorelay demonstrated that the temporal dynamics of gene expression determine a decision to sporulate (de Jong *et al.*, 2010a).  $P_{kinA-gfp}$ ,  $P_{kinB-gfp}$ ,  $P_{spo0F-gfp}$ ,  $P_{spo0B-gfp}$  and  $P_{spo0A-gfp}$  reporters did not show higher overall fluorescence levels in sporulating cells compared to non sporulating cells. In both groups cell fluorescence increased from the initiation of tracking. Those cells destined for spore formation had increased promoter activity in  $P_{kinA-gfp}$ ,  $P_{spo0F-gfp}$ ,  $P_{spo0B-gfp}$  and  $P_{spo0A-gfp}$  earlier than non spore formers. This early promoter activity is more significant than the magnitude of gene expression. In non sporulating cells gene expression in the phosphorelay is also increased. The commitment to sporulation is delayed by the upregulation of *rapA*. RapA removes phosphate from the relay reducing Spo0A activation which might otherwise lead to sporulation (de Jong *et al.*, 2010b). As with competence the temporal regulation of gene expression provides a time window in which the sporulation phenotype can develop. Unlike ComK, the master regulator of sporulation, Spo0A, is upregulated gradually. Sporulation is a heterogeneous response without the bimodal switch. The heterogeneous phenotype is determined by timely activation of genes both activating and repressing this phenotype.

## **1.2. Stochasticity in gene expression is a source of heterogeneity**

Stochastic responses to homogenous environmental conditions generate distinct phenotypes in *B. subtilis* (Cozy *et al.*, 2012; Gamba *et al.*, 2015). Motility and competence in *B. subtilis* are examples of an organism gaining selective advantage from the utilisation of inherent stochasticity in gene expression. *B. subtilis* employs mechanisms to fine tune and amplify the stochastic process of gene expression leading to bistability, critical for differentiation (Maamar and Dubnau, 2005). Fundamental to this is the ability to manage stochastic noise such that it can be utilised or minimised.

### **1.2.1. Noise in gene expression**

Gene expression is an inherently noisy process (Gillespie, 1977). The DNA coding for a protein is first copied into mRNA before it is translated into a polypeptide chain. The polypeptide chain must then correctly fold to provide the three dimensional structure which enables it to function. This process involves discrete molecules interacting to synthesise the protein. The nature of these interactions are affected by the random movement of small particles in a liquid (Einstein, 1956). Molecules within the cell are constantly moving. They move in different directions and have different velocities. The random movement and

collision of molecules is described as stochastic (Wilkinson, 2009). For a chemical reaction to take place, molecules must physically interact with each other. The stochastic movement of the reacting molecules generates the stochastic synthesis of the final product. The time it takes for a reaction to take place has a random element to it. Where the number of reacting molecules are large the individual variation in the reactions will be obscured. Where the numbers of reacting molecules are small the differences between them is more apparent. Small molecule numbers therefore increase the effects of stochasticity (McAdams and Arkin, 1999). Gene expression in *B. subtilis* is predominantly initiated from a single molecule of chromosomal DNA. Reactions originating from a single molecule increases the variation in output. This variation in output is described as noise. Gene expression can be initiated from more than one site, preceding single or multiple genes (Micka and Marahiel, 1992; Wu *et al.*, 1992). Physical contact between transcription factors, DNA and polymerases are required for gene expression. These low copy number interactions increase the noise, contributing towards cell to cell differences. The process and efficiency of ribosomal binding, translation into polypeptides and the correct folding into proteins require numerous chemical reactions. These too require molecules to be in the correct place, at the correct time and in the necessary concentrations. Then these molecules must physically interact with enough energy to achieve the transition state for a reaction to take place. Noise in this process affects expression of the gene itself and the expression of downstream genes and associated regulatory networks. This noise can ultimately be utilised to determine cell fate through differentiation (Pedraza and van Oudenaarden, 2005; Suel *et al.*, 2006).

The source of noise within the cell contains both intrinsic and extrinsic components (Elowitz *et al.*, 2002). Extrinsic noise is produced through variation in concentration, location and state of the components required for gene expression. As a result of extrinsic noise there will always be a level of cell to cell differences in the concentration of transcription factors. Where transcription factors require activation this will further propagate noise (Fengos and Iber, 2013). These constitute extrinsic sources of noise. By contrast noise intrinsic to gene expression is generated specifically during the biochemical reactions during the process of transcription and translation. The inherent stochasticity dictates that the time frame for synthesis of each correctly folded protein will be different on each occurrence. In the absence of extrinsic noise there will still be variation in the process of gene expression on a cell to cell level.

### 1.2.2. *Measuring intrinsic and extrinsic noise*

Noise in gene expression is quantified by comparison of the individual cell with others in the population. Both the relative standard deviation and the relative variance are used to calculate noise (Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002; Wilkinson, 2009). In the first instance the ratio of the standard deviation ( $\sigma$ ) to the mean ( $\mu$ ) states that noise =  $\sigma/\mu$  (Elowitz *et al.*, 2002). This is known in statistics as the coefficient of variation (CoV). A practical advantage of the CoV is that it is dimensionless. It remains independent of the units of measurement. The second method to define noise is the relative variance, stating that noise strength =  $\sigma^2/\mu$  (Ozbudak *et al.*, 2002). In statistics this is known as the Fano Factor. A feature of  $\sigma/\mu$  data in a Poisson distribution is the dependence on the mean. An increase in mean will result in a decrease in the noise. By comparison  $\sigma^2/\mu$  calculations of noise strength are independent from the mean. In situations where it is possible to measure absolute molecule numbers  $\sigma^2/\mu$  noise calculations will measure deviations from Poisson (Ozbudak *et al.*, 2002). Landmark papers have justifiably used both measurements to describe noise in biological systems (Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002).

The first landmark paper discussed in this thesis defined noise as the relative standard (Elowitz *et al.*, 2002). The methodology has subsequently been used in publications for over a ten year period (Singh and Soltani, 2013). Elowitz's research defined experimental methods to distinguish between intrinsic and extrinsic noise. Two reporter genes driven by identical lactose inducible promoters were integrated into the chromosome of *E. coli*. Reporters were different derivatives of the gene coding for green fluorescent protein (*gfp*). Both alleles are distinguishable by colour, one coding for cyan fluorescent protein (*cfp*) and the other for yellow fluorescent protein (*yfp*). They were placed equidistant from the origin of replication on different strands of the chromosome. Measurements of intrinsic noise were based on the extent to which fluorescence in the two reporters correlate (Figure 6). Fluorescence in each reporter is modelled in the absence of intrinsic noise (Figure 6 A). Here the relative protein levels match over time. In the presence of intrinsic noise fluctuations in fluorescence do not correlate (Figure 6 B). Calculating intrinsic noise using this system requires that the cellular concentrations of molecules needed for expression (transcription factors, polymerases etc.) are uniformly diffused. A concentration gradient in the molecules required for expression may favour one reporter over the other. The effects of a concentration gradient will change with movement of DNA during the cell cycle. Gene dosage effects can also impact expression in this construct (Sauer *et al.*, 2016). The folding



time of reporter proteins is also presumed to be identical. Since these are virtually identical

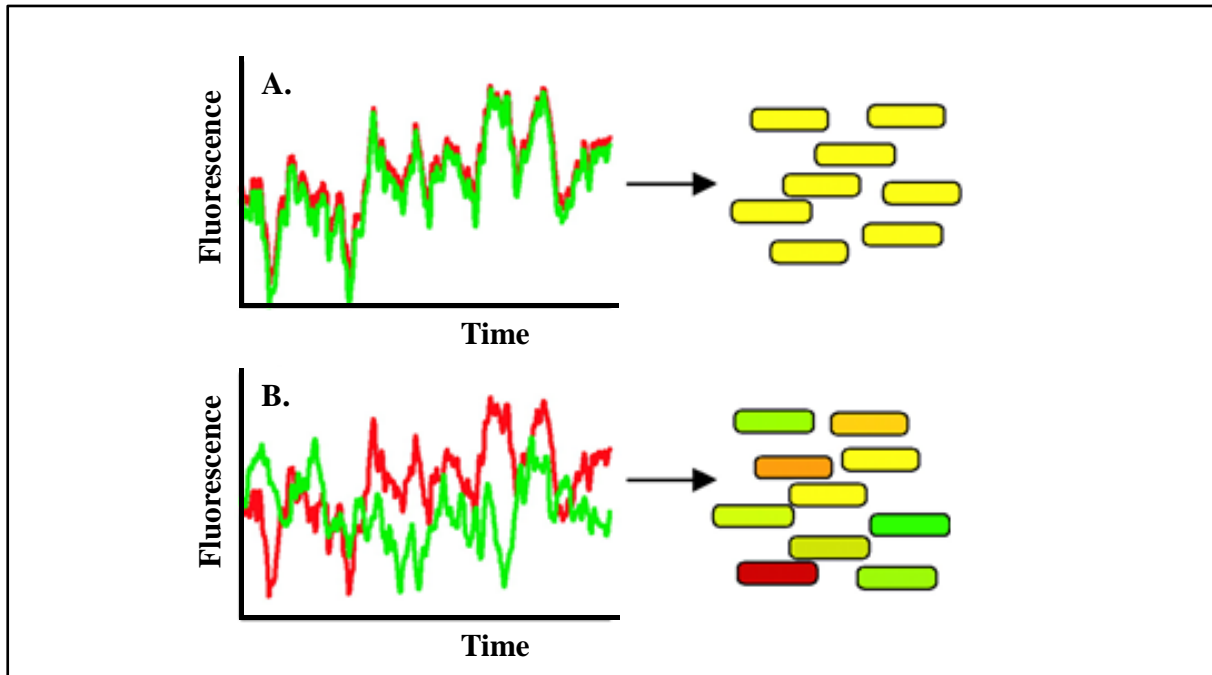


Figure 6. Two fluorescent reporter genes are used to quantify intrinsic noise. Green and red fluorescent genes are under the control of identical promoters. **A**, in the absence of intrinsic noise fluorescence in both reporters is identical. **B**, intrinsic noise alters the expression of the two genes. Adapted from Elowitz *et al.*, 2002.

proteins it might be reasonable to assume this is the case. However any alteration to the DNA sequence will require alternative amino acids and alternative transfer RNAs to deliver them to the ribosome. In this context tRNA availability is a source of extrinsic noise which has not been accounted for (Zhang and Ignatova, 2009). Consequently it may contribute to what is seen as intrinsic noise.

### 1.2.3. *Bursting patterns of transcription and translation*

Consistent with a stochastic model there is evidence for a bursting pattern in transcription. The synthesis of new mRNA molecules does not occur at regular, predictable time intervals. In eukaryotes this has been explained in terms of chromatin remodelling. DNA in eukaryotes is packaged by chromatin, which provides structural organisation to the genetic material. In addition to this it provides a structural obstacle for the transcriptional machinery to access the DNA and initiate transcription. The process of unpacking and repacking DNA from chromatin is dynamic, occurring over hours and is dependent on histone modification enzymes (Tumbar *et al.*, 1999; Hublitz *et al.*, 2009). This chromatin remodelling is a

stochastic process responsible for generating a bursting pattern of transcription (Rybakova *et al.*, 2015).

Alternative approaches have been used to provide evidence for transcriptional bursts in prokaryotes. Critical to each is the ability to distinguish individual mRNAs at the single cell level. One method involves labelling mRNA with binding sites for the bacteriophage MS2d coat protein (Johansson *et al.*, 1998). Quantifying transcriptional bursts utilises a synthetic target mRNA containing the gene coding for monomeric red fluorescent protein (*mrfp1*) followed by 96 binding sites for the MS2 coat protein. A translational fusion to the MS2d coat protein and GFP labels the target mRNA (Golding *et al.*, 2005). The fluorescent intensity of GFP correlates to the number of mRNA molecules in the cell (Figure 7). A drawback of this technique is the alteration required of the target mRNA. Labelling in such a

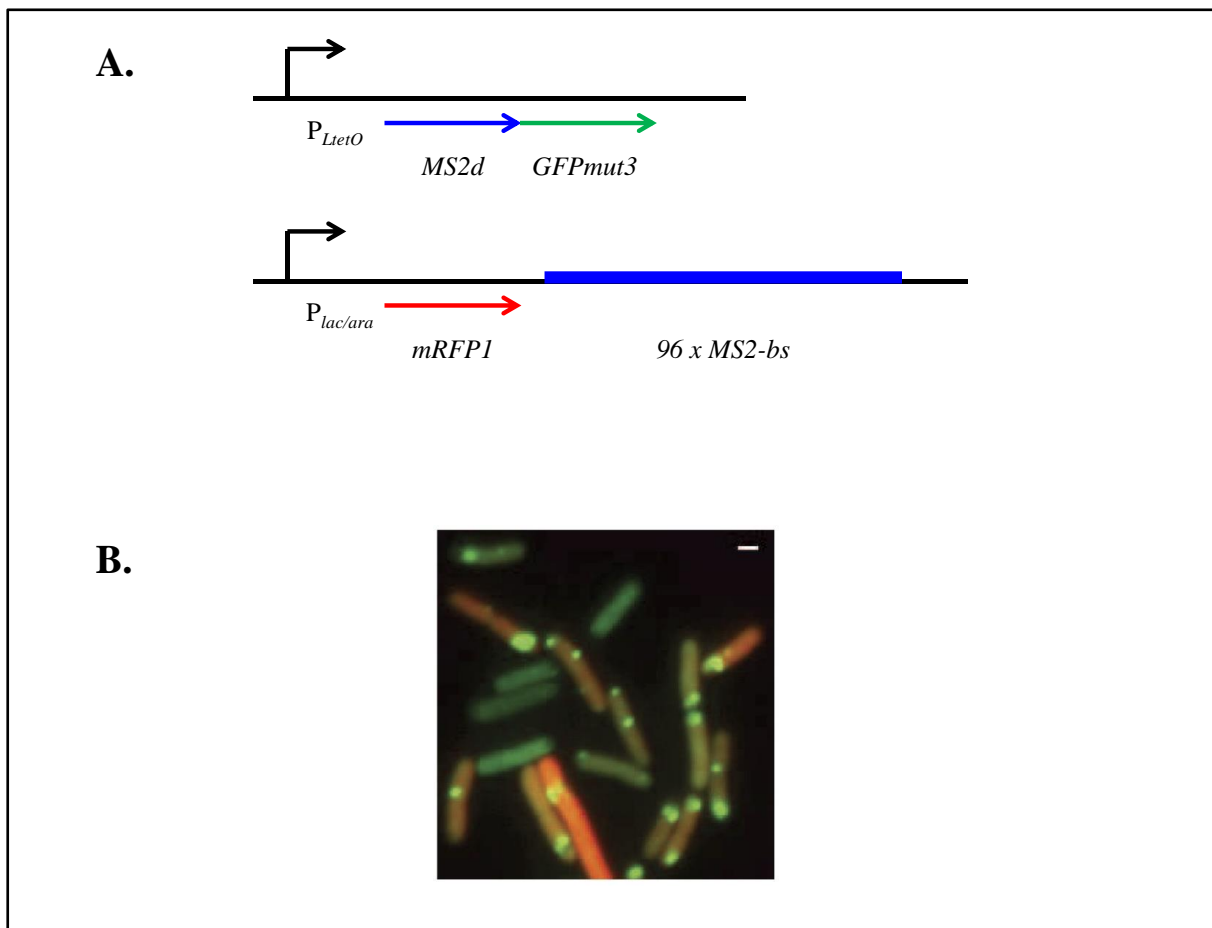


Figure 7. Fluorescent intensity of labelled mRNA corresponds to the number of mRNA molecules. **A**, the translational fusion of the MS2d coat protein with green fluorescent protein binds to the target mRNA. The mRNA target contains the gene coding for red fluorescent protein and 96 binding sites for the MS2d coat protein. **B**, the fluorescent intensity of GFP correlates to the number of mRNA molecules. The intensity of red fluorescence correlates to the rate of translation. Adapted from Golding *et al.*, 2005.

way alters the secondary structure of the transcript and may produce altered mRNA stability effects. This limitation is not a feature of fluorescent *in situ* hybridisation (Femino *et al.*, 1998). The target mRNA in this technique is detected with a fluorescently labelled probe. The intensity of the transcript can then be quantified and the fluorescent foci is proportional to the labelled mRNA. In both techniques a stochastic bursting of transcription is observed (Skinner *et al.*, 2013).

Evidence for a bursting pattern of protein synthesis from single mRNA molecules was produced using a yellow fluorescent protein reporter with a *tsr* tag (Yu *et al.*, 2006). The tag targets the fluorescent reporter to the membrane. This circumvents the problem of distinguishing between diffuse reporters in an auto-fluorescing background. The repressed state of the *P<sub>lac</sub>* promoter was used to drive expression of *tsr-yfp*. *P<sub>lac</sub>* is repressed by tetrameric LacI binding at two sites, looping the DNA (Choi *et al.*, 2008). Repressor affinity for the DNA allows stochastic release and occasional transcription of *tsr-yfp*. With individual fluorophores detectable on the membrane Yu *et al.* were able to quantify the stochastic bursts of translation over cell cycles

#### **1.2.4. Noise in transcription and translation**

Stochastic noise is a feature of both transcription and translation (Einstein, 1956; Ozbudak *et al.*, 2002; Wilkinson, 2009). Under different conditions these two process can generate different noise profiles (Oudenaarden, 2008)(Figure 8). A high rate of transcription coupled to a low rate of translation results in a low noise state. There is little variation in protein numbers over time (Figure 8A). Conversely, a low transcription rate coupled to a high translation rate will result in greater variation over time (Figure 8 B). Low copy number mRNA increases noise in the system. High rates of translation maintain the noise generated during transcription. Both conditions can produce the same mean with very different noise profiles.

Experimentally defining the separate contribution of noise in transcription and translation was published in the second landmark paper by Ozbudak *et al.* (2002). The IPTG inducible *P<sub>spac</sub>* was used to drive expression of *gfp* in *B. subtilis*. Transcriptional efficiency was altered through induction with varied IPTG concentration. The noise strength in GFP fluorescence was calculated as  $\sigma^2/\mu$ . To alter the translational noise, strains with point mutations in the RBS and start codon of *gfp* were used. Ozbudak *et al.* (2002) found that translational efficiency had more effect on noise than transcriptional efficiency. A 15% increase in

translational efficiency increased the noise strength by approximately 9%. To produce the same increase in transcriptional noise required a 100% increase in transcriptional efficiency. These data were consistent with their Monte Carlo simulation of translation where proteins are synthesised in bursts from single mRNA molecules. The model defined mean burst size ( $b$ ) from a single mRNA molecule as dependent on the rate of protein synthesis ( $K_P$ ) and the time taken for mRNA degradation ( $Y_R$ ).  $b = K_P/Y_R$  where the rates of mRNA synthesis ( $K_R$ ) and protein degradation ( $Y_P$ ) are constant. Where  $K_P$  is increased there will be a greater burst size before the mRNA is degraded. Conversely a low  $K_P$  will reduce protein variation over time. The simulation identifies heterogeneity in protein number over time with varied parameters. Higher noise is generated where  $K_R$  is low and  $K_P$  is increased in comparison to high  $K_R$  and low  $K_P$ .

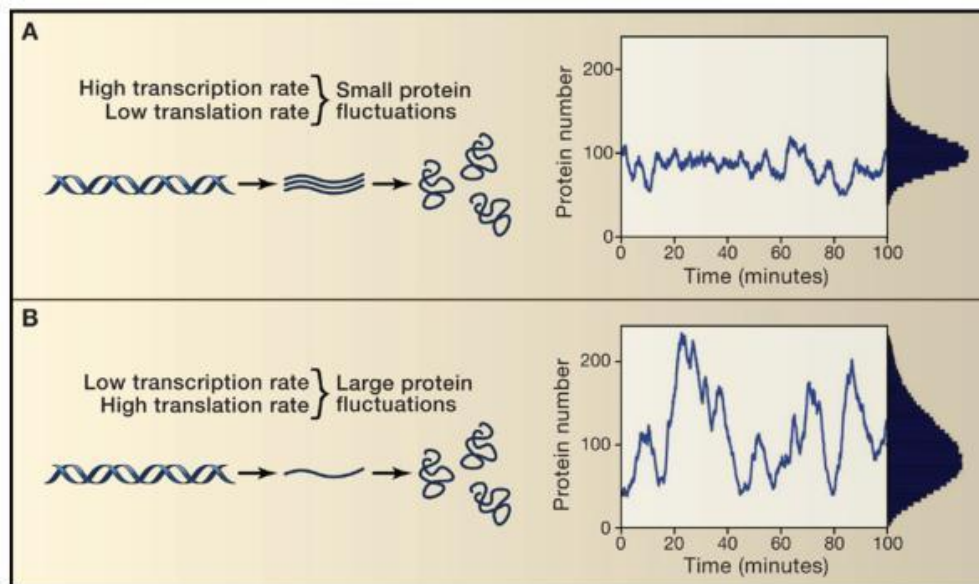


Figure 8. Noise in gene expression results from a combination of transcription and translation. **A**, high rates of transcription coupled to low rates of translation produce little noise. **B**, low rates of transcription coupled to high translation rates result in high noise. Reproduced from Oudenaarden *et al.*, 2008.

### 1.2.5. Noise in gene networks

The stochastic nature of gene expression determines that there will always be intrinsic noise. In the expression of a single gene this noise may not generate significant differences in protein numbers. However, the protein produced in gene expression also has a function within a cell. Consider a simple circuit where protein A negatively regulates expression of gene  $b$ . Intrinsic noise will result in cell to cell differences in the levels of protein A. Protein

A therefore becomes a source of extrinsic noise for the expression of gene *b*. In this hypothetical regulatory network protein B modulates expression of gene *c*. The intrinsic noise in the expression of gene *a* is propagated throughout the network and has its effects on the expression levels of both genes *b* and *c*. This hypothetical network was constructed to quantify noise propagation from one gene to the next (Pedraza and van Oudenaarden, 2005). Pedraza's construct includes two reporter genes used to monitor expression and noise (Figure 9 A). Gene one in the network is repressed by the constitutively expressed LacI protein. Transcription of gene one is controlled by addition of IPTG. Gene one is a negative repressor of gene two. The gene for cyan fluorescent protein (*cfp*) is co-transcribed with gene one correlating CFP fluorescence to the expression levels of gene one. Gene two is itself a reporter gene coding for a yellow fluorescent protein (*yfp*). Without addition of IPTG, gene one is repressed and gene two is active. There is low CFP and high YFP fluorescence (Figure 9 B & C). As the concentration of IPTG is increased a transition occurs where gene one expression increases and gene two expression decreases. At high concentrations of IPTG

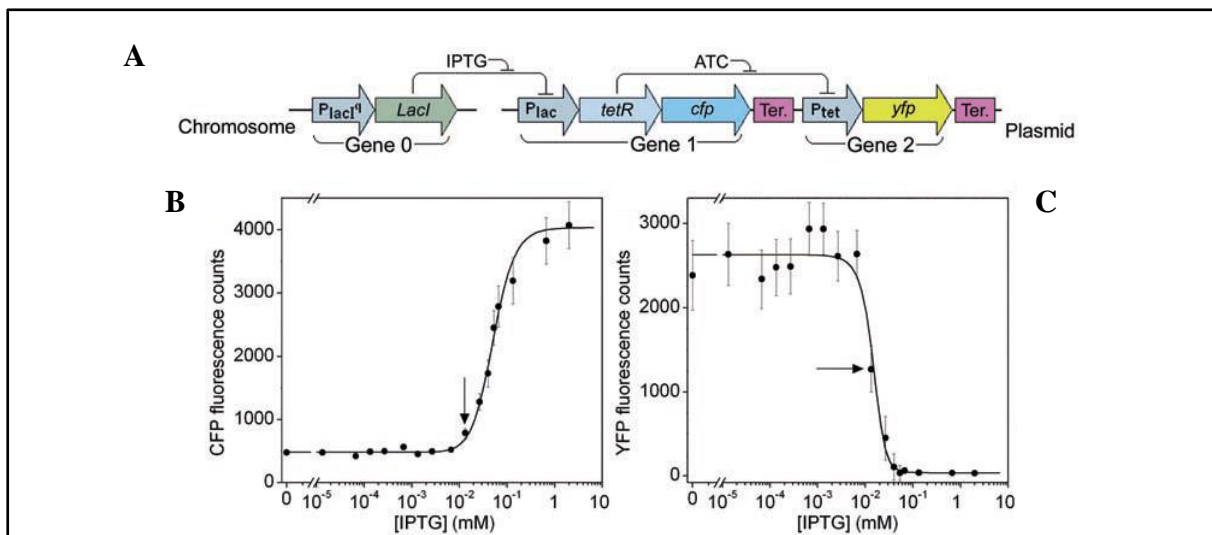


Figure 9. Propagation of noise can be investigated using a synthetic gene network. **A**, gene one is negatively repressed by LacI and gene two is repressed by TetR. The constitutive expression of *lacI* maintains repression of gene one and constitutive expression of gene two. **B**, the expression of gene one is modulated by addition of IPTG and reported by the fluorescent activity of CFP. **C**, the expression of the downstream gene two is determined by the IPTG concentration and reported with the fluorescent activity of YFP. Adapted from Pedraza *et al.*, 2005.

there is high CFP fluorescence and low YFP fluorescence. In the repressed state gene one exhibits high noise (Figure 10 A). Increasing levels of IPTG induction reduces noise in gene one and causes a transient increase in noise levels of gene two (Figure 10 B). The

experiments demonstrate there is a range of inducer concentrations capable of eliciting increased noise levels in the expression of downstream genes. These data have implications for the model organism *B. subtilis*. *B. subtilis* utilises increased noise to drive differentiation into distinct subpopulations (Suel *et al.*, 2007). The system employed by Pedraza *et al.* define parameters in which noise can be increased.

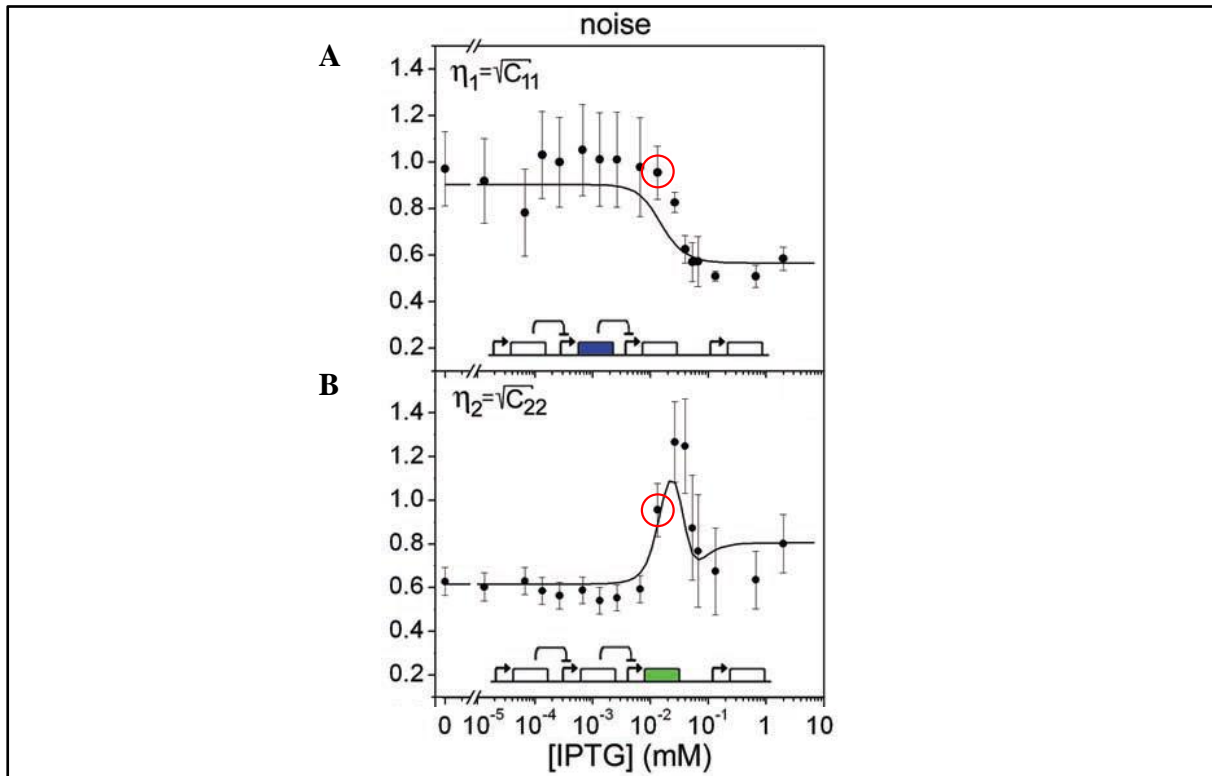


Figure 10. Noise propagation in gene networks. **A**, the activity *cfp* is modulated by addition of IPTG. Noise is calculated under different induction conditions as the coefficient of variation (standard deviation ( $\sigma$ ) / mean ( $\mu$ )). **B**, Noise in the downstream *yfp* gene is indirectly modulated by varied inducer concentration. Adapted from Pedraza *et al.*, 2005.

### 1.3. Synthetic Biology tools and applications

The goal of synthetic biology is to engineer purposeful biological circuits in living organisms (Khalil and Collins, 2010). Control over the synthetic circuit can be required for a variety of purposes. The solution to unique challenges drives innovation in synthetic biology. It encourages multidisciplinary team working to engineer functional solutions to diverse problems. It can be used to develop our understanding of biological systems. Permutations in the structure of regulatory networks can provide insight into their function (Sprinzak and Elowitz, 2005). The components of a synthetic circuit can be selected from a variety of organisms and optimised for expression in alternative hosts. The diversity of available parts

enable the design architecture to produce systems with wide functionality. The biofilm of *B. subtilis* can be utilised to fill cracks in concrete (Koh *et al.*, 2010). Swarming motility directs *B. subtilis* into the damaged areas and quorum sensing initiates biofilm formation. The production of extracellular levans glue seals the cracks, protecting the structure. Synthetic biology approaches to engineer bacteria offer alternatives to fossil fuels (Howard *et al.*, 2013). Howard *et al.* integrated metabolic pathways from non-native organisms into *E. coli* producing variable chain length hydrocarbons. Controlling exogenous free fatty acid controls the hydrocarbon chain length. Synthetic gene circuits can be structured to function as logic gates and combined into computational networks (Friedland *et al.*, 2009). Networks can be engineered to act as biosensors for external stimulus (Kobayashi *et al.*, 2004). UV radiation damages DNA, acting as the positive stimulus for the biosensor. The genetic circuit responds to this with expression of *gfp*, a visual output identifying the damage.

### **1.3.1. Synthetic gene circuits**

Increasing the complexity of the synthetic construct increases the complexity of the data, and potentially, our subsequent understanding. Isogenic cells in a population respond to a range of internal and external signals in complex regulatory networks (Arrieta-Ortiz *et al.*, 2015). The ability to sense, process and respond to stimulus is fundamental to all life forms. The ideal response to environmental conditions must be calculated to provide the optimal outcome. An objective in synthetic biology is to use bacteria for complex calculations (Siuti *et al.*, 2013). Logic gates are the basic building block of digital systems. The simplest forms of logic gates have one or two input signals and a single output signal (Figure 11). The NOT gate has no output when there is an input signal. It produces an output in the absence of an input. The AND gate requires both input signals to produce an output. The OR gate requires only one input signal to produce an output. If an OR gate has both input signals it will produce an output signal. Arrangement of these simple logic gates can generate increasingly complex input/output relationships. The combination of the AND followed by NOT gate reverses the output of the AND gate, producing the NAND gate. Layered logic gates are the foundation from which complex calculations can be made. Each of these logic gates have been generated in synthetic biology approaches to produce digital signals responding to a

range of stimuli (Tabor et al., 2009; Moon et al., 2012; Siuti et al., 2013; Wang et al., 2013; Stanton et al., 2014).

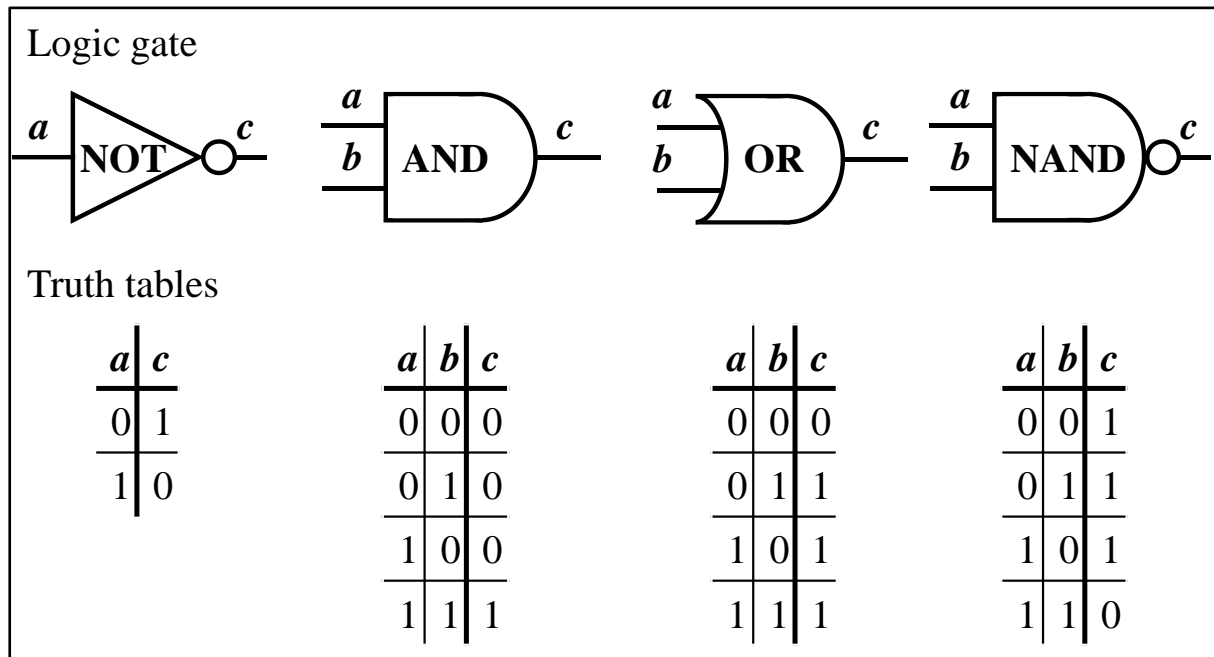


Figure 11. Logic gates are the simplest functional unit of digital circuits. Input signals *a/b* produce a binary on (1) or off (0) output signals (*c*). Truth tables for each logic gate determine the input/output relationship.

In biological terms a negatively regulated gene is functioning as a NOT gate. Tetracycline resistance is one such example, resistance is mediated by a NOT gate. The TetR homodimer bound to the operator site is the input for this NOT gate. The output is off until TetR is removed from the operator. Tetracycline changes the affinity of TetR for the DNA binding region, removing it from the promoter and allowing transcription to take place (Hillen and Berens, 1994). TetR based NOT gates are abundant in nature. These functionally similar homologs have sufficient differences in their structure and operator sites to make them a useful synthetic biology tool. Manipulation of the operator position enabled the construction of 16 TetR based NOT gates (Stanton *et al.*, 2014). A critical feature of these logic gates is their orthogonal function. This refers to the absence of cross-talk between the individual components. Each NOT gate should respond to one unique signal, the repressor protein. An orthogonal repressor is able to switch “OFF” gene expression through binding at one unique region of DNA.

The AND gate requires two signals to produce an output. Transcription factors and chaperones can be engineered to produce this functionality (Moon *et al.*, 2012). A



tetracycline inducible promoter drives expression of *invF*, the transcription factor. A second, anhydrous tetracycline inducible promoter drives expression of *sicA*, the chaperone. Only in the presence of both transcription factor and chaperone can the output be switched “ON”.

Moon *et al.* (2002) increased the complexity of the circuit with multiple AND gates. Four sensors control the input of two AND gates, which in turn control the input of a third AND gate (Figure 12). Each AND gate utilises the same signal/response architecture. The inputs are promoters driving expression of either a transcription factor or chaperone protein. The output of each gate is a promoter driving gene expression in the third AND gate. Signal output in this system is visualised by the expression of a red fluorescent protein.

Engineering an OR gate requires a gene regulated by two upstream promoters. One example of a naturally occurring OR gate is found in the regulation of the *fla/che* operon in *B. subtilis*. Two upstream promoters:  $P_{sigD}$  and  $P_{sigA}$  both positively regulate expression of the operon. The precise nature of *fla/che* regulation is more complex than a simple OR gate (Mordini *et al.*, 2013). There are multiple regulatory elements producing the gene products only when needed (see section 1.1.1 page, 2). One aim in synthetic biology is to integrate multiple logic gates to produce wide ranging input/output signals. This can be used to perform increasingly complex calculations, increase understanding or to monitor environmental signals.

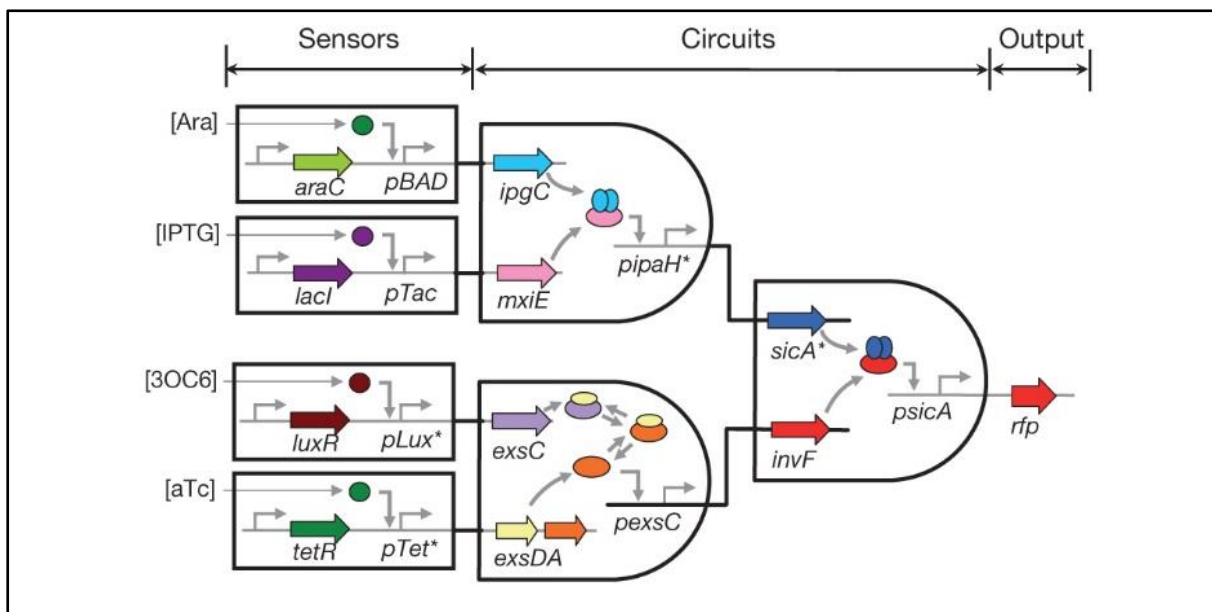


Figure 12. Layered AND gates respond to four input signals with a binary on/off response. Inducible promoters form the input signals in two primary AND gates. All four input signals are required to induce expression of *rfp*, the output signal of the secondary AND gate. Figure reproduced in part from Moon *et al.* (2012).

### 1.3.2. Biosensors

Naturally occurring signal transduction pathways can be manipulated to report the presence of environmental signals (Scheller *et al.*, 2001). Bacterial periplasmic binding proteins (bPBPs) conjugated to fluorophores have been used to report the presence of sugars, dipeptides, anions, cations and amino acids (De Lorimier *et al.*, 2002). Fluorophores sensitive to conformational change in bPBPs, alter their fluorescent intensity in response to the signal. Synthetic biology approaches have been used to programme a defined response to light. A bacterial lawn, grown on a petri dish has been modified to compute the edge between light and dark (Tabor *et al.*, 2009). A mask covering one half of the petri dish exposed half of the plate to light (Figure 13 A). A synthetic NOT gate responds to “NOT” light with production of an extracellular signalling molecule, able to diffuse from the dark areas of the petri dish (Figure 13 B). The genetic circuit is constructed where the output of the light dependent NOT gate is the first input in a secondary AND gate (Figure 13 C). The second input in the AND gate is an additional NOT gate, inverting the light dependent response of the first NOT gate. The AND gate is switched to the “ON” position in cells which are in the light and responding to the extracellular signalling molecule. The final

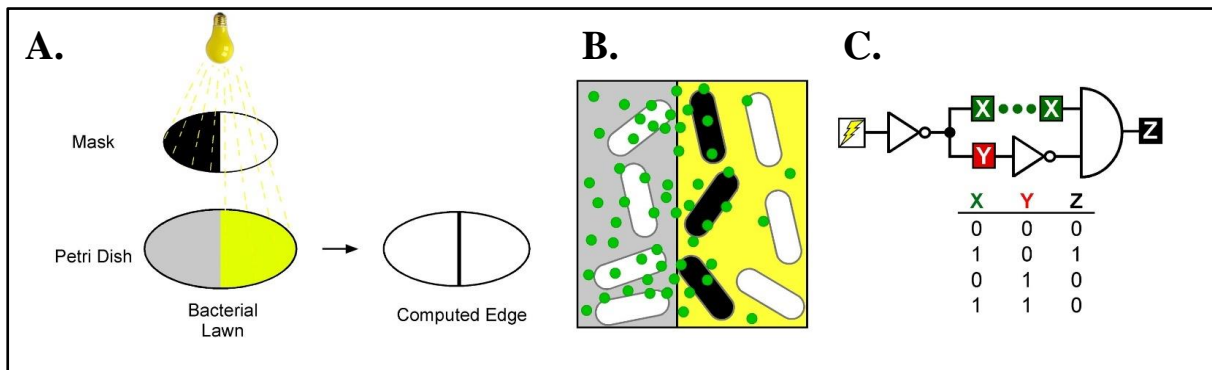


Figure 13. Synthetic logic gates designed to compute the edge between light and dark. **A**, a mask is used to grow a lawn of bacteria, half in darkness, half in the light. The digital circuit computes the light boundary and reports with production of a black pigment. **B**, Cells in the dark area respond with production of a diffusible signal (green circles). Bacteria in the light area respond to the signal with production of a black pigment. **C**, The digital architecture computing the edge between light and dark. The primary NOT gate produces an output in the absence of light. The output of the primary NOT gate is inverted by a secondary NOT gate, producing a signal where light is present. Output signals from both NOT gates are the input signals for the AND gate, producing the colour change signal. A cell must be in the light and close enough to receive signals from cells in the dark in order to activate production of the colour change pigment. Figure adapted from Tabor *et al.* (2009)

output of the AND gate is a black pigment. Diffusion limits the functional range of the extracellular signalling molecule, producing black cells only at the edge of the light and dark areas.

Multiple AND gates have been used to generate discrete responses within subsections of bacterial communities (Wang *et al.*, 2013). In one subpopulation the primary AND gate responds to arsenic and mercury (Figure 14 **A**). The output signal of the primary AND gate, LuxI is the first input signal for the secondary AND gate. The secondary AND gate is present within a different subpopulation of bacteria. Copper and LuxI are required for a signal output in the secondary AND gate, activating expression of the gene coding for a red fluorescent protein. This system extends the calculations made by logic gates outside the boundaries of a single organism. Exposure to a combination of arsenic, mercury and copper produces a red fluorescent signal, specific to a subpopulation of bacteria (Figure 14 **B**).

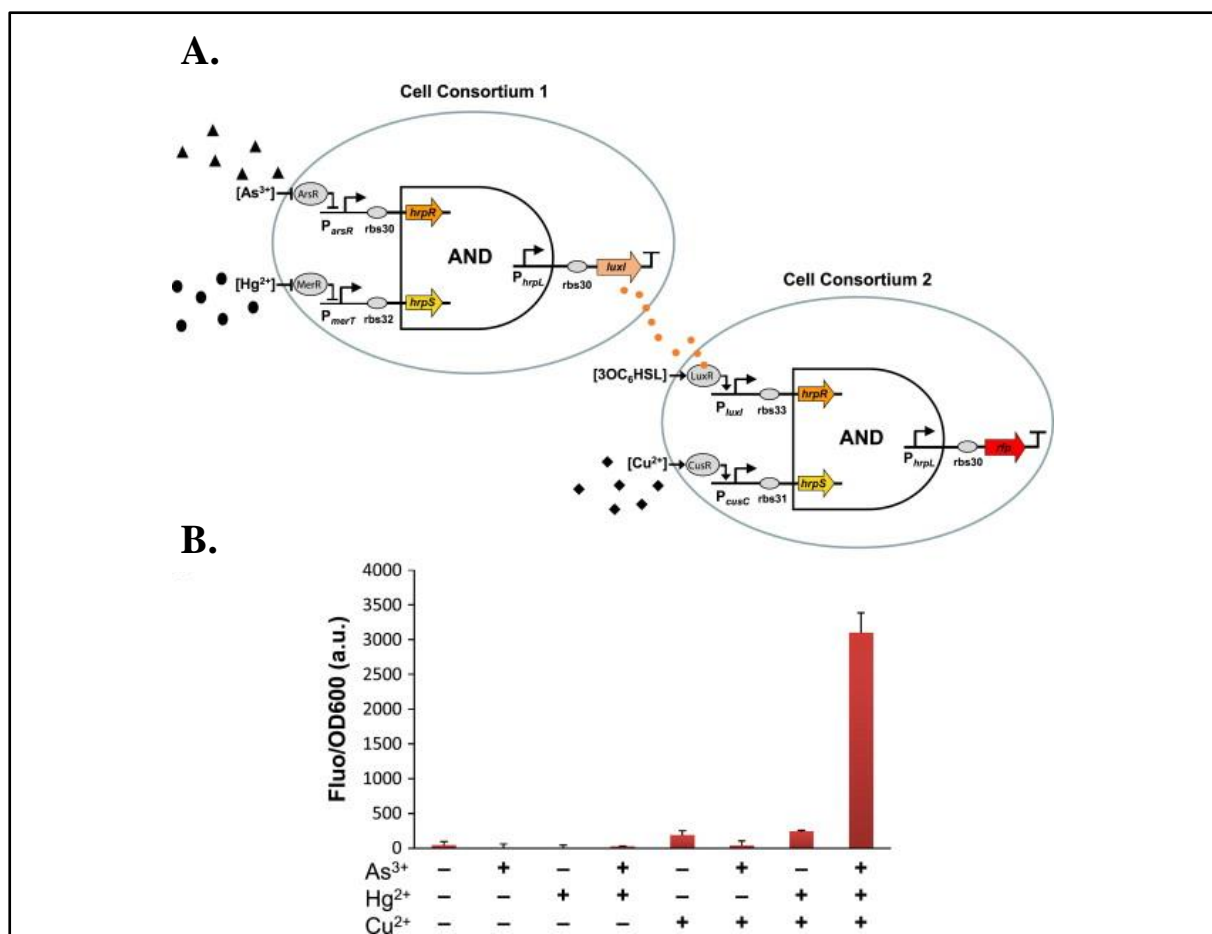


Figure 14. Biosensor communication between cell lines. **A**, The extracellular output signals from one cell are required as an input signal for a different cell type. **B**, Fluorescent activity in response to the input signalling molecules, 10 hours post exposure. Figure adapted from Wang *et al.* (2013).

### 1.3.3. *Fluorescent reporters*

The design of synthetic gene networks often requires the use of a fluorescent molecule as the signal output to monitor the system's behaviour (De Lorimier *et al.*, 2002; Rosenfeld *et al.*, 2005; Rosenfeld *et al.*, 2007; Palani and Sarkar, 2011; Wang *et al.*, 2013). One such molecule is the green fluorescent protein (GFP). The protein originates from the *Aequorea victoria* jellyfish. It was first observed in 1962 and the gene (*gfp*) first cloned in 1992 (Prasher *et al.*, 1992). The 2008 Nobel Prize in Chemistry was awarded to Shimomura, Chalfie and Tsien "for the discovery and development of the green fluorescent protein, GFP". The protein becomes active after folding into the correct structure and maturation of the light emitting fluorophore. Maturation requires molecular oxygen in a three step process involving a cyclization, dehydration and oxidation reaction (Tsien, 1998). The wild-type *gfp* has been altered in multiple processes to improve its function as a reporter. Amino acid substitutions F64L & S65T produced a ~ 6 fold brighter protein than the wild type, referred to as enhanced GFP (EGFP) (Yang *et al.*, 1996). Three rounds of mutation and screening resulted in the cycle three substitutions: F99S, M153T, and V163A. The cycle three mutations substituted hydrophobic residues for hydrophilic ones, reducing aggregation and enabling the protein to remain soluble (Cramer *et al.*, 1996). Substitution Y66W alters the emission spectra to produce a cyan fluorescent protein (CFP) and substituting wild type T203Y produces a yellow fluorescent protein (YFP) (Day and Davidson, 2009). Fluorescent reporter genes predate synthetic biology and have been extensively used to understand biological phenomena. Translational fusions have been used to elucidate the function of proteins of interest (Sun and Margolin, 1998). The temporal expression and location of the protein can be visualised using a GFP reporter, providing a valuable insight into the function of cellular components (Gamba *et al.*, 2009). One limitation of GFP reporters concerns their ability to correctly fold and mature when linked in translational fusions. A widely used EGFP (folding reporter) with the cycle 3 mutations in *E. coli* was found to misfold when the fused protein of interest does not fold correctly (Pedelacq *et al.*, 2006). Further mutations and screening lead to substitutions: S30R, Y39N, N105T, Y145F, I171V, and A206V. These alterations allowed the protein to fold correctly when fused to poorly folded proteins and the GFP was termed superfolder GFP (sfGFP). After denaturation in urea, 100% of sfGFP refolded, and at over 3 times the rate of the folding reporter.

The mCherry reporter is another well characterised addition to the fluorescent toolbox. There is minimal overlap in the excitation and emission wavelengths of mCherry and sfGFP.

sfGFP has maximal excitation at 450 nm and maximal emission at 508 nm (Andrews *et al.*, 2007). mCherry excites at 587 nm and emits at 610 nm (Shaner *et al.*, 2004). This allows their combined use single cells to visualise more than one process at a time. The mCherry protein originates from DsRed, termed for its host organism the corral *Discosoma* species (Shaner *et al.*, 2004). As with GFP variants amino acid substitutions in DsRed were required to prevent dimerization. Substitutions Q66M and T147S produced the monomeric red fluorescent protein (mRFP). To improve its use as a fusion tag, N and C terminal additions from GFP were included. Substitution M163Q removed the ~510 nm absorbance peak found in GFP. This significant mutation allows the combined use of GFP and mCherry reporters with reduced overlap in excitation/emission spectra. The final alterations: N6aD, R17H, K194N, T195V and D196N complete the mCherry protein. The *mCherry* gene was originally used for expression in eukaryotic HeLa cells. The codon sequence has been altered for prokaryotic use and termed *mCherry2* in this thesis.

The maturation time of mCherry and sfGFP have not been established in *B. subtilis*, the model organism used in this thesis. In yeast sfGFP has a maturation  $T_{0.5} \sim 6$  minutes and mCherry2  $T_{0.5}$  is  $\sim 16$  minutes (Khmelninskii *et al.*, 2016). The maturation of both these proteins is rate limited by reactions forming the fluorophore. In each case availability of molecular oxygen is the rate limiting step. The available data for mCherry maturation in *E. coli* ranges between  $\sim 15$  minutes and 40 minutes (Shaner *et al.*, 2004; Merzlyak *et al.*, 2007). The refolding kinetics of sfGFP were determined by the group responsible for its construction (Pedelacq *et al.*, 2006). However the refolding times cannot be meaningfully compared to the maturation time. Refolding occurs after the fluorophore has matured. Denaturation does not return the fluorophore to its original state (Andrews *et al.*, 2007). A difference in the folding and maturation of these two proteins may be an important design consideration. Quantifying gene expression based on the activity of different fluorescent reporters would need to consider their relative maturation times. In addition to this there are reported cell cycle effects on mCherry and GFP in *B. subtilis*. Sporulation stages 2 & 3 reduced the detectable levels of mCherry. In late sporulation GFP was diminished while mCherry was not (Doherty *et al.*, 2010).

## 1.4. Aims

The primary aim in this thesis is to investigate stochastic noise in gene expression, through exponential growth, transition phase and stationary phase. Stochastic noise is a feature of all biochemical interactions and the model organism *B. subtilis* utilises this noise to drive differentiation into distinct phenotypes. Previous research has established the use of the fluorescent reporter proteins to investigate noise in gene expression. Fluorescent activity is used to establish noise in the process of transcription and translation. Typically, noise investigations take place in exponentially growing cultures of bacteria. *B. subtilis* has mechanisms to increase gene expression noise at specific times outside exponential growth. The question is therefore asked: is gene expression noise growth phase dependent? The single time point assay will be extended into the physiology of the bacterial growth curve.

Design features of regulatory gene circuits have been considered within the context of gene expression noise. The hypothesis is that choice of promoter, reporter and integration loci will alter the output of the system. The suitability of the component parts is therefore also an important consideration. A secondary aim of this thesis is to characterise and evaluate regulatory gene circuits used to quantify heterogeneity in gene expression. To what extent do the design characteristics of a synthetic circuit alter the output of the system?

## Chapter 2. Materials and Methods

### 2.1. Bacterial Growth Conditions

#### 2.1.1. *General growth conditions*

Unless stated otherwise the following growth conditions were adhered to. Bacterial strains were stored at -80 °C in 10% DMSO. Strains were activated on LB agar plates at 37 °C overnight with appropriate antibiotic. Pre-cultures were produced using cells from a single colony, inoculated in 5 ml LB and grown overnight at 37 °C for between 14 and 16 hours. Antibiotics were added to select for plasmids in *E. coli* and reduce sporulation in *B. subtilis*.

#### 2.1.2. *Standard assay conditions*

Pre-cultures were grown overnight at 30 °C on an orbital shaker with the appropriate antibiotic. 400 µl of pre-culture was diluted into 20 ml pre-warmed LB (37 °C) without antibiotics and grown for approximately two hours in a shaking water bath. The Optical Density at 600 nm (hereafter referred to as OD) was confirmed to be between 0.2 and 0.6 and the cells diluted back to an OD = 0.05 in 20 ml pre-warmed LB (37 °C) without antibiotics. The second dilution corresponds to zero minutes in the standard assay. Unless stated otherwise the reporter genes were induced from the first dilution and induction was maintained at a constant level throughout the assay. Culture was removed from the flask every 30 minutes to measure the OD for the entire assay. Between zero and 90 minutes 1 ml of culture was removed to measure the OD. To measure the OD at 120 minutes, 0.5 ml of culture was removed and diluted up to 1 ml in LB. From 150 minutes until the final time point 0.2 ml of culture was removed and diluted into 1 ml. Samples for microscopy were taken from the cuvettes used for OD measurements.

#### 2.1.3. *Preparation of chemically competent cells*

1 ml of *E. coli* pre-culture was diluted into 200 ml LB at 37 °C on the orbital shaker. The culture was grown to an OD between 0.1 and 0.2 then transferred to 50 ml flacon tubes and centrifuged for 10 minutes at 4 °C and 4500 rpm. Each pellet was re-suspended in 40 ml pre-chilled 100 mM CaCl<sub>2</sub> and held on ice for 40 minutes. Centrifugation was repeated and each pellet re-suspended in 1 ml pre-chilled 100 mM CaCl<sub>2</sub> with 0.1% glycerol. 100 µl aliquots in 1.5 ml centrifuge tubes were flash-frozen in liquid nitrogen and stored at -80 °C until required.

#### 2.1.4. Transformation by heat shock

Chemically competent *E. coli* DH5 $\alpha$  cells were thawed on ice before adding approximately 50 ng of purified plasmid. Plasmid volume never exceeded 10% of the total volume. DNA was mixed by gently tapping the tube before leaving on ice for 30 minutes. Cells were heat shocked for 50 seconds in a 42 °C water bath and left on ice for 2 minutes. 900  $\mu$ l LB was added and the tubes placed on an orbital shaker at 37 °C for 45 minutes to 1 hour. The transformed cells were selected on LB agar with appropriate antibiotic at 37 °C overnight.

5-alpha Competent *E. coli* cells were used to transform Gibson ligations. These cells were provided by New England Biolabs, stored at -80 °C. Cells were thawed on ice before transformation by heat shock according to the manufacturer's instructions.

#### 2.1.5. Transformation into competent *B. subtilis*

Cells from a single *B. subtilis* colony were inoculated into 10 ml MM competence medium and grown overnight at 37 °C on an orbital shaker for between 14 and 16 hours. 600  $\mu$ l of the overnight culture was transferred into 10 ml room temperature MM and shaken for three hours at 37 °C. After three hours 10 ml of pre-warmed starvation media was added and the cells were grown for a further 2 hours to develop competence. 400  $\mu$ l of competent culture was transferred to a 1.5 ml micro centrifuge tube and mixed with 5  $\mu$ l DNA with an approximate 100 ng/ml concentration. Competent cells were shaken with DNA for 45-60 minutes and 100  $\mu$ l plated onto selective media.

#### 2.1.6. Antibiotic concentrations

Antibiotic	Working Concentration ( $\mu$ g/ml)	
	<i>Bacillus subtilis</i>	<i>Escherichia coli</i>
Ampicillin	50	100
Chloramphenicol	5	12.5
Erythromycin	2	150
Kanamycin	2	50
Spectinomycin	100	100
Tetracycline	10	12.5

Table 1. Antibiotic concentrations used with *B. subtilis* and *E. coli*.



## **2.2. Molecular Cloning**

### **2.2.1. *Preparation of plasmid DNA***

Plasmid DNA was prepared from cultures using a GenElute Plasmid Miniprep kit (Sigma Aldrich) according to the manufacturer's instructions. A standard 5 ml overnight pre-culture was set up and 3 ml of culture used to isolate plasmid DNA. Plasmids were eluted into 30 – 100 µl filter sterilised dH<sub>2</sub>O depending on the required volume and concentration of DNA. DNA quality and concentration was confirmed on a NanoDrop NA-1000.

### **2.2.2. *Preparation of chromosomal DNA***

Genomic DNA was prepared using a GenElute Bacterial Genomic DNA kit (Sigma Aldrich) according to the manufacturer's instructions. A standard 5 ml overnight pre-culture was set up and 1 ml of culture used to isolate genomic DNA. Genomic DNA was eluted into 50 µl filter sterilised dH<sub>2</sub>O. DNA quality and concentration was confirmed on a NanoDrop NA-1000.

### **2.2.3. *Polymerase chain reaction***

Primers for polymerase chain reactions (PCR) were synthesised by Eurogentec. PCR reactions took place using Q5 High-Fidelity DNA Polymerase (NEB) according to the manufacturer's instructions. Primer annealing temperatures were calculated using the T<sub>m</sub> Calculator tool provided by NEB (<http://tmcalculator.neb.com/#/>). 50 µl reactions were set up with thermocycler times at the maximum recommended by NEB and for 35 cycles. DNA was held at 4 °C after the final extension. Where necessary the template DNA was removed by a 30 minute incubation with 2 µl DpnI (Promega) at 37 °C. PCR products were purified using a GenElute PCR Clean-Up kit (Sigma Aldrich) according to the manufacturer's instructions.

### **2.2.4. *DNA agarose gel electrophoresis***

The length and quality of DNA fragments were checked on 1% agarose gels made with TAE running buffer (Sigma) and Ethidium Bromide (0.4 µg/ml). DNA was diluted 5:1 in 6 X Loading Dye (Promega) with 6 – 20 µl run per lane. Samples were run at 120 V alongside 6 µl Promega 1 Kb or 100 bp ladder. Gels were imaged in a Syngene InGenius Transilluminator using Genesnap software.

### **2.2.5. *Gel extraction***

Where required DNA was cut from agarose gels using a scalpel and purified using a GenElute Gel Extraction kit (Sigma Aldrich) according to the manufacturer's instructions. Correct fragments were identified using a Syngene Transilluminator.

### **2.2.6. *Digestion***

Plasmid DNA and PCR products were digested with Promega enzymes according to the manufacturer's instructions. 20 µl test digests were incubated for 1 hour before running the reaction on an agarose gel. 50 µl digests for subsequent ligation were incubated for 3 hours. Where two restriction enzymes were used DNA was purified between steps using a PCR Clean-Up kit (Sigma Aldrich). Where one enzyme was used the vector was treated with Thermosensitive Alkaline Phosphatase (TSAP) and heat inactivated according to the manufacturer's instructions (Promega). Digestion was confirmed through gel electrophoresis. Where necessary gel extraction was used to purify the required DNA. In all other cases a PCR Clean-Up kit (Sigma Aldrich) was used to purify the digested DNA. The digested product was checked on an agarose gel before ligation.

### **2.2.7. *Ligation of DNA***

Ligation of DNA digested with restriction enzymes used T4 DNA Ligase (Promega) in a 25 µl reaction with 2 µl enzyme. A standard ligation used 50 ng of digested vector with approximately 3-fold molar excess of insert. The ligation was incubated at room temperature for 20 minutes before heat inactivating the enzyme at 70 °C for 10 minutes.

### **2.2.8. *Gibson Assembly***

Gibson assembly exclusively used PCR products together with the Gibson Assembly Cloning Kit (NEB), according to the manufacturer's instructions. PCR reactions were treated with *DpnI* for 30 minutes at 37 °C and purified with a PCR Clean-Up kit. DNA was checked on an agarose gel and the purified DNA was quantified using the NanoDrop NA-1000. The size of DNA fragment, concentration and molarity were used to determine the volumes of each PCR product to add, in accordance with the manufacturer's instructions.

### **2.2.9. Screening**

Plasmids transformed into *E. coli* DH5 $\alpha$  were purified and test digested with an enzyme producing different length fragments from the digested vector. The size of DNA fragments were confirmed on an agarose gel before the plasmids were sent for sequencing with either GATC Biotech or Source Bioscience. Sequencing primers were designed according to the sequencing company's requirements and synthesised by Eurogentec. Once the sequence was confirmed plasmids were transformed into *B. subtilis* at either the *amyE* or *aprE* locus. Double crossover at the *amyE* locus was confirmed though growth on LB agar supplemented with 0.5% starch and staining with iodine. Integration at the *aprE* locus required preparation of genomic DNA and PCR using primers oTE120 and oTE137. The size of PCR product confirmed the successful integration.

## **2.3. Northern Blotting**

### **2.3.1. Generating single stranded RNA probes**

Linear template PCR products between 200 and 220 bp were generated using primers *gfp\_qpcr\_F* and oTE77 for the *sfGFP* template. Primers oTE131 and oTE132 were used to generate the *mCherry* template. RNA probes were generated using a DIG RNA labelling kit (Roche) with T7 polymerase according to the manufacturer's instructions. Probes were tested by crosslinking serial dilutions onto a membrane, bound to *Anti-Digoxigenin-AP*, Fab fragments (Roche) and visualising with addition of Tropix CDP-Star reagent (Applied Biosystems) in an ImageQuant LAS4000 system.

### **2.3.2. Standard RNA purification**

Strains were grown for Northern blotting in standard assay conditions. 2 ml of exponentially growing cultures was pelleted at 4 °C for one minute at 12000 rpm before flash freezing in liquid nitrogen. 1 ml of stationary phase culture was pelleted and frozen in the same manner. RNA was subsequently purified using Macherey-Nagel's NucleoSpin RNA (MN) kit according to the manufacturer's instructions. RNA concentration and quality were checked using a Nano Drop ND-100 spectrophotometer.

### **2.3.3. RNA stability purification**

With the exception of exponentially growing cells, strains were grown for RNA stability assays according to the standard RNA purification protocol. In exponentially growing

cultures strains were diluted into 30 ml LB at time point zero. Strains were induced with 1 mM IPTG from the first dilution. Transcription was blocked at the specified time by the addition of Rifampicin (150 µg/ml) during exponential growth and for separate samples during stationary phase. 2 ml of culture was removed during exponential growth and 1 ml of culture was removed during stationary phase. The removed culture was pelleted at 4 °C for one minute at 12000 rpm before flash freezing in liquid nitrogen. The sampling time points were at: 0, 5, 10, 15, 30, 45 and 60 minutes after Rifampicin treatment. Additional culture was removed for OD measurements during the one minute pelleting time and the OD quantified after the pellets had been flash frozen. Four cultures were assayed for RNA stability in each of the three experimental repeats. Cell pellets were prepared for lysis on ice by resuspension in 450 µl lysis buffer from the Macherey Nagel, NucleoSpin® RNA kit (MN kit) with 4.5 µl β-Mercaptoethanol. Silica beads were used to lyse the sample (lysing matrix B, MP Biomedicals) in a Precellys 24 tissue homogeniser at 4 °C. From this point forward the lysed samples were purified using the MN kit according to the manufacturer's instructions. The RNA concentration and purity was calculated using the Nano Drop ND-100 spectrophotometer.

#### **2.3.4. RNA Agarose gel electrophoresis**

2.5 µg total RNA was made up to a final volume of 10 µl in 1 X MOPS (page 147). The RNA sample was mixed by pipette with 10 µl RNA loading dye (page 147) and run alongside 3 µl DIG-labelled RNA Molecular Weight Marker I (Roche) in a formaldehyde gel (page 152). The gel was run in 1 X MOPS running buffer at 52 V for 3 hours. RNA in the gel was stained with Ethidium Bromide (5 µg/ml) in sterile dH<sub>2</sub>O by gentle agitation on the giro rocker (Stuart Scientific) for 5 minutes. Excess stain was removed with three washes in sterile dH<sub>2</sub>O on the giro rocker for 20 minutes. The gel was then left overnight in clean sterile dH<sub>2</sub>O at 4 °C. Gels were imaged in a Syngene InGenius Transilluminator using Genesnap software.

#### **2.3.5. Blotting**

RNA was transferred from the gel onto positively charged nylon membrane (Roche) using a Biorad Vacuum Blotter and mBAR vacuum pump. The gel was treated by addition of with denaturation solution to the top of the gel for 5 minutes. After this time any excess solution was removed from the top of the gel with a sterile glove and pushed to the edge of the vacuum blotter. The excess liquid was absorbed with tissue paper and removed. This

process was repeated with a 5 minute treatment with neutralisation solution. RNA was vacuum blotted onto the membrane over 6 hours. The gel was prevented from drying with a pool of 20 X SSC transfer solution maintained on top of the gel throughout the transfer. The membrane with transferred RNA was gently rinsed in 20 X SSC transfer solution before RNA was cross-linked to the membrane by UV radiation. Crosslinking was performed using the auto-crosslink function in the Stratagene UV Stratalinker 1800.

### **2.3.6.        *Hybridisation and visualisation***

Membranes were washed at 65 °C in 30 ml of pre-hybridisation solution in a Biometra OV3 rotating oven for 1 hour. RNA probes were denatured in hybridisation solution at 90 °C for 15 minutes followed by 15 minutes on ice. Pre-hybridisation solution was removed and 10 ml of hybridisation solution added. Hybridisation took place overnight at 65 °C in the rotating oven. The membranes were washed for 5 minutes in solution 1 at room temperature followed by three 15 minutes washes in solution 2 at 65 °C. Membranes were then transferred to the giro rocker (Stuart Scientific) and washed for 5 minutes in 30 ml of tween wash buffer, followed by a 30 minute incubation in 30 ml of blocking buffer. The blocking buffer was removed and 20 ml fresh blocking buffer added, containing 2 µl *Anti-Digoxigenin-AP*, Fab fragments (Roche). The *Anti-Digoxigenin-AP*, Fab fragments were incubated with the membrane for 20 minutes followed by a 60 minute wash in Tween wash buffer. The membrane was washed a second time in Tween wash buffer for 30 minutes then incubated in alkaline phosphatase buffer for 5 minutes. The membrane was treated with 5 µl of CDP-Star reagent (Roche) diluted into 1 ml of alkaline phosphatase buffer and incubated in the dark for 5 minutes. Images of the membranes were acquired with varied exposure times using an ImageQuant LAS4000 system (GE Healthcare).

### **2.3.7.        *mRNA quantification***

Relative mRNA levels were quantified using Image J. Equal sized rectangular sections were drawn around each band, together with one empty section of the image. The location of each band was stored in the ROI manager and the “measure” function used to calculate the intensity within the band, defined by Image J as RawIntDen. The intensity from the empty section of the image was then subtracted from the intensity of the bands. The relative intensity of the bands were defined by calculating the intensity of the mRNA band divided by the intensity of mRNA in the brightest band.

## **2.4. Protein Immunoblotting**

### **2.4.1. Preparation of whole cell lysate**

Cells were grown for the required time according to the standard growth protocol. The pellet from a volume of culture equivalent to 1ml at OD=1 was flash frozen in liquid nitrogen then re-suspended in 200 µl chilled 1 x SDS-PAGE buffer and kept at 4 °C. Cells were lysed by sonication for 15 seconds at 25% of maximum power using a Sonics Vibra-Cell (model number). Sonication was repeated in stationary phase cells for a total of three times. The lysate was heated to 100 °C for 10 minutes, vortexed for 10 seconds and centrifuged for 1 minute at 1100 rpm before loading 20 µl on a gel.

### **2.4.2. Protein gel electrophoresis**

Protein samples were separated through a 1 mm 12% Tricine SDS polyacrylamide gel (section 10.1.29 & 10.1.30) using BioRad mini-PROTEAN equipment. Samples were run through the stacking gel at 25 mA and 40 mA through the separating gel. The molecular weight of samples were estimated against 6 µl SeeBlue® Plus2 (Invitrogen) standard.

### **2.4.3. Coomassie stain**

Gels were stained with coomassie blue solution on a giro rocker (Stuart Scientific) for 30 minutes with gentle agitation. The background stain was removed in destaining solution for 30 minutes with gentle agitation. Destaining was repeated until protein bands were clearly distinguishable. Gels were left in dH<sub>2</sub>O overnight before imaging on an Epson perfection 3490 PHOTO scanner. Protein concentration was estimated using ImageJ v1.50a software.

### **2.4.4. Membrane transfer**

Protein samples were transferred onto a 0.45 µm Polyvinylidene Difluoride (PVDF) blotting membrane (GE Healthcare). The membrane was activated in methanol for 15 seconds before placing into cold 1 x transfer buffer. Transfer cassettes were assembled containing: sponge, 2 x Whatman Filter paper, membrane, gel, 2 x Whatman filter paper and sponge. The cassettes were loaded into the transfer tank with the membrane closer to the positive electrode and filled with chilled 1 x transfer buffer. Protein samples were transferred to the membrane at 120 V for 2 hours

#### **2.4.5. Protein detection**

Membranes were blocked in PMT overnight on a Stuart Roller Mixer at 30 rpm and 4 °C. Two washes in room temperature PMT were followed by addition of primary antibody (Abcam Rabbit polyclonal to GFP or mCherry) in PMT (1:1000 dilution) for two hours. Non-specifically bound antibody was removed with five washes in PMT. The membrane was then incubated in PMT containing secondary Goat- $\alpha$ -Rabbit IgG antibody conjugated to Horse Radish Peroxidase (Gibco BRL) at 1:10 000 dilution for one hour at room temperature. Four washes in 1 x PBS (pH 7.4) removed non-specifically bound secondary antibody and the GE Healthcare ECL Plus Chemiluminescent Western Blot detection kit was used according to the manufacturer's instructions. Images were obtained with varying exposure times using a GE Healthcare ImageQuant LAS4000 system.

#### **2.5. Fluorescence Microscopy**

0.75  $\mu$ l of cultured *B. subtilis* cells were fixed onto multispot microscope slides. Slides were prepared with 500  $\mu$ l of molten 1% agarose dissolved in dH<sub>2</sub>O containing 0.1  $\mu$ g/ml 4',6-diamidino-2-phenylindole (DAPI) and 5 $\mu$ l Benzyl Alcohol Anhydrous (BA). Use of BA reduced the concentration of DAPI required for nucleoid staining. Images were acquired using a Nikon Eclipse Ti with Nikon PLAN FLUOR 100X/1.30 oil infinity 0.17 WD 0.16 objective lens, CoolSNAP HQ CCD camera (Photometrics) and MetaMorph v7.7.8.0 software (Molecular Devices). Phase contrast images were exposed for 100 ms. sfGFP images were exposed for 1000 ms using a Nikon GFP-B 505 nm long pass dichroic, mCherry images were exposed for 1000 ms using a Chroma 41043 610 nm long pass dichroic and DAPI images used a Nikon UV-2E/C 400 nm long pass dichroic with 500 ms exposure.

##### **2.5.1. Cell Profiler analysis of microscope images**

16-bit images analysed with Cell Profiler (version 2.1.1) required cells with DAPI stained nucleoids to define a region of interest (ROI). The Cell Profiler workflow was initiated by loading the microscope images within a folder and identifying them according to their file name. Background light from the images in the red and green channels was subtracted and the modified images temporarily saved for subsequent analysis. Definition of the DAPI fluorescence in the nucleoid was enhanced by Cell Profiler. The enhanced image was used to identify the outline of fluorescence in the nucleoid. The size of the outline was reduced by one pixel before measurement of the object's size and shape. The outline of each nucleoid

was superimposed onto the modified images in the red and green channels. The pixel intensity of fluorescent light within each nucleoid was then calculated with both the data and metadata exported to a csv format. The nucleoid outlines were overlaid onto the original phase contrast and DAPI images, saved and visually checked for quality control purposes. The DAPI staining method was used to identify individual cells within growing chains of *B. subtilis*. The following settings within the Cell Profiler analysis modules remained constant across all microscopy assays:

**LoadImages** identified images based on an exact text match with the file naming structure determined when acquiring images using MetaMorph.

**Apply Threshold** removed background light from fluorescent images in the red and green channels. The “global threshold” strategy was selected together with the “Otsu” method to produce three classes of threshold. Pixels in the middle intensity were assigned to background and the entropy was minimized. No smoothing method was selected, the threshold correction factor was set to 1 and the lower and upper bounds on the threshold set to 0.000000 and 1.000000. Pixels below the threshold were set to zero and the threshold subtracted from the remaining pixel intensities.

**Enhance Or Suppress Features** was applied to the DAPI images. The image was “enhanced” with the “speckles” feature and the feature size set to 20.

**Identify Primary Objects** defined ROIs from the output of the Enhance Or Suppress Features module. The threshold settings were identical to the Apply Threshold module with the “threshold correction factor” set to 2.35. Objects touching the border of the image were discarded. No objects within the image were discarded based on size. Clumped objects within the image were distinguished by shape and the pixel intensity was used to draw dividing lines between clumped objects. Settings were adjusted to automatically calculate the size of smoothing filter for declumping, the minimum allowed distance between two local maxima and to speed up the process by using lower resolution images to find local maxima. Holes identified in objects were filled after thresholding and declumping. The object outlines were retained for use in the following module. The module was set to continue processing if excessive numbers of objects were identified.



**Expand Or Shrink Objects** was applied to the output of the Identify Primary Objects module. Each object in the image was shrunk in size by one pixel. The “fill holes...” radio button was set to “no” and the object outlines retained.

**Measure Object Size Shape** was set to measure the objects from the Expand Or Shrink Objects module, without calculating the Zernike features.

**Measure Object Intensity** selected the object outlines produced in the Expand Or Shrink Objects module to measure. Measurements were made on the image outputs of the Apply Threshold modules.

**Export To Spreadsheet** had the column delimiter set to comma and the representation of Nan/Inf set to NaN. Radio buttons set to “yes” were: overwrite without warning, add image metadata columns to your object data file and export all measurement types.

**Overlay Outlines** produced an image of the original DAPI fluorescence overlaid with the outline generated in the Expand Or Shrink Objects Module. The outline display mode was set to grayscale and the width of outlines set to 4. The brightness of the outlines was set to the max of the image and the outlines loaded from an image. This module was repeated a second time to draw outlines onto the original phase contrast image.

**Save Images** generated a 16-bit tif file of the image produced in the Overlay Outlines Module. The colormap was set to bone and no radio buttons were set to “yes”. The module was repeated to save each of the images produced in the Overlay Outlines modules.

### **2.5.2. *MicrobeTracker analysis***

Analysis with MicrobeTracker used the diffuse sfGFP signal to identify cells with the batch processing tool. Meshes outlining individual cells were confirmed by eye and where necessary adjusted using the manual tools within the software. Meshes not accurately defining cells were removed from the analysis. The data was exported using the exportcells2xls function and two further columns were added. The experimental time point was added and signal1 (total per cell fluorescence) / area used to determine the mean per cell fluorescence. The parameters used in the MicrobeTracker analysis are listed in Table 5 within the appendix. This method is not suitable to analyse cells within growing chains of bacteria, such as those found during exponential growth in *B. subtilis*.

### 2.5.3. *Quality control*

Images were visually inspected in MetaMorph during acquisition. Poor quality images were removed pre-analysis under the following criteria: image unfocused, high DAPI background fluorescence, DAPI signal where no cell is visible. On occasions both Microbe Tracker and Cell Profiler were unable to distinguish between the signal used to define a ROI and background signal. In MicrobeTracker this was due to high levels of background fluorescence. In Cell Profiler this was due to high DAPI fluorescence outside of a living cell. Small splits in the agarose occasionally produced high DAPI fluorescence. Images where cells had recently lysed produced increased DAPI fluorescence. To avoid background fluorescence being identified as a cell, images were visually inspected both before and after analysis. Images containing anomalous ROIs were removed from the analysed data. Further quality control was required of data produced by Cell Profiler. This is discussed in Chapter 3.

## 2.6. Table of strains

Strain	TPA number	Relevant Genotype	Construction or Reference
168CA	3763	<i>trpC2</i> wild-type <i>B. subtilis</i>	Laboratory strain
DH5 $\alpha$		<i>E. coli</i> F- $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> ) U169 <i>recA1 endA1 hsdR17</i> (rK-, mK+) <i>phoA supE44</i> $\lambda$ - <i>thi-1 gyrA96 relA1</i>	Commercial (Invitrogen)
5-alpha	3899	<i>E. coli fhuA2</i> $\Delta$ ( <i>argF-lacZ</i> )U169 <i>phoA glnV44</i> $\Phi$ 80 $\Delta$ ( <i>lacZ</i> )M15 <i>gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Commercial (NEB)
AH5	3559	168CA <i>amyE::(P<sub>hyperspank-sfGFP*-ssrA</sub> (AAAV) lacI spc)</i>	Alex Henderson MRes
AH6	3560	168CA <i>amyE::(P<sub>hyperspank-sfGFP*-ssrA</sub> (ADAV) lacI spc)</i>	Alex Henderson MRes
AH7	3561	168CA <i>amyE::(P<sub>hyperspank-sfGFP*</sub> lacI spc)</i>	Alex Henderson MRes
CJ1	4093	168CA <i>amyE::(P<sub>hyperspank-sfGFP-ilvD(terminator) spc</sub></i>	Transformation of pCJ1 into competent 168CA
CJ3	4101	168CA <i>amyE::(P<sub>hyperspank-sfGFP-ilvD(terminator) spc</sub> aprE::(lacI ery)</i>	Transformation of pCJ2 into competent CJ1
TE13	3720	168CA <i>amyE::(P<sub>hyperspank-lacZ-sfGFP* lacI</sub> spc)</i>	Transformation of pTE13 into competent 168CA
TE47	3909	168CA <i>amyE::(P<sub>hyperspank-sfGFP*-ilvD(terminator) lacI(cis) spc</sub></i>	Transformation of pTE47 into competent 168CA
TE48	3918	168CA <i>amyE::(P<sub>hyperspank-ilvD(terminator) lacI(cis) spc</sub></i>	Transformation of pTE48 into competent 168CA

TE49	3919	168CA <i>amyE::(P<sub>hyperspank</sub>-sfGFP*-mCherry2-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE49 into competent 168CA
TE50	3920	168CA <i>amyE::(P<sub>hyperspank</sub>-mCherry2-sfGFP*-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE50 into competent 168CA
TE51	3921	168CA <i>amyE::(P<sub>hyperspank</sub>-mCherry2-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE51 into competent 168CA
TE52	3922	168CA <i>amyE::(P<sub>xyIR</sub>-sfGFP*-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE52 into competent 168CA
TE53	3923	168CA <i>amyE::(P<sub>veg</sub>-sfGFP*-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE53 into competent 168CA
TE58	4146	168CA <i>amyE::(P<sub>hyperspank</sub>-sfGFP-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE58 into competent 168CA
TE59	4121	168CA <i>amyE::(P<sub>hyperspank</sub>-mCherry2*-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE59 into competent 168CA
TE78	4457	168CA <i>amyE::(P<sub>hyperspank</sub>-sfGFP(RiboTempo)-ilvD(terminator) lacI(cis) spc)</i>	Transformation of pTE67 into competent 168CA

Table 2. Strains used in this thesis. The asterisk denotes presence of the N-terminal translational linker MEFLQ.

## 2.7. Table of plasmids

Plasmid Name	TPA number	Vector	Genotype	Description	Reference
pAH5	3540	pDR111	<i>amyE::(P<sub>hyperspank</sub>-sfGFP*-ssrA (AAAV) lacI spc) bla</i>		Alex Henderson unpublished
pAH6	3541	pDR111	<i>amyE::(P<sub>hyperspank</sub>-sfGFP*-ssrA (ADAV) lacI spc) bla</i>		Alex Henderson unpublished
pAH7	3542	pDR111	<i>amyE::(P<sub>hyperspank</sub>-sfGFP* lacI spc) bla</i>		Alex Henderson unpublished
pCS30		Unknown	<i>aprE::(kan ery) bla</i>	<i>aprE</i> integration vector	Chris Sauer unpublished
pPG35		Unknown	<i>amyE::(P<sub>comG</sub>-lacZ-gfp+ spc) bla</i>	Used for <i>lacZ</i> PCR	Pamela Gamba (2015)
pSG1154		Unknown		Used for <i>P<sub>xyIR</sub></i> PCR	Lewis and Marston (1999)

pSS125		Unknown	<i>erm bla himar9 Tn YLB-1::(kan t<sub>erzA</sub>/t<sub>braB</sub> P<sub>veg</sub>sfGFP P<sub>spac</sub>-lacZ t<sub>tywoG</sub>/t<sub>tywoF</sub>)</i>	Used for <i>P<sub>veg</sub></i> PCR	Simon Syvertsson unpublished
pSS153		Unknown	<i>aprE::(mCherry2 lacI cat) bla</i>	Used for <i>mCherry2</i> PCR	Simon Syvertsson unpublished
pCJ1	4095	pTE19	<i>amyE::(P<sub>hyperspank</sub>-sfGFP-ilvD(terminator) spc) bla</i>	<i>ilvD</i> terminator added from <i>B. subtilis</i> genomic DNA using oTE121 oTE122. <i>BamHI NotI</i> digest. Sequenced with oTE7	Chris Jones unpublished
pCJ2	4100	pCS30	<i>aprE: lacI ery</i>	Gibson assembly. Vector amplification with oTE134 and oTE135. <i>lacI</i> cloned from pTE47 with oTE136 and oTE137. Sequenced with oTE9 and oTE61.	Chris Jones unpublished
pTE13	3720	pAH7	<i>amyE::(P<sub>hyperspank</sub>-lacZ-sfGFP* lacI spc) bla</i>	<i>lacZ</i> amplified from pPG35 using oTE18 and lacZ_R, digested with HindIII and ligated into pAH7. Sequenced with oTE15, oTE47 and gfp_qpcr_F.	This work
pTE19	3898	pAH7	<i>amyE::(P<sub>hyperspank</sub>-sfGFP* ΔlacI spc) bla</i>	<i>lacI</i> removed by amplification of pAH7 using primers oTE31 and oTE32, NotI digest and relegation. Sequenced with oTE19 and oTE16.	This work

pTE34	3820	pAH7	<i>amyE::(P<sub>hyperspank</sub>-sfGFP*-mCherry2 lacI spc) bla</i>	Gibson assembly to introduce <i>mCherry2</i> . Contains 35bp intergenic region. Identical copies of RBS 7 bp from <i>sfGFP*</i> and <i>mCherry2</i> . Vector amplified with goTE2 and goTE3, insert from pSS153 with goTE1 and goTE4. Sequenced with oTE7, oTE9 and oTE15.	This work
pTE35	3822	pAH7	<i>amyE::(P<sub>hyperspank</sub>- mCherry2- mCherry2-sfGFP* lacI spc) bla</i>	Gibson assembly to introduce <i>mCherry2</i> . Contains 35bp intergenic region. Identical copies of RBS 7 bp from <i>sfGFP*</i> and <i>mCherry2</i> . Vector amplified with goTE14 and goTE15, insert from pSS153 with goTE13 and goTE16. Sequenced with oTE9, oTE62 and oTE19.	This work
pTE36	3849	pTE35	<i>amyE::(P<sub>hyperspank</sub>- mCherry2-sfGFP* lacI spc) bla</i>	Duplicate <i>mCherry2</i> removed via AgeI digestion and relegation. Sequenced with oTE9, oTE62 and oTE19.	This work
pTE37	3888	pTE34	<i>amyE::(P<sub>hyperspank</sub>- mCherry2 lacI spc) bla</i>	<i>sfGFP</i> removed from pTE34 with NheI digestion and relegation. Sequenced with oTE19.	This work
pTE40	3851	pAH7	<i>amyE::(P<sub>xylR</sub>-sfGFP* lacI spc) bla</i>	Gibson assembly to swap <i>P<sub>hyperspank</sub></i> with <i>P<sub>xylR</sub></i> . Vector amplified with oTE69 and oTE70, insert amplified from pSG1154 with oTE67 and oTE68. Sequenced with oTE17.	This work

pTE43	3890		<i>amyE::(P<sub>veg</sub>-sfGFP* lacI spc) bla</i>	Gibson assembly used to swap <i>P<sub>hyperspank</sub></i> with <i>P<sub>veg</sub></i> . Vector amplified with oTE103 and oTE104, insert amplified from pSS125 with oTE105 and oTE106. Sequenced with primers oTE9 and oTE16.	This work
pTE46	3900	pTE19	<i>amyE::(P<sub>hyperspank</sub>-sfGFP* lacI(cis) spc) bla</i>	Gibson assembly to introduce <i>lacI</i> adjacent to <i>sfGFP</i> with divergent orientation. Vector amplified using oTE118 and oTE119. Insert amplified from pAH7 using oTE117 and oTE120. Sequenced with oTE61.	This work
pTE47	3908	pTE46	<i>amyE::(P<sub>hyperspank</sub>-sfGFP*-ilvD(terminator) lacI(cis) spc) bla</i>	<i>ilvD</i> terminator introduced as with pCJ1. Sequenced using oTE7, oTE9 and oTE15.	This work
pTE48	3910	pTE47	<i>amyE::(P<sub>hyperspank</sub>-ilvD(terminator) lacI(cis) spc) bla</i>	Gibson assembly used to ligate <i>P<sub>hyperspank</sub></i> from pDR111 into pTE47 with <i>P<sub>hyperspank</sub>-sfGFP*</i> removed. Insert amplified with oTE126 and oTE127. Vector amplified with oTE125 and oTE128. Sequenced using oTE7, oTE9 and oTE15.	This work
pTE49	3911	pTE47	<i>amyE::(P<sub>hyperspank</sub>-sfGFP*-mCherry2-ilvD(terminator) lacI(cis) spc) bla</i>	Gibson assembly used to ligate <i>P<sub>hyperspank</sub>-sfGFP*-mCherry2</i> from pTE34 into pTE47. Insert amplified with oTE126 and oTE127. Vector amplified with oTE125 and oTE128. Sequenced using primers oTE7, oTE9, oTE15 and oTE62.	This work



pTE50	3912	pTE47	<i>amyE::(P<sub>hyperspank-mCherry2-sfGFP*</sub>-ilvD(terminator) lacI(cis) spc) bla</i>	Gibson assembly used to ligate <i>P<sub>hyperspank-mCherry2-sfGFP*</sub></i> from pTE36 into pTE47. Insert amplified with oTE126 and oTE127. Vector amplified with oTE125 and oTE128. Sequenced using primers oTE7, oTE9, oTE15 and oTE62.	This work
pTE51	3913	pTE47	<i>amyE::(P<sub>hyperspank-mCherry2-ilvD(terminator) lacI(cis) spc) bla</sub></i>	Gibson assembly used to ligate <i>P<sub>hyperspank-mCherry2</sub></i> from pTE37 into pTE47. Vector amplified with oTE125 and oTE128. Sequenced using primers oTE9, oTE15 and oTE62.	This work
pTE52	3914	pTE47	<i>amyE::(P<sub>xyIR-sfGFP*-ilvD(terminator) lacI(cis) spc) bla</sub></i>	Gibson assembly used to swap <i>P<sub>hyperspank-sfGFP*</sub></i> with <i>P<sub>xyIR-sfGFP*</sub></i> . Vector amplified with oTE125 and oTE128, insert amplified from pTE40 using oTE126 and oTE127. Sequenced with oTE9 and oTE16.	This work
pTE53	3915	pTE47	<i>amyE::(P<sub>veg-sfGFP*-ilvD(terminator) lacI(cis) spc) bla</sub></i>	Gibson assembly used to swap <i>P<sub>hyperspank-sfGFP*</sub></i> with <i>P<sub>veg-sfGFP*</sub></i> . Vector amplified with oTE125 and oTE128, insert amplified from pTE43 using oTE126 and oTE127. Sequenced with oTE9 and oTE16.	This work

pTE58	4110	pTE47	<i>amyE::(P<sub>hyperspank</sub>-sfGFP- ilvD(terminator) lacI(cis) spc) bla</i>	Gibson assembly to remove translated linker (MEFLQ) from 5' end of <i>sfGFP*</i> . Primers oTE140 and oTE141 were used to amplify and pTE47 and remove the linker. Sequenced using primers oTE16 and oTE19.	This work
pTE59		pTE51	<i>amyE::(P<sub>hyperspank</sub>-mCherry2*- ilvD(terminator) lacI(cis) spc) bla</i>	Gibson assembly to add translated linker (MEFLQ) to 5' end of <i>mCherry2</i> . Primers oTE142 and oTE143 were used to amplify and insert the linker. Sequenced using oTE19.	This work
pTE66	4454	pEX-A2	<i>sfGFP(RiboTempo)</i>	Codon optimised <i>sfGFP</i> supplied by Eurofins Genomics in pEX-A2 vector.	Order number: 3663741
pTE67	4455	pTE47	<i>amyE::(P<sub>hyperspank</sub>-sfGFP(RiboTempo)- ilvD(terminator) lacI(cis) spc) bla</i>	Gibson assembly to replace <i>sfGFP*</i> with codon optimised <i>sfGFP(RiboTempo)</i> . Plasmid backbone amplified with oTE125 and oTE156, insert amplified from pTE66 using oTE155 and oTE126. Sequenced using oTE9 oTE14 and oTE16.	This work

Table 3. Plasmids used in this thesis.

## 2.8. Table of oligonucleotides

Name	Sequence (5' to 3')	Details	Description
gfp_qper_F	TCCAATTCTTGTTGAATTAGATGG	Northern blotting primer	<i>sfGFP</i> (35-59) forward
lacZ_R_Final	GCATCACGAAGCTTGCTAGCATATCTCCTTCCGGCCGATC	<i>HindIII</i> site	<i>lacZ</i> reverse
goTE1	TAGATAGATAGTCATTCGGCAGATAAGGAGGAGCTAGCATGGT CAGCAAGGGAGAGG	Gibson primer	<i>mCherry2</i> insert forward
goTE2	ATGCTAGCTCCTCCTTATCTGCCGAATGACTATCTATCTATTTG TAGAGCTCATCCATG	Gibson primer	pAH7 vector reverse
goTE3	GGAATGGACGAATTATACAAATAAGCATGCAAGCTAATTCGGT G	Gibson primer	pAH7 vector forward
goTE4	CACCGAATTAGCTTGCATGCTTATTTGTATAATTCGTCCATTCC	Gibson primer	<i>mCherry2</i> insert reverse
goTE13	GCTTTAAGGAGGAGCTAGCATGGTCAGCAAGGGAGAGG	Gibson primer	<i>mCherry2</i> insert forward
goTE14	CCTCTCCCTTGCTGACCATGCTAGCTCCTCCTTAAAGC	Gibson primer	pAH7 vector reverse
goTE15	AATAAATAGATAGTCATTCGGCAGATAAGGAGGAGCTAGCATG GAATTCCTGCAGATGAGC	Gibson primer	pAH7 vector forward
goTE16	GCTAGCTCCTCCTTATCTGCCGAATGACTATCTATTTATTTGTAT AATTCGTCCATTCC	Gibson primer	<i>mCherry2</i> insert reverse
oTE7	CCAGACAACCATTACCTGTC	Sequencing primer	<i>sfGFP</i> 585-605

oTE9	TACGTACGATCTTTCAGCCG	Sequencing primer	<i>lacI</i> promoter forward
oTE15	ATGTGAGCACTCACAATTC	Sequencing primer	<i>P<sub>hyperspank</sub></i> forward
oTE16	TTGACACCAGACCAACTG	Sequencing primer	Intergenic reverse. 55 bp upstream of <i>amyE</i> front
oTE17	ATATCTCTTGCCAGTCACG	Sequencing primer	Intergenic forward. 326 bp upstream from <i>P<sub>xyIR</sub></i>
oTE18	CGCGAAGCTTAAGGAGGCCCTACTGTGGAAGTTACTGACG	<i>HindIII</i>	<i>lacZ</i> forward
oTE19	GATGAAGATCTTCCCGATG	Sequencing primer	intergenic forward. upstream from <i>P<sub>hyperspank</sub></i>
oTE31	GCGGCGGCCGCGGATTCATTAATGCAGC	<i>NotI</i>	<i>lacI</i> deletion forward
oTE32	CGCGCGGCCGCGCAACCGTTTTTTCGGAAGG	<i>NotI</i>	<i>lacI</i> deletion reverse
oTE47	TCTGCCATTGTCAGACATG	Sequencing primer	<i>lacZ</i> +2744-2763 forward
oTE61	AAGACGGTACGCGACTG	Sequencing primer	<i>lacI</i> + 491 – 507
oTE62	CCTGGAGCATAACAACGTC	Sequencing primer	<i>mCherry2</i> +582 - 600
oTE67	CAATGGCAAGAACGTTGCTCGATATTGAAAATACTGATGAGG	Gibson primer	<i>P<sub>xyIR</sub></i> insert forward
oTE68	TCCATGCTAGCTCCTCCTTATCTAGATGCATTTTATGTCATATTG	Gibson primer	<i>P<sub>xyIR</sub></i> insert reverse

oTE69	CAATATGACATAAAATGCATCTAGATAAGGAGGAGCTAGCATG GAATTC	Gibson primer	$P_{xyIR}$ vector forward
oTE70	CCTCATCAGTATTTTCAATATCGAGCAACGTTCTTGCCATTG	Gibson primer	$P_{xyIR}$ vector reverse
oTE71	CAATGGCAAGAACGTTGCTCGAAATTTGTCAAATAATTTTAT TGACAACG	Gibson primer	$P_{veg}$ insert forward
oTE72	CTTAATTGTTATCCGCTCACAATTACATTTATTGTACAACACGA GC	Gibson primer	$P_{veg}$ insert reverse
oTE73	GCTCGTGTTGTACAATAAATGTAATTGTGAGCGGATAACAATT AAG	Gibson primer	$P_{veg}$ vector forward
oTE74	CGTTGTCAATAAAATTATTTTGACAAAATTTTCGAGCAACGTTCT TGCCATTG	Gibson primer	$P_{veg}$ vector reverse
oTE77	<u>TAATACGACTCACTATAGGGAGAGTCATGCCGTTTCATGTGAT</u> CC	Northern blotting primer	<i>sfGFP(226-247)</i> T7 reverse
oTE103	GCTCGTGTTGTACAATAAATGTATAAGGAGGAGCTAGCATG	Gibson primer	$P_{veg}$ vector $\Delta lacI O1$ forward
oTE104	CAATAAAATTATTTTGACAAAATTTTCGAGCAACGTTCTTGCC	Gibson primer	$P_{veg}$ vector $\Delta lacI O1$ reverse
oTE105	GGCAAGAACGTTGCTCGAAATTTGTCAAATAATTTTATTG	Gibson primer	$P_{veg}$ insert $\Delta lacI O1$ forward
oTE106	CATGCTAGCTCCTCCTTATACATTTATTGTACAACACGAGC	Gibson primer	$P_{veg}$ insert $\Delta lacI O1$ reverse

oTE117	GCCAGTCACGTTACGTTATTAGCTCACTGCCCGCTTTCCAG	Gibson primer	<i>lacI</i> forward
oTE118	CTGGAAAGCGGGCAGTGAGCTAATAACGTAACGTGACTGGC	Gibson primer	Vector reverse
oTE119	CATATGTAAGATTTAAATGCAACCGGAACGAAAATCGCCATTC G	Gibson primer	Vector forward
oTE120	CGAATGGCGATTTTCGTTCCGGTTGCATTTAAATCTTACATATG	Gibson primer	<i>lacI</i> reverse
oTE121	CGCGCGGCCCGCGTGACAAGTGCCAACACCG	<i>NotI</i>	<i>ilvD</i> terminator forward
oTE122	CGCGGATCCGCGTACGACAAGCTGGCGCATG	<i>BamHI</i>	<i>ilvD</i> terminator reverse
oTE125	CGGTGGAAACGAGGTCATCATTTCCCTCCG	Gibson primer	Vector forward
oTE126	CGGAAGGAAATGATGACCTCGTTTCCACCG	Gibson primer	Insert reverse
oTE127	GGTGAACGCTCTCCTGAGTAGGACAAATCC	Gibson primer	Insert forward
oTE128	GGATTTGTCCTACTCAGGAGAGCGTTCACC	Gibson primer	Vector reverse
oTE131	AGATTTAAGGTGCACATGGAAG	Northern blotting primer	<i>mCherry2</i> (52-73) forward
oTE132	<u>TAATACGACTCACTATAGGGAGATCAGGAATATCCGCAGGATG</u>	Northern blotting primer	<i>mCherry2</i> (238-257) T7 reverse
oTE134	CTGGAAAGCGGGCAGTGAGAAACCGTGTGCTCTACGACC	Gibson primer	pCS30 vector forward
oTE135	GGGTAAATGTGAGCACTCACCTAGCTATCGCCATGTAAGC	Gibson primer	pCS30 vector reverse
oTE136	GCTTACATGGCGATAGCTAGGTGAGTGCTCACATTTACCC	Gibson primer	<i>lacI</i> insert forward
oTE137	GGTCGTAGAGCACACGGTTTCTCACTGCCCGCTTTCCAG	Gibson primer	<i>lacI</i> insert reverse

oTE140	TTTAAGGAGGAGCTAGCATGAGCAAAGGAGAAGAAGCTTTTC	Gibson primer	<i>sfGFP</i> * linker deletion forward
oTE141	TCCTTTGCTCATGCTAGCTCCTCCTTAAAG	Gibson primer	<i>sfGFP</i> * linker deletion reverse
oTE142	CATGGAATTCCTGCAGATGGTCAGCAAGGGAGAGGAAG	Gibson primer	<i>mCherry2</i> * linker addition forward
oTE143	GACCATCTGCAGGAATTCCATGCTAGCTCCTCCTTAAAG	Gibson primer	<i>mCherry2</i> * linker addition reverse
oTE155	CAATTAAGCTTTAAGGAGGAGCTAGCATGTC	Gibson primer	Optimised <i>sfGFP</i> insert forward
oTE156	GACATGCTAGCTCCTCCTTAAAGCTTAATTG	Gibson primer	Optimised <i>sfGFP</i> vector reverse

Table 4. Oligonucleotides used in this thesis. Recognition sites for enzymes and the T7 polymerase are underlined.

## Chapter 3. Quality control

### 3.1. Introduction

In this thesis gene expression is manipulated using a number of synthetic constructs. The constructs contain fluorescent reporter genes and are cultured under a variety of conditions. Typically the fluorescent output is measured at the single cell level using a microscope. Images are subsequently processed using Cell Profiler software, unless specifically stated otherwise. Cell Profiler has limitations on the image composition it can process to generate valid data. To overcome the limitations, a quality control pipeline was required to assess the validity of Cell Profiler generated results. The pipeline is an extension to the image analysis software, published in its own right (Carpenter *et al.*, 2006). For this reason the novel code in the R based programming language is considered a result.

The process of image acquisition generates background fluorescence. An ideal image would only capture the fluorescent light emitted from the cells. In practice it is not possible to produce such an image. Light from the fluorescent cells illuminates the agarose in the immediate surroundings. Areas of agarose distant from any cells also direct light into the microscope camera (Figure 15). This background light can result from ambient light in the microscope room, light from the excitation source or noise generated from the set-up of the microscope equipment. These sources of background light are minimised, however it is not possible to remove them altogether. For the purpose of image analysis it is important to differentiate between the background light in the image and the fluorescence produced by the cells. Images are typically analysed using Cell Profiler software. The image analysis modules in this software are selected by the user and can be adapted to produce the most accurate results. The 'Apply Threshold' module is used to determine the background light in the image. Cell Profiler uses an adapted Otsu method to separate light into three levels: foreground, mid-levels and background (Otsu, 1979). These light levels are comprised of the cells, the surrounding area illuminated by the cells and the 'empty' areas of agarose, distant from any fluorescent cells. Cell profiler uses the intensity of every individual pixel to define the foreground, mid-level and background light levels in the image. It combines the mid light levels and the background to define an overall background level of light in the image. This is the threshold and the value is subtracted from the intensity of every pixel in the image. The fluorescence of each cell in the image is calculated as:



$$\text{Fluorescence of cell} = \frac{\text{total fluorescence in cell} - \text{threshold of image}}{\text{area of cell}}$$

The strategy used to calculate the threshold of each image generally works well. However, under certain conditions Cell Profiler is not able to accurately calculate background fluorescence. This results in the production of anomalous data. To avoid these data invalidating the results a quality control pipeline was required.

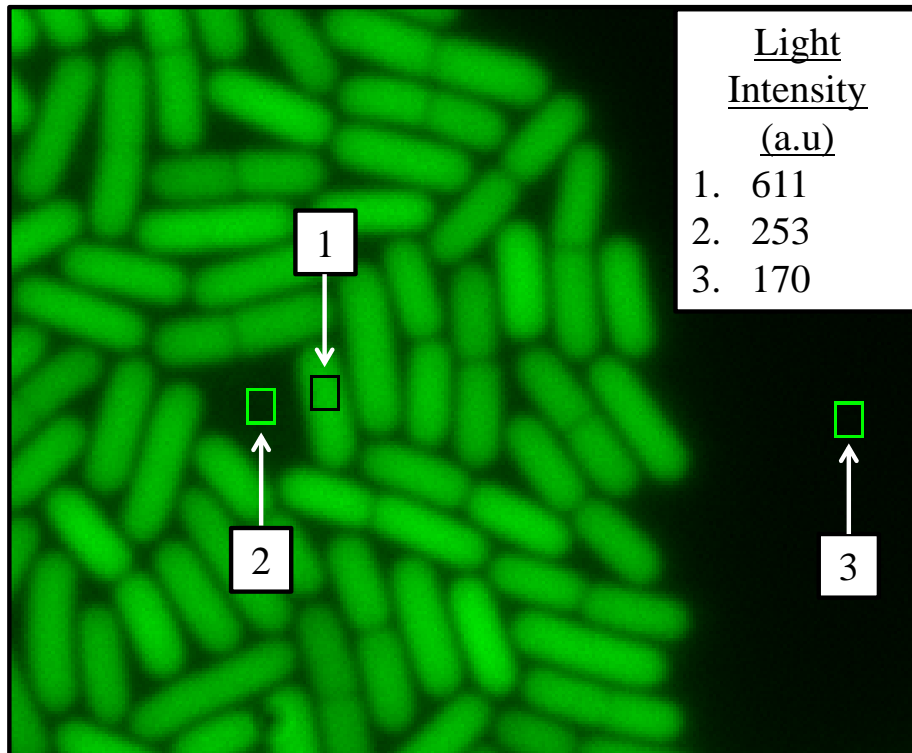


Figure 15. Fluorescent cells illuminate the surrounding agarose. The mean light intensity has been calculated within the three boxes using Image J. Box 1, light is captured from within the cell. Box 2, light is captured from an “empty” area of the slide, close to brightly fluorescing cells. Box 3, light is captured from an area of agarose distant from any brightly fluorescing cells.

### 3.2. Results and discussion

Anomalous microscope data is produced where the threshold value of the image is set too high. This occurs in conditions where most of the image is comprised of cells rather than empty agarose and the tightly packed cells are brightly fluorescent. The consequences of these conditions are two-fold. Firstly, there is less empty space from which to calculate the background fluorescence. Secondly, the fluorescent cells are illuminating the empty space around them. This combines to increase the threshold level calculated by Cell Profiler. The

result is illustrated from standard assay conditions using the *sfGFP\** strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*). Fluorescence is low during exponential growth and high by the final time point in stationary phase (Figure 16 **A**). However, from 180 - 240 minutes fluorescence appears to reduce. This reduction is not consistent with expectations of the system or with other data sets. It is not genuine. Where cells are not densely packed (Figure 16 **B & C**) Cell profiler is able to accurately calculate and remove the background fluorescence (Figure 16 **D**). Densely packed and brightly fluorescing cells (Figure 16 **E & F**) lead to an increased threshold calculation. When the background fluorescence is removed from the image much of the cells fluorescence is also lost (Figure 16 **G**). The mean fluorescence of cells in Figure 16 **D** is ~ 380 a.u and the mean in **G** is ~ 120 a.u. Prior to subtraction of the threshold the mean fluorescence is ~ 590 a.u for cells in both images.

Criteria must be established to determine what an acceptable threshold calculation is. Threshold values vary from image to image. In most instances this variation is normal. On some occasions Cell Profiler generates anomalous data. The question is therefore, what criteria can be used to define threshold anomalies? The threshold errors are essentially outlying data points. A simple and robust strategy for defining outliers is used to generate box-and-whisker plots. Outliers are defined as any point below the first quartile minus 1.5 x the interquartile range, or any point above the third quartile plus 1.5 x the interquartile range. This is the basis of the quality control pipeline used to define and omit calculations which have been based on outlying threshold values. This leaves one more question. How should the threshold values in a dataset be grouped to calculate the outliers? Each data set is treated individually using the pipeline. Where conditions in the assay have been changed the threshold values are grouped separately. Threshold values for each strain, induction condition and experimental repeat are all quality controlled separately. Datasets are not combined before they are quality controlled. A script using the R programming language was developed to identify and omit outlying data (Figure 17). In addition to this the script produces boxplots of the threshold values grouped together and a csv file containing the image data removed by the quality control pipeline (Figure 18 **A & B**). Prior to quality control the threshold value calculated by Cell Profiler results in lower fluorescence at the 240 minute time point (Figure 18 **C**). Noise calculations based on these data show a peak in noise at the 240 minute time point (Figure 18 **E**). Post quality control the dip in fluorescence at 240 minutes is no longer present (Figure 18 **D**) and there is no increase in noise at the same time point (Figure 18 **E**).

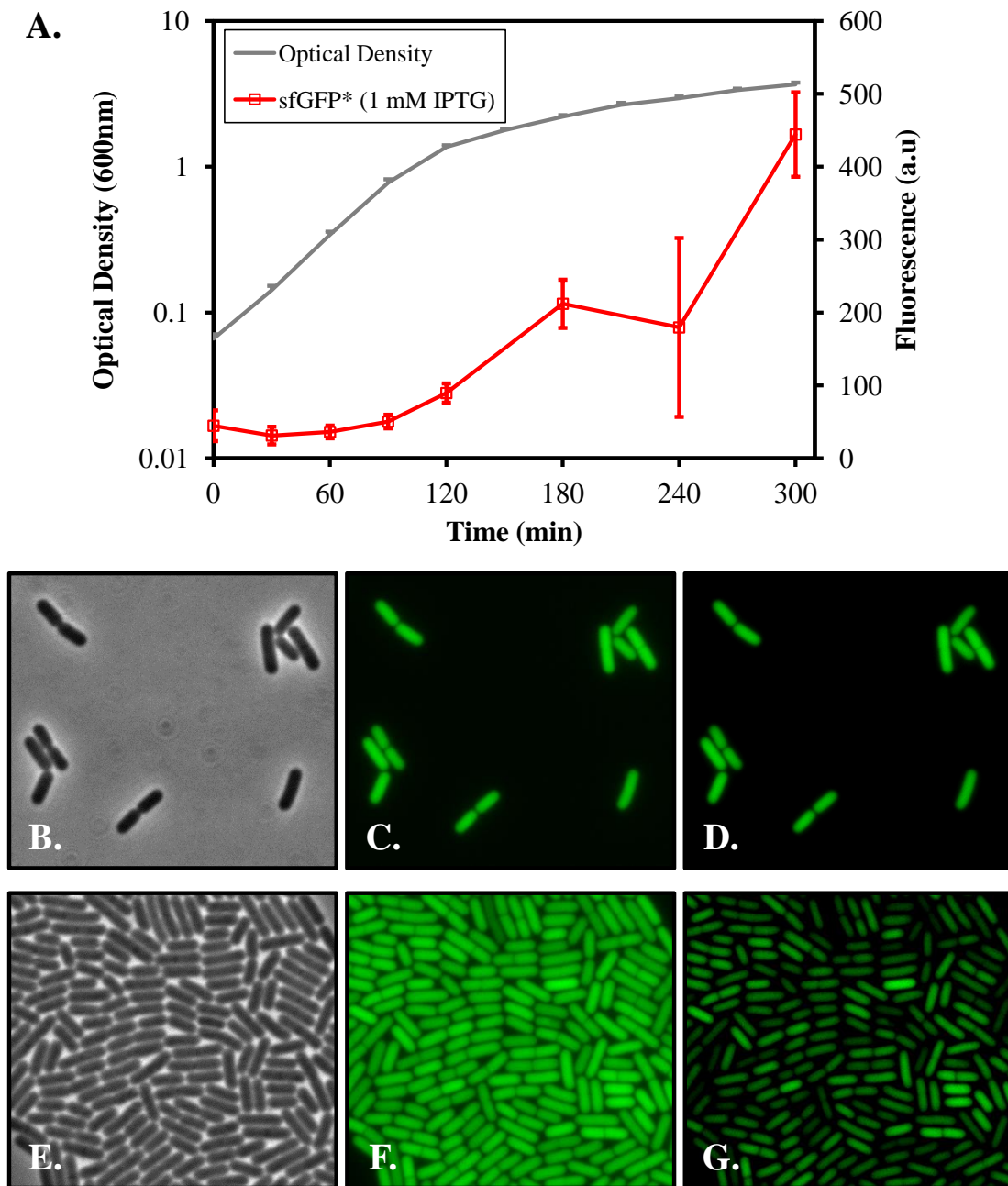


Figure 16. Limitations of the image analysis software Cell Profile. Background light in each images is calculated by Cell Profiler. This threshold value is then subtracted from every pixel within the image. **A**, growth rate and fluorescence in the *sfGFP\** reporter strain (AH7, *Phyperspank-sfGFP\*-lacI*) induced with 1 mM IPTG. The data represents one experimental repeat of cells cultured under standard assay conditions. Cell Profiler has generated threshold anomalies at 240 minutes. Grey line, optical density. Red line, fluorescence. **B**, phase contrast image of cells at 240 minutes. **C**, microscope image of *sfGFP\** fluorescence in the cells shown in **B**. **D**, Cell Profiler generated image of the cells shown in **C** after the background light has been subtracted. **E-G**, Cells as in **B-D** where the threshold value has produced anomalous results. Cell numbers for the single dataset are recorded in Table 14, page 162.

```

# Sets input folder, reads data file and sets folder to output data #
setwd("E:/Heterogeneity/TPE_125 AH7 AH6 AH5 induced uninduced/CP output
4th May 2014")
t = read.csv("Image.csv")
setwd("E:/Thesis temp/Results chapter quality control")

# Outputs two boxplots of the threshold values calculated by Cell
Profiler in the +/- IPTG conditions #
tiff("plus minus IPTG Boxplots.tiff", width = 8, height = 8, units =
"cm", pointsize = 14, res = 600, family="Times New Roman")
par(mar=c(2,2,0.5,0.25))
boxplot(t$Threshold_FinalThreshold_ThreshGreen[t$Metadata_induction ==
0], t$Threshold_FinalThreshold_ThreshGreen[t$Metadata_induction == 1],
      lwd=1.5, main = "", ylab = "", xlab = "", col="green",
      names = c("0 mM IPTG", "1 mM IPTG"))
box(lty = "solid", lwd=1.5)
dev.off()

# Splits data into +/- IPTG conditions #
t0 = subset(t, (t$Metadata_induction == 0))
t1 = subset(t, (t$Metadata_induction == 1))

# Calculates the upper and lower limits of accepted threshold values #
t0_i3 = quantile(t0$Threshold_FinalThreshold_ThreshGreen, 0.75)
t0_i1 = quantile(t0$Threshold_FinalThreshold_ThreshGreen, 0.25)
t0_upper = t0_i3 + 1.5 * IQR(t0$Threshold_FinalThreshold_ThreshGreen)
t0_lower = t0_i1 - 1.5 * IQR(t0$Threshold_FinalThreshold_ThreshGreen)

t1_i3 = quantile(t1$Threshold_FinalThreshold_ThreshGreen, 0.75)
t1_i1 = quantile(t1$Threshold_FinalThreshold_ThreshGreen, 0.25)
t1_upper = t1_i3 + 1.5 * IQR(t1$Threshold_FinalThreshold_ThreshGreen)
t1_lower = t1_i1 - 1.5 * IQR(t1$Threshold_FinalThreshold_ThreshGreen)

# Returns data frames only containing the outliers #
t0f <- t0[t0$Threshold_FinalThreshold_ThreshGreen > t0_upper |
t0$Threshold_FinalThreshold_ThreshGreen < t0_lower, ]
t1f <- t1[t1$Threshold_FinalThreshold_ThreshGreen > t1_upper |
t1$Threshold_FinalThreshold_ThreshGreen < t1_lower, ]

# Subsets outlier data frames to only include columns which identify the
image. Outputs a csv file #
chk0 <- t0f[, c("Metadata_Strain", "Metadata_induction", "Metadata_time",
"Metadata_Image", "Threshold_FinalThreshold_ThreshGreen")]
chk1 <- t1f[, c("Metadata_Strain", "Metadata_induction", "Metadata_time",
"Metadata_Image", "Threshold_FinalThreshold_ThreshGreen")]
comb01 <- rbind(chk0, chk1)
write.csv(file="outliers.csv", comb01)

# Identifies the Image Number of outlying data #
frame0 <- t0f$ImageNumber
frame1 <- t1f$ImageNumber

# Sets folder from which to read microscope image analysis #
setwd("E:/Heterogeneity/TPE_125 AH7 AH6 AH5 induced uninduced/CP output
4th May 2014")

# Reads data file and removes outlying data before final analysis #
rd = read.csv("ShrunkenNuclei.csv", header = TRUE)
rd <- subset(rd, !(rd$ImageNumber %in% frame0 | rd$ImageNumber %in%
frame1))

```

Figure 17. R script used to identify and omit outlying microscope data from the final analysis. Cell Profiler software calculates the background light in each microscope image. A box-and-whiskers criteria is used to define outliers in these threshold values. The script produces a boxplot of the threshold values and a csv file identifying the images from which the outliers originate. The script then produces a data frame for subsequent analysis without the outliers.

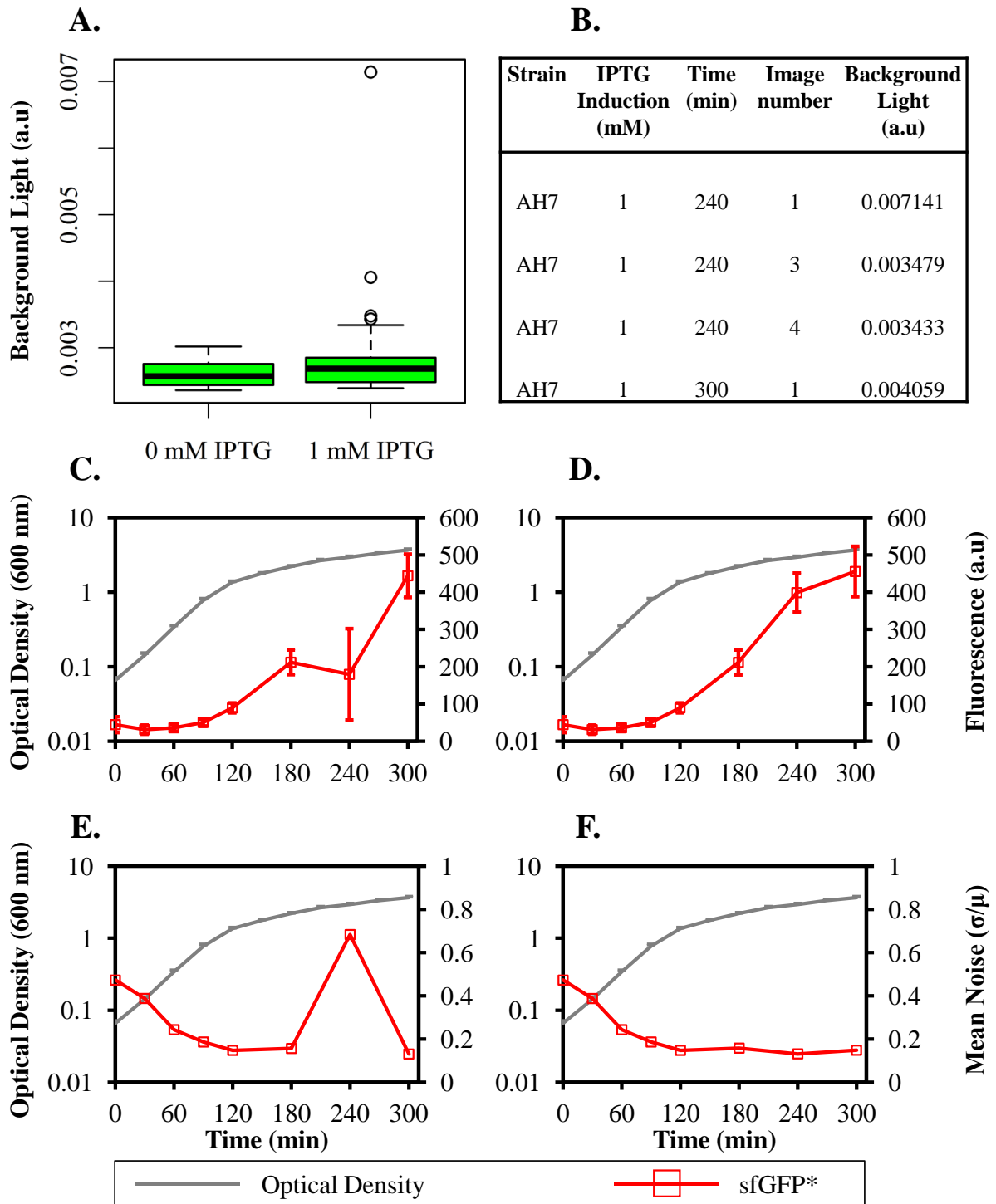


Figure 18. Outlying data is omitted using a quality control pipeline (Figure 17). **A**, background light in each microscope image is calculated by Cell Profiler software. These threshold values are plotted for each experimental condition specified in the script. **B**, image data omitted from the analysis is recorded in a csv file. **C**, fluorescence in  $n=1$  datasets for the *sfGFP\** reporter has been analysed without quality control. **D**, fluorescence in the same data set, analysed after quality control. **E**, noise calculations based on fluorescence data in **C**, prior to quality control. **F**, noise calculations based on fluorescence data in **D**, after quality control. Grey lines, optical density. Red lines, fluorescence. Cell numbers for the single dataset are recorded in Table 14, page 162.

### **3.3. Conclusions**

The quality control pipeline identifies and omits anomalous data from the Cell Profiler analysis. Accurate fluorescence data is vital for noise calculations. The background light in the image must be subtracted before calculating noise. The required calculation must be based on the fluorescence emitted from within the cell. A calculation based on the fluorescence of a cell + the background light will artificially reduce any cell to cell differences and result in anomalous noise data. The quality control pipeline is rapid and provides a robust criteria for defining acceptable image analysis parameters. It removes any subjectivity in defining what an acceptable threshold calculation is and what is not. The script generates two visual outputs. Boxplots of the threshold values in each experimental condition provide an immediate impression of the dataset. The csv file provides details of the images producing anomalous results. The user can immediately locate the microscope images to visually assess what the pipeline has removed. The operator time required to produce valid, high volume data is minimised through use of this quality control pipeline

## Chapter 4. Growth phase dependent heterogeneity in the expression of a single gene

### 4.1. Introduction

Noise in gene expression is a stimulus for heterogeneity in *B. subtilis* (Leisner *et al.*, 2008). This heterogeneity leads to differentiation resulting in distinct phenotypes (Errington, 2003; Leisner *et al.*, 2008; Vlamakis *et al.*, 2013; Mukherjee and Kearns, 2014). It is therefore important to understand the biochemical origins of noise which can ultimately generate pleiotropic responses in isogenic populations of bacteria. As previously discussed the process of both transcription and translation are subject to the stochastic nature of interacting molecules (section 1.2.1). A landmark study defined experimental conditions to isolate transcriptional and translational noise (Ozbudak *et al.*, 2002). Through varying the efficiency in both these processes Ozbudak *et al.* were able to conclude that translational efficiency was the greatest source of gene expression noise. Their research used *gfpmut2* coding for a green fluorescence protein as their reporter for noise in both transcription and translation. The  $P_{\text{spac}}$  promoter controlled expression of *gfpmut2*.  $P_{\text{spac}}$  is repressed by LacI with *lacI* downstream of the constitutive  $P_{\text{pen}}$  promoter. Tetrameric LacI binds to the operator DNA preventing transcription. Expression of *gfpmut2* occurs on addition of the allosteric inhibitor of LacI, IPTG. IPTG binding causes disassociation of LacI from DNA (Choi *et al.*, 2008). Disassociation allows sigma-factors to bind and recruit the RNA polymerase initiating transcription. The construct was integrated at the *amyE* locus on the chromosome of *B. subtilis* 168 rather than expressing it from a plasmid (Figure 19 A). This avoids potential cell to cell differences in plasmid copy numbers. The fluorescent signal may depend on the copy number of *gfpmut2* reducing the validity of the noise data.

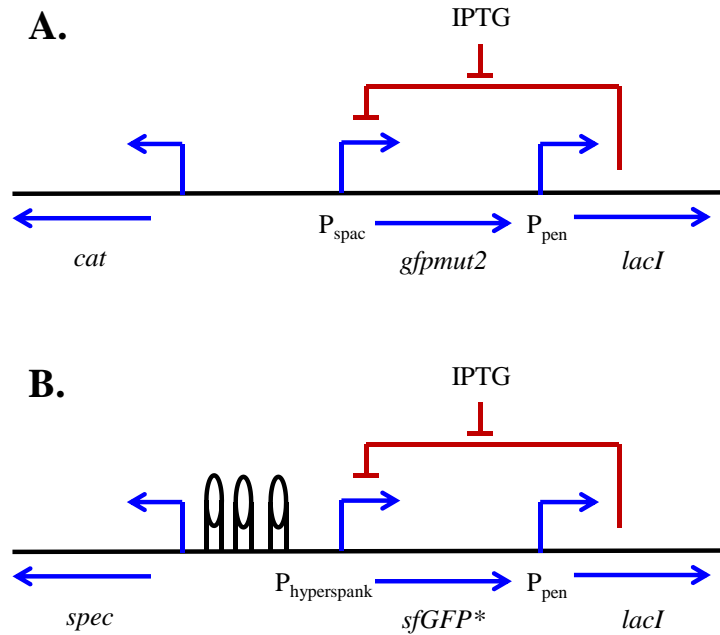


Figure 19. Schematic drawings of the constructs used to quantify noise in gene expression. Right-angled arrows are promoters, straight arrows are genes and T-bars indicate repression. Transcriptional terminators are raised black ovals. Both constructs are integrated at the *amyE* locus in *B. subtilis*. **A**, represents the strain used by Ozbudak et al. (2002). **B**, represents the *sfGFP\** reporter strain used in this thesis.

The experiments of Ozbudak *et al.* used strains with mutations in the RBS or start codon to vary the fluorescent activity. This was quantified using flow cytometry and the mean signal plotted against the noise strength (Figure 20). The difference in noise was defined as the result of varied translational efficiency. Varied inducer concentrations were used with a single strain to determine the noise generated in transcription (Figure 20 **A**). Varying the inducer concentration was defined as varying the transcriptional efficiency. Noise strength was quantified when the cells were in late exponential growth phase at an  $OD_{600} \approx 1.0$ . Use of a single OD provides a snap shot window of noise in the system. By collecting data at a single OD any growth phase effects on gene expression will be obscured. This raises the question of how growth phase effects noise?

The *sfGFP\** reporter used in this thesis (strain AH7,  $P_{\text{hyperspank}}\text{-sfGFP*}\text{-lacI}$ ) has similar properties to the construct used by Ozbudak et al. (Figure 19 **B**).  $P_{\text{hyperspank}}$  is a strong IPTG inducible promoter driving expression of a gene coding for a superfolder green fluorescence protein, *sfGFP\** (Pedelacq *et al.*, 2006). Flow cytometry was used by Ozbudak *et al.* This



powerful tool generates large amounts of statistically significant data. One contention with using flow cytometry to analyse *B. subtilis* stems from the chaining phenotype of non-motile cells during exponential growth. For this reason fluorescence microscopy was used for analysis through exponential, transition phase and into stationary growth phase. Visual images allow Cell Profiler software to analyse single cells within chains. These images can be assessed by eye to validate the data. Whilst Ozbudak *et al.* found noise to be a feature of translational efficiency the initial experiments presented in this results chapter do not address the contribution of translational noise. Instead the focus establishes the contribution of growth phase to gene expression noise. Expectations are that noise will be higher in transition phase than exponential or stationary phase. This is due to the difference in size between exponentially growing cells and stationary phase cells. Exponentially growing cells are larger than stationary phase cells. Transition into stationary phase requires a shift in the global gene expression profile required to achieve this phenotypic change. Global noise contributes to the noise profile of unrelated genes (Pedraza and van Oudenaarden, 2005). The prediction that transition phase will produce the largest noise takes into account the well documented stationary phase heterogeneity (Leisner *et al.*, 2008; Veening *et al.*, 2008). *B. subtilis* has mechanisms to fine tune and amplify noise during stationary phase (Gamba *et al.*, 2015). However the fluorescent reporters used in this project are regulated in isolation by LacI and the gratuitous inducer IPTG. LacI is constitutively expressed from the non-native  $P_{pen}$  promoter. With no metabolism of IPTG the concentration in the batch culture is expected to remain constant throughout the experiment. The theory would argue that promoter activity in this synthetic operon will also be constant. This theory must be experimentally tested.

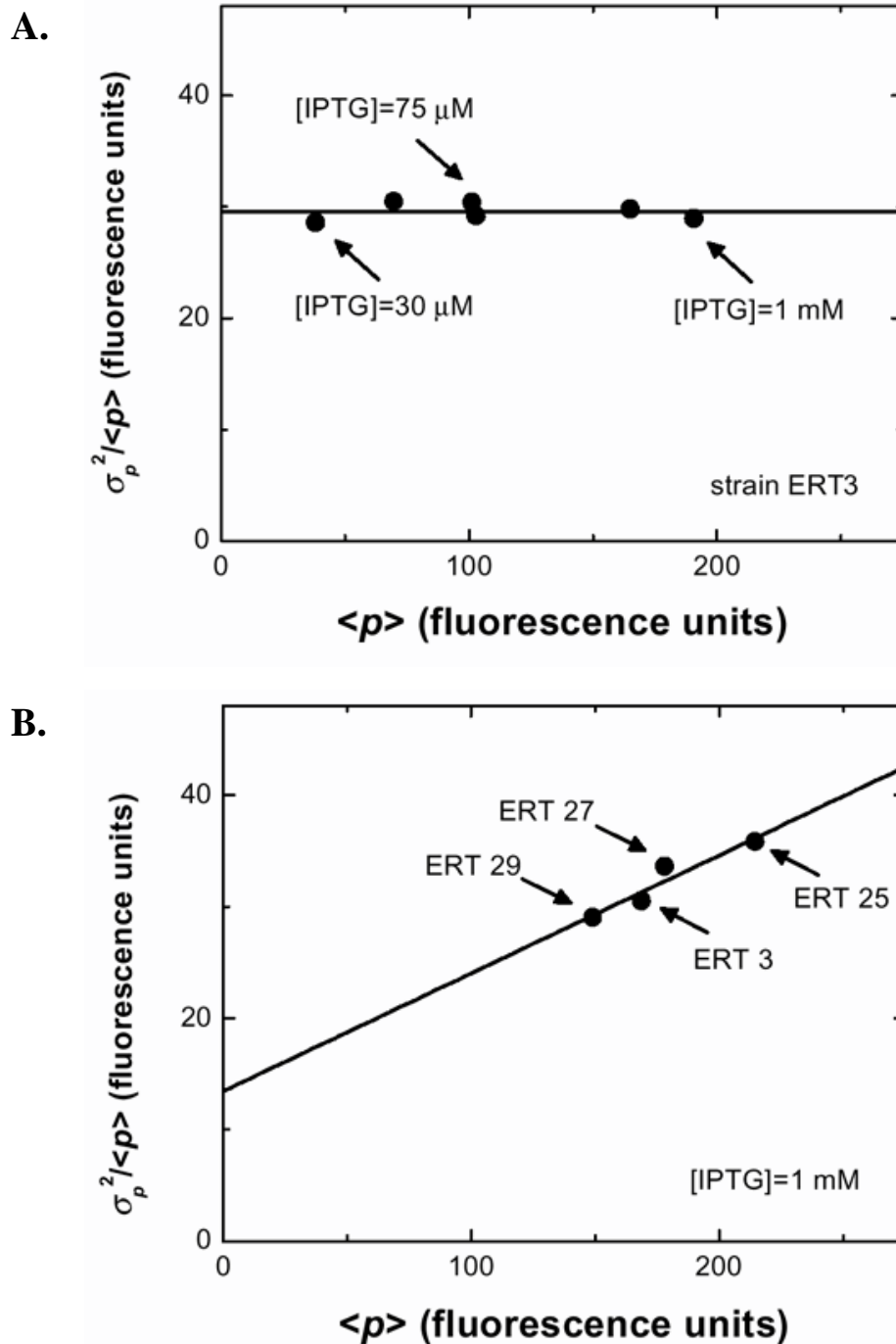


Figure 20. Phenotypic noise isolated from varied transcriptional efficiency and varied translational efficiency. Noise strength is calculated as the standard deviation squared divided by the mean ( $\sigma^2 P / \langle p \rangle$ ). **A**, a single strain was used to calculate noise strength with varied transcriptional efficiency. Induction of *gfpmut2* with varied IPTG concentrations alters the fluorescence output with little effect on the noise strength. **B**, ERT numbered strains contain mutations in the RBS or start codon altering the translational efficiency. Fluorescence output has increased correlation to noise strength. Figure reproduced from Ozbudak *et al.*, 2002.

## 4.2. Results and discussion

### 4.2.1. Growth phase dependent induction *sfGFP\**

To test if maximal noise occurs in transition phase a *sfGFP\** reporter ( $P_{hyperspank}$ -*sfGFP\**-*lacI*) was observed from exponential growth over a 300 minute time course. The standard assay conditions used in this thesis grow an overnight culture from a single colony at 30 °C with antibiotic. Incubation at 30 °C as opposed to 37 °C, together with addition of antibiotic reduce the growth rate. Slow growth reduces sporulation in an overnight culture. The pre-culture allows time for outgrowth of spores, removing this phenotype from the assay culture. The pre-cultures are diluted 50-fold from the overnight cultures into LB at 37 °C containing the required concentration of IPTG. The second dilution to an OD  $\approx$  0.05 corresponds to zero minutes in the standard assay. It is important to note that induction of reporter gene expression occurs on dilution from the overnight culture. This contains cells with varied levels of metabolic activity. The cells are grown through the lag phase and into exponential phase before a second dilution step and the start of assay measurements. Induction at the first dilution avoids potential differences in the uptake of inducer from presenting as gene expression noise.

Preliminary data established that full induction of *sfGFP\** occurs with addition of 1 mM IPTG (Figure 54, page 154). In the *sfGFP\** reporter strain an unexpected growth phase dependent phenotype was observed. Microscope images acquired during the assay showed brightly fluorescing cells at every time point. Cells in exponential phase (Figure 21 **A & B**) appear to have similar fluorescence as those in transition (Figure 21 **C & D**) and those in stationary phase (Figure 21 **E & F**). The analysis software used in this thesis displays the image data with the brightness and contrast automatically adjusted for each image. This facilitates the user to differentiate between cells with similar light intensities. On screen images can appear very similar whilst masking large differences in the actual fluorescence within cells. This is also translated into the print image where further resolution is lost.

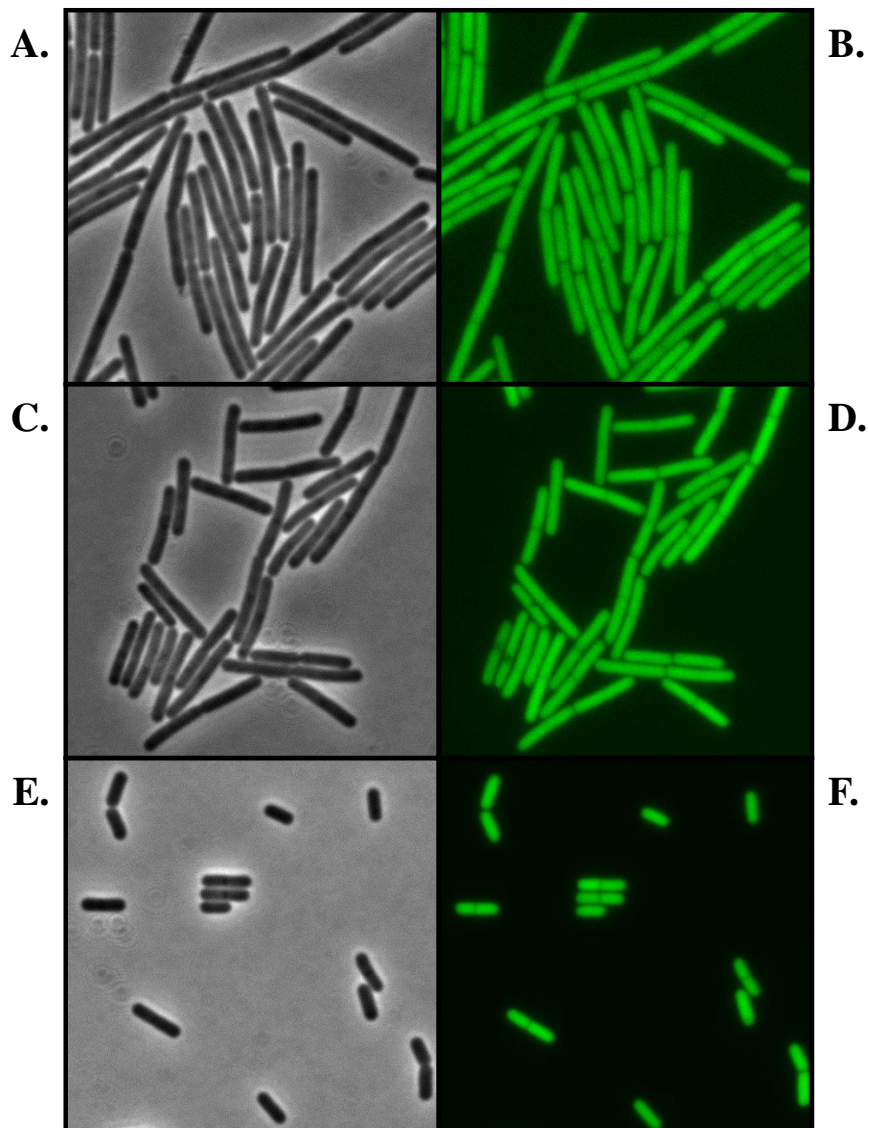


Figure 21. Phase contrast and fluorescent microscope images of the *sfGFP\** strain induced with 1 mM IPTG in a standard assay. Microscope images are taken of the strain (AH7, *Phyerspank-sfGFP\*-lacI*) at 90 minutes (**A & B**), 120 minutes (**C & D**) and 300 minutes (**E & F**). Fluorescent images are false coloured green. Data is representative of n=3 independent repeats.

Fluorescence levels are calculated using Cell Profiler analysis of the standard assays where the *sfGFP\** strain has full or no IPTG induction (Figure 22). The cultures are growing exponentially from the start of measurements at zero minutes until the 90 minute time point (Figure 22 **A**, grey line). This corresponds to a low level of fluorescence in the fully induced condition (Figure 22 **A**, red line). Between 90 minutes and 120 minutes the growth rate reduces as the cultures enter transition phase. Growth rate after 120 minutes continues to

reduce as the cells progress into stationary growth phase. Transition into stationary phase corresponds to the largest increase in fluorescence with full IPTG induction. From 90 to 240 minutes fluorescence increases exponentially and the data points correlate with  $R^2 > 0.99$  (Figure 22 A, inset). From 240 to 300 minutes fluorescence increases, although at a reduced rate. It is not possible to observe these differences under the microscope (Figure 21). Image analysis software is required. These data identify a growth phase dependent sfGFP\* activity when the system is fully induced. Growth phase dependent induction of sfGFP\* is also a feature of the un-induced condition. Without IPTG induction *sfGFP\** cells have a mean fluorescence of 7.4 – 7.6 (a.u) throughout exponential growth, up until stationary phase at 180 minutes (Figure 22 A, blue line). The fluorescence without IPTG induction was not expected to be zero. *B. subtilis* auto fluoresces with light in the GFP channel (see section 4.2.2). In addition to this, promoters using *lac* operator sites are known to be leaky (Choi *et al.*, 2008). This is an expected phenotype of  $P_{\text{hyperspank}}$ . The un-induced conditions at the last two time points have an increase in both the mean fluorescence and standard deviation. This growth phase dependent loss of sfGFP\* regulation was not expected and has implications for the noise profile.

Under full induction conditions there is a lag in fluorescence occurring during exponential growth (Figure 22). Only on transition into stationary phase growth does the fluorescence begin to increase. This suggests a limitation in the experiments of Ozbudak *et al.* They defined an OD  $\approx 1.0$  to be late exponential growth and used it for their noise measurements. Strains in both Ozbudak's experiments and in this thesis were grown in LB at 37 °C. Data in this thesis (Figure 22) identifies that transition phase starts at an OD between 0.769 and 1.354 (90 and 120 minutes). Standard assay conditions used in this thesis determine an OD  $\approx 1.0$  to be within a period of significant change in reporter expression. Ozbudak *et al.*'s use of a single time point does not capture this. Changes in gene expression correspond to an increase in global noise (Pedraza and van Oudenaarden, 2005). Increased global noise may have unforeseen effects, specific to Ozbudak *et al.*'s snapshot of bacterial gene expression. Assays of cultures in exponential, transition and stationary phase provide greater understanding of these growth phase effects.

Noise calculations throughout the standard assay translate heterogeneity into unit free numerical data. Noise is defined as the standard deviation of fluorescence ( $\sigma$ ) divided by the mean fluorescence ( $\mu$ ). When the system is fully induced noise is highest at the zero minute time point (Figure 22 B, red line). As the system stabilises through exponential growth noise

reduces. There was a predicted increase in noise during transition phase which did not occur in these assay conditions. Noise has stabilised by 120 minutes and remains low thereafter. The noise profile without IPTG induction is similar, until the final two time points (Figure 22 **B**, blue line). From zero minutes the system is stabilising and noise reduces, remaining low until 180 minutes. At the final two time points there are high levels of noise and the heterogeneity is visible under the microscope (Figure 22 **C & D**). There is a sub-population of fluorescent cells with the loss of tight LacI regulation. A loss of LacI regulation is an observed feature of the *lac* operon in *E. coli* (Choi *et al.*, 2008). However, any translation to *B. subtilis* is hypothetical. The organisms are different, as are the promoters contained within them. It is currently unclear why the noise observed without IPTG induction is limited to stationary phase.

The noise data without IPTG induction suggests a reason for the increased noise at the start of the assay. Heterogeneity in fluorescence can be expected in the overnight culture. These cells are in stationary phase. They have ~ 2 hours of growth in which this fluorescence can be diluted out to reduce cell to cell differences. In addition to this each dilution step requires the bacteria to respond to new environmental conditions. This is expected to produce a global spike in noise. The second dilution step will produce noise and the heterogeneity in fluorescence will be diluted out through multiple rounds of cell division.

Data presented in this section is contrary to the prediction that high noise would be a feature of transition phase. The critical aspect of noise in this system occurs without IPTG induction. At the latter stages of the assay, noise is generated through a loss of sfGFP\* regulation. Loss of regulation appears to be related to the time a culture spends in stationary phase. It is important to consider this in the context of the fluorescence during the whole of the assay. It is not yet clear if the low levels of fluorescence at other time points in the assay are caused by a loss of sfGFP\* regulation or if it is caused by auto fluorescence of *B. subtilis*.

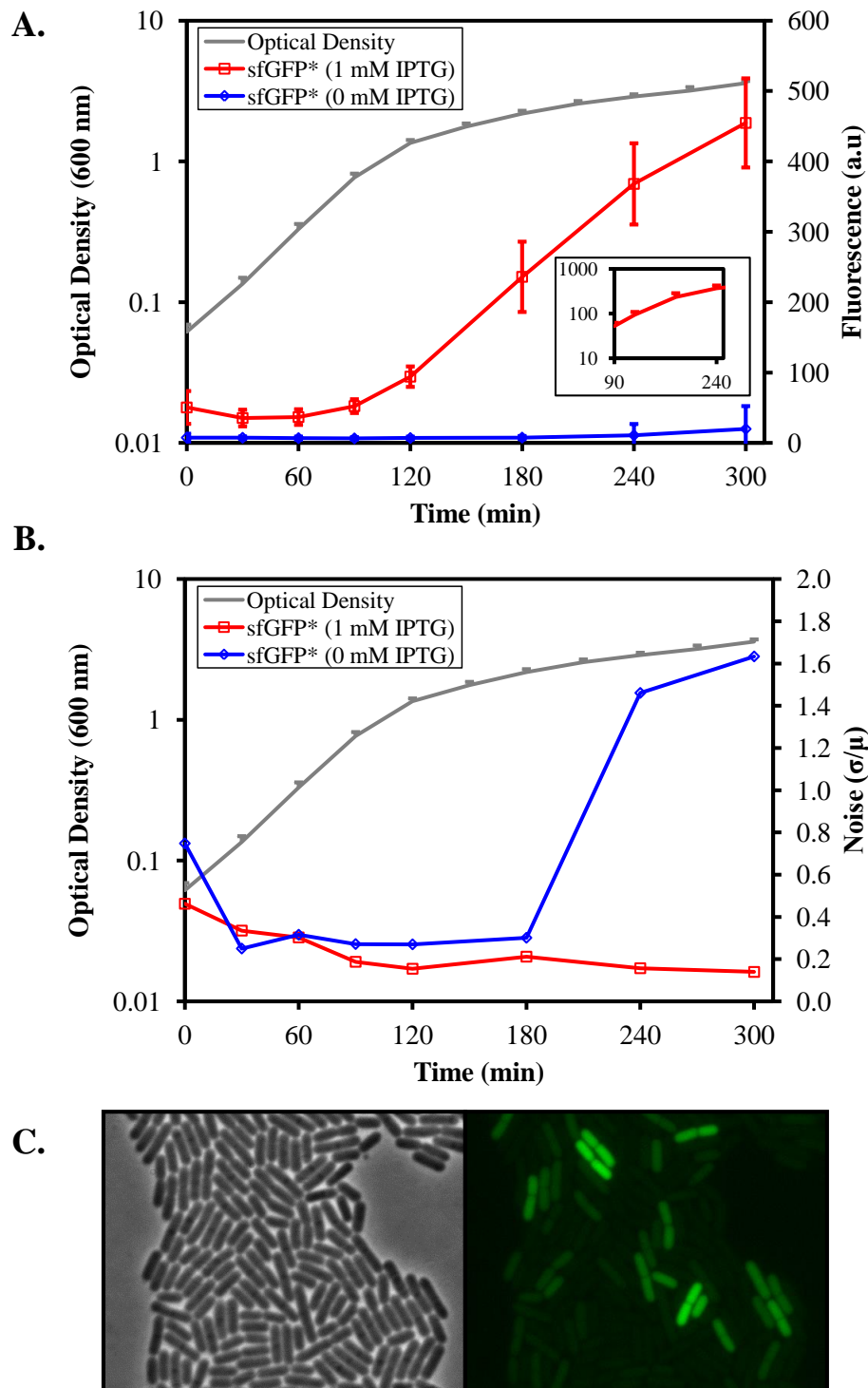


Figure 22. Growth phase effects fluorescence and heterogeneity in the expression of a single gene. **A**, the *sfGFP\** reporter strain (AH7,  $P_{hyperspank}$ -*SfGFP\**-*lacI*) was induced with 1 mM IPTG (red line) and without IPTG induction (blue line) in standard assays. Grey line, mean optical density of bacterial cultures with error bars of the standard deviation. Inset, increase in *sfGFP\** fluorescence is exponential between 90 and 240 minutes. **B**, heterogeneity in the population is quantified as noise. Noise is calculated as the standard deviation ( $\sigma$ ) divided by the mean ( $\mu$ ). **C**, phase contrast image of the *sfGFP\** strain at the 300 minute time point without IPTG induction. **D**, *sfGFP\** fluorescence in the cells corresponding to C. Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 9, page 160.

#### 4.2.2. *Background fluorescence in Bacillus subtilis*

Control strain  $\Delta sfGFP$  (TE48,  $P_{hyperspank}-\Delta sfGFP lacI$ ) was used to identify the background fluorescence of *B. subtilis*. The  $\Delta sfGFP$  strain was also used to identify background noise levels. Background noise will be generated through auto fluorescence of *B. subtilis*, the process of image acquisition and the process of image analysis. These components can be grouped together and described as experimental noise. Assays of the  $\Delta sfGFP$  strain containing 1 mM IPTG were compared to cultures with no IPTG induction (Figure 23). Any fluorescence in the  $\Delta sfGFP$  strain cannot be sfGFP\* fluorescence and must be due to *B. subtilis* auto fluorescing or noise in the acquisition/analysis of the image data. Mean background fluorescence ranged from 5.8 (a.u) to 9.9 (a.u) in both conditions (Figure 23 **A**). For the purposes of subsequent image analysis sfGFP\* fluorescence values  $\leq 10$  will be considered background. The noise in background fluorescence was relatively constant throughout the assays with increased variability on addition of 1 mM IPTG (Figure 23, **B**). It is very likely the highest noise level measured was generated, for the most part, through image acquisition and analysis rather than reflecting the contribution of *B. subtilis* auto fluorescence (Figure 23. Red line, 240 minutes). The mean noise over the entire assays was 0.34 with 1 mM IPTG and 0.34 without IPTG. These data provide a range of fluorescence and noise values which can be expected when a *sfGFP\** reporter strain is not induced with IPTG. Fluorescence  $> 10$  a.u is not due to *B. subtilis* auto fluorescence. Noise levels close to 0.34 can be attributed to system noise, not gene expression noise.



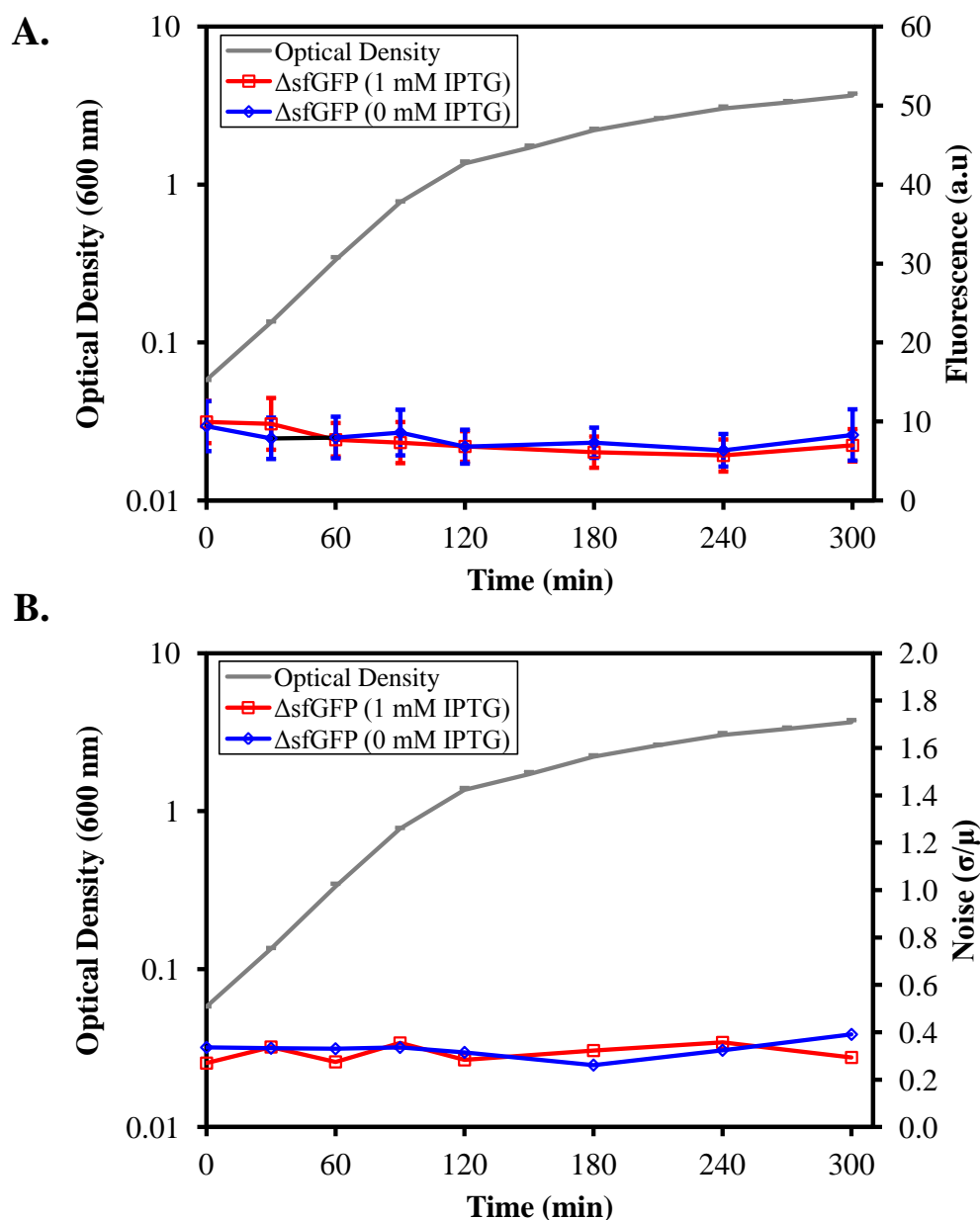


Figure 23. Background fluorescence and noise in the  $\Delta sfGFP$  control strain (TE48, *Phyperspank-ΔsfGFP\* lacI*). Strains were induced in standard assay conditions (page 30) with 1 mM IPTG (red lines) and 0 mM IPTG (blue lines). Growth of the cultures are measured by the optical density (grey line). Error bars are the standard deviation. **A**, mean fluorescence of control strains with and without the addition of IPTG. **B**, heterogeneity in strain TE48 +/- 1 mM IPTG induction. Heterogeneity in the population is quantified as noise. Noise is calculated as the standard deviation ( $\sigma$ ) divided by the mean fluorescence ( $\mu$ ). Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 10, page 161.

Data from the  $\Delta sfGFP$  control strain can be used to identify the overlap between experimental noise and gene expression noise. The mean fluorescence in the *sfGFP\** strain without IPTG induction has been plotted against the  $\Delta sfGFP$  control strain (Figure 24 A). As the data for  $\Delta sfGFP$  was comparable with and without IPTG an  $n=6$  mean was used to

represent the  $\Delta sfGFP$  strain. Both the  $sfGFP^*$  and  $\Delta sfGFP$  strains produced similar levels of fluorescence up until the 180 minute time point (Figure 24, red line and blue lines respectively). From zero to 180 minutes the mean fluorescence is within the range expected for background values ( $\leq 10$ ). For this reason it is not appropriate to use statistical analysis in comparison of the two data sets. Analysis would only identify if experimental noise was different in the two populations. Only at the last two time points does the mean increase beyond 10. Welch two-sample t tests were used to confirm a difference in the mean values. The assay conditions have eight time points where fluorescence is calculated. For this reason a  $p$ -value  $< 0.00625$  ( $0.05/8$ ) would be sufficient to reject the null hypothesis at a 95% confidence interval. The actual  $p$ -values for the 240 minute and 300 minute time point were both  $< 2.2^{16}$ . In isolation this fluorescence data suggests deregulation of  $sfGFP^*$  only occurs at the last two time points. However, through comparison of the noise data it becomes clear there is also gene expression noise at the zero minute time point (Figure 24, **B**). The  $\Delta sfGFP$  strain has similar noise levels throughout the assays. By comparison the  $sfGFP^*$  strain has increased noise at the zero minute time point. The noise data from 30 minutes to 180 minutes is similar in both the  $\Delta sfGFP$  strain and the  $sfGFP^*$  strain. Noise in the  $sfGFP^*$  strain is within the background range expected for an un-induced condition. Loss of  $sfGFP^*$  regulation is clearly apparent at the last two time points. Use of the  $\Delta sfGFP$  control strain provides clarity in expected levels of background fluorescence and noise. Together these data validate the phenotype of the  $sfGFP^*$  strain without IPTG induction. Noise is a genuine feature of the un-induced condition at zero, 240 and 300 minutes.

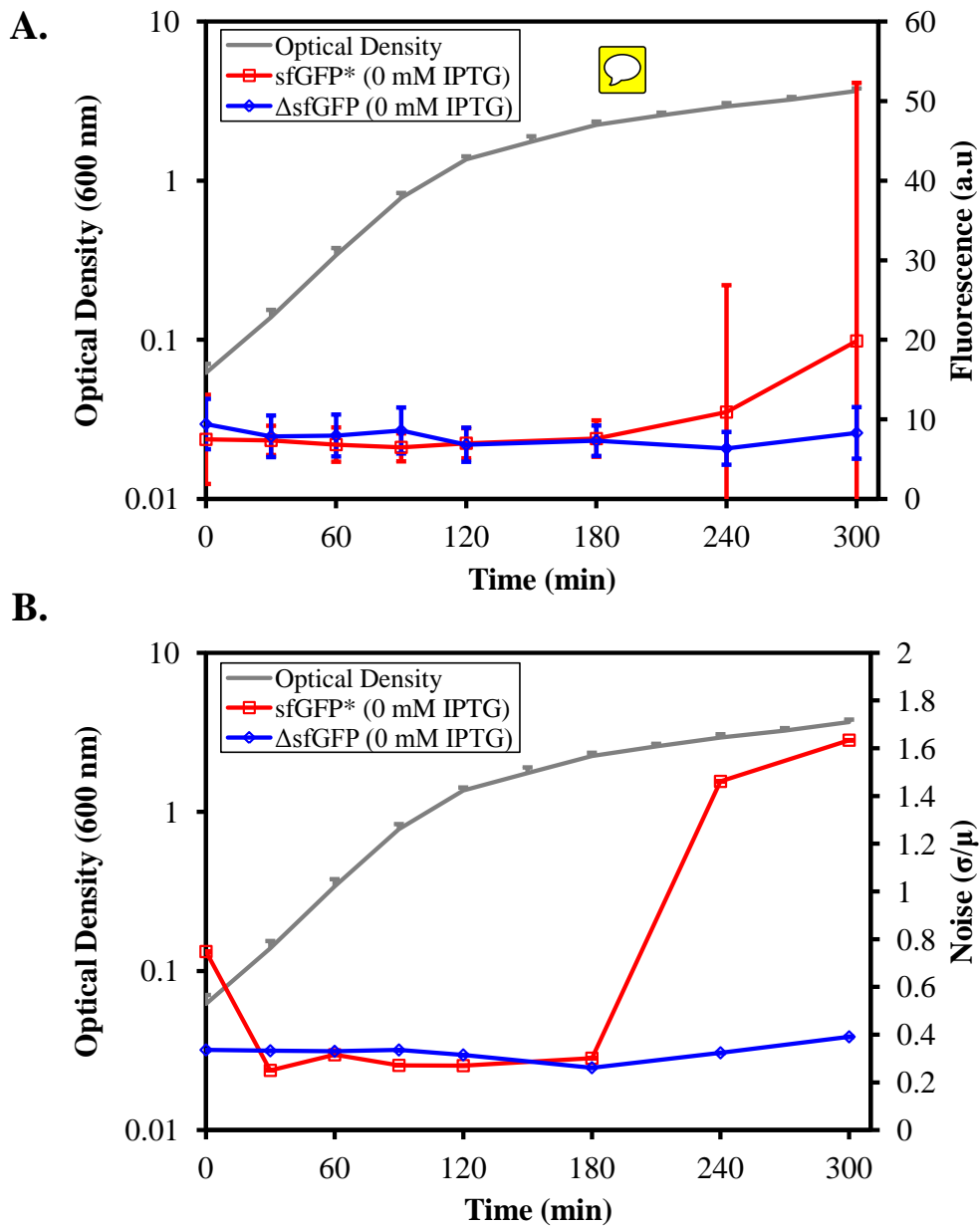


Figure 24. Fluorescence and noise in the sfGFP\* reporter strain. The sfGFP\* reporter (AH7, *P<sub>hyperspank</sub>-sfGFP\* lacI*) is compared to the  $\Delta$ sfGFP\* control strain (TE48, *P<sub>hyperspank</sub>- $\Delta$ sfGFP\* lacI*). The sfGFP\* strain has no IPTG induction and  $\Delta$ sfGFP\* strain combines data from inducing and non-inducing conditions (Figure 23). Grey lines, growth rate of the cultures measured by the optical density. Error bars are the standard deviation from the mean. **A**, fluorescence in strain AH7 and control strain TE48. **B**, heterogeneity in AH7 and TE48 are quantified as noise. Noise is calculated as the standard deviation ( $\sigma$ ) divided by the mean fluorescence ( $\mu$ ). Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 11, page 161.

### 4.2.3. *Inducer concentration affects fluorescence levels and noise*

Previous modelling and experimental data have suggested that little noise is generated in transcription (Ozbudak *et al.*, 2002). Data presented in this thesis (section 4.2.1) identifies the largest source of heterogeneity in sfGFP\* expression occurs in stationary phase without IPTG induction (Figure 22). Transcription in this condition is a source of gene expression noise. Stationary phase cells (at 240 and 300 minutes) have 4-5 fold higher noise than basal levels. One of the questions addressed in this section is what happens to the heterogeneity at intermediate rates of transcription? The *sfGFP\** reporter strain has been fully induced with 1 mM IPTG. Varying the concentration of IPTG varies the rate of transcription. This data will also address the exponential phase lag in fluorescence (Figure 22, red line). More specifically, how is the lag effected by the concentration of IPTG? Fluorescence levels increase significantly between 90 minutes and 120 minutes with full induction. Without IPTG induction there is an increase in fluorescence between 180 minutes and 240 minutes. Therefore, will the lag time be affected by intermediate levels of IPTG induction?

The intermediate concentrations of IPTG used in these experiments were selected to provide data on full induction (1 mM IPTG), two conditions between basal levels and full induction (0.1 mM and 0.05 mM IPTG) and one condition below basal levels of induction (0.01 mM IPTG). This provides a: 10-fold, 20-fold and 100-fold reduction in IPTG concentration. At all induction conditions the *sfGFP\** strain exhibited a lag in fluorescence during exponential growth (Figure 25, **A**). The mean fluorescence did not increase above that of the zero minute time point until after 90 minutes. By 120 minutes the lag in fluorescence has ended and fluorescence increases at all levels of induction. The 90 – 120 minute time points correspond to the end of exponential growth and transition phase. At the lowest IPTG concentration fluorescence was close to background and ~ 11 a.u at 120 minutes. By 300 minutes fluorescence had increased to ~ 45 a.u. Varied IPTG concentration has a non-linear effect on the output of the system (Figure 25 **C**). The magnitude of fluorescence is altered without removing the growth phase dependent activity.

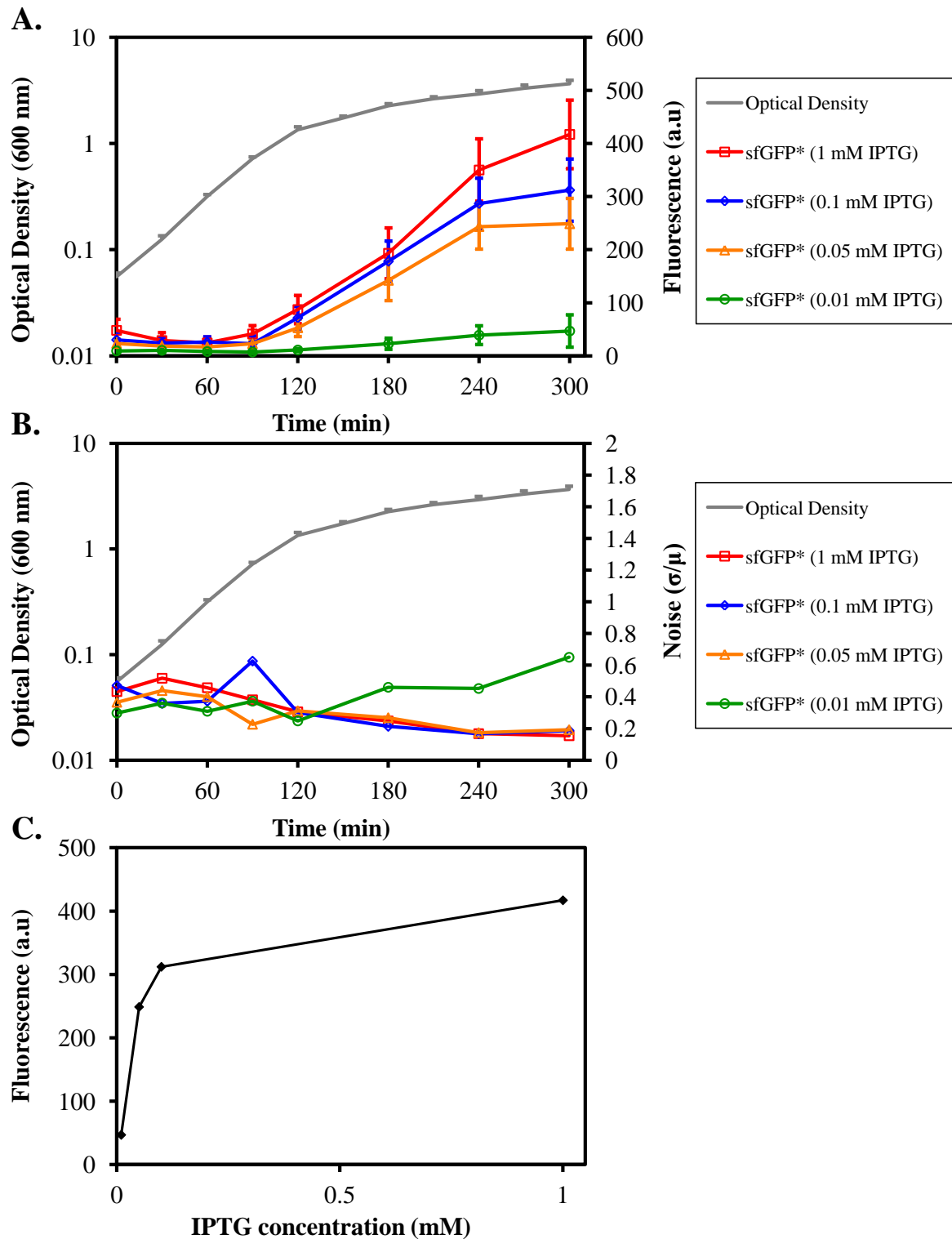


Figure 25. Transcription rate alters the fluorescence and noise in the *sfGFP\** reporter strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*). Varied IPTG concentrations were added to *sfGFP\** cultures in standard assay conditions. **A**, mean fluorescence and optical density. **B**, noise, calculated as the standard deviation ( $\sigma$ ) divided by the mean fluorescence ( $\mu$ ) and optical density. **C**, IPTG effects on fluorescence at 300 minutes. Red lines, 1 mM IPTG. Blue lines, 0.1 mM IPTG. Orange lines, 0.05 mM IPTG. Green lines, 0.01 mM IPTG. Grey lines, optical density. Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 12, page 161.

Varying the concentration of IPTG alters the rate of transcription, resulting in a difference in the fluorescent activity of the reporter. With the exception of the 0.01 mM IPTG condition varying the rate of transcription makes little difference to the noise profile (Figure 25 B). In every condition noise stabilises through exponential growth. At 120 minutes noise is below background levels (0.34) for all IPTG concentrations assayed. It is only the 0.01 mM condition where noise increases from 120 minutes until the final time point. At other IPTG concentrations noise is reduced below background levels. The noise profile at 0.01 mM IPTG is closest to the 0 mM IPTG condition (Figure 25 B & Figure 22 B respectively). The stationary phase noise is present in both conditions. Addition of 0.01 mM IPTG reduces noise by ~ 50% by the 300 minute time point. At intermediate levels of IPTG induction the noise phenotype is lost altogether.

#### **4.2.4. Does protein stability affect fluorescence levels and noise in sfGFP\* reporter strains?**

Data presented in this chapter (section 4.2.1) has established two growth phase dependent phenotypes in strain AH7. A lag in fluorescence occurs during exponential growth. An increase in stationary phase noise is observed without IPTG induction. This section will address the effects of protein stability on both of these phenotypes. Strains with unstable sfGFP\* were used for comparison with the sfGFP\* reporter. The translated sfGFP\* includes a tag targeting the protein for degradation with ClpXP (Gottesman *et al.*, 1998; Wiegert and Schumann, 2001). The sequence, known as an SsrA-tag is comprised of an 11 amino acid carboxy-terminal peptide. Use of the native *ssrA* prevents any visible fluorescent signal during any time in a standard assay with full induction (Syvertsson, 2013). Proteolysis mediated by this SsrA tag is too effective for use in this system. Mutations in the final four residues of the tag alter the efficiency of degradation. The low stability sfGFP\* strain (AH5, *P<sub>hyperspank</sub>-sfGFP\*-ssrA(AAAV)-lacI*) contains the mutated *ssrA* tag with the carboxy-terminal amino acids AAAV and the medium stability sfGFP\* strain (AH6, *P<sub>hyperspank</sub>-sfGFP\*-ssrA(ADAV)-lacI*) translates ADAV at the caboxy-terminus.

Under full induction conditions the growth phase dependent phenotype was lost in strains with reduced sfGFP\* stability (Figure 26 A). In the sfGFP\* reporter (Figure 26 A, red line) fluorescent activity reduces from the zero to 30 minute time point. It remains low during exponential growth, increasing in transition phase and continuing to increase throughout stationary phase. The medium stability sfGFP\* strain exhibited the same profile of

fluorescence during exponential growth, with a reduced magnitude of activity (Figure 26 A, green line). The medium stability *sfGFP\** strain produced a different profile in stationary phase. A relatively constant level of activity was maintained at ~ 6-fold higher than during exponential growth. The low stability *sfGFP\** strain also exhibited the same profile of fluorescence during exponential growth, with a further reduced magnitude of activity (Figure 26 A, purple line). Stationary phase activity in this strain was not maintained. Fluorescence in the low stability *sfGFP\** strain peaked at 120 minutes and slowly reduced to background levels by 300 minutes. These data provide evidence of a growth phase dependent phenotype which is partially dependent on protein stability. Throughout exponential growth and into transition phase all three strains produced the same fluorescence profile. The exponential phase lag was still present, however, the measurable increase in fluorescence was limited to a single time point in transition phase. These data suggest that in stationary phase the increased fluorescence in the original *sfGFP\** reporter strain is due to an accumulation of a stable fluorescent protein.

The stability of *sfGFP\** makes little difference to the noise produced under full induction conditions (Figure 26, C). In all strains noise reduces through exponential growth and remains low throughout transition and stationary phase. Noise is only above background levels at two points. In the medium stability *sfGFP\** strain noise above background at 180 minutes can be attributed to noise in image acquisition and analysis, rather than genuine gene expression noise (Figure 26, C, green line). In the *sfGFP\** reporter strain noise is above background at the zero minute time point (Figure 26, C, red line). This strain produces a stable *sfGFP\**. Any heterogeneity at the zero minute time point must be diluted out. Strains with mutated *ssrA*-tags are able to remove heterogeneity by degrading *sfGFP\** in addition to diluting it. Higher noise levels at the start of the assay would therefore be expected with the *sfGFP\** reporter strain.

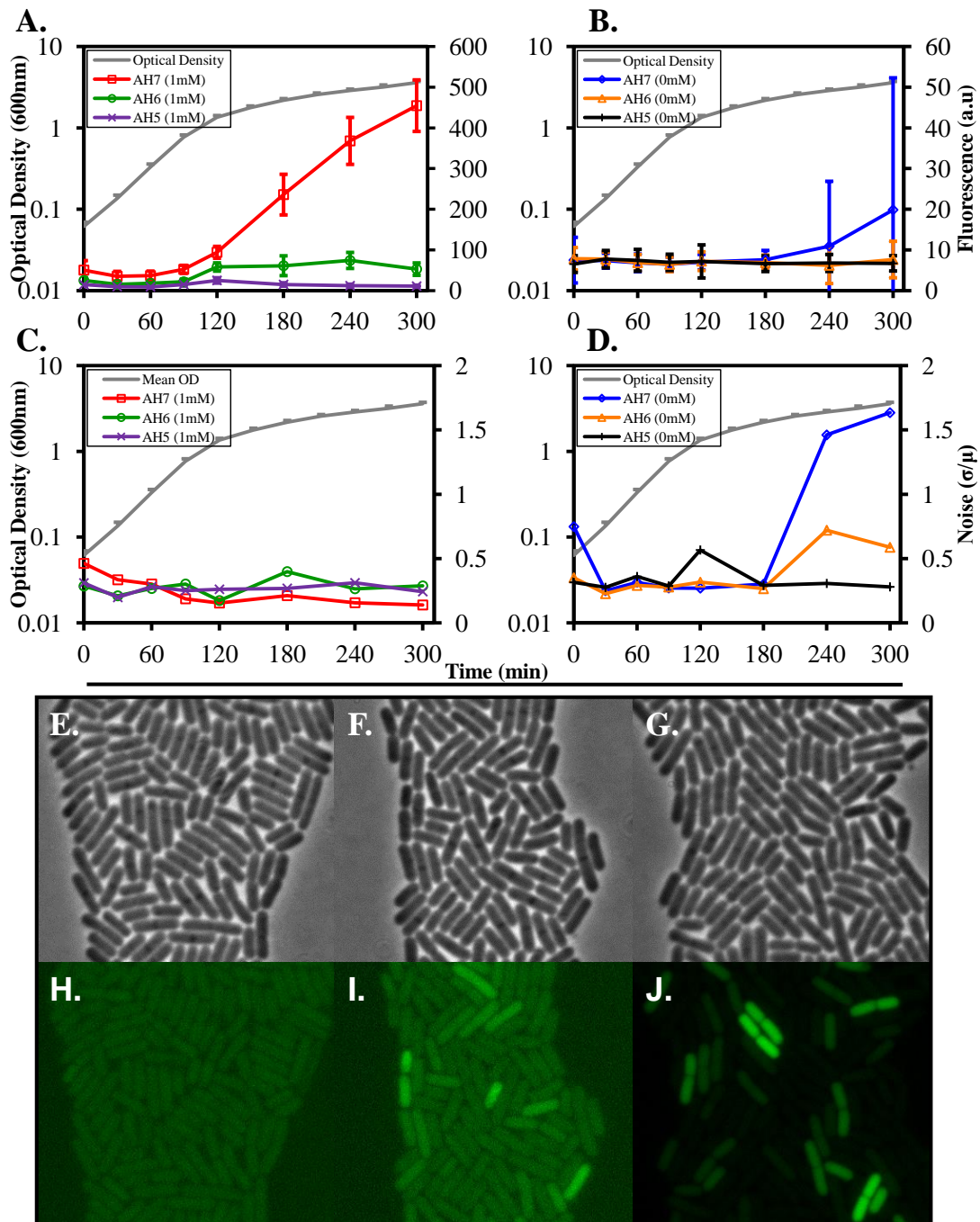


Figure 26. Protein stability alters sfGFP\* activity and noise. The *sfGFP\** reporter strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*) is compared to strains with mutated *ssrA*-tags, targeting sfGFP\* for proteolysis. **A**, fluorescence and growth in strains fully induced with 1 mM IPTG. **B**, Fluorescence and growth is strains with no IPTG induction. **C**, Noise in fully induced strains. **D**, noise in strains with no IPTG induction. Red and blue lines, the *sfGFP\** reporter strain. Green and orange lines, the medium stability *sfGFP\** (AH6, *P<sub>hyperspank</sub>-sfGFP\*-ssrA(ADAV)-lacI*) strain. Purple and black lines, the low stability *sfGFP\** strain (AH5, *P<sub>hyperspank</sub>-sfGFP\*-ssrA(AAAV)-lacI*). Grey lines, mean optical density of all strains. **E-F**, phase contrast images of the low stability, medium stability and *sfGFP\** reporter strains respectively. **H-J**, fluorescent images corresponding with **E-F**. False coloured green. Images E-J were captured at 300 minutes. Cell numbers corresponding to n=3 independent repeats are recorded in Table 13, page 162.



Without IPTG induction the stability of sfGFP\* affects the fluorescent profile in standard assay conditions (Figure 26, **B**). All strains have background levels of fluorescence ( $\leq 10$  a.u) from zero to 180 minutes. The *sfGFP\** reporter has increased mean fluorescence with large standard deviations at the final two time points. Mean fluorescence in the medium stability *sfGFP\** strain does not rise above background levels, however the size of standard deviation bars indicates that some of the cells are fluorescing. At no point does the low stability *sfGFP\** strain produce measurable fluorescence. This induction condition was used to establish the effects of protein stability on gene expression noise. In the *sfGFP\** reporter strain there is expected noise at the zero minute time point in the assay (Figure 26 **D**, blue line). At 120 minutes there appears to be high noise in the low stability *sfGFP\** strain (Figure 26 **D**, black line). No fluorescence was visible at this time point. This noise is therefore system noise, rather than genuine gene expression noise. Excluding these two specific points, noise is at or below background levels for all strains from zero minutes until 180 minutes. At the final two time points the low stability *sfGFP\** strain produced no noise (Figure 26 **D**, black line). The noise phenotype was present in the medium stability strain, with a  $> 2$ -fold reduction compared to the *sfGFP\** reporter (Figure 26 **D**, orange and red lines respectively). Heterogeneity in fluorescence is clearly absent from the low stability *sfGFP\** strain at 300 minutes (Figure 26 **E & H**). It is visible in the medium stability *sfGFP\** strain and well defined in the *sfGFP\** strain (Figure 26 **F - J**). Together these data identify a contributing factor in the stationary phase noise phenotype. Protein stability determines the extent of quantifiable noise observed in *sfGFP\** gene expression.

### 4.3. Conclusions

This chapter investigated heterogeneity through exponential, transition and stationary phase growth. Two phenotypes were identified which had an apparent growth phase dependence in the *sfGFP\** strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*). When the system was fully induced with 1 mM IPTG there is a lag in fluorescence during exponential growth. Fluorescence only increases significantly from transition phase and into stationary phase, where it continues to increase. The second phenotype occurred without IPTG induction. Heterogeneity was observed in stationary phase at the final two time points of the assay. The loss of regulation is observable as a fluorescent subpopulation under the microscope and with the high noise values calculated from the data. Caution is therefore advised before use of IPTG inducible

promoters in research. Quantifying the output of the system based on expression levels must first consider the effects of growth. It is also important to recognise the limitations of a system where the “OFF” state is not fully achieved during stationary phase. Toxicity of an expressed protein may determine the suitability of use with an IPTG inducible system.

To investigate the two phenotypes induction was altered through addition of varied IPTG concentrations. This varied the rate of transcription and the fluorescent output of the system. These data confirmed that the transcription rate does not affect the growth phase dependent lag in fluorescent activity. At each induction condition there was low levels of fluorescence during exponential growth followed by an increase in transition phase, with all induction conditions there was an increase in fluorescence throughout stationary phase. Transcription rate altered the magnitude of this increase and intermediate levels of IPTG induction did not alter the noise profile. Only at the lowest IPTG concentration (0.01 mM) was noise in the system increased. Addition of 0.01 mM IPTG produced increased stationary phase noise at a ~ 2-fold reduction compared to the 0 mM IPTG condition. Thus increasing the rate of transcription through addition of varied IPTG concentrations reduces or removes the stationary phase noise phenotype.

The contribution of sfGFP\* stability on the two phenotypes was addressed through use of *ssrA*-tags with altered capacity to target sfGFP\* for proteolysis. Protein stability was demonstrated to alter the heterogeneity found without IPTG induction at the later time points in the assay. The stable *sfGFP\** strain produced the highest noise. The medium stability *sfGFP\** strain produced intermediate levels of noise and the low stability *sfGFP\** strain produced background levels of fluorescence and noise. The contribution of protein stability on the lag in fluorescence was less clear. Protein stability affected the magnitude of fluorescence throughout the assays. However, the fluorescence profile during exponential growth up to transition phase at 120 minutes was remarkably similar, regardless of protein stability. During stationary phase there was no significant increase in fluorescence with either of the *ssrA*-tagged *sfGFP\** strains. The conclusion from this data is that protein stability determines the accumulation of fluorescence in stationary phase and the level of stationary phase noise in the system. It does not explain if the two phenotypes occur due to choice of regulator, promoter or reporter so further investigation of this system is required.

## Chapter 5. LacI regulation

### 5.1. Introduction

In this thesis a negatively regulated system is used to investigate heterogeneity in gene expression. The repressor protein LacI binds to two operator regions in the  $P_{\text{hyperspank}}$  promoter of the *sfGFP\** reporter strain (AH7,  $P_{\text{hyperspank}}\text{-sfGFP*}\text{-lacI}$ ) (Figure 27 A). LacI binding obstructs sigma-factor binding and/or recognition thus preventing subsequent recruitment of RNA polymerase. LacI binds as a tetramer of two identical dimers (Figure 27 B). Each of the four LacI molecules contains one allosteric binding site, accommodating one molecule of IPTG (Choi *et al.*, 2008). LacI production in this system is under the control of the  $P_{\text{pen}}$  promoter, originally cloned from *Bacillus licheniformis*. In its native configuration  $P_{\text{pen}}$  is negatively regulated by the repressor protein PenI (Wittman and Wong, 1988). Absence of PenI from *B. subtilis* 168 CA strains results in the constitutive expression of *lacI* in the *sfGFP\** reporter strain. The strain used by Ozbudak *et al.* to investigate transcriptional and translational noise also utilised  $P_{\text{pen}}$  to drive expression of *lacI* (see introduction and section 4.1). In order to make a direct comparison the *sfGFP\** reporter strain was designed to express *lacI* from the same locus as Ozbudak's strain: downstream of the *sfGFP\** reporter gene. There is an absence of a known transcriptional terminator between the 3' end of the reporter gene and the  $P_{\text{pen}}$  promoter driving *lacI* expression. This makes it likely that some read through transcription of *lacI* from the reporter will occur. In such case the addition of IPTG will control both the expression of the reporter and the repressor simultaneously. This construct thus generates negative auto-regulation (Figure 28 A) that can reduce extrinsic sources of noise in transcription (Dublanche *et al.*, 2006). The advantage of negative auto-regulation is that it allows stricter control of the off state. Unwanted transcription will increase the repressor concentration, providing tighter control of gene expression. However, this system generates unwanted problems in the *sfGFP\** reporter strain as the *sfGFP\** activity is linked to growth phase (Figure 22, page 70). There are low levels of fluorescence during exponential growth. Fluorescence then increases dramatically during transition phase and continues to increase during stationary phase. It would theoretically follow that there will be lower levels of LacI during exponential growth and higher levels in stationary phase. The question must be asked, does LacI contribute to the growth phase dependent activity observed with the *sfGFP\** reporter strain?

**A.**

```

1   GGGTAAATGTGAGCACTCACAATTCATTTTGCAAA
           O2
36  AGTTGTTGACTTTATCTACAAGGTGTGGCATAATG
           -35                               -10
71  TGTGTAATTGTGAGCGGATAACAATTAAGCTTTAA
           +1                               O1   RBS
106 GGAGGAGCTAGCATG
           RBS                               Start

```

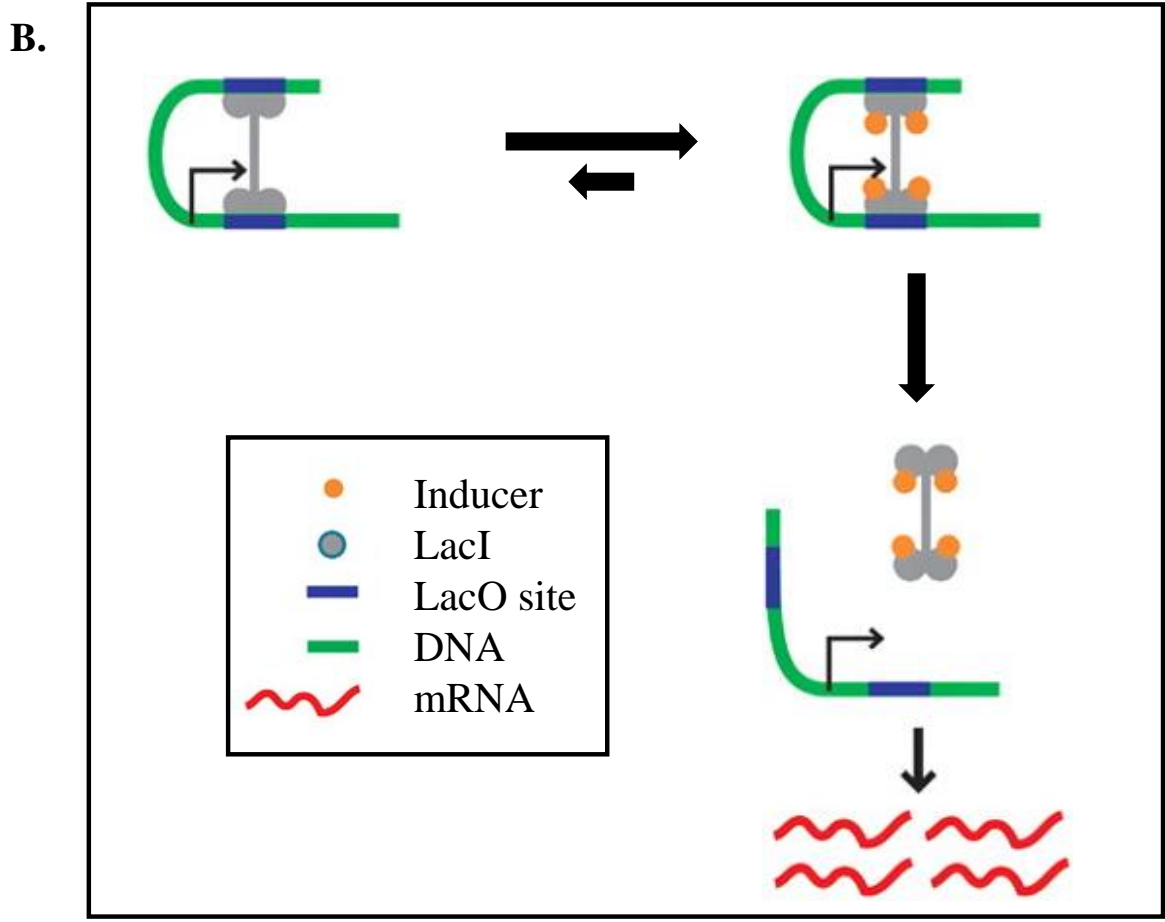


Figure 27. Tetrameric LacI binding to operator sites in  $P_{\text{hyperspank}}$  regulates transcription. **A**,  $P_{\text{hyperspank}}$  promoter sequence. Features are underlined and the transcriptional start site is in bold. RBS and start codon are also included. **B**, Inducer (IPTG) binding at the allosteric sites reduces LacI affinity for the operators causing disassociation from DNA. mRNA synthesis then occurs. Adapted from Choi *et al.*, 2008.

A simple strategy to investigate LacI regulation is to express the reporter on a separate transcript from the repressor. This will establish the contribution of repressor location on both the growth phase dependent activity and noise phenotypes in the *sfGFP\** strain. To accommodate new strains containing the *sfGFP\** gene the nomenclature must be changed. In this chapter the original *sfGFP\** reporter strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*) will be referred to as the *lacI* strain (Figure 28 A). A second strain, referred to as *lacI-cis*, (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) was constructed to remove any possibility of reporter/repressor co-transcription (Figure 28 B). In this strain *lacI* is separated from *sfGFP\** by three transcriptional terminators and expressed divergently from the reporter. The construct contains the *ilvD* transcriptional terminator downstream of *sfGFP\** (de Hoon *et al.*, 2005). This is designed to prevent read-through transcription into native *B. subtilis* DNA, providing tight control over the synthetic construct. This construct was then used to identify if negative auto-regulation is responsible for the growth phase dependent fluorescence observed in 0.

The regulator proximity to its target is also an important consideration for an investigation into heterogeneity in gene expression. There are a number of examples in prokaryotic cell biology where trans-acting regulators control gene expression (Henkin *et al.*, 1991; Fajardo-Cavazos and Nicholson, 2000; Inacio and de Sa-Nogueira, 2007). This has particular relevance to *B. subtilis*. *B. subtilis* utilises temporal changes in noise to determine cell fate (section 1.1.3). A commitment to differentiate into a genetically competent state is triggered by a fine tuned stochastic process (Leisner *et al.*, 2008; Veening *et al.*, 2008). There are many effector molecules contributing to this bistable phenotype. ComS is one such protein, responsible for the up-regulation of the master regulator of competence, ComK (Hamoen *et al.*, 2003b). ComS is a trans-acting molecule located approximately 726 Kb distant from ComK (Kunst *et al.*, 1997). Since diffusion is a stochastic process, trans-acting regulation may increase gene expression noise to facilitate stochastic differentiation. The *lacI-trans* strain (CJ3, *P<sub>hyperspank</sub>-sfGFP\* lacI\_trans*) was constructed to investigate how proximity of the regulator gene affects heterogeneity in the reporter activity (Figure 28 C). In this strain *P<sub>pen-lacI</sub>* was integrated at the *aprE* locus approximately 776 Kb from the reporter, *sfGFP\**. To remain consistent the reporter transcript is identical to the *lacI-cis* strain. It contains three transcriptional terminators upstream of the *P<sub>hyperspank</sub>* promoter and the *ilvD* transcriptional terminator downstream of *sfGFP\**.

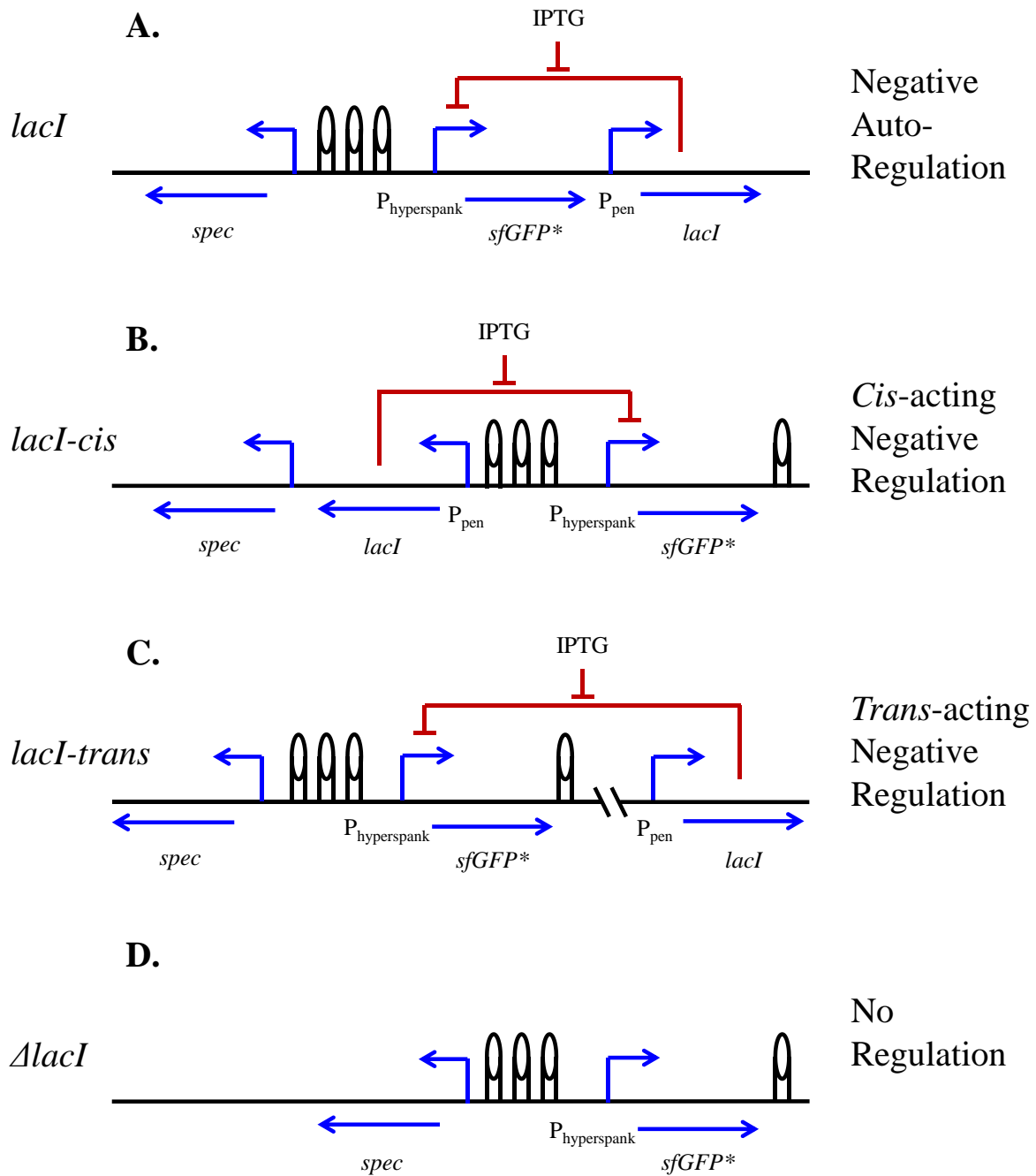


Figure 28. Schematic drawing of the constructs used to investigate repressor location effects. Right-angled arrows are promoters, straight arrows are genes and T-bars indicated repression. Transcriptional terminators are raised black ovals. Constructs are integrated at the *amyE* locus with the exception of *P<sub>pen</sub>-lacI* in C. Parallel break lines denote an approximate 776 Kb distance where *P<sub>pen</sub>-lacI* is integrated at the *aprE* locus. **A**, the *lacI* strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*). **B**, the *lacI-cis* strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*). **C**, the *lacI-trans* strain (CJ3, *P<sub>hyperspank</sub>-sfGFP\* lacI\_trans*). **D**, the unregulated  $\Delta lacI$  strain (CJ1, *P<sub>hyperspank</sub>-sfGFP\*  $\Delta lacI$* ).

To complete the investigation into repressor location it is important to know what happens when there is no regulation (Figure 28 D). Expression of the regulator gene *lacI* is driven by the constitutive  $P_{pen}$  promoter. Whilst expression of  $P_{pen}$  is constitutive the variation in LacI molecule numbers has not been established and cannot be presumed. As discussed there may be *lacI* expression that is growth phase dependent. Theoretically, the magnitude of reporter output could be explained by the concentration of LacI. The  $\Delta lacI$  control strain (CJ1,  $P_{hyperspank}$ -*sfGFP\**  $\Delta lacI$ ) was constructed to report on the maximal *sfGFP\** activity reached when no negative repression is present. This strain can be used to establish if the growth phase dependent *sfGFP\** activity is LacI dependent (Figure 22, page 70).

In this chapter there are therefore four strains used to determine the effects of *lacI* location on  $P_{hyperspank}$ -*sfGFP\** reporter activity (Figure 28). These strains were designed to investigate the two growth phase dependent phenotypes observed with the *lacI* strain. Firstly, with full IPTG induction there is low fluorescence during exponential growth, increasing in transition and stationary phase. Secondly, in the absence of IPTG there is high stationary phase noise. Based on the information available the following predictions can be made:

1. The *lacI* strain (AH7,  $P_{hyperspank}$ -*sfGFP\**-*lacI*) will exhibit negative auto-regulation.
2. Negative auto-regulation will result in the lowest fluorescent output and the lowest noise phenotype.
3. Trans-acting negative regulation will produce the highest noise phenotype.

## 5.2. Results and discussion

### 5.2.1. Transcript lengths determined by Northern Blotting

It is important to know where transcriptional termination takes place in the strains used in 0. Negative auto-regulation is expected to reduce the output of *sfGFP\** and also to reduce the noise. Identifying the transcript length will determine if negative auto-regulation is a feature of the *lacI* strain. The question is, does the *lacI* strain produce one long transcript? Northern blotting was used to confirm the transcript length of both the *lacI* and *lacI-cis* strains (Figure 29 A). The *lacI* strain produces a band (lane 4) similar in size to the 2.9 Kb of the 23S rRNA (Kunst *et al.*, 1997). By comparison the *lacI-cis* strain (lane 2) produces a transcript smaller than the 16S rRNA (1.6 Kb). The transcript in *lacI* contains the *sfGFP\** reporter, the *lacI* repressor and part of the *amyE* gene native to *B. subtilis* (Figure 29 B). This establishes that negative auto-regulation is a feature of the *lacI* strain. The reduced transcript length in *lacI-cis* provides evidence that the *ilvD* transcriptional terminator is functioning as expected. The

conclusion from this Northern Blot is that *sfGFP\** in *lacI-cis* is indeed controlled by *cis*-acting negative regulation and there is no read-through into native *B. subtilis* DNA.

### 5.2.2. *sfGFP\** activity is dependent on the location of the repressor

Northern blotting confirmed the prediction that negative auto-regulation is a feature of the *lacI* strain. The next question is therefore, how does negative auto-regulation effect the fluorescent activity and noise in *sfGFP\**? The *lacI-cis* strain expresses the repressor from the same locus on a separate, divergent transcript. Expression of *lacI* in *lacI-trans* occurs from a distance of approximately 776 Kb at the *aprE* integration locus. The *lacI*, *lacI-cis* and *lacI-trans* strains were compared in standard assay conditions (Figure 30 A). In all strains there is a reduction in fluorescence from the 0 – 30 minute time points. Fluorescence remains low during exponential growth, increasing in transition phase and throughout stationary phase. As expected the *lacI* strain co-expressing the reporter and repressor produced the lowest fluorescence yield under full induction conditions (Figure 30 A, purple line). Negative auto-regulation has reduced the output of the system when compared to *cis* and *trans*-acting negative repression. At the 300 minute time point fluorescence was ~ 410 a.u, ~ 600 a.u and ~ 550 a.u for the *lacI*, *lacI-cis* and *lacI-trans* strains respectively. When comparing the data from n=3 repeats the difference in fluorescence at 300 minutes is statistically significant. Comparison of the *lacI* and *lacI-cis* strains returned a p-value of 0 with 2657 df using a Welch two-sample t-test. The same test comparing *lacI-cis* and *lacI-trans* returned a p-value of  $2.3 \times 10^{-24}$  with 2990 df. A comparison of fluorescence at each time point has been included in the appendix as Table 29 (page 167). These data show that negative auto-regulation reduces the maximum output of the system. In addition to this, *cis*-regulation produced more fluorescence than *trans*-regulation. No predictions were made regarding the output of *cis* vs. *trans*-regulated systems. It is known that the genetic locus of integration can alter the output of a reporter gene (Thompson and Gasson, 2001). It is theoretically possible that *lacI* integration at the *aprE* locus may result in a different LacI concentration than for strains where *lacI* is integrated at the *amyE* locus. Data presented in section 5.2.5 provides an explanation for why LacI concentration would not explain the growth phase dependent phenotypes. For this reason LacI concentration is not considered further.



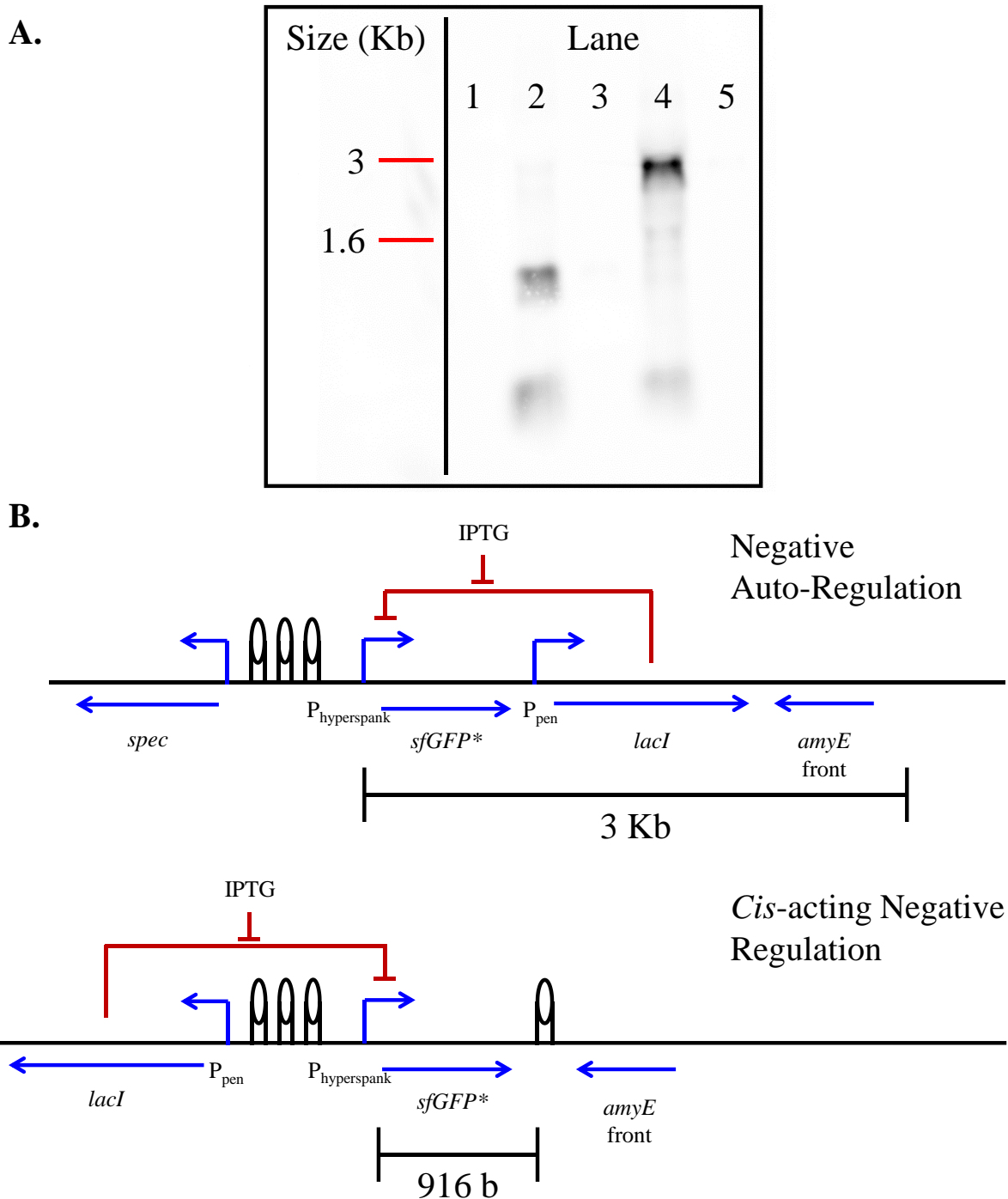


Figure 29. Co-transcription of reporter/repressor is a feature of the *lacI* strain (AH7,  $P_{hyperspank}$ -*sfGFP\**-*lacI*). **A**, Northern blot of *sfGFP\** mRNA, performed on cells grown with +/- 1 mM IPTG induction at 120 minutes into a standard assay. Size markers are approximated from ethidium bromide labelled 23S and 16S ribosomal RNA, visible on the nitrocellulose membrane under UV radiation. Lane 1, strain 168ca produces no visible bands. The culture was grown in media containing 1 mM IPTG. Lane 2, the *lacI-cis* strain (TE47,  $P_{hyperspank}$ -*sfGFP\**-*lacI-cis*) under full induction conditions (1 mM IPTG). Lane 3, the *lacI-cis* strain with no IPTG induction. Lane 4, the *lacI* strain with 1 mM IPTG induction. Lane 5, the *lacI* strain without IPTG induction. **B**, scaled representations of the *lacI* and the *lacI-cis* strains.

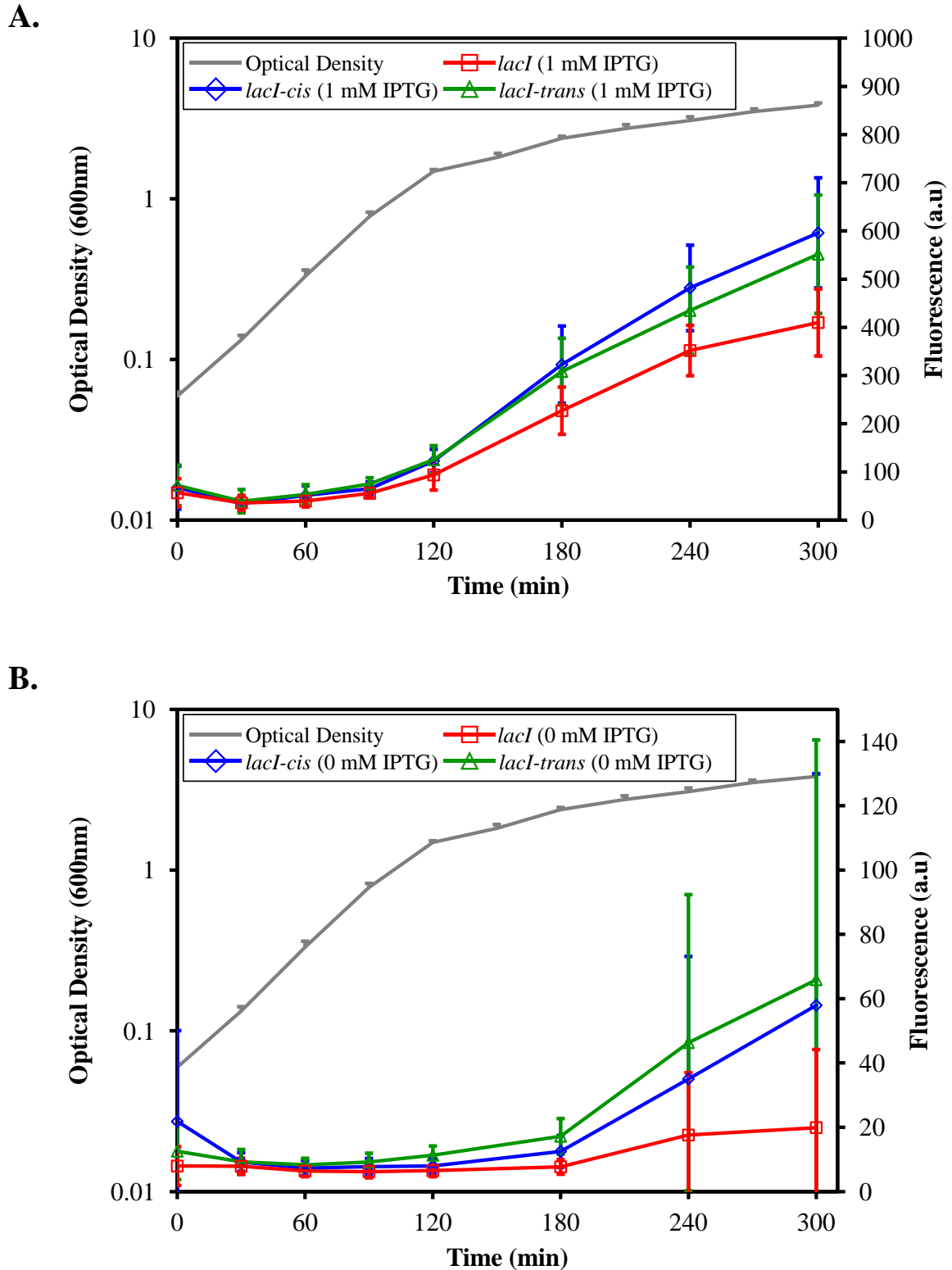


Figure 30. Regulator location affects sfGFP\* activity. **A**, fluorescence in the negative autoregulating *lacI* strain (AH7,  $P_{hyperspank}$ -sfGFP\**-lacI*), *lacI-cis* (TE47,  $P_{hyperspank}$ -sfGFP\**lacI\_cis*) and the *lacI-trans* strain (CJ3,  $P_{hyperspank}$ -sfGFP\**lacI\_trans*) induced with 1 mM IPTG. **B**, fluorescence in *lacI*, *lacI-cis* and *lacI-trans* strains without IPTG induction. Cell numbers corresponding to n=3 independent repeats are recorded in Table 15, page 162.

When comparing the expression of strains without IPTG induction a co-transcription effect is also apparent (Figure 30 **B**). The difference between strains is most visible at the 240 and 300 minute time points. At 300 minutes the mean fluorescence is ~ 20, 58 and 66 a.u in strains with auto, *cis* and *trans*-acting regulation respectively. The *lacI* strain, with negative auto-regulation is most efficient at keeping expression switched off. Leaky expression will be repressed by negative feedback. There is ~ 3-fold reduction in mean fluorescence in the *lacI* strain. The differences between strains have been analysed using Welch two-sample t-tests (Table 30, page 168). At 300 minutes a true difference of fluorescence in the *lacI* and *lacI-cis* strains was observed with a p-value of  $2.0 \times 10^{-75}$  and 1847 df. The difference in *lacI-cis* and *lacI-trans* fluorescence was also significant with a p-value of 0.00173 and 3245 df.

### 5.2.3. *Noise in gene expression is not determined by lacI location*

Data in Chapter 3 observed growth phase dependent noise in the *lacI* strain without IPTG induction (Figure 22, page 70). The phenotype of interest occurs at the last two time points. During exponential growth, noise in gene expression stabilises. In transition phase and early stationary phase noise in the system remains relatively low. At the last two time points there is heterogeneous gene expression resulting in high noise. Some cells in the population are expressing sfGFP\*. This phenotype is also observed when *lacI* is expressed from a different location (Figure 31). There is little difference between the noise phenotypes of the negative auto-regulating strain (Figure 31, black line), the *cis*-acting negatively regulated strain (blue line) and the strain with *trans*-acting negative regulation (orange line). The magnitude of fluorescence is ~ 3-fold lower in the negative auto-regulating strain without IPTG induction (Figure 30 **B**). This difference does not alter the noise phenotype. This is clear evidence that *lacI* location is not the source of noise. This is contrary to predictions that *trans*-acting negative regulation will result in the highest noise. The predication that negative auto-regulation will exhibit the least noise was also not true. During exponential growth noise is similar to the strain with *cis*-acting regulation. At the last two time points noise is the same in strains with both the *cis* and *trans*-acting regulation strains. *lacI* location therefore does not explain the noise phenotype of *P<sub>hyperspank</sub>-sfGFP\** strains, meriting further investigation.

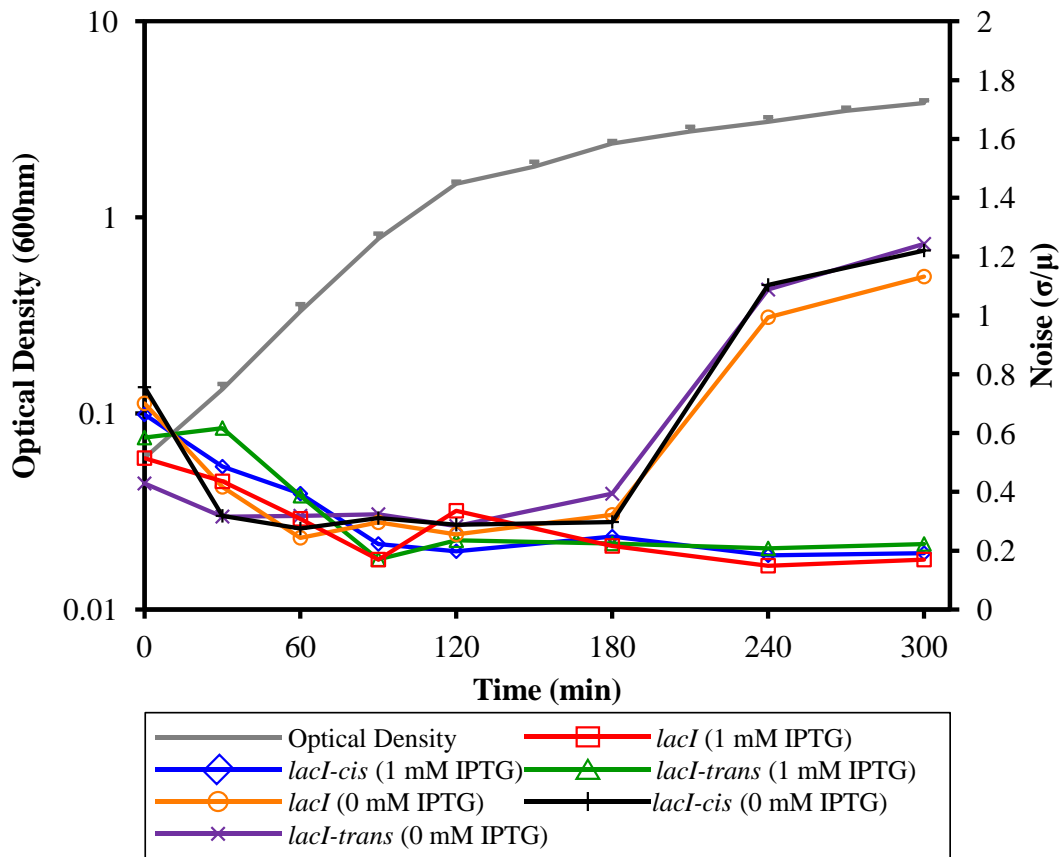


Figure 31. Regulator location makes little difference to the noise in gene expression. Three strains express *lacI* from different loci, induced with 1 mM or 0 mM IPTG. Red and orange lines, *lacI* strain (AH7,  $P_{hyperspank}$ -*sfGFP*\*-*lacI*). Blue and black lines, *lacI-cis* strain (TE47,  $P_{hyperspank}$ -*sfGFP*\* *lacI\_cis*) Green and purple lines, *lacI-trans* strain (CJ3,  $P_{hyperspank}$ -*sfGFP*\* *lacI\_trans*) Grey lines, mean optical density of all strains. Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 15, page 162.

#### 5.2.4. Repressor location alters the efficiency of response in the system

Data in the previous chapter showed that inducing *sfGFP*\* with varied IPTG concentrations altered the activity of *sfGFP*\* (Figure 25, page 76). The question raised in this section is how does the location of *lacI* alter the efficiency of response in the system? An efficient system will be sensitive to induction. Gene expression will be upregulated to maximal levels across a narrow range of inducer concentrations (Figure 54, page 154). When the system is “OFF” there will be no expression. When “ON” there will be full expression. The *lacI* strain produces ~ 80%, 63% and 19% of maximal fluorescence with a 10-fold, 20-fold and 100-fold reduction in inducer concentration (Table 6, page 160). The distinction between fluorescent activity in these induction conditions is visually apparent in stationary phase (Figure 32 A).

The difference is statistically significant at all time points between the 1 mM and 0.1 mM, the 0.1 mM and 0.05 mM and the 0.05 mM and 0.01 mM conditions (Table 31, page 169). The *lacI-cis* strain is more responsive to IPTG than the *lacI* strain. *lacI-cis* produces higher levels of fluorescence than the *lacI* strain (Figure 32 **B** & **A** respectively). There is less distinction between the induction conditions in the *lacI-cis* strain. *lacI-cis* produces ~ 98%, 81% and 31% of maximal fluorescence with a 10-fold, 20-fold and 100-fold reduction in inducer concentration (Table 7, page 160). The difference in fluorescence between the induction conditions have been compared using Welch two-sample tests (Table 32, page 170). These differences are statistically significant in 21 out of the 24 comparisons. When there is trans-acting negative regulation in the system the efficiency of response to inducer concentration is further increased (Figure 32 **C**). *lacI-trans* produces ~ 100%, 92% and 38% of maximal fluorescence with a 10-fold, 20-fold and 100-fold reduction in inducer concentration (Table 8, page 160). The differences in mean fluorescence between the 1 mM and 0.1 mM IPTG conditions were statistically significant in only two out of the 8 time points compared (Table 33, page 171). Of the two time points which are significantly different one of them has the highest mean output at 0.1 mM induction rather than 1 mM. This is strong evidence that the system is fully induced at 0.1 mM IPTG. The *lacI-trans* strain has the most efficient response to the inducer. Positioning *lacI* in *cis* produced the second most sensitive response and the negative auto-regulated strain was the least sensitive to IPTG. The difference in fluorescent output between the *lacI* strains is most noticeable at the 0.01 mM IPTG condition (Figure 32). At 300 minutes the *lacI* strain produced fluorescence of ~ 50 a.u. By comparison both the *lacI-cis* and *lacI-trans* strains were both ~ 4-fold higher at ~ 200 a.u.

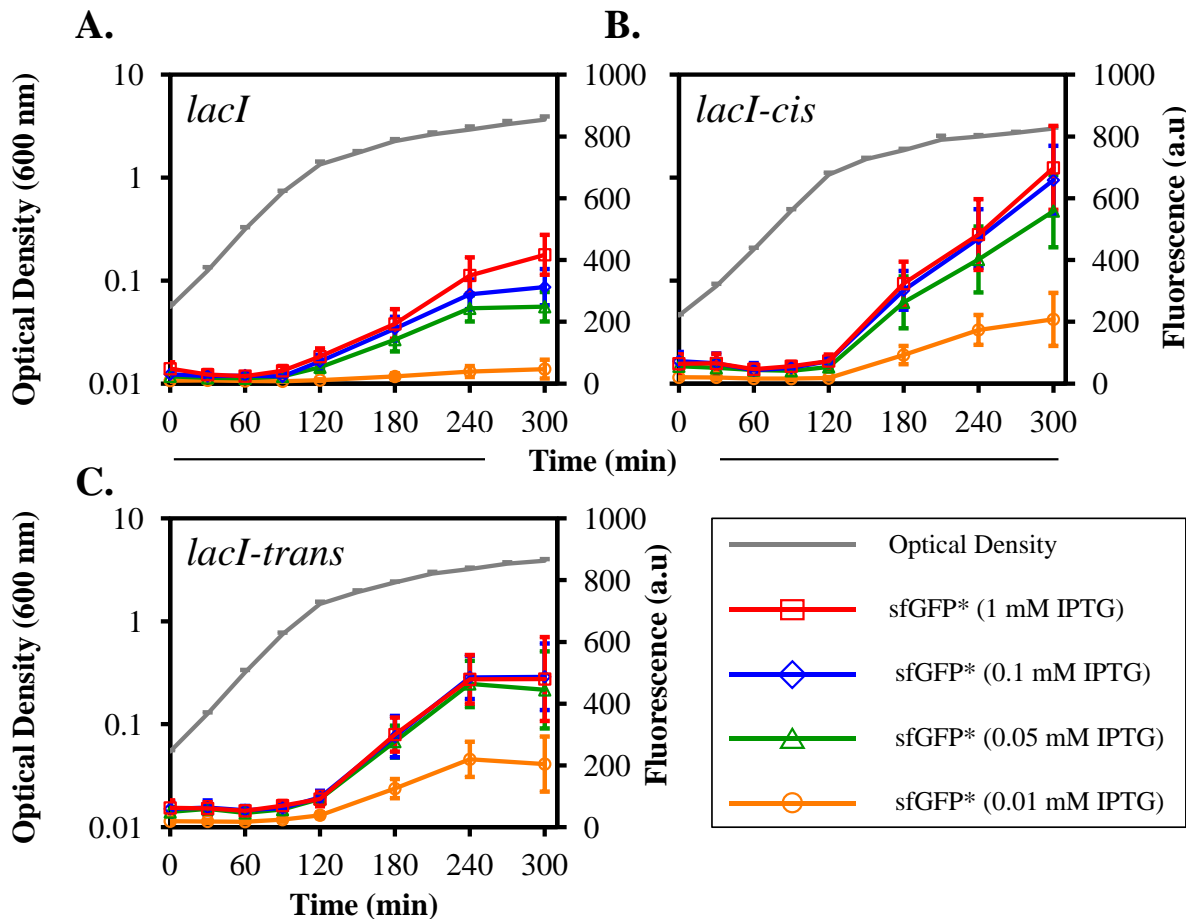


Figure 32. *lacI* location alters the response to varied inducer concentrations. Fluorescence is induced with varied IPTG concentrations and compared to growth in three strains. **A**, *lacI* strain (AH7,  $P_{hyperspank}$ -sfGFP\*-*lacI*). **B**, *lacI-cis* strain (TE47,  $P_{hyperspank}$ -sfGFP\* *lacI\_cis*). **C**, *lacI-trans* strain (CJ3,  $P_{hyperspank}$ -sfGFP\* *lacI\_trans*). Red lines, mean fluorescence of strains induced with 1 mM IPTG. Blue lines, 0.1 mM IPTG. Green lines, 0.05 mM IPTG. Orange lines, 0.01 mM IPTG. Grey lines, mean optical density. Cell numbers corresponding to n=3 independent repeats are recorded in Table 16, page 163. A secondary analysis of these data is presented in Figure 33.

When compared in the same experimental repeats there was little difference in fluorescence between the *lacI-cis* and *lacI-trans* strains (Figure 30). Under full induction conditions the *lacI-cis* strain produced the highest fluorescence at the final two time points. This was surprising, yet could be a result of a difference in LacI concentration. It could also be a result of a growth phase dependent difference in the two systems. This raises a question. Does the efficiency of response in these systems change with growth phase? During exponential growth there is little difference in the *lacI-cis* and *lacI-trans* strain's sensitivity to inducer (Figure 33 A). *lacI-cis* and *lacI-trans* both respond differently from the *lacI* strain. The *lacI*

strain reaches 80% of maximal expression at 0.1 mM IPTG. By comparison, *lacI-cis* and *lacI-trans* reach 98% and 100% of maximal expression at 0.1 mM IPTG (Table 6, Table 7 & Table 8 respectively, page 160. Statistical analysis in Table 31, Table 32 and Table 33, page 169 - 171). In transition phase there is a clear difference in the sensitivity to inducer at lower IPTG concentrations (Figure 33 **B**). At the 0.1 mM concentration, *lacI-cis* has identical fluorescence to the 1mM condition and the *lacI-trans* strain exceeds the fluorescent activity in the 1 mM IPTG condition. 0.1 mM IPTG is therefore considered to produce maximal fluorescence in both the *lacI-cis* and *lacI-trans* strains in transition phase. In stationary phase the difference in sensitivity is also evident in the 0.1 mM IPTG condition (Figure 33 **C**). These data show a growth phase dependent sensitivity to inducer concentration.

#### 5.2.5. *ΔlacI used to investigate the growth phase dependent phenotype in sfGFP\**

Data obtained with strains expressing *lacI* from different loci all result in growth phase dependent activity of sfGFP\* (Figure 30). Under full induction conditions fluorescence is low during exponential growth, increasing through transition phase and stationary phase. Data presented up to this point has not established if LacI concentration varies during the different growth phases. It is reasonable to predict that LacI concentration will alter the activity of sfGFP\*. It is therefore important to investigate the phenotype with known LacI concentrations. The most straightforward way to accomplish this is by removing LacI regulation altogether, allowing a comparison between none and some. This will answer an important question: Does LacI contribute to the growth phase dependent phenotype? Comparison was made between the *lacI-cis* strain and the unregulated *ΔlacI* strain (Figure 34). Low levels of fluorescence are observed during exponential growth in both of these strains, followed by an increase through transition phase and into stationary phase. Therefore LacI does not contribute to this growth phase dependent phenotype.

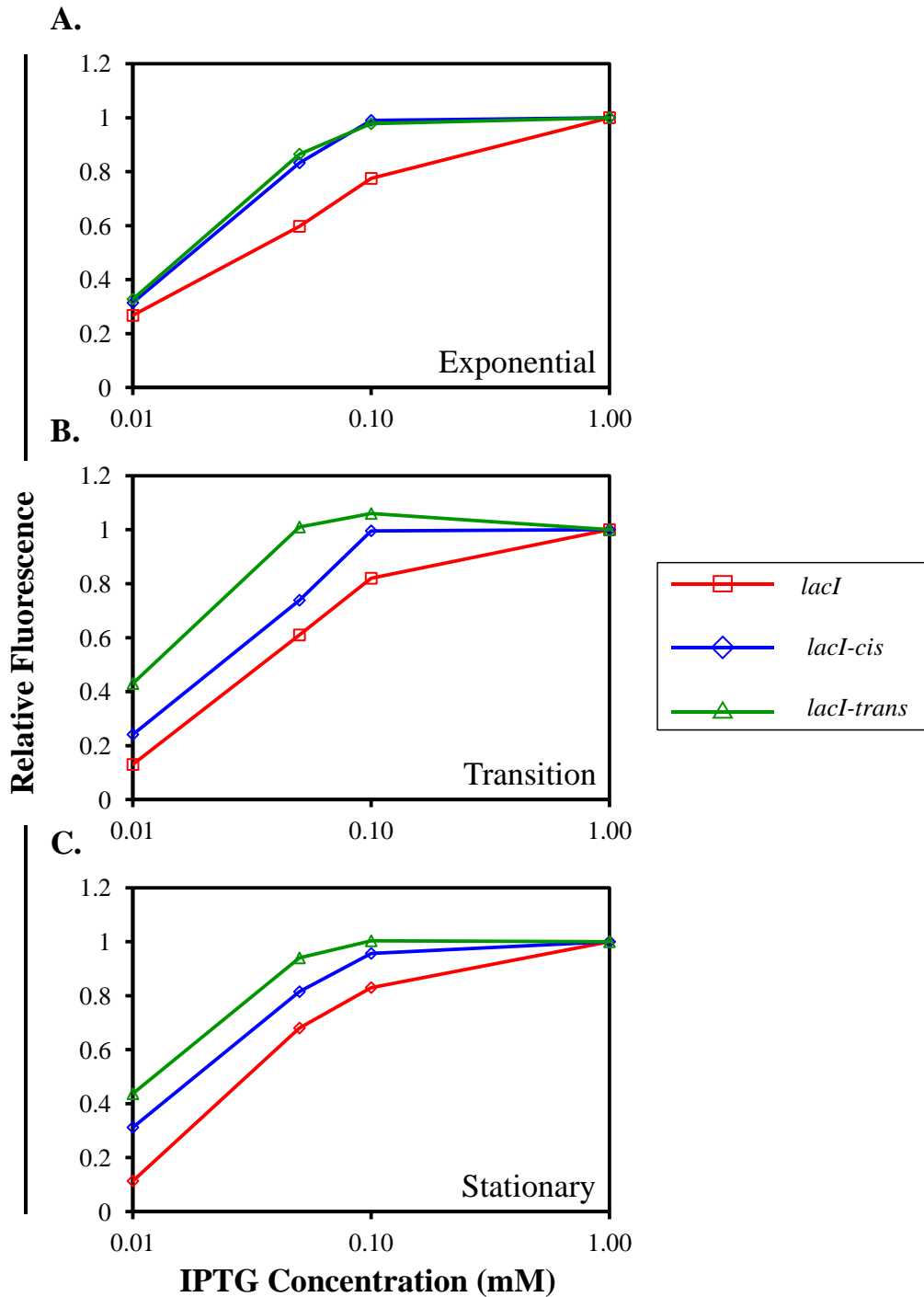


Figure 33. *lacI* location and growth phase alter the system sensitivity to inducer concentrations. The relative fluorescence for strains: *lacI* (AH7,  $P_{hyperspank}$ -sfGFP\**-lacI*), *lacI-cis*, (TE47,  $P_{hyperspank}$ -sfGFP\* *lacI\_cis*) and *lacI-trans* (CJ3,  $P_{hyperspank}$ -sfGFP\* *lacI\_trans*) are calculated. The relative fluorescence = mean fluorescence at inducer concentration / mean fluorescence at 1 mM IPTG. **A**, relative fluorescence during exponential growth, calculated as the mean relative fluorescence at 0, 30, 60 and 90 minutes. **B**, relative fluorescence during transition phase at 120 minutes. **C**, relative fluorescence during stationary phase, calculated as the mean fluorescence at 180, 240 and 300 minutes. Red lines, *lacI* strain. Blue lines, *lacI-cis* strain. Green lines, *lacI-trans* strain. Cell numbers corresponding to n=3 independent repeats are recorded in Table 16, page 163. An alternative analysis of these data is presented in Figure 32.



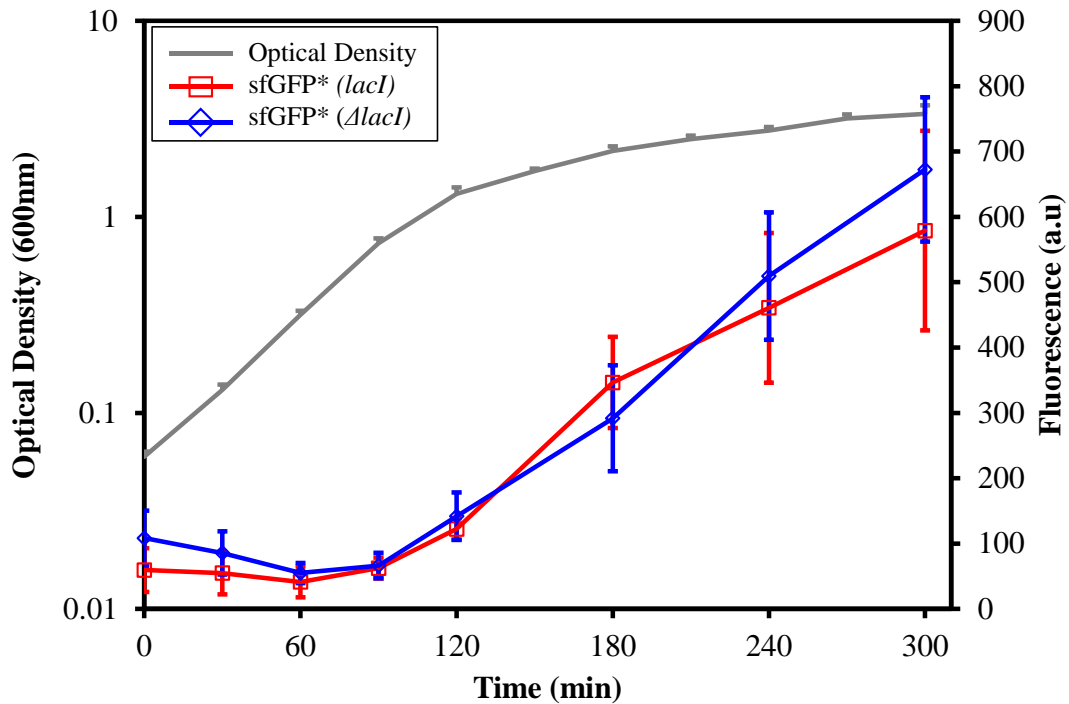


Figure 34. Growth phase dependent sfGFP\* activity is not caused by LacI regulation. Mean fluorescence of two sfGFP\* strains in standard assay conditions. Red line, *lacI-cis* strain (TE47,  $P_{hyperspank}$ -sfGFP\* *lacI\_cis*) fully induced with 1 mM IPTG. Blue line,  $\Delta lacI$  strain (CJ1,  $P_{hyperspank}$ -sfGFP\*  $\Delta lacI$ ) without IPTG induction. Grey line, mean optical density. Cell numbers corresponding to n=3 independent repeats are recorded in Table 17, page 163.

### 5.3. Conclusions

This chapter aimed to establish the effects of repressor location on sfGFP\* activity. There are four main conclusions to be drawn from the data presented:

1. *lacI* location makes no difference to the noise phenotype observed without IPTG induction.
2. Auto-regulation reduces the output of the system.
3. *lacI* location alters the sensitivity to inducer, with the exception of *lacI-cis* and *lacI-trans* during exponential growth.
4. LacI is not needed to produce the growth phase dependent activity phenotype.

As predicted Northern blotting confirmed there is negative auto-regulation in the *lacI* strain. It proved the *ilvD* transcriptional terminator in the *lacI-cis* and *lacI-trans* strains was functional in prevention of read through transcription into the native *B. subtilis* DNA.

Negative auto-regulation was predicted to reduce the activity of sfGFP\* and also the noise. As expected *lacI* co-transcription with the reporter was found to reduce the activity of sfGFP\*. An unexpected finding was the virtually identical noise phenotype with negative auto-regulation, *cis* and *trans*-acting regulation. Heterogeneity levels, quantified by noise ( $\sigma/\mu$ ) are not affected by the magnitude of fluorescence. It would be tempting to utilise negative auto-regulation to repress expression from a leaky promoter such as  $P_{\text{hyperspank}}$ . Data in this chapter has shown this reduces the magnitude of fluorescence by approximately 3-fold. However, it must be recognised that the underlying biology driving the noise phenotype is not affected. There is still expression and the resulting heterogeneity is identical.

The *lacI-cis* and *lacI-trans* strains were both found to exhibit a growth phase dependent difference in the sensitivity to IPTG. During exponential growth *lacI-cis* and *lacI-trans* responded to varied inducer concentration with very similar efficiency. In transition and stationary phase the *lacI-cis* strain was less sensitive to varied inducer levels than the *lacI-trans* strain. This highlights the importance of thoroughly testing gene circuits designed for an investigation. Similar functionality during exponential growth does not mean there will be similar functionality in transition or stationary phase.

The final section in this chapter concerned the growth phase dependent sfGFP\* activity phenotype observed with the *lacI*, *lacI-cis* and *lacI-trans* strains. In all of these strains there is low sfGFP\* activity during exponential growth, increasing during transition phase and stationary phase. Use of the  $\Delta lacI$  strain established this phenotype is not LacI dependent. The phenotype occurred without LacI regulation, arguing the need for further investigation.

## Chapter 6. Transcription and translation effects in the reporter system

### 6.1. Introduction

Use of the  $P_{\text{hyperspank}}$  promoter driving expression of sfGFP\* has produced two phenotypes of interest:

- i. There is growth phase dependent sfGFP\* activity when fully induced.
  - Fluorescence is low during exponential growth.
  - Fluorescence increases in transition phase and throughout stationary phase.
- ii. There is growth phase dependent noise in sfGFP\* without induction.

These phenotypes are affected by protein stability. Fusing sfGFP\* to mutated SsrA-tags targets the proteins for degradation (Figure 26, page 79). In these strains sfGFP\* fluorescence does no longer continues to increase through stationary phase. The profile of activity in exponential and transition phase remains whilst stationary phase fluorescence is dependent on the half-life of sfGFP\* (Figure 26 A). Without induction, stationary phase noise is reduced or not quantifiable, depending on the efficiency of the tag (Figure 26 B). Protein stability is the only variable investigated to impact the two phenotypes.

Negative regulation in the system was investigated by moving the location of the repressor gene, *lacI*. *lacI* was moved from the original negative auto-regulation module to create *cis*-acting negative regulating, *trans*-acting negative regulating and no regulation modules. The maximum output of each system was different, however the above phenotypes were not affected. There was still growth phase dependent activity under full induction conditions (Figure 30, page 89). Use of the  $\Delta lacI$  strain proved that this activity is independent of LacI regulation (Figure 34, page 96). With respect to the stationary phase noise without IPTG induction, *lacI* location was found to alter the magnitude of fluorescent activity. However, the resulting noise levels were not affected by this difference in mean fluorescent activity (Figure 31, page 91).

#### 6.1.1. Transcription

This chapter continues to investigate these two phenotypes. To what extent are the observed phenotypes dependent on transcription and/or translation? Or, is it a combination of both of these processes? Considering transcription, experiments in this thesis have exclusively used  $P_{\text{hyperspank}}$  to drive gene expression.  $P_{\text{hyperspank}}$  is a synthetic promoter originally derived from

$P_{\text{spac}}$ .  $P_{\text{spac}}$  combined a single *lac* operator site from *E. coli* with the promoter from the *B. subtilis* phage SPO-1 (Yansura and Henner, 1984).  $P_{\text{hyperspank}}$  has been altered to include a second operator site and includes a single nucleotide change from G to T at the -1 position. Is it possible that the two phenotypes result from the use of this synthetic promoter?

Therefore, we asked what impact alternative, native promoters might have upon the observed phenotypes? The *cis*-acting negatively regulated strain (TE47,  $P_{\text{hyperspank}}\text{-sfGFP}^*\text{-lacI}_{\text{cis}}$ ) has been used for comparison in this chapter. Hereafter this is referred to as the  $P_{\text{hyperspank}}$  strain.

### 6.1.2. Translation

Data in Chapter 3 identified the impact of translated SsrA-tags, targeting sfGFP\* for degradation (Figure 26, page 83). These C-terminal linkers affected the stability of sfGFP\*, altering its half-life. Protein degradation can occur at both the N-terminus and C-terminus (Gottesman *et al.*, 1998; Butz *et al.*, 2011). In this thesis a single derivative of the gene coding for green fluorescent protein (*sfGFP*\*) has been used. A specific feature of this *sfGFP*\* is the N-terminal translational linker coding for the amino acids MEFLQ (presence of the linker is denoted by the asterisk). The original strain used in this thesis, AH7, was designed by Alex Henderson (Henderson, 2012). The master's project did not specify a rationale for inclusion of the linker. It was not included to investigate a specific question and may have resulted from a cloning error. The gene is functional and it is important to establish the effects of this linker on sfGFP\* activity. Does the translation linker contribute to the growth phase dependent activity or noise observed with *sfGFP*\* strains?

Another aspect of translation important in this system is codon usage. The *sfGFP*\* used in this thesis is thought to have advantages over other *gfp* variants. It was termed “superfolder green fluorescence protein” (*sfGFP*) due to its high folding fidelity and quick refolding (>95% fluorescence recovered by 4 minutes) after denaturation *in vitro* (Pedelacq *et al.*, 2006). However, this robust reporter was designed and tested for use in *E. coli*. Codon bias is different in *E. coli* and *B. subtilis* (Zhang and Ignatova, 2009). In *B. subtilis* altering the codon usage of *sfGFP* results in varied activity of the protein (Overkamp *et al.*, 2013). Overkamp *et al.* used the OPTIMIZER software (Puigbo *et al.*, 2007) to create *sfGFP* variants codon optimised for different organisms. The rationale behind Overkamp's strategy did not produce the results they expected. In *B. subtilis* the *sfGFP* gene optimised for expression in *S. pneumoniae* was brighter than the *sfGFP* gene optimised for expression in *B.*

*subtilis*. A single time point was used to define expression levels of the gene, thus any growth phase dependent variation was not established. In *E. coli*, codon usage patterns differ according to growth rate (Dong *et al.*, 1996), whereas in *B. subtilis* the composition of the translational machinery also varies over time (Rosenberg *et al.*, 2012). Theoretically, the rate of a gene's translation could change depending on the growth phase. It is possible that a gene may have codon usage optimised for expression in stationary phase rather than exponential growth. It is therefore appropriate to ask: will codon usage affect the growth phase dependent activity? An alternative codon optimisation strategy was used to answer this question.

A final consideration of translational effects in this chapter concerns the reporter gene choice. *gfp* variants are not the only fluorescent reporters available for gene expression studies. Translation of *sfGFP* may have unique features in *B. subtilis*. Therefore, does the choice of reporter affect the growth phase dependent activity or noise phenotypes observed with *sfGFP*? To investigate these two questions the alternative fluorescent reporter *mCherry2* was used.

## 6.2. Results

### 6.2.1. *Is transcription from P<sub>hyperspank</sub> responsible for the observed phenotypes?*

In previous experiments transcription rate has been altered through induction with varied concentrations of IPTG (Figure 25, page 76 & Figure 32, page 93). At all induction levels above 0 mM IPTG there is growth phase dependent *sfGFP*\* activity. Without IPTG induction there is high stationary phase noise at the final two time points (Figure 25 & Figure 31, page 91). These experiments have all used  $P_{\text{hyperspank}}$  to drive expression of *sfGFP*\*. To test the contribution of  $P_{\text{hyperspank}}$  on the phenotypes, two alternative promoters,  $P_{\text{xyI}}$  and  $P_{\text{veg}}$  were used to drive expression of the *sfGFP*\* reporter.  $P_{\text{xyI}}$  and  $P_{\text{veg}}$  are both native to *B. subtilis*, allowing transcription to occur under different conditions.  $P_{\text{xyI}}$  is negatively regulated by XylR and the promoter is induced on addition of xylose (Gartner *et al.*, 1988; Gartner *et al.*, 1992). By comparison the  $P_{\text{veg}}$  control is constitutively active during both the exponential and stationary growth phase (Fukushima *et al.*, 2003).  $P_{\text{veg}}$  is not expected to produce a growth phase dependent phenotype. Expression of *sfGFP*\* from both of these promoters produced up to 10-fold lower activity than from  $P_{\text{hyperspank}}$  (Figure 35 A). There is a similar profile in the  $P_{\text{hyperspank}}\text{-sfGFP}$ \* strains (Figure 30, page 89) and the  $P_{\text{xyI}}\text{-sfGFP}$ \* strain (Figure 35 A). Therefore, transcription from these different promoters alters the

magnitude of fluorescence without changing the growth phase dependent phenotype on induction.

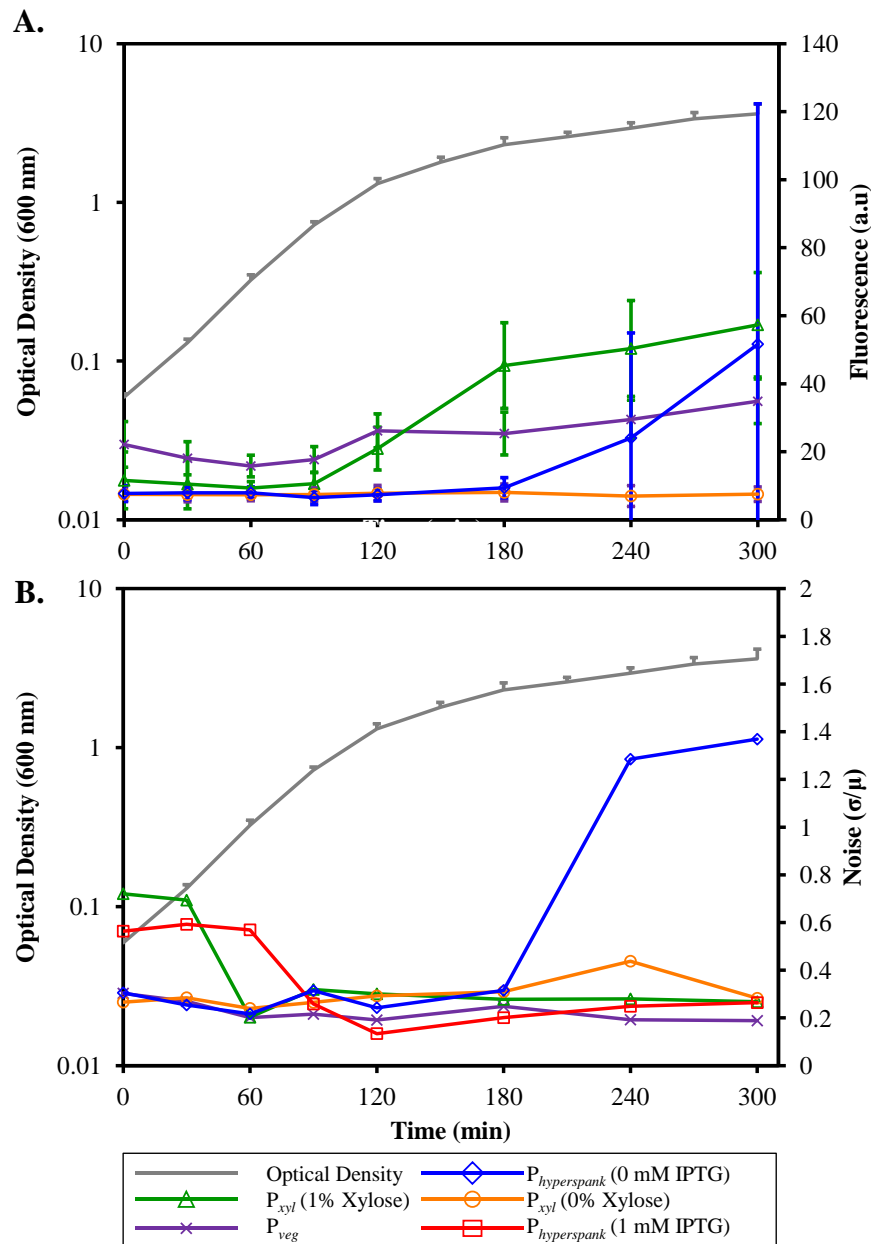


Figure 35. Growth phase dependent sfGFP\* activity is a feature of the  $P_{hyperspank}$  and  $P_{xyl}$  promoters while noise is restricted to  $P_{hyperspank}$ . **A**, fluorescence (coloured lines) and optical density (grey line) of strains grown in standard assay conditions. **B**, noise in gene expression calculated by the standard deviation ( $\sigma$ ) divided by the mean fluorescence ( $\mu$ ). Blue lines, strain TE47 without IPTG induction. Green lines, strain TE52 ( $P_{xyl}$ -sfGFP\*) induced with 1% xylose. Orange lines, TE52 without xylose induction. Purple lines, strain TE53 ( $P_{veg}$ -sfGFP\*) without addition of xylose or IPTG. Data from strain TE47 with 1 mM IPTG has been omitted from **A**. Representative data for strain TE47 can be seen in Figure 30, page 89. Cell numbers corresponding to n=3 independent repeats are recorded in Table 18, page 164.

Contrary to expectations a growth phase dependent phenotype was also apparent with transcription of *sfGFP\** driven from the  $P_{veg}$  control (Figure 35 A). In both the  $P_{veg}$  and the  $P_{hyperspank}$  strains there is a reduction in fluorescence from the 0 minute time point until 60 minutes. Fluorescence then increases in both strains until transition phase at 120 minutes. With  $P_{hyperspank}$  there is a large, continual increase in *sfGFP\** fluorescence from transition phase throughout stationary phase. By comparison, *sfGFP\** activity in the  $P_{veg}$  strain decreases between transition phase and the first time point in stationary phase, followed by a moderate increase during stationary phase. Activity during exponential growth is lower than during stationary phase. This is a surprising result in the  $P_{veg}$  control strain. The constitutive promoter was not expected to exhibit distinct differences in *sfGFP\** activity dependent on growth phase. In fact, the  $P_{veg}$  control would suggest a revised description of the growth phase dependent activity phenotype. It is accurate to report lower fluorescence in exponential growth phase with all promoters, rather than terming it low fluorescence during exponential growth. It is not accurate to say that fluorescence increases from transition phase throughout stationary phase, as stated in the introduction to this chapter. It is more accurate to say that the increase in fluorescence during stationary growth phase still remains, however the magnitude of the increase is dependent on the promoter.

The second phenotype of interest is the stationary phase noise generated without IPTG induction. This phenotype was observed in the auto-regulated, *cis*-regulated and *trans*-regulated  $P_{hyperspank}$ -*sfGFP\** strains (Figure 31, page 91). This noise is not observed in the  $P_{xyl}$  strain without xylose induction (Figure 35 C, purple line). Noise in the  $P_{xyl}$  strain is not expected since there was no fluorescence from which it could occur. Noise is not a feature of the constitutively active  $P_{veg}$  strain. It is only in the  $P_{hyperspank}$  strain without IPTG induction (orange line) where the mean fluorescence increases above background levels (Figure 35 A). Any noise in  $P_{xyl}$  can therefore be attributed to experimental noise rather than a reflection of *sfGFP\** expression. This is good evidence that the stationary phase noise phenotype is promoter dependent. Stochastic disassociation of LacI from the operators is a viable explanation for this phenotype. LacI is native to *E. coli*. In *E. coli* tetrameric LacI binds to two operator sites in  $P_{lac}$  (Figure 27, page 83). Without inducer bound to the allosteric sites LacI still disassociates from the promoter (Choi *et al.*, 2008). These stochastic disassociation events occur for a limited time before the promoter is repressed again. Transcription of the *lac* operon is short-lived, yet sufficient to produce functional Lac permease in the membrane of the cell. This adaption confers the ability to internalise lactose and rapidly upregulate

genes in response to it. The system has evolved to allow some expression of *lacZY* through functional derepression without the presence of an inducer. Stationary phase noise in *P<sub>hyperspank</sub>-sfGFP\** strains may result from the same process.

### **6.2.2. Translation of the five amino acid N-terminal linker, MEFLQ**

Translation time and folding of sfGFP may be altered by the presence of the N-terminal translational MEFLQ linker present in sfGFP\* (presence of the linker is denoted by the asterisk). It is important to confirm the effect this linker has on the activity phenotype. Does the translational linker affect the growth phase dependent activity of the reporter? Two strains were used to answer this question. The  $\Delta$ linker strain (TE58, *P<sub>hyperspank</sub>-sfGFP lacI\_cis*) is an identical replica of the *P<sub>hyperspank</sub>* strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) without the translational linker. Both of these strains produced the same activity phenotype under full induction conditions (Figure 36). There is low fluorescence during exponential growth, increasing in transition phase and throughout stationary phase. The magnitude of fluorescence is reduced by the presence of the linker. The difference in fluorescent output of these two strains is statistically significant at every time point in the assays (Table 34, page 171). Importantly it confirms the growth phase dependent activity phenotype is not affected by the translational linker.



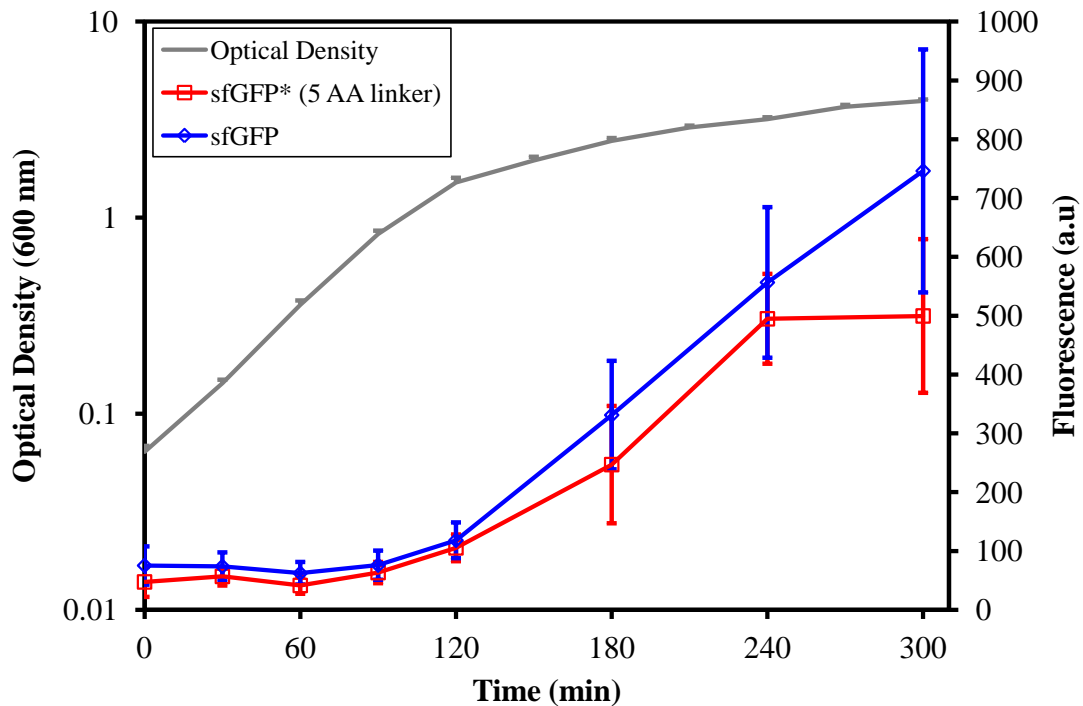


Figure 36. The N-terminal translational linker present on *sfGFP\** is not responsible for the fluorescent profile in *P<sub>hyperspank</sub>-sfGFP* strains. Mean fluorescence (coloured lines) of strains fully induced (1 mM IPTG) in standard assay conditions are compared to the mean optical density (grey line). Red line, strain TE47 (*P<sub>hyperspank</sub>-sfGFP\** *lacI<sub>cis</sub>*) containing the linker coding for amino acids MEFLQ. Blue line, strain TE58 (*P<sub>hyperspank</sub>-sfGFP* *lacI<sub>cis</sub>*) without the linker. Cell numbers corresponding to n=3 independent repeats are recorded in Table 19, page 164.

### 6.2.3. Codon optimisation of *sfGFP*

The GFP variant used in this thesis has been mutated and tested for desired attributes in *E. coli* (Pedelacq *et al.*, 2006). In *E. coli* the tRNA availability alters as the growth rate is changed (Dong *et al.*, 1996). Since tRNA availability is assumed to be the limiting step in translation (Zhang and Ignatova, 2009) it may be an important consideration for this thesis. It is possible that tRNA availability in *B. subtilis* is also linked to growth phase. This could impact on the observed growth phase dependent phenotypes. It leads to the question: will changing the codons of *sfGFP* alter the growth phase dependent activity? To address this question two identical *sfGFP* proteins were required with different DNA sequences. Maintaining a strong signal from the reporter was a desired outcome. Ideally *sfGFP* activity would be different between the two genes. To achieve this a codon optimised version of *sfGFP* was required to produce increased fluorescence in *B. subtilis*. Previous attempts to

optimise *sfGFP* for *B. subtilis* (Overkamp *et al.*, 2013) employed the use of OPTIMIZER (Puigbo *et al.*, 2007). OPTIMIZER has an algorithm based on highly expressed genes, specific to individual organisms. It is thought the selection pressures of these highly expressed genes will ensure the codon bias is optimal for translation. This can be integrated with the codon bias in ribosomal genes and with the gene copy number of tRNAs. The codon sequence of a gene entered into OPTIMIZER will be changed in order to optimise the expression levels for the prokaryotic organism of choice. Use of this tool did not produce the expected outcome where *sfGFP* was optimised for *B. subtilis* (Overkamp *et al.*, 2013). For this reason an alternative strategy was sought. A more recent algorithm than OPTIMIZER identifies charged tRNA abundance as the rate limiting step in translation (Zhang and Ignatova, 2009). The speed of codon translation is not only determined by the concentration of amino acid bound tRNAs. The algorithm of Zhang *et al.* predicts that the proximity of codons will also affect the translation speed. A cluster of “slow” codons can have a synergistic effect and stall translation. Use of the RiboTempo tool from Prof. Zoya Ignatova’s laboratory (Zhang and Ignatova, 2009) has been used to compare the predicted translation speed of *sfGFP* variants with the data of Overkamp *et al.* (Figure 37). The difference in predicted translation speed correlates to differences in the experimentally defined activity levels. The quicker translated gene has the highest activity. For this reason the RiboTempo tool was used to identify *sfGFP* codons with the highest translation speed. Every codon in *sfGFP* was manually checked and where possible changed to increase the translation speed. The optimised gene sequence was then run through the RiboTempo tool to predict the translation speed of the whole gene. Comparison of *sfGFP\**, the original *sfGFP* and the optimised *sfGFP* predicted that the translation rate would be increased in the optimised gene (Figure 37). This correlates to the actual fluorescence observed in the strains (Figure 38). Translation of *sfGFP\** containing the MEFLQ linker produced the lowest fluorescence while the codon optimised *sfGFP* produced the highest fluorescence. Analysis of the data confirmed a statistically significant difference in the mean fluorescence of these strains at all time points (Table 34, page 171 and Table 35, page 172). However, the original questions asked of these strains regarded the growth phase dependent activity of *sfGFP*. Both the translational linker and the codon usage did not affect the phenotype. The output of the system was affected without changing the lower fluorescence in exponential phase, followed by an increase in transition phase and throughout stationary growth phase.

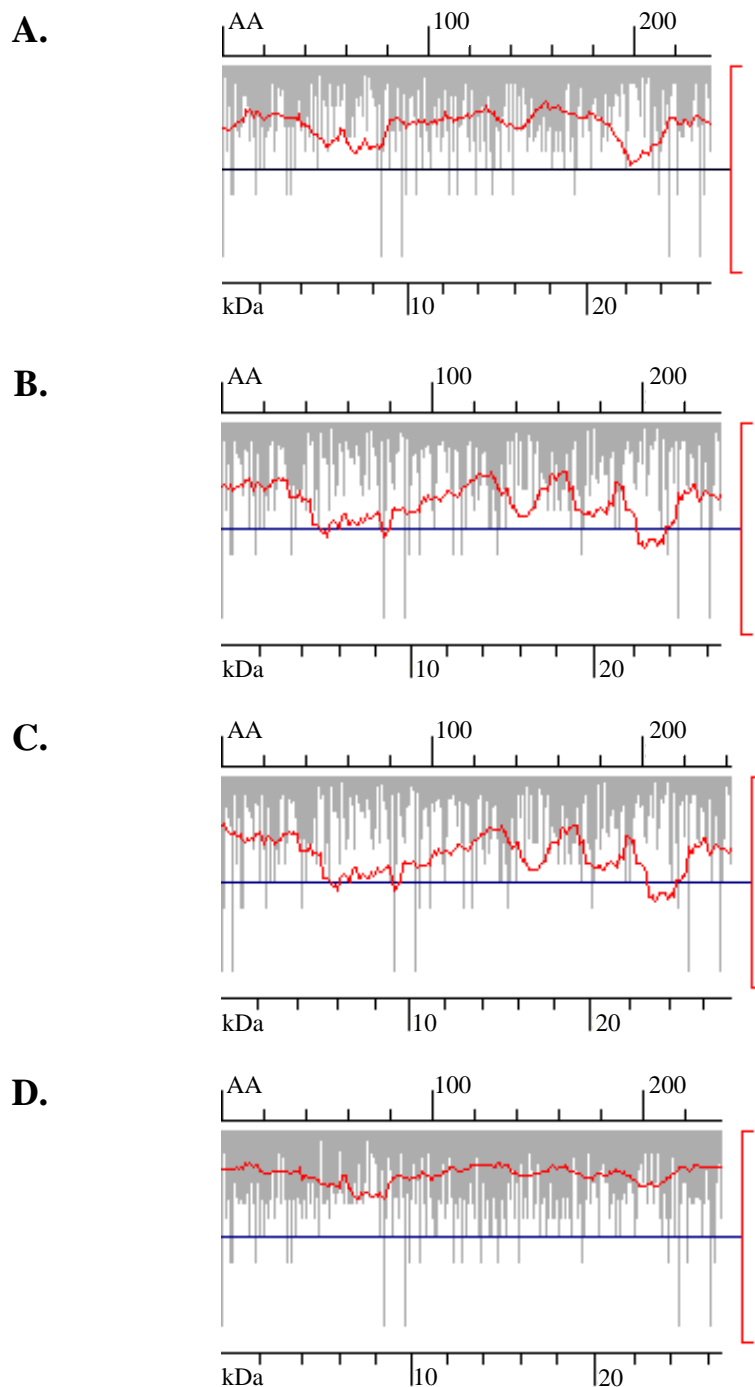


Figure 37. Predicted translation speed in *sfGFP* variants. The bar charts show the translation speed of individual codons. The height of the white bar denotes the speed of translation. A short bar is a quickly translated codon. A long white bar is a slowly translated codon. The red line predicts the overall translation speed, determined by the codons themselves and their proximity to other codons. Where the translation speed drops below the middle line a halt in translation is predicted to occur. **A**, *sfGFP* codon optimised for *S. pneumonia* (Overkamp *et al.*, 2013). **B**, original *sfGFP* gene (Pedelacq *et al.*, 2006). **C**, *sfGFP\** (asterisk denotes presence of the MEFLQ translational liner). **D**, codon optimised *sfGFP* using the algorithm provided by Zoya Ignatova's laboratory (Zhang and Ignatova, 2009). Bar charts produced using the RiboTempo tool downloaded from (<https://www.chemie.uni-hamburg.de/bc/ignatova/tools-and-algorithms.html>).

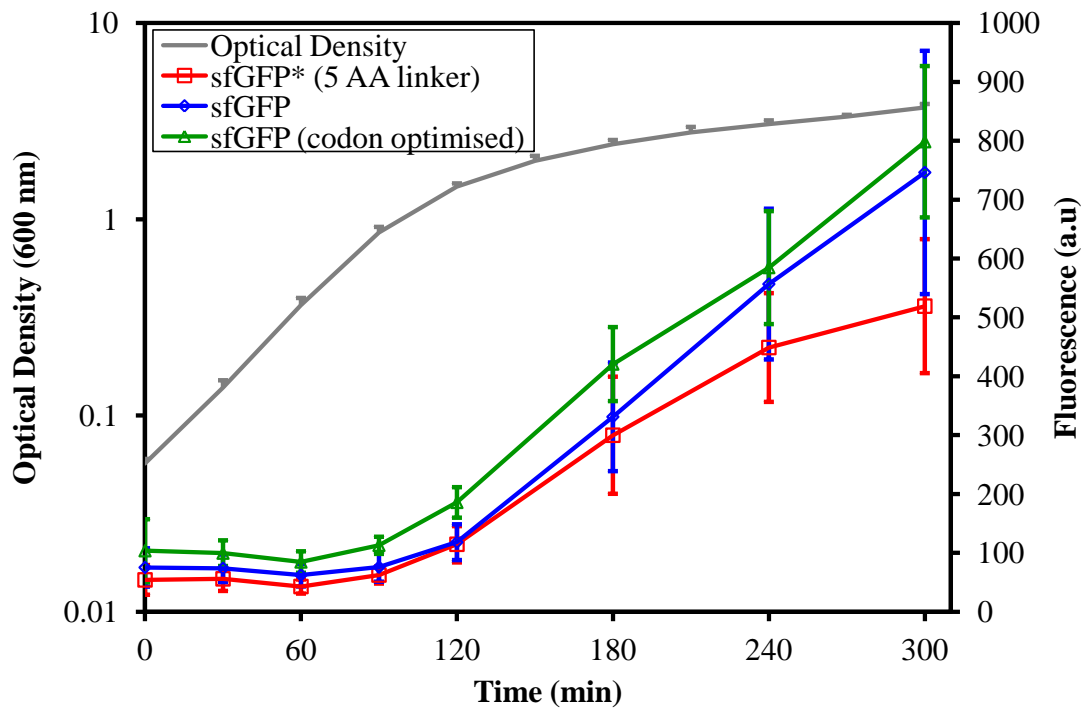


Figure 38. Codon optimisation of *sfGFP* increases the fluorescent output of the gene without altering the growth phase dependent activity. Red line, sfGFP\* strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) containing the MEFLQ translational linker. Blue line, sfGFP strain (TE58, *P<sub>hyperspank</sub>-sfGFP lacI\_cis*) containing the original *sfGFP* (Pedelacq *et al.*, 2006). Green line, optimised-sfGFP strain (TE76, *P<sub>hyperspank</sub>-sfGFP\_RiboTempo lacI\_cis*) containing the codon optimised *sfGFP*. Grey lines, mean optical density. Cell numbers corresponding to  $n \geq 3$  independent repeats are recorded in Table 20, page 164.

#### 6.2.4. Alternative reporters

Altering transcription and translation of *sfGFP* has not removed the growth phase dependent activity of the reporter. The activity of the promoter can be established by measuring the fluorescent output of the system during growth. Data presented in this thesis suggests there are limitations to this approach. The activity of sfGFP\* is not solely determined by the promoter. It is important to know if this is a specific feature of sfGFP. Does the growth phase dependent activity occur with an alternative fluorescent reporter? If the answer to this question is no, a number of errors could be expected. GFP reporters will suggest a promoter is most active during stationary growth phase. This may not reflect a true difference in the rate of transcription. In addition a growth phase dependence in sfGFP activity could confuse the results of protein localisation studies. Translational GFP fusions have been used to

visualise the location of proteins (Webb *et al.*, 1995). Data in this chapter shows a ~ 10-fold difference in fluorescence between exponential growth and stationary phase growth (Figure 35 A). Hypothetically, the concentrations of proteins fused to GFP in exponential phase may be under-detected if based on the fluorescent activity, which is growth phase dependent. There may be correlation in protein concentration without correlation in the functional state of the proteins. This thesis therefore asks, will there be growth phase dependent activity in the expression of the alternative fluorescent *mCherry2* reporter?

To test the effects of the reporter on this system the *sfGFP\** open reading frame (ORF) was swapped for *mCherry2* using Gibson cloning (Gibson *et al.*, 2009). The *mCherry2* strain (TE51, *P<sub>hyperspank</sub>-mCherry2 lacI cis*) is identical to the *sfGFP\** strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI cis*) except for the reporter. The different reporters produced very different fluorescent profiles. The *sfGFP\** reporter produces the expected growth phase dependent phenotype (Figure 30 A, red line. Page 89). In contrast the *mCherry2* reporter appears to have very low levels of fluorescence independent of growth phase (Figure 39 A). There is no characteristic reduction in fluorescence from the 0 – 30 minute time points. *sfGFP* fluorescence increases from the 60 – 90 minute time points while *mCherry2* fluorescence reduces. In stationary phase between 180 and 300 minutes *mCherry2* fluorescence reduces while *sfGFP\** fluorescence increases. Microscope images illustrate the difficulty in distinguishing between the low levels of fluorescence during exponential growth and in stationary phase (Figure 39 B - E). It has been a continuing assertion in this thesis that the difference in *sfGFP\** fluorescence is dependent on growth phase. The *P<sub>veg</sub>-sfGFP\** control strain (Figure 35) had the most similar levels of fluorescence during exponential growth and stationary phase. This has been used as a comparison tool to illustrate the differences between *sfGFP\** and *mCherry2* reporters producing similar levels of fluorescence (Figure 40). There is a noticeable difference in *P<sub>veg</sub>* driven *sfGFP\** fluorescence between exponential growth and stationary phase (Figure 40 A). The mean stationary phase fluorescence in the *P<sub>veg</sub>-sfGFP\** strain is higher than the upper quartile fluorescence during exponential growth. By comparison it is very difficult to distinguish *mCherry2* fluorescence in exponential and stationary phase growth (Figure 40 B). The mean values, standard deviations and interquartile ranges are almost identical. Both of these data sets are contrary to expectations. *P<sub>veg</sub>* is a constitutive promoter. No growth phase dependent difference was expected in *sfGFP\** fluorescence under *P<sub>veg</sub>* control. *sfGFP\** fluorescence is correlated to the activity of *P<sub>hyperspank</sub>*. For this reason *P<sub>hyperspank</sub>* driving expression of *mCherry2* was also expected to

produce a growth phase dependent activity phenotype. In reality there is no difference in mCherry2 fluorescence between exponential growth and stationary phase. Taken together the promoter exchange data (Figure 39) and the reporter exchange data (Figure 40) suggest there may be intrinsic properties of *sfGFP\** leading to false assumptions based on the growth phase activity. Growth phase dependent activity is lost with the alternative, mCherry2 reporter.

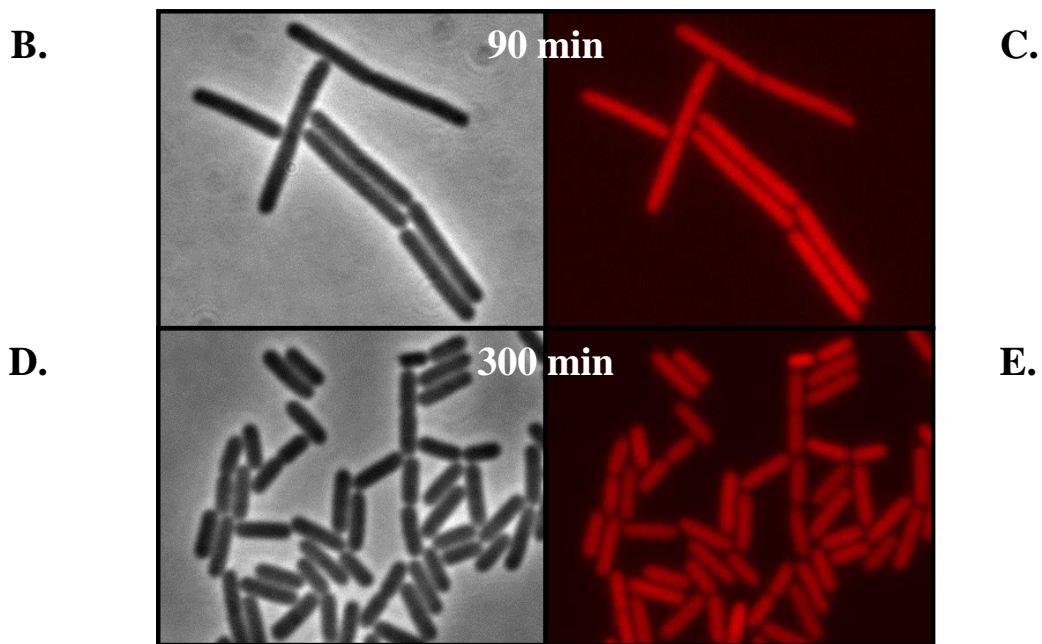
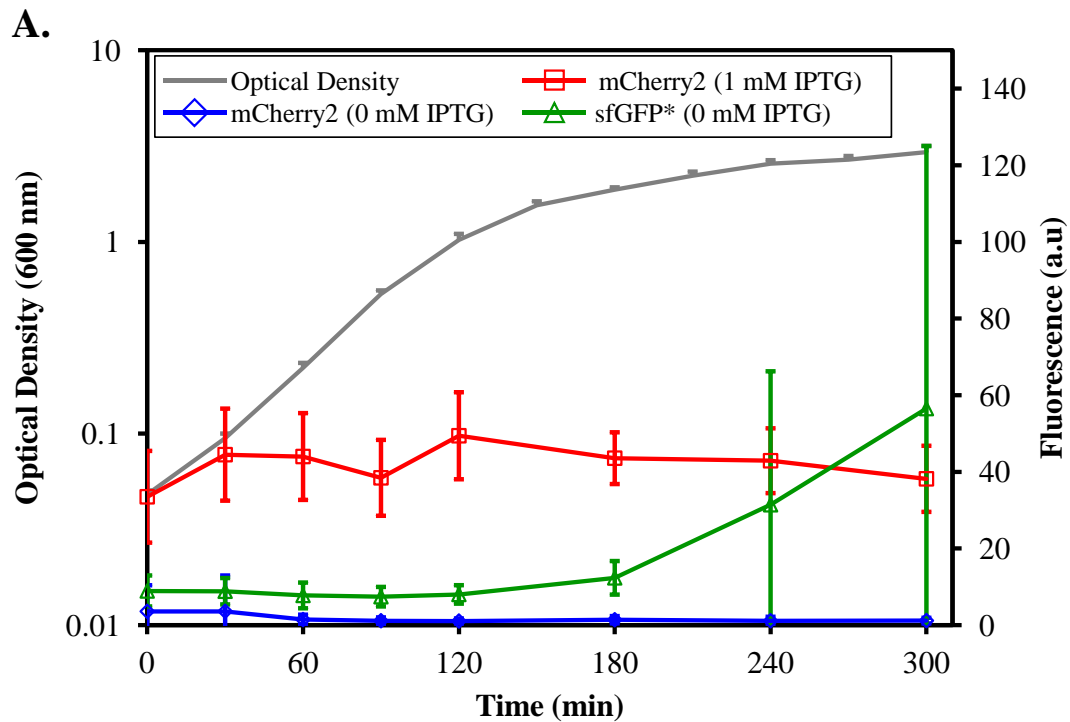


Figure 39. Growth phase dependent fluorescence does not occur with an mCherry2 reporter. **A**, growth and fluorescence in standard assay conditions. **B**, phase contrast image of mCherry2 strain (TE51, *P<sub>hyperspank</sub>-mCherry2 lacI<sub>cis</sub>*) cells at 90 minutes. **C**, fluorescent image cells as in **B**. **D**, phase contrast image of strain TE51 cells at 300 minutes. **E**, fluorescent image of cells as in **D**. Cells in **B** – **E** have been induced with 1mM IPTG. Red fluorescence is false coloured from greyscale. Blue line, mCherry2 fluorescence induced with 1 mM IPTG. Orange line, mCherry2 fluorescence without IPTG induction. Green line sfGFP\* fluorescence without IPTG induction. Grey lines, mean optical density. sfGFP\* fluorescence in the 1 mM IPTG condition has been omitted from this figure. For representative data see Figure 30, page 89. Cell numbers corresponding to  $n \geq 3$  independent repeats are recorded in Table 21, page 165.

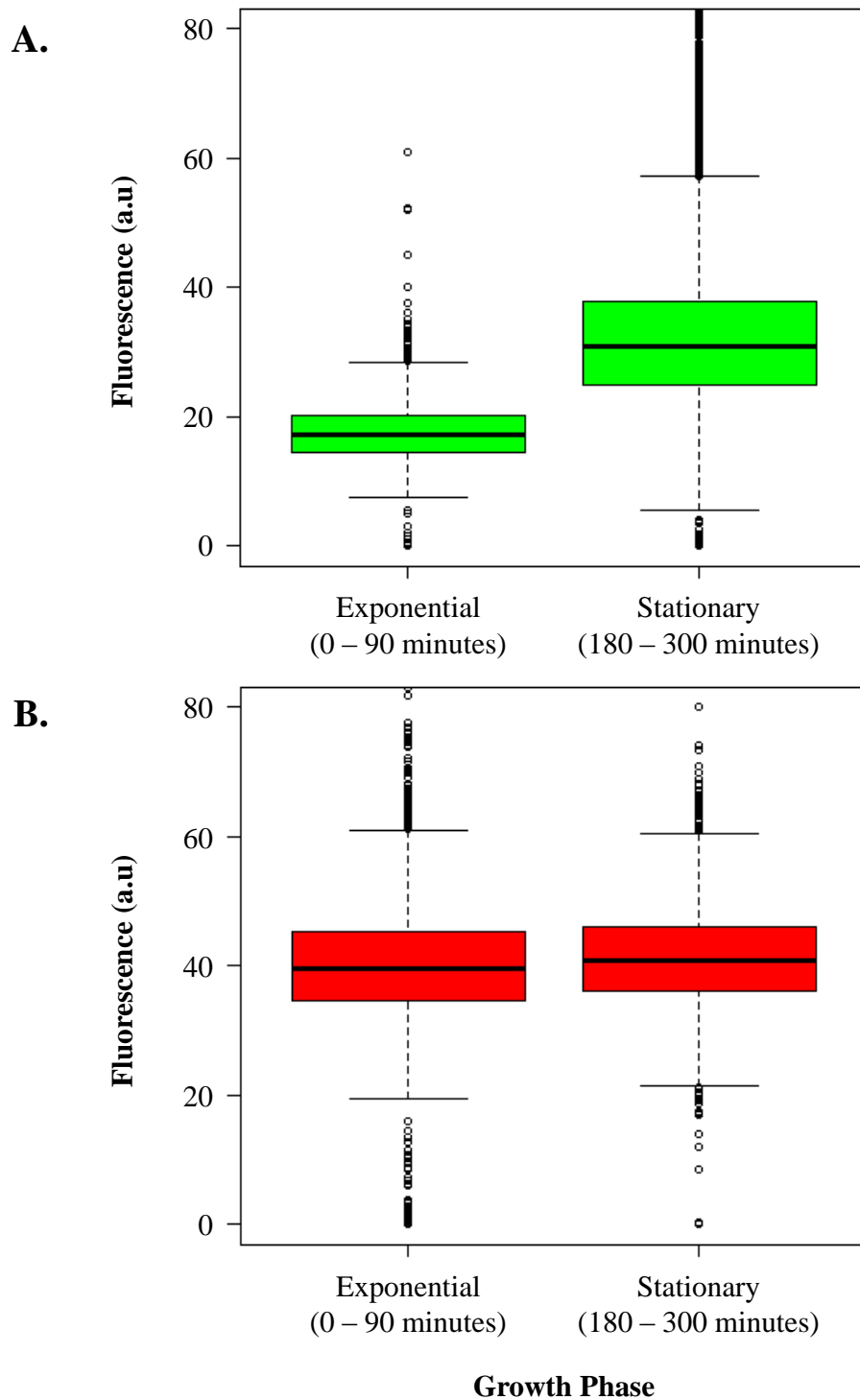


Figure 40. Growth phase dependent differences in sfGFP\* and mCherry2 fluorescence. **A**, sfGFP\* fluorescence in strain TE52 ( $P_{veg}$ -sfGFP\*) during exponential and stationary phase growth. Mean fluorescence in exponential growth is 17.6 a.u and 34.9 a.u in stationary phase. **B**, mCherry2 fluorescence in strain TE51 ( $P_{hyperspank}$ -mCherry2  $lacI_{cis}$ ) during exponential and stationary phase growth under full induction conditions. Mean fluorescence during exponential growth is 39.9 a.u and 41.1 a.u in stationary phase. The y-axis limits exclude data points from the image. Data represents n=3 independent repeats.



Without IPTG induction the mCherry2 reporter also produced a different phenotype. There is no visible increase in the mean mCherry2 fluorescence at the last two time points (Figure 39 **B**, orange line). The standard deviation bars remain tightly compact suggesting there will also be little noise. However, when the noise is calculated it appears that the un-induced mCherry2 strain has very high noise to begin with, followed by consistently high noise levels thereafter (Figure 41). This is an artefact resulting from low signal to noise ratio in the red channel of the microscope. The sfGFP\* strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) has been used as a control to define the levels of auto-fluorescence (Figure 41 **B**). At the start of the assay there are low levels of fluorescence in the mCherry2 strain with heterogeneity evident from the large standard deviations (Figure 41 **B**, orange line). The system has stabilised by 60 minutes and there is little difference between red channel fluorescence in the control strain (Figure 41 **B**, purple line) and in the strain containing the *mCherry2* gene (orange line). This confirms there is no mCherry2 fluorescence from exponential growth at 60 minutes until the final time point in stationary phase. Any associated noise is experimentally produced, rather than reflecting heterogeneity in gene expression. Auto-fluorescence in the red channel is < 2.0 a.u (Figure 41 **B**). By comparison auto-fluorescence in the green channel is  $\leq 10$  a.u (Figure 23, page 72). This difference in auto-fluorescence is a technical limitation of the experiment. The signal to noise ratio in the red channel is lower than the green channel. This results in higher background mCherry2 noise compared to sfGFP\*.

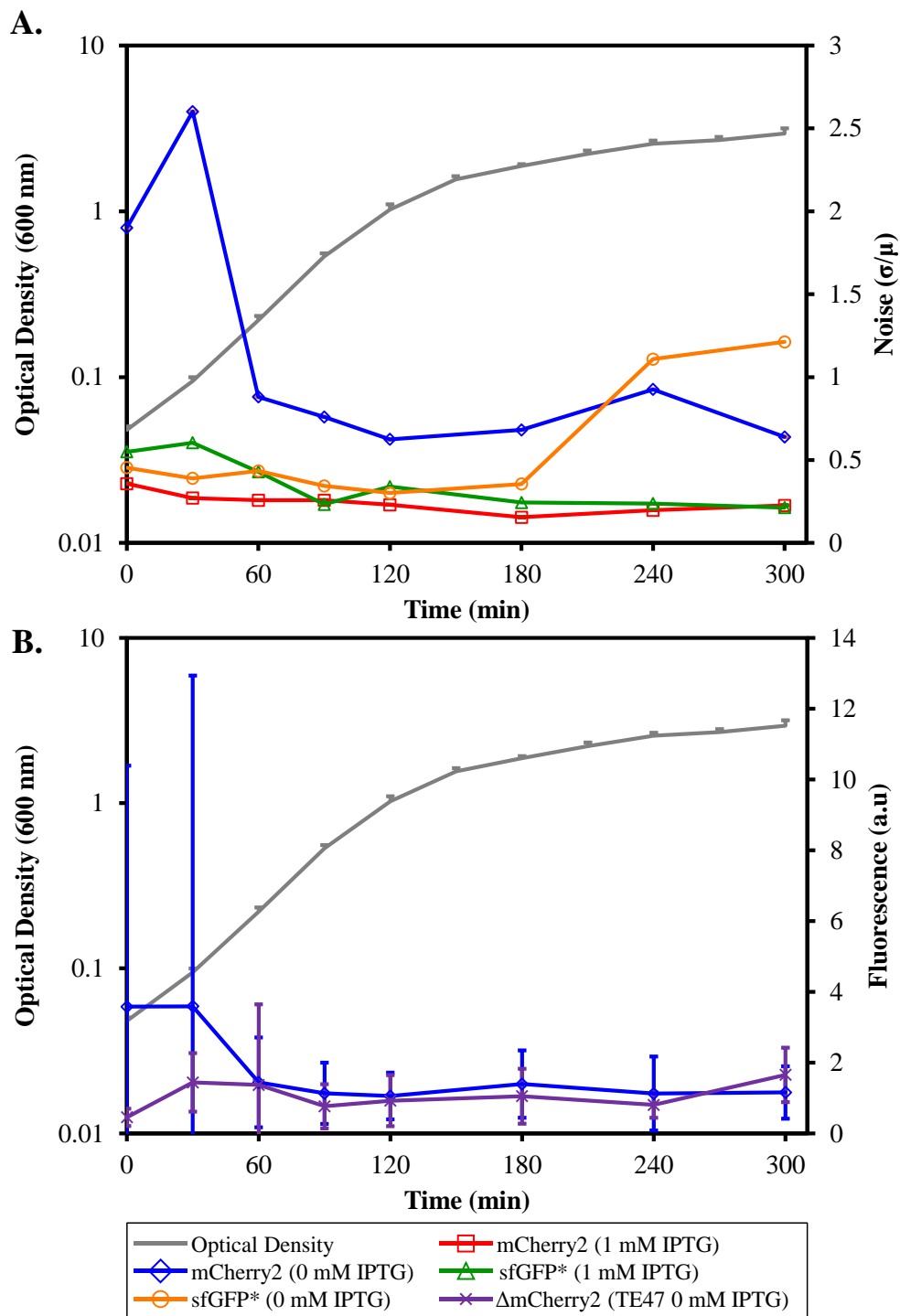


Figure 41. An artefact generates high noise with an mCherry2 reporter when compared to sfGFP\*. **A**, noise in the fluorescent activity of the sfGFP\* strain (TE47,  $P_{hyperspank}$ -sfGFP\*  $lacI$ -cis) and the mCherry2 strain (TE51,  $P_{hyperspank}$ -mCherry2  $lacI$ -cis) in standard assay conditions. **B**, red channel fluorescence without IPTG induction. The mCherry2 strain is compared to the sfGFP\* strain, acting as a negative control. Red line, noise in the mCherry2 strain induced with 1 mM IPTG. Blue line, noise the mCherry2 strain without IPTG induction. Green line, noise in the sfGFP\* strain induced with 1 mM IPTG. Orange lines, noise in the sfGFP\* strain without IPTG induction. Purple line, red channel fluorescence in the sfGFP\* strain without IPTG induction. Grey lines, mean optical density. Cell numbers corresponding to  $n \geq 3$  independent repeats are recorded in Table 21, page 165.

### 6.3. Conclusions

This chapter investigated the input of transcription and translation into the two phenotypes observed with  $P_{\text{hyperspank}}$  driving *sfGFP\** expression. Transcription of the reporter was modified through use of alternative promoters. Translation was modified by a 5 amino acid deletion at the N-terminal terminus of the sfGFP reporter and through codon optimisation of the gene. An alternative fluorescent reporter, mCherry2 was then compared to the sfGFP\* reporter.

Transcription of the gene using alternative promoters produced interesting results. Growth phase dependent activity of the reporter was still evident irrespective of the promoter used to drive gene expression. All three promoters show a transition between exponential growth and stationary phase. However all three promoters had a different profile.  $P_{\text{hyperspank}}$  was the most active promoter.  $P_{\text{xyl}}$  had similar kinetics to  $P_{\text{hyperspank}}$  but the activity, as measured by sfGFP\* activity, was reduced by approximately 10-fold. Stationary phase fluorescence increased to a lesser extent with  $P_{\text{xyl}}$  compared to  $P_{\text{hyperspank}}$ . The constitutive  $P_{\text{veg}}$  promoter was not expected to result in any growth phase dependent activity. A flat line of sfGFP\* activity was expected with this weak promoter. During exponential growth there was a characteristic dip in fluorescence, similar to  $P_{\text{hyperspank}}$ . Unlike any other promoter there was a small reduction in fluorescence from transition phase to stationary phase. The stationary phase sfGFP\* activity of  $P_{\text{veg}}$  increased very slightly. These data provide evidence that the growth phase activity is not solely dependent on the promoter. Using sfGFP\* as a reporter produced some form of growth phase dependent activity in all three of the promoters investigated.

The promoter exchange data also provided important information on the second phenotype of interest.  $P_{\text{hyperspank}}$  driving expression of *sfGFP\** resulted in a fluorescent sub-population without IPTG induction. This was evident at the last two time points during stationary phase. With  $P_{\text{xyl}}$  driving expression of *sfGFP\** there was no increase in fluorescence without xylose induction. It appears that this phenotype is specific to  $P_{\text{hyperspank}}$ . This is consistent with the known function of LacI regulation in its native organism, *E. coli*.

Two different alterations were made to the sfGFP reporter. The translational linker was removed and the codon sequence optimised. These changes fine-tuned fluorescence without

removing the growth phase dependent activity of the reporter. By replacing *sfGFP\** with the *mCherry2* reporter significant changes were observed. There was no growth phase dependent activity in the *mCherry2* reporter under full induction conditions. *mCherry2* fluorescence during exponential growth was very difficult to distinguish from stationary phase fluorescence. This suggests an intrinsic component in the nature of *sfGFP*, resulting in the growth phase dependent activity. In support of this, stationary phase heterogeneity was removed when  $P_{\text{hyperspank}}$  drives expression of *mCherry2*. Stationary phase heterogeneity is therefore specific to the promoter/reporter combination in these assay conditions. Heterogeneity was lost with  $P_{\text{hyperspank}}$  driving expression of *mCherry2* and lost with  $P_{\text{xyI}}$  driving expression of *sfGFP\**.

## Chapter 7. Dual gene reporter constructs

### 7.1. Introduction

Data in the previous chapter established that our observed phenotypes are specific to sfGFP. Swapping the *sfGFP\** gene for the *mCherry2* reporter gene removed a growth phase dependent activity and stationary phase noise in standard assay conditions (Figure 39, page 110 and Figure 41, page 113). Both the sfGFP\* strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) and the mCherry2 strain (TE51, *P<sub>hyperspank</sub>-mCherry2 lacI\_cis*) used in this thesis were designed to be identical, with the exception of the open reading frame (Figure 42 A & B). The gene lengths are similar, 713 and 708 bp for *sfGFP\** and *mCherry2* respectively. The intention was to have no differences in transcription levels when using the *sfGFP* or *mCherry2* reporter genes. If transcription is not altered by the ORF swap then further investigation would logically include translation, folding and maturation of the proteins. Prior to this it is necessary to establish if the transcription levels between the constructs are actually identical. A straightforward method to accomplish this is to express both reporter genes on a single transcript. To this end two further strains were produced: the green-red strain (TE49, *P<sub>hyperspank</sub>-sfGFP\*-mCherry2 lacI\_cis*) and the red-green strain (TE50, *P<sub>hyperspank</sub>-mCherry2-sfGFP\* lacI\_cis*) (Figure 42 C & D). In all four constructs: the sfGFP\* strain, mCherry2, green-red and the red-green strain, DNA is identical in the promoter up to the start codon of the first reporter gene. It is identical from the stop codon of the last reporter gene past the *ilvD* transcriptional terminator. The intergenic region in both of the dual reporter gene constructs is identical. It comprises a 35 bp region, containing an RBS and 7 bp spacer preceding the start codon of the second gene. This RBS and spacer is identical to the RBS and spacer preceding the first reporter gene. In short, the dual reporter constructs are identical, with the exception of the ORF swaps. The question these constructs were designed to answer is: will the sfGFP\* and mCherry2 phenotypes remain the same if the reporters are on the same transcript? If transcriptional differences are responsible for the sfGFP\* (Figure 30, page 89) and mCherry2 phenotypes (Figure 39, page 110) the following predictions can be made: sfGFP\* and mCherry2 activity in the dual reporter constructs will correlate. Either mCherry2 will exhibit growth phase dependent activity and noise or sfGFP\* will no longer exhibit growth phase dependent activity or noise. A third possibility also exists. Both phenotypes may be present irrespective of the gene order.

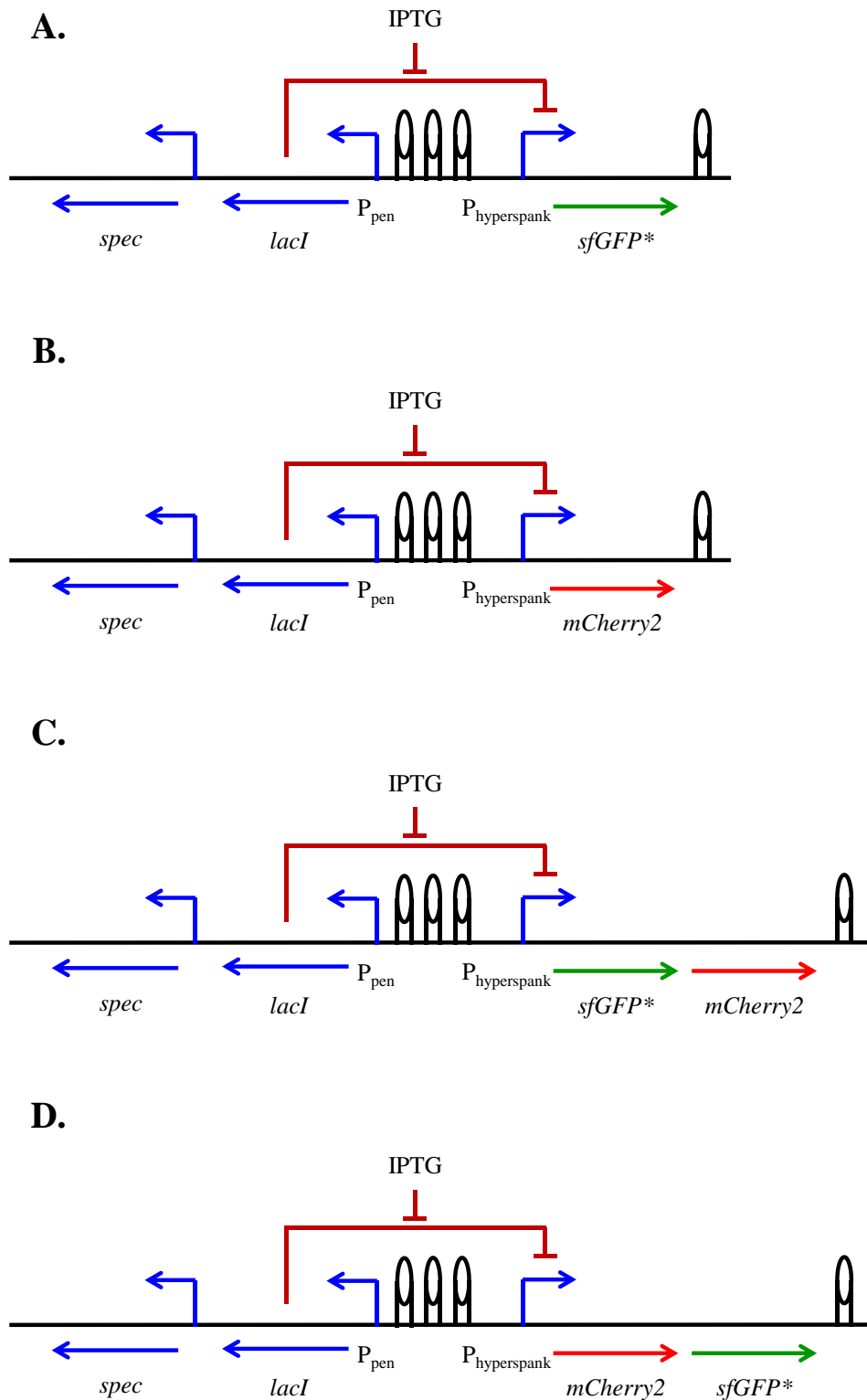


Figure 42. Monocistronic and bicistronic reporter constructs. Right-angled arrows are promoters, straight arrows are genes and T-bars indicated repression. Transcriptional terminators are raised black ovals. **A**, *sfGFP\** strain (TE47,  $P_{hyperspank}$ -*sfGFP\** *lacI\_cis*). **B**, *mCherry2* strain (TE51,  $P_{hyperspank}$ -*mCherry2* *lacI\_cis*). **C**, Green-red strain (TE49,  $P_{hyperspank}$ -*sfGFP\**-*mCherry2* *lacI\_cis*). **D**, Red-green strain (TE50,  $P_{hyperspank}$ -*mCherry2*-*sfGFP\** *lacI\_cis*).

## 7.2. Results and discussion

### 7.2.1. *Does expression from a monocistronic or bicistronic operon alter the activity of sfGFP\* or mCherry2?*

To investigate transcription effects in the monocistronic sfGFP\* and mCherry2 strains were assayed together with the green-red and red-green bicistronic strains (Figure 43). The bicistronic operons were designed to avoid any possible transcriptional differences between the monocistronic reporter-gene constructs. In both of these strains the characteristic growth phase dependent activity phenotypes in sfGFP\* fluorescence are present (Figure 43 A & B). Neither of the bicistronic reporter strains produced growth phase dependent mCherry2 fluorescence (Figure 43 C & D). Without IPTG induction, mCherry2 fluorescence has stabilised by the 60 minute time point in all three strains (Figure 43 C). There is no increase in stationary phase fluorescence. Mean mCherry2 fluorescence in stationary phase (180 – 300 minutes) is ~ 1.7 a.u., ~ 2.1 a.u and ~ 1.5 a.u for the monocistronic mCherry2, the green-red and the red-green strains respectively. As defined in section 6.2.4 this is background fluorescence. When expressed on the same transcript the difference in sfGFP\* and mCherry2 phenotypes was maintained. This was true irrespective of gene order, suggesting that the phenotypic differences are not due to a difference in transcription.

Transcription effects cannot be completely ruled out due to a difference in the magnitude of fluorescence in the bicistronic reporters. When induced, sfGFP\* fluorescence in the green-red strain was very similar to fluorescence in the monocistronic sfGFP\* strain (Figure 43 A, blue and green lines respectively). Under the same conditions, mCherry2 fluorescence in the red-green strain was very similar to the monocistronic mCherry2 strain (Figure 43 C, blue and red lines respectively). A striking difference occurred with the reporter gene transcribed from the second position in the operon. In the red-green strain sfGFP\* fluorescence during exponential growth (0 – 90 minutes) was ~ 232 a.u. (Figure 43 A). By comparison, sfGFP\* fluorescence in the monocistronic sfGFP\* strain was ~ 60 a.u. In stationary phase at the 300 minute time point, sfGFP\* fluorescence in the red-green strain was ~ 1221 a.u. and ~ 632 a.u for the monocistronic sfGFP\* strain. When expressed as the second gene in an operon sfGFP\* fluorescence was ~ 4-fold higher during exponential growth and ~ 2-fold higher in stationary phase. This is very different from the mCherry2 activity (Figure 43 C). In the green-red strain the mean mCherry2 fluorescence over the entire assay was ~ 11 a.u. In the

monocistronic mCherry2 strain the mean fluorescence was  $\sim 42$  a.u. over the entire assay. When expressed as the second gene in an operon mCherry2 fluorescence was  $\sim 4$ -fold lower.

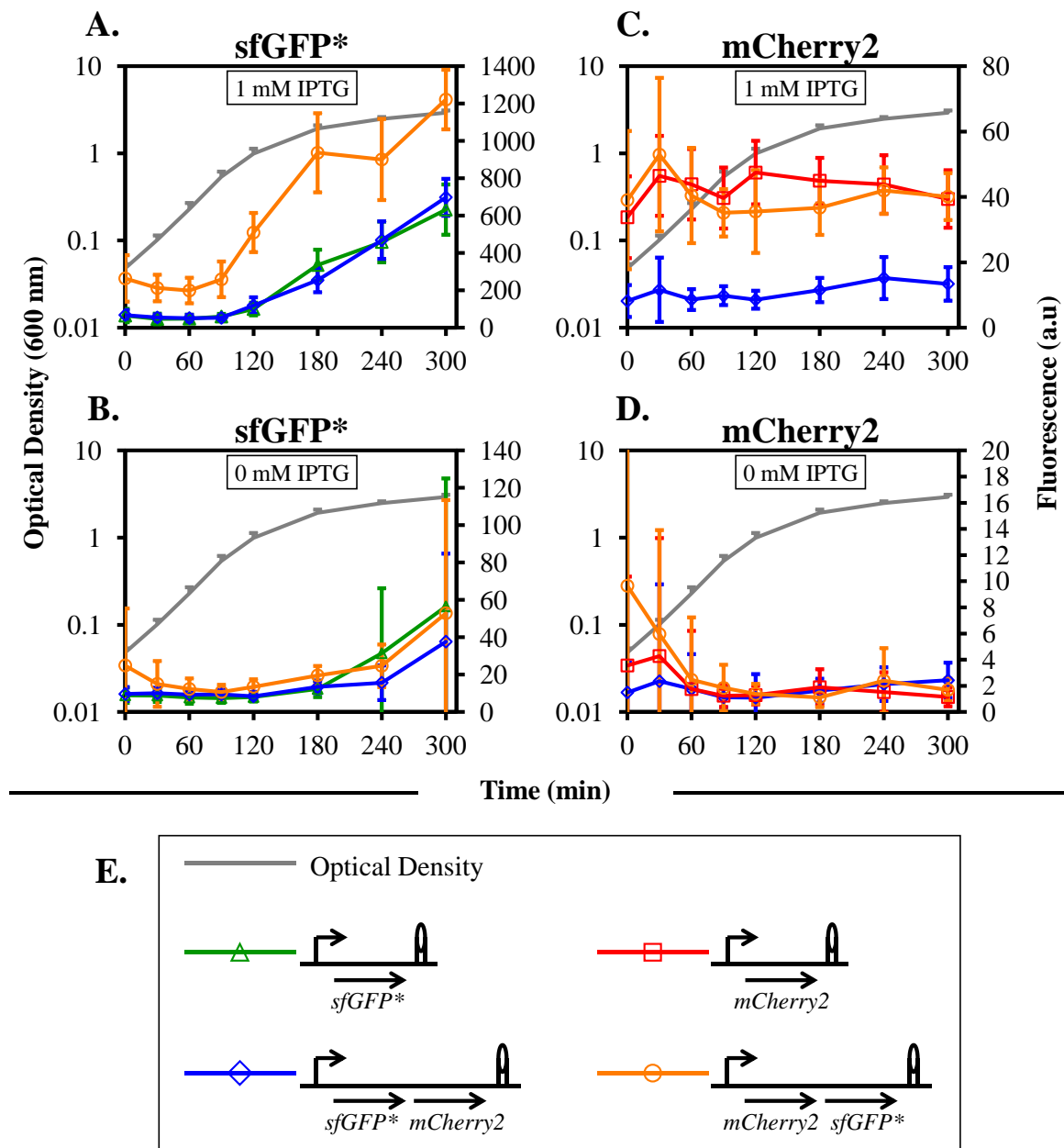


Figure 43. Gene order in bicistronic operons alters the fluorescent output of the system. **A**, sfGFP\* fluorescence induced with 1 mM IPTG. **B**, sfGFP\* fluorescence without IPTG induction. **C**, mCherry2 fluorescence induced with 1 mM IPTG. **D**, mCherry2 fluorescence without IPTG induction. **E**, schematic key of the strains/colours in A – D. Green lines, sfGFP\* strain (TE47,  $P_{hyperspank}$ -sfGFP\* lacI\_cis). Red lines, mCherry2 strain (TE51,  $P_{hyperspank}$ -mCherry2 lacI\_cis). Blue lines, green-red strain (TE49,  $P_{hyperspank}$ -sfGFP\*-mCherry2 lacI\_cis). Orange lines, red-green strain (TE50,  $P_{hyperspank}$ -mCherry2-sfGFP\* lacI\_cis). Cell numbers corresponding to  $n \geq 3$  independent repeats are recorded in Table 22, page 165.



### 7.2.2. *Does expression from a monocistronic or bicistronic operon alter the noise produced with sfGFP\* or mCherry2 reporters?*

With respect to noise the original phenotype of interest is observed with the sfGFP\* strain, in stationary phase, without IPTG induction. It was not clear why this phenotype was absent from the mCherry2 strain. Expressing both reporters on the same transcript was designed to remove differences in transcription between the reporters. It is now clear that gene order in the bicistronic operons alters the fluorescent output of the system (Figure 43). For this reason the associated noise must be considered in all four reporter strains, in both the induced and un-induced conditions. Does expression from a bicistronic operon alter the noise phenotypes? Under full induction conditions there is very little difference in sfGFP\* noise between the monocistronic and bicistronic reporters (Figure 44, **A**). The system stabilises during exponential growth and noise remains low through stationary phase. There is little difference in the stationary phase sfGFP\* noise without IPTG induction (Figure 44, **B**). The noise phenotype is evident at the final two time points in all strains. Where strains are induced with 1 mM IPTG there is little difference in mCherry2 noise (Figure 44, **C**). Noise in the green-red strain stabilises at a slightly higher level in stationary phase. The mean fluorescence in this strain was ~ 11 a.u throughout the assay. This 4-fold reduction compared to the other two mCherry2 strains results in a lower signal to noise ratio in the green-red strain. For this reason noise is expected to be higher. Without IPTG induction there was no mCherry2 noise in stationary phase (Figure 44, **D**). As previously discussed, the fluorescence in all three *mCherry2* strains was at background levels (section 7.2.1). Any stationary phase noise is therefore technical noise and can be disregarded. The data has been included as mCherry2 fluorescence at the 0 minute and 30 minute time points were above background levels for the mCherry2 and red-green strain (Figure 43 **D**). This results in high noise at the start of the assay (Figure 44, **D**. Note difference in scale). It should also be noted that the red-green strain produced higher sfGFP\* noise than the other two *sfGFP\** strains at the start of the assay (Figure 44, **B**).

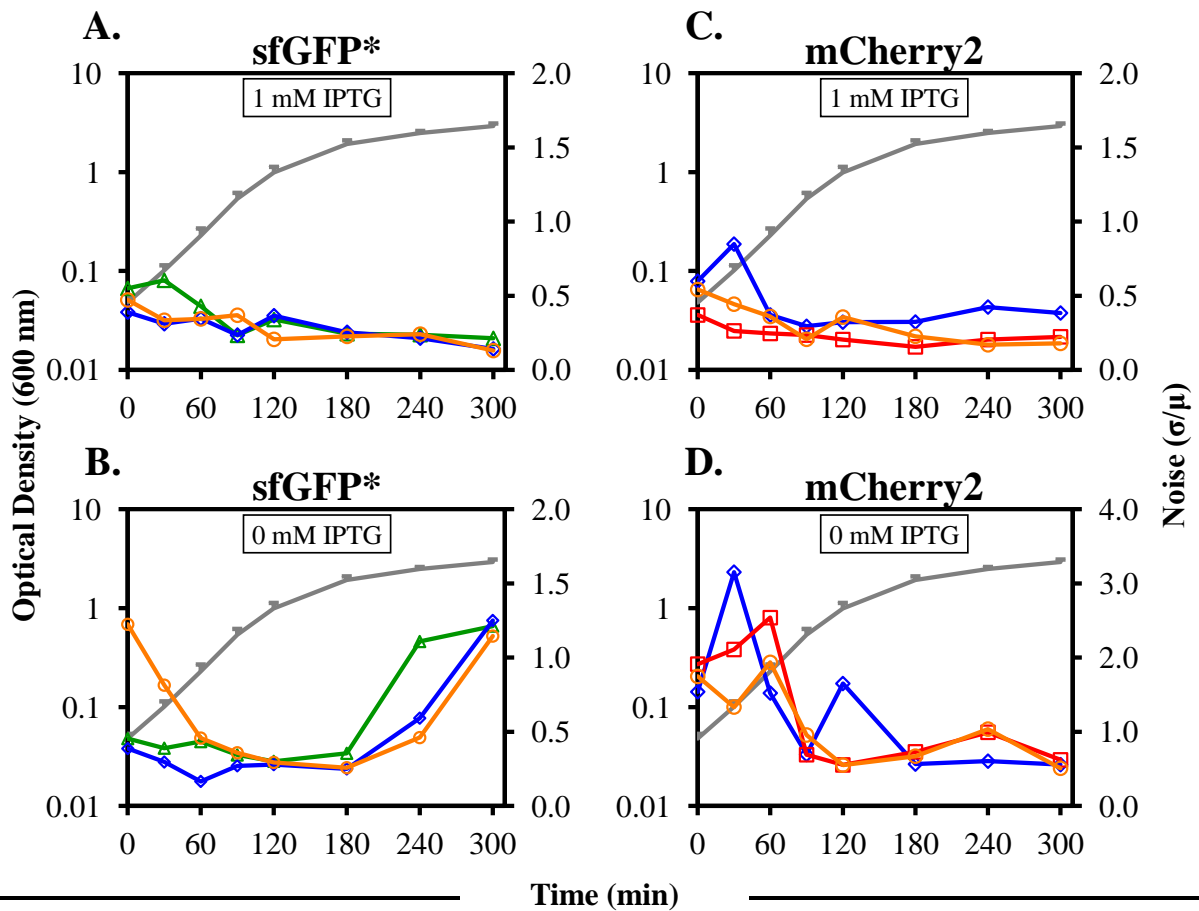


Figure 44. Gene order in bicistronic operons generates little difference to the noise profiles. **A**, sfGFP\* noise under full induction conditions. **B**, mCherry2 noise under full induction conditions. **C**, sfGFP\* noise without IPTG induction. **D**, mCherry2 noise without IPTG induction. **E**, schematic key of the strains/colours in A – D. Green lines, sfGFP\* strain (TE47,  $P_{hyperspank}$ -sfGFP\* *lacI\_cis*). Red lines, mCherry2 strain (TE51,  $P_{hyperspank}$ -mCherry2 *lacI\_cis*). Blue lines, green-red strain (TE49,  $P_{hyperspank}$ -sfGFP\*-mCherry2 *lacI\_cis*). Orange lines, red-green strain (TE50,  $P_{hyperspank}$ -mCherry2-sfGFP\* *lacI\_cis*). Cell numbers corresponding to  $n \geq 3$  independent repeats are recorded in Table 22, page 165.

### 7.3. Conclusions

Use of bicistronic reporter gene constructs was designed to remove any transcriptional difference between *mCherry2* expression and *sfGFP\**. With *sfGFP\** and *mCherry2* expressed on the same transcript the difference in reporter phenotypes still occurred. This strongly suggests a difference in transcription was not responsible for the difference in *sfGFP\** and *mCherry2* phenotypes. However, the magnitude of fluorescent activity was dramatically altered by gene position in the operon. As the first gene in a bicistronic operon both *sfGFP\** and *mCherry2* behaved like the monocistronic *sfGFP\** and *mCherry2* strains. *sfGFP\** as the second gene in the operon led to a ~ 4-fold increase in fluorescence during exponential growth. *mCherry2* as the second gene in the operon reduced the fluorescence by ~ 4-fold over the course of the assay. This unexpected behaviour of the bicistronic reporters does not allow transcriptional effects to be ruled out. Currently all *sfGFP\** strains exhibit growth phase dependent activity and stationary phase noise. No *mCherry2* strains produced growth phase dependent activity or stationary phase noise. Where *mCherry2* was expressed as the first gene, or only gene in an operon there was increased heterogeneity at the start of the assay. This translated as very high noise, due to the difference in *B. subtilis* auto-fluorescence in the green and red channels. It is necessary to further investigate both transcript and protein levels to explain the difference in *sfGFP\** and *mCherry2* phenotypes.

## Chapter 8. Correlating fluorescent activity with RNA and protein levels

### 8.1. Introduction

This chapter further investigates the difference in phenotypes between *sfGFP\** and *mCherry2* strains. The growth phase dependent *sfGFP\** phenotypes were absent when the reporter gene was substituted for *mCherry2*. A possible difference in transcription between these two strains was investigated by expressing the two reporter genes on a single mRNA transcript. In one dual reporter strain, *sfGFP\** is transcribed before *mCherry2*. In the other dual reporter strain, gene order is reversed and *mCherry2* is transcribed first. In all of these strains *sfGFP\** and *mCherry2* activity are very different. There is always a growth phase dependent activity with the *sfGFP\** reporter and never with the *mCherry2* reporter. Without IPTG induction there is always stationary phase noise in the *sfGFP\** reporter and never with the *mCherry2* reporter. There were also individual differences between the two dual reporter constructs. The activity of the first gene transcribed is very similar to the activity of the corresponding single gene construct. As the second gene in an operon *sfGFP\** and *mCherry2* behave differently. *sfGFP\** as the second gene leads to a ~ 4-fold increase in exponential phase *sfGFP\** activity. *mCherry2* as the second gene reduces *mCherry2* activity by ~ 4-fold over the entire assay. The phenotypic differences between these four constructs cannot be explained simply by quantifying the fluorescent activity. To fully understand the system it is necessary to correlate the processes involved in gene expression with the observed fluorescent output. Therefore, relative levels of mRNA transcript, mRNA stability and protein levels were investigated to correlate with the fluorescent activity.

### 8.2. Results and discussion

#### 8.2.1. *Do transcription levels match fluorescent activity?*

$P_{\text{hyperspank}}$  driving expression of either *sfGFP* or *mCherry2* produced different activity phenotypes. To assess if transcription levels match the fluorescent activity, standard assay conditions were modified to enable collection of samples for Northern blotting, microscope data and growth rate. Samples were taken for Northern blotting in exponential growth (90 minutes), transition phase (120 minutes) and two time points in stationary phase (210 and 300 minutes). The change in assay conditions did not alter the phenotypes observed in Chapter 7.

Northern blots were used to evaluate the relative transcript levels between the different growth phases (Figure 45). Under full induction conditions *sfGFP\** mRNA levels in the single gene *sfGFP\** strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) do not match the fluorescent activity (Figure 45 **A**). Fluorescent activity increased while transcript levels remained high throughout the assay (Figure 55, page 155). Without IPTG induction the *sfGFP\** transcript levels and the fluorescent activity correlate more closely. During exponential growth and in transition phase there are very low levels of transcript and background levels of fluorescence (Figure 55, page 155). At 210 minutes the highest transcript levels are detected. This correlates with the first increases in fluorescence, ~ 10 a.u. above background levels. A large increase in fluorescence then follows at the final, 300 minute time point and is matched by high levels of mRNA.

The single gene *mCherry2* strain (TE51, *P<sub>hyperspank</sub>-mCherry2 lacI\_cis*) produces a difference in both the fluorescent activity and the transcript levels, when compared to the *sfGFP\** strain. Under inducing conditions *mCherry2* fluorescent activity remains relatively constant (Figure 56, page 156). This is not matched by similar levels of mRNA. Unlike *sfGFP\**, as time progresses through the assay *mCherry2* mRNA levels continually reduce (Figure 45 **B**). The *mCherry2* reporter strain requires less and less mRNA to maintain similar levels of fluorescence in the system. The reduction in mRNA levels are statistically significant with a 95% confidence interval and p-value < 0.05 (Table 36, page 172). Without IPTG induction *mCherry2* transcript is only visible at the final time point (Figure 45 **B**). This increase in transcript does not produce heterogeneity within the population and there is no associated increase in noise (Figure 56 **B**, page 156).

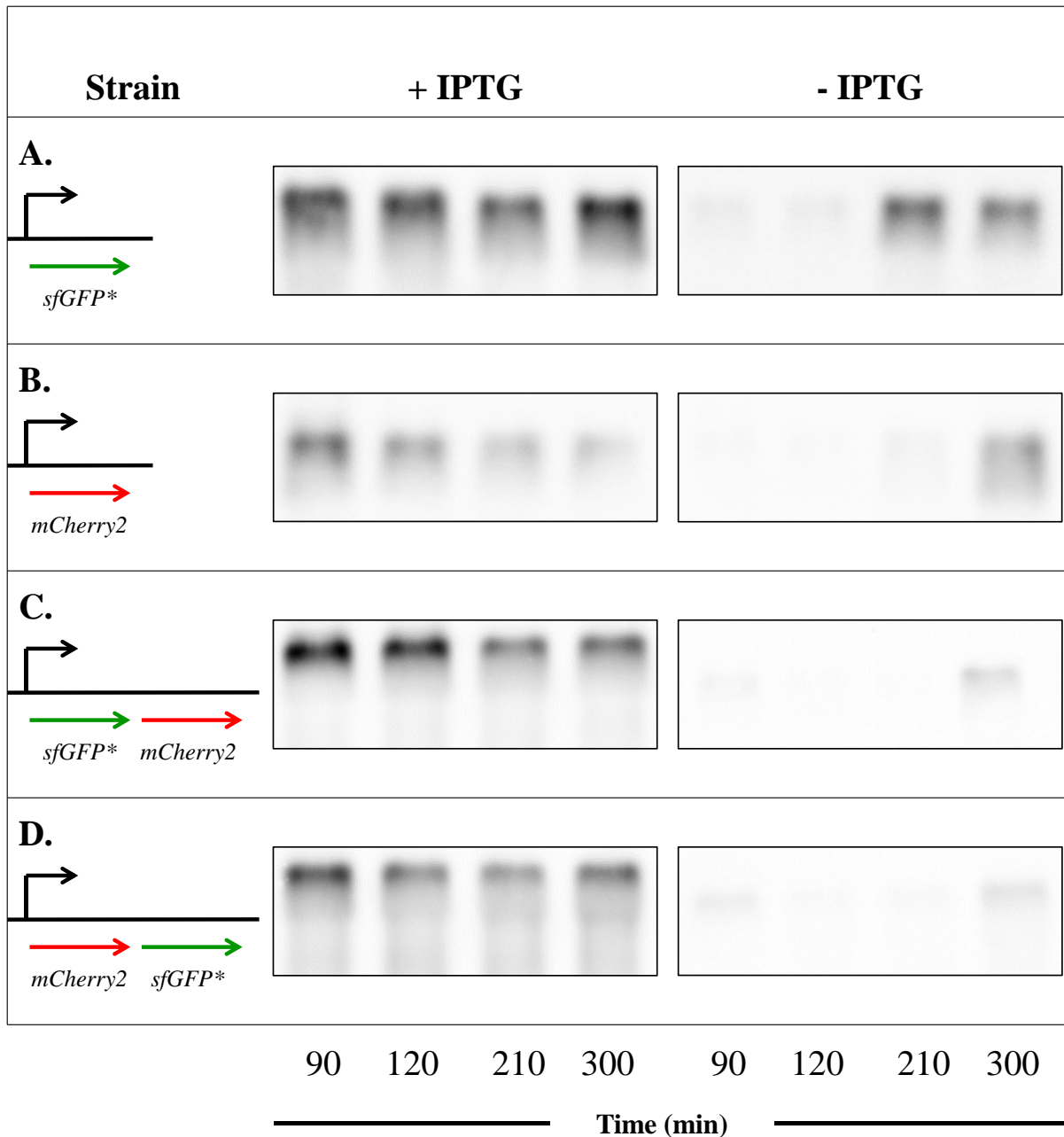


Figure 45. Transcript variation in  $P_{\text{hyperspank}}$  strains. Northern blotting was performed at the specified times on cultures grown in 1 mM IPTG and without IPTG. **A**, *sfGFP\** strain (TE47,  $P_{\text{hyperspank}}\text{-}sfGFP^* lacI_{cis}$ ) using an *sfGFP\** probe. **B**, *mCherry2* strain (TE51  $P_{\text{hyperspank}}\text{-}mCherry2 lacI_{cis}$ ) using an *mCherry2* probe. **C**, green-red strain ( $P_{\text{hyperspank}}\text{-}sfGFP^*\text{-}mCherry2 lacI_{cis}$ ) using an *sfGFP\** probe. **D**, red-green strain ( $P_{\text{hyperspank}}\text{-}mCherry2\text{-}sfGFP^* lacI_{cis}$ ) using an *sfGFP\** probe. See appendix: Figure 55, page 155 - Figure 58, page 158 for associated growth, fluorescence and noise data. Data is representative of  $n=3$  independent repeats.

The fluorescent activity phenotypes and noise in the dual reporter gene constructs were as expected (Figure 57, page 157). In the green-red strain (TE49, *P<sub>hyperspank</sub>-sfGFP\*-mCherry2 lacI\_cis*) there was only growth phase dependent activity and noise with the sfGFP\* reporter. Under inducing conditions there were high levels of transcript throughout the assay (Figure 45 C). Without IPTG induction the increase in transcript occurred only at the final time point. Transcript levels in both dual reporter gene constructs were similar. Inducing conditions in the red-green strain (TE50, *P<sub>hyperspank</sub>-mCherry2-sfGFP\* lacI\_cis*) resulted in high transcript levels throughout the assay (Figure 45 D). Without IPTG induction the increase in transcript was only observed at the final time point. In both dual gene reporter constructs, transcript levels without IPTG induction were most similar to the single gene mCherry2 construct. The increase in transcript occurred only at the 300 minute time point. In the single gene sfGFP\* strain transcript levels increased by 210 minutes.

Irrespective of the strain, transcript levels did not match the sfGFP\* activity when induced with 1 mM IPTG. This confirms that the growth phase dependent activity phenotype observed with sfGFP\* is not transcription dependent. The difference between *sfGFP\** and *mCherry2* RNA levels in the single gene constructs has not been explained. Why, with identical 1 mM IPTG induction conditions, is there a difference in *mCherry2* mRNA compared to *sfGFP\** mRNA? Without IPTG induction the transcription levels matched the sfGFP\* activity. The stationary phase noise observed with the sfGFP\* reporters is therefore transcription dependent. This also raises the question of why there is no noise with the mCherry2 reporters when there is clearly transcript present.

### **8.2.2. Does the reporter gene alter the transcript stability?**

Under full induction conditions fluorescence and mRNA levels in the single gene reporter constructs are significantly different. There is least sfGFP\* fluorescence in exponential growth and high levels of *sfGFP\** mRNA (Figure 55, page 155 and Figure 45 A). By comparison mCherry2 fluorescence remains constant while mRNA levels drop (Figure 56, page 156 and Figure 45 B). The expectation that mRNA levels in both of these constructs would remain similar in both of these strains was not observed. To test if mRNA stability could explain this difference, transcription in these strains was blocked with rifampicin and Northern blotting was used to calculate the half-lives of the transcripts (Table 37, page 173). During exponential growth *mCherry2* is more stable than *sfGFP\**. The initial degradation of both transcripts is similar (Figure 46 A & B). Initially there is a sharp drop in the transcript

levels of both strains. *sfGFP\** transcript levels continue to drop sharply compared to the *mCherry2* transcript. The *mCherry2* mRNA has a rapid initial reduction followed by a slow decline. In stationary phase the stability of both transcripts appears reduced (Figure 46 C & D). The mean half-life of *sfGFP\** during exponential growth is 8.3 minutes and 4.8 minutes in stationary phase (Table 37, page 173). The *sfGFP\** differences were only significant with a p-value < 0.074. By comparison the mean *mCherry2* half-life was 14.0 minutes during exponential growth and 4.2 minutes in stationary phase. The *mCherry2* differences were significant with a p-value < 0.053. These data strongly suggest that *mCherry2* mRNA stability has a growth phase dependency.

A growth phase dependent difference in mRNA stability raises questions when considering the dual reporter gene constructs. In the dual reporter strains the fluorescent activity of the first gene transcribed was always most similar to that of the corresponding single gene construct. Would the transcript stability also differ according to the order of the reporter genes? The green-red and red-green strains were assayed for mRNA degradation using probes for both *sfGFP\** and *mCherry2* (Figure 47). During exponential growth transcript stability in the green-red strain appeared lower than the red-green strain. Degradation of the *sfGFP\** and *mCherry2* transcripts in the green-red strain (Figure 47 A, blue and orange lines) appear similar to the single gene *sfGFP\** transcript (Figure 46 A, red line). Degradation of the *sfGFP\** and *mCherry2* transcripts in the red-green strain (Figure 47 A, green and purple lines) appear similar to the single gene *mCherry2* construct (Figure 46 A, black line). This suggests that gene order in the bicistronic operons affects mRNA stability during exponential growth.

As with the single gene constructs the transcript stability in the dual reporter gene constructs is reduced in stationary phase (Figure 47 D - F). Analysis reveals little difference in the mRNA levels between the strains assayed and probes used (Figure 47 D). There is no lingering decline in transcript levels during stationary phase. Visual inspection of the data gives good cause to assert there is a growth phase dependent difference in transcript stability. However, the calculated difference in transcript half-lives between exponential growth and in stationary phase was not statistically significant for either of the dual reporter strains (Table 37, page 173).



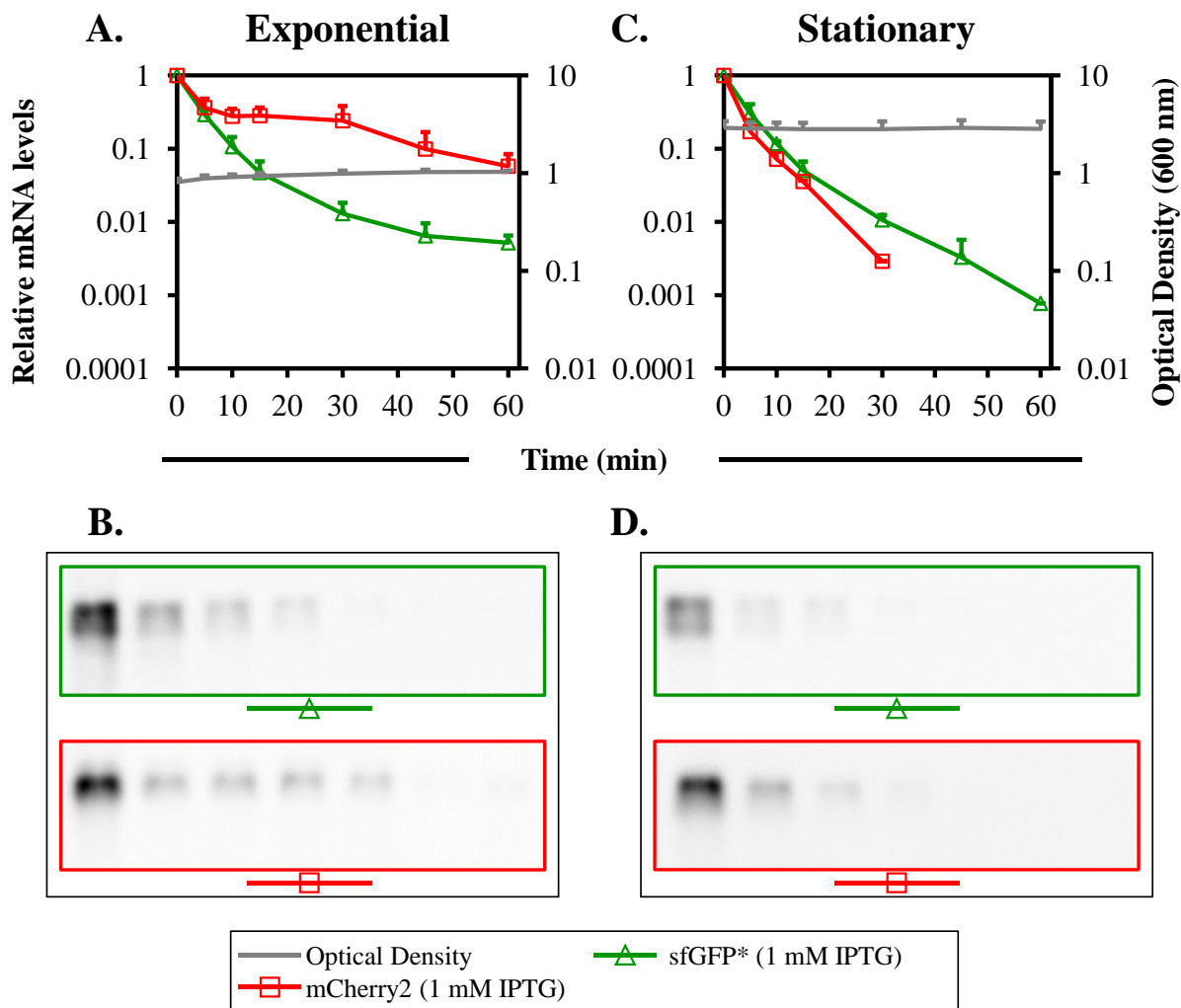


Figure 46. mRNA degradation in the single gene *sfGFP\** and *mCherry2* strains. mRNA stability is compared during exponential growth (90 minutes) and in stationary phase (300 minutes) in the *sfGFP\** (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) and *mCherry2* (TE51, *P<sub>hyperspank</sub>-mCherry2 lacI\_cis*) strains. Rifampicin was used to block transcription and Northern blotting used to calculate the relative *sfGFP\** and *mCherry2* transcript levels. **A**, mRNA degradation during exponential growth. **B**, Northern blots taken from samples during exponential growth. **C**, mRNA degradation in stationary phase. **D**, Northern blots taken from samples during stationary phase. Green lines/boxes, *sfGFP\** strain. Red lines/boxes, *mCherry2* strain. Half-life calculations and statistical analyses are recorded in Table 37, page 173. Data is representative of n=3 independent repeats.

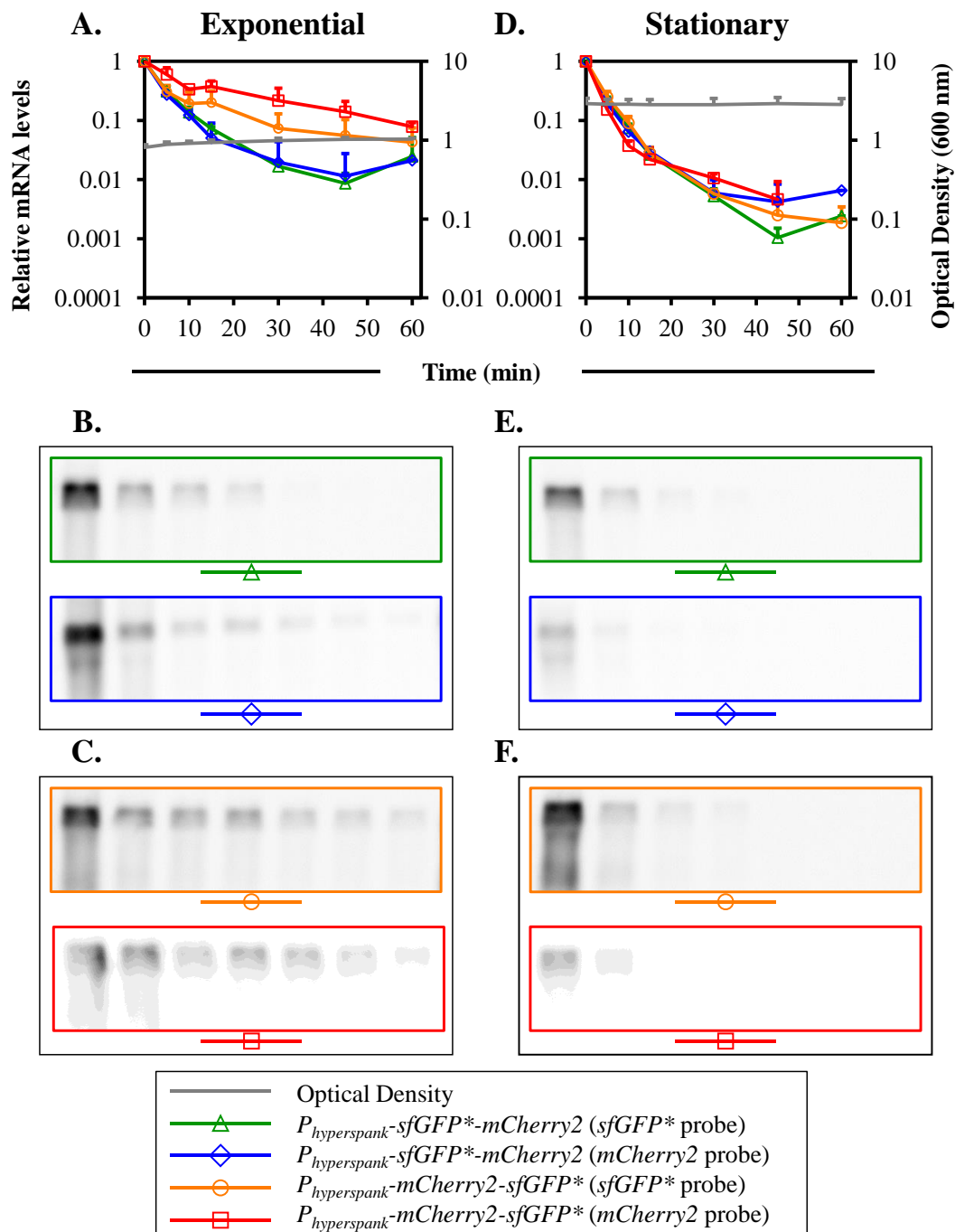


Figure 47. mRNA degradation in the bicistronic *sfGFP\*-mCherry2* and *mCherry2-sfGFP\** reporter strains. mRNA stability is compared during exponential growth (90 minutes) and in stationary phase (300 minutes) in the green-red (TE49, *P<sub>hyperspank</sub>-sfGFP\*-mCherry2 lacI\_cis*) and red-green (TE50, *P<sub>hyperspank</sub>-mCherry2-sfGFP\* lacI\_cis*) strains. Rifampicin was used to block transcription and Northern blotting used to calculate the relative *sfGFP\** and *mCherry2* transcript levels. **A**, mRNA degradation during exponential growth. **B**, Northern blots of the green-red strain during exponential growth. **C**, Northern blots of the red-green strain during exponential growth. **D**, mRNA degradation in stationary phase. **E**, Northern blots of the green-red strain during exponential growth. **F**, Northern blots of the red-green strain during exponential growth. Half-life calculations and statistical analyses are recorded in Table 37, page 173. Data is representative of n=3 independent repeats for *sfGFP\** probes and n=2 independent repeats for *mCherry2* probes.

### 8.2.3. *Do protein levels match the activity profile?*

The time taken for translation of mRNA into an active protein has not been quantified in this thesis. Correct folding and maturation of the fluorescent reporters takes time. In the dynamic process of gene expression there is likely to be sfGFP\* and mCherry2 protein at different stages of synthesis and activity. For this reason it is important to establish a correlation between protein levels and fluorescent activity. There is the possibility that protein levels will not match either the mRNA abundancies or fluorescent activity of the reporter. Any difference could provide insight into the phenotypes observed with the sfGFP\* and mCherry2 reporters. The question must therefore be asked, will the level of protein match the mRNA transcript levels and the fluorescent output in the system?

Samples were taken for Immunoblotting during modified standard assay conditions (see section 8.2.1). A typical increase in fluorescence was observed during the growth phases in the single reporter gene sfGFP\* strain (Figure 48 A). There are notable similarities between the relative levels of fluorescence and the relative protein levels (Figure 48 B & C). There is an increase in both sfGFP\* protein levels and fluorescence over the course of the assay. This suggests that protein in the sfGFP\* strain is mostly active. This is very different from the single gene mCherry2 strain. The assay conditions produced little change in mCherry2 fluorescence, ranging between 36 a.u and 56 a.u (Figure 48 D). By comparison mean sfGFP\* fluorescence ranges between 63 a.u and 505 a.u (Figure 48 A). The relatively stable mCherry2 activity is observed concurrently with a reduction in mCherry2 protein levels (Figure 48 E & F). There is most similarity between mCherry2 protein levels (Figure 48 E) and *mCherry2* transcript levels (Figure 45 B). When compared to the sfGFP\* strain, mCherry2 translation is more closely correlated to the mRNA levels and there is less activity in the pool of translated proteins. The mCherry2 strain contains a higher proportion of inactive reporter protein. It is noted that there are two bands present on the mCherry2 Immunoblots (Figure 48 E). There is no visible difference in the intensity of the two bands. For this reason it is considered unlikely to be the source of disparity between the relative protein levels and relative fluorescence levels. If the bands were separated the relative protein levels would not change. As such identifying an RBS and start codon downstream of the ORF start codon is not a priority.

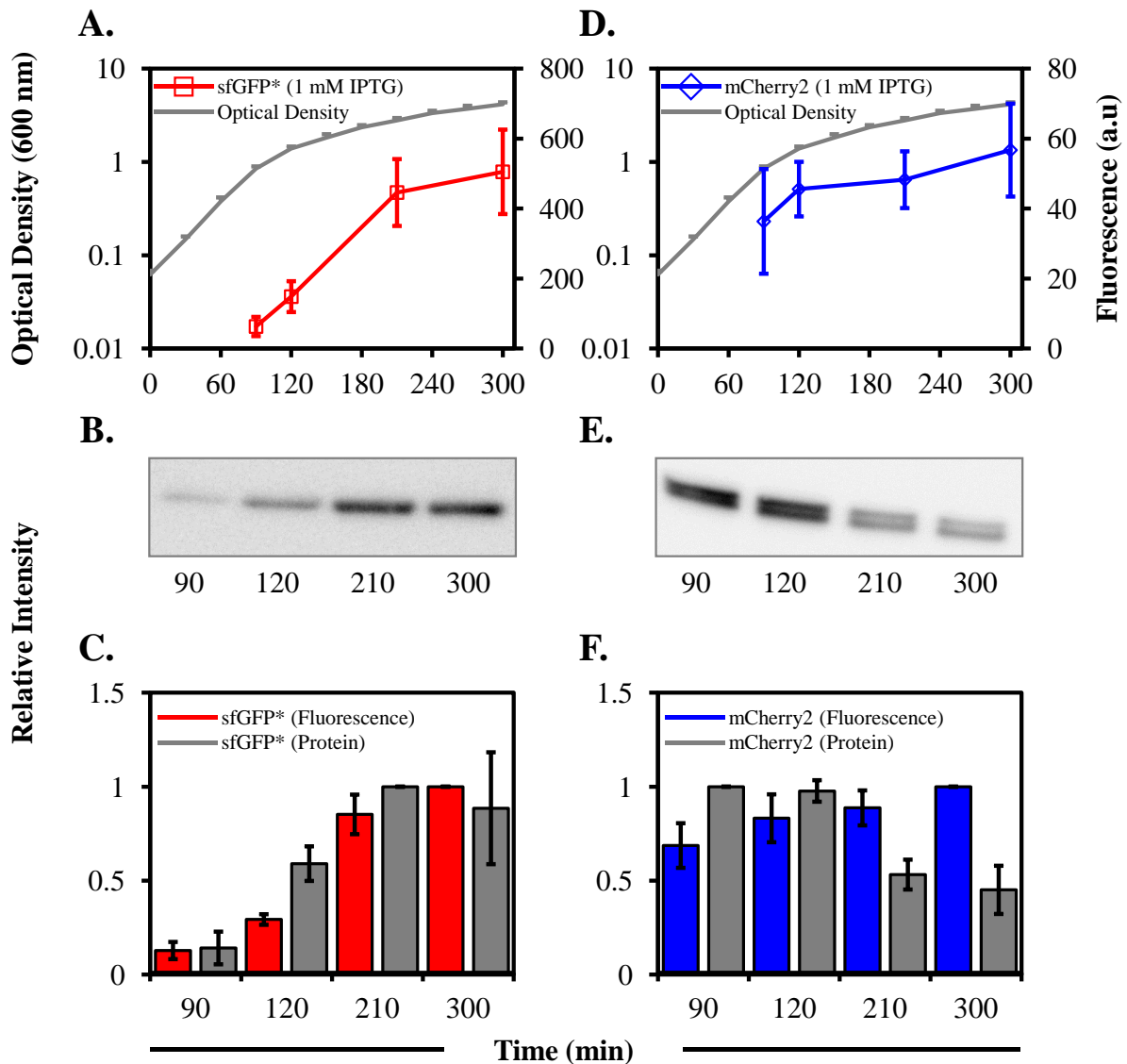


Figure 48. Protein levels and fluorescent activity are monitored concurrently with growth in the single reporter gene *sfGFP\** and *mCherry2* constructs. The *sfGFP\** (TE47,  $P_{hyperspank}$ -*sfGFP\** *lacI\_cis*) and *mCherry2* (TE51,  $P_{hyperspank}$ -*mCherry2* *lacI\_cis*) strains were induced for gene expression with 1 mM IPTG. **A**, fluorescence and growth in the *sfGFP\** strain (red and grey lines respectively). **B**, Immunoblots of the *sfGFP\** strain. **C**, relative fluorescence and relative protein levels in the *sfGFP\** strain (red and grey bars). **D**, fluorescence and growth in the *mCherry2* strain (blue and grey lines). **E**, Immunoblots of the *mCherry2* strain. **F**, relative fluorescence and relative protein levels in the *mCherry2* strain (blue and grey bars). Relative values are calculated as the value at  $T_n/T_{max}$ , where  $T = time (minutes)$ . Data is representative of  $n=3$  independent repeats. Cell numbers recorded in Table 23 page 165.

When both reporter genes are transcribed on a single mRNA, gene order affects the fluorescent activity (Figure 43, page 119). With standard assay conditions modified to remove samples for immunoblotting this difference in activity remains (Figure 49). The

green-red strain displays the characteristic stationary phase increase in sfGFP\* fluorescence (Figure 49 A, green line). As expected, mCherry2 fluorescence remains low, increasing from ~ 10 a.u during exponential growth to ~ 15 a.u in stationary phase at 300 minutes (Figure 49 A, red line). Both sfGFP\* and mCherry2 protein levels in this strain are very similar to each other, increasing throughout the assay (Figure 49 B). When expressed as the first gene in an operon sfGFP\* protein levels are similar to the single gene sfGFP\* strain, suggesting a high proportion of active protein (Figure 48 B). mCherry2 protein levels in the green-red strain increase through the assay, without a corresponding increase in fluorescent activity (Figure 49 A & B). This indicates the presence of inactive mCherry2 protein in the system. In the green-red dual reporter strain there is inactive mCherry2 protein during stationary phase. In the single mCherry2 strain the inactive protein is observable during exponential growth and in transition phase (Figure 48 E).

The dual reporter red-green strain produced similar levels of mCherry2 fluorescence as the single gene mCherry2 construct (Figure 43, page 119). When transcribed as the second gene in the operon, sfGFP\* fluorescence increased in comparison to the single gene sfGFP\* strain. Under modified assay conditions for immunoblotting there were expected levels of fluorescence in the red-green strain (Figure 49 C). Protein levels in the red-green strain were high throughout the assay (Figure 49 D). The increase in sfGFP\* fluorescence in the red-green strain is consistent with the observable increase in sfGFP\* protein levels (Figure 43, page 119 and Figure 49, page 133 ). When expressed as the second gene in an operon sfGFP\* fluorescence during exponential growth was ~ 4-fold higher than in the single gene sfGFP\* strain. Stationary phase sfGFP\* fluorescence was ~ 2-fold higher in the dual reporter red-green strain. A corresponding increase in sfGFP\* protein levels was observed in the red-green strain (Figure 49 C & D). mCherry2 fluorescence in the red-green strain is relatively stable and present alongside high levels of mCherry2 protein (Figure 49 C & D). The fluorescence in this dual reporter strain is similar to the single gene mCherry2 strain (Figure 48 D). However, the mCherry2 protein levels in the red-green strain do not match the mCherry2 protein levels in the single gene mCherry2 strain. In the red-green strain mCherry2 protein levels are higher in stationary phase than in the single gene mCherry2 strain. Increased mCherry2 protein levels do not lead to an increase in the fluorescent activity of the red-green strain. In all strains assayed for immunoblotting and fluorescence, there is greater correlation between sfGFP\* protein levels and fluorescence than mCherry2

protein levels and fluorescence. The sfGFP\* reporter produces a higher proportion of active protein than the mCherry2 reporter.

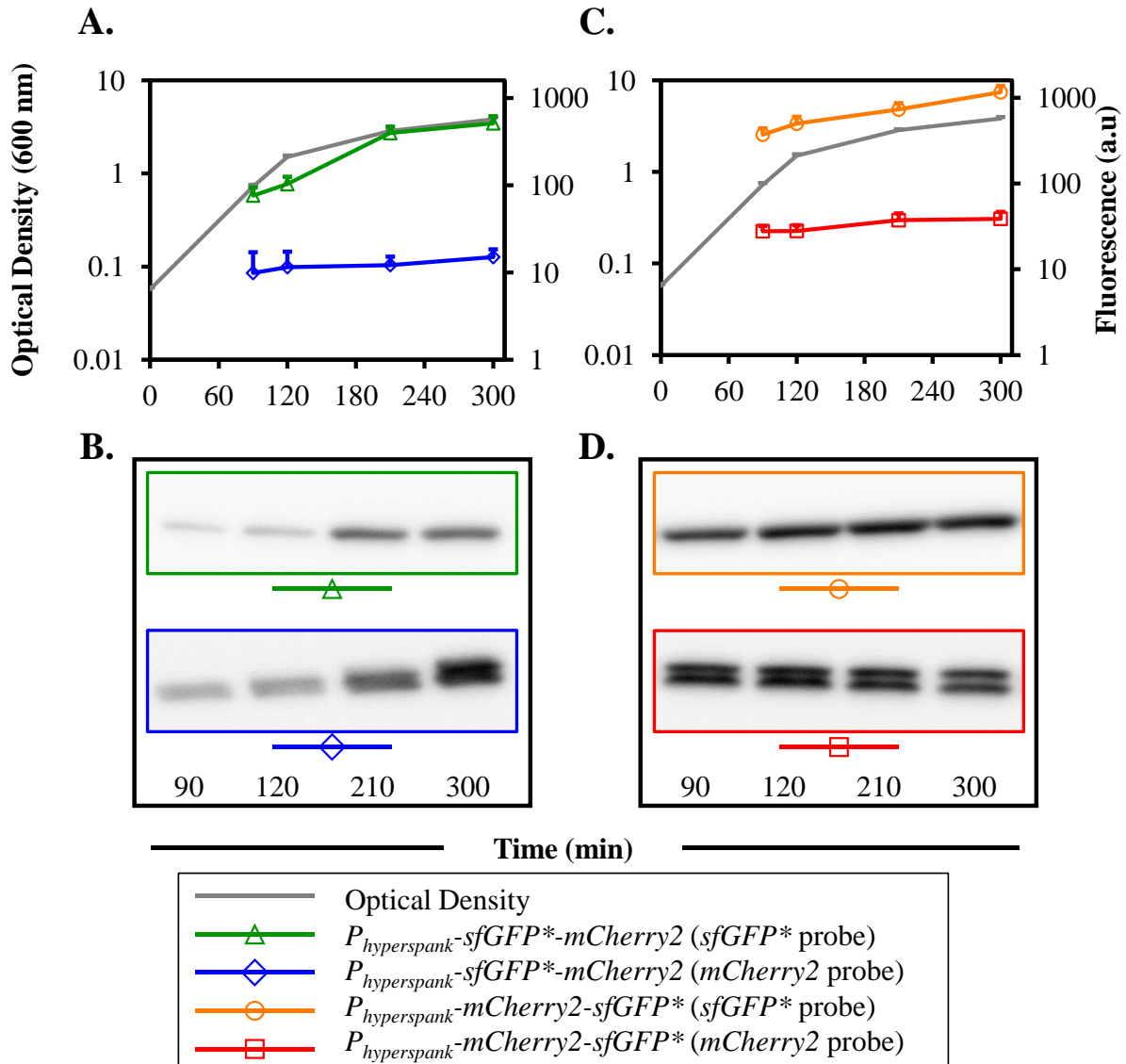


Figure 49. Comparison of protein levels and fluorescent activity in the bicistronic *sfGFP\*-mCherry2* and *mCherry2-sfGFP\** reporter constructs. The green-red (TE49, *P<sub>hyperspank</sub>-sfGFP\*-mCherry2 lacI\_cis*) and red-green (TE50, *P<sub>hyperspank</sub>-mCherry2-sfGFP\* lacI\_cis*) strains are induced for gene expression with 1 mM IPTG. **A**, fluorescence (coloured lines) and growth (grey line) in the green-red strain. **B**, Immunoblots of fluorescent reporters in the green-red strain. **C**, fluorescence and growth in the red-green strain. **D**, Immunoblots of fluorescent reporters in the red-green strain. Data is representative of n=3 independent repeats. Cell numbers are recorded in Table 24, page 166.

#### 8.2.4. *Conditions resulting in mCherry2 heterogeneity*

The use of standard assay conditions with the single gene mCherry2 strain did not produce either of the two phenotypes associated with *P<sub>hyperspank</sub>-sfGFP\**. There was no growth phase dependent increase in fluorescence and there was no stationary phase heterogeneity (Figure 39, page 110 & Figure 41, page 113). When transcript levels were monitored in the 0 mM IPTG condition, *mCherry2* mRNA was visible at the final time point (Figure 45 **B**) without any associated increase heterogeneity (Figure 56 **B**, appendix page 156). All strains containing *mCherry2* are now known to contain inactive protein during standard assay conditions (Figure 48 & Figure 49). It is possible that folding and maturation time of mCherry2 contributes to the discrepancy between mCherry2 protein levels and fluorescent activity. It is also plausible that mCherry2 misfolding causes the discrepancy. Both possibilities raise the question of what, if any, conditions there are which will produce the sfGFP\* phenotypes in the mCherry2 reporter strain? Preliminary data was produced where the single gene sfGFP\* and mCherry2 strains were left to grow for 24 hours under inducing conditions and without IPTG induction. Microscope images reveal a similar phenotype in both strains. Under inducing conditions there is fluorescence in both the sfGFP\* and mCherry2 strains (Figure 50 **A & B**). Without induction there is visible heterogeneity in both strains (Figure 50 **C & D**). Analysis of the data confirms that under inducing conditions the sfGFP\* fluorescence (Figure 51 **A**) and the mCherry2 fluorescence (**B**) are similar. The fluorescent intensity of mCherry2 is ~ 84% of sfGFP\*. Mean sfGFP\* fluorescence is ~ 500 a.u and mCherry2 fluorescence is ~ 420 a.u. By comparison, the final time point in the standard assay conditions produced mCherry2 fluorescence at ~ 6% of sfGFP\* fluorescence (Figure 40).

Without IPTG induction the 24 hour growth assay reduced both the mean fluorescence and noise in the sfGFP\* strain (Figure 51 **C**). Mean fluorescence was 15.4 a.u and noise calculated to be 0.62. At the 300 minute time point in standard assay conditions sfGFP\* fluorescence was 56.5 a.u with noise levels of 1.3 (Figure 43, page 119 and Figure 44, 121). The 24 hour assay produced mCherry2 fluorescence and noise above background levels (Figure 51 **D**). Mean fluorescence was 11.8 a.u with noise levels of 1.5. It is not possible to compare this to the mCherry2 noise values observed in standard assay conditions. mCherry2 fluorescence did not increase above background levels and the associated noise is technical noise rather than gene expression noise (Figure 43, page 119 and Figure 44, 121). Only with mCherry fluorescence above background levels is the comparison between sfGFP\* and

mCherry2 noise appropriate. These data reveal that growth conditions can determine the presence or absence of the two phenotypes in the mCherry2 strain.

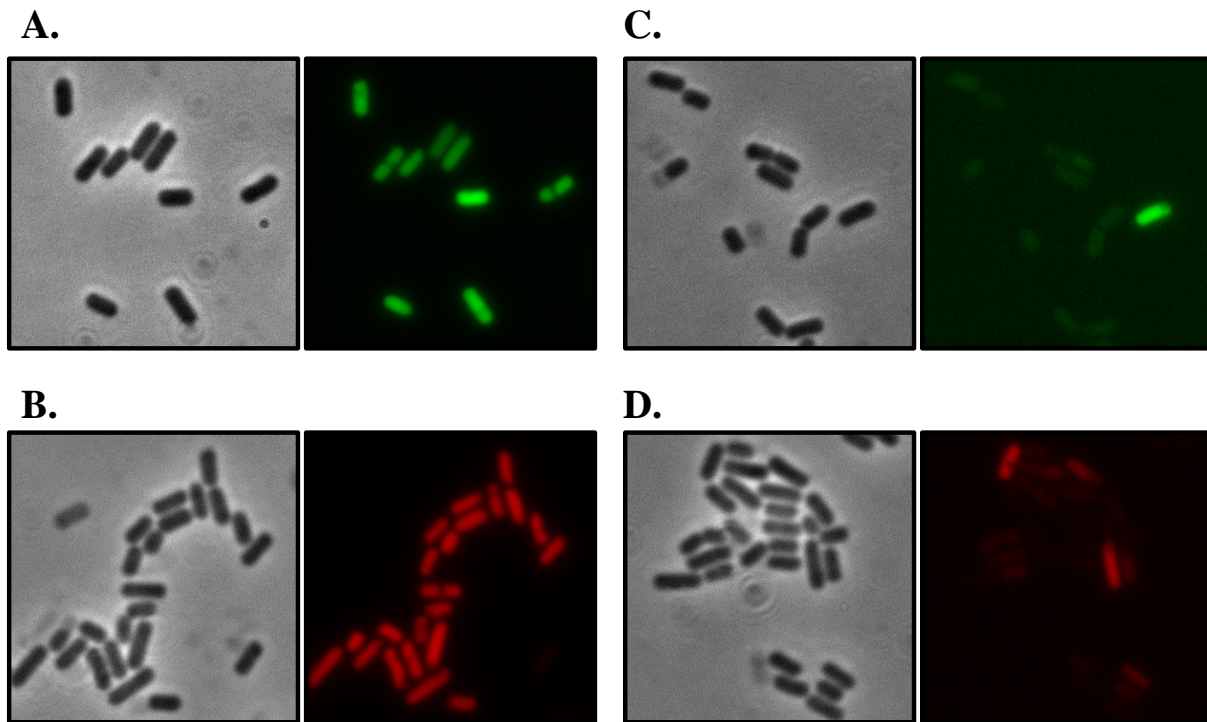


Figure 50. Microscope images of the sfGFP\* and mCherry2 strains after 24 hours growth. Phase contrast and fluorescent images were taken of strains grown under inducing conditions (1 mM IPTG) or without induction (0 mM IPTG). **A**, sfGFP\* strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) with induction. **B**, mCherry2 strain (TE51, *P<sub>hyperspank</sub>-mCherry2 lacI\_cis*) with induction. **C**, sfGFP\* strain without induction. **D**, mCherry2 strain without induction.



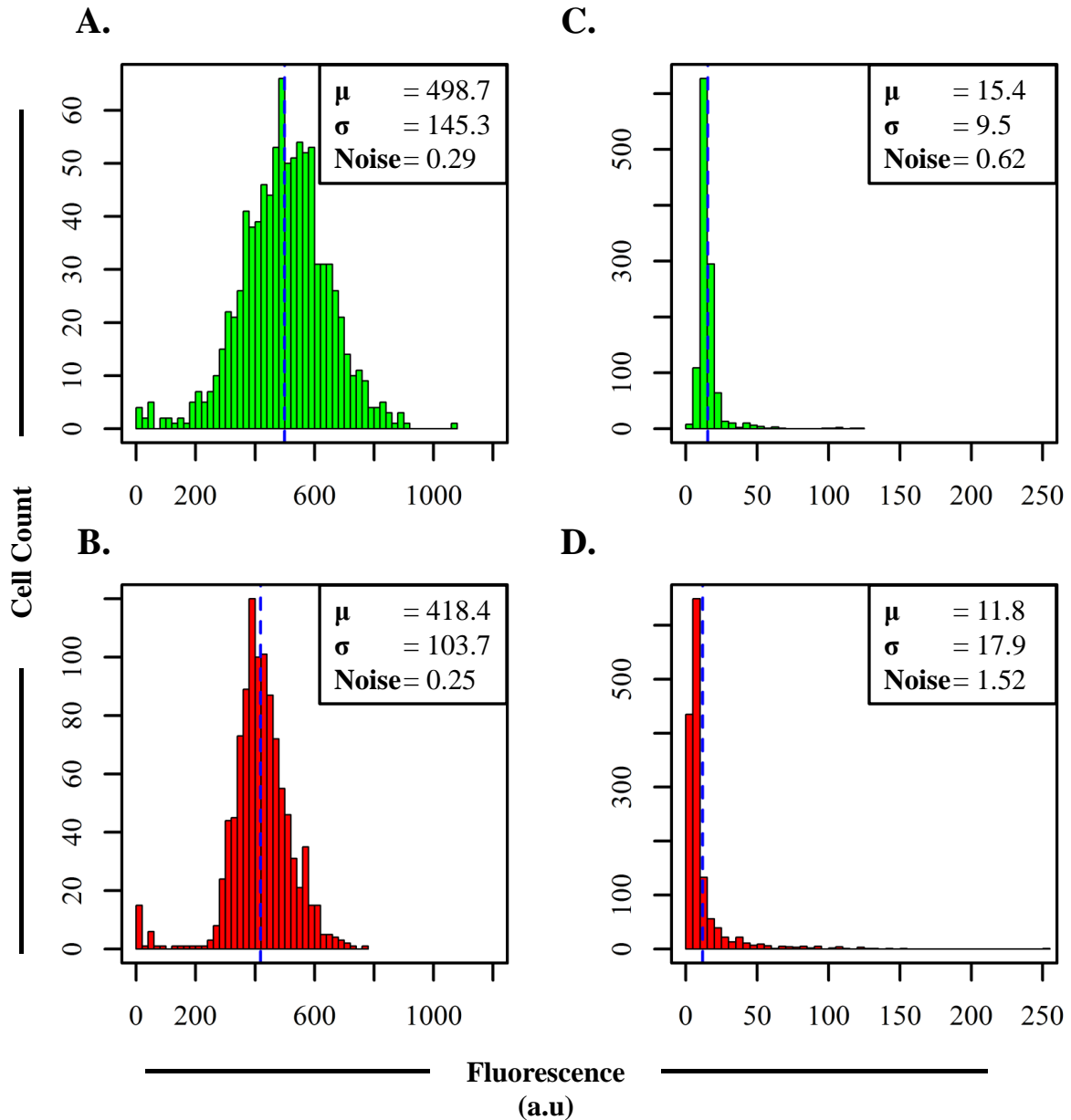


Figure 51. Fluorescence in the sfGFP\* and mCherry2 strains after 24 hours growth. Strains were grown under inducing conditions (1 mM IPTG) or without induction (0 mM IPTG). **A**, sfGFP\* strain (TE47,  $P_{hyperspank}$ -sfGFP\*  $lacI$ -cis) with induction. **B**, mCherry2 strain (TE51,  $P_{hyperspank}$ -mCherry2  $lacI$ -cis) with induction. **C**, sfGFP\* strain without induction. **D**, mCherry2 strain without induction. Figure legend gives the mean fluorescence ( $\mu$ ) and standard deviation ( $\sigma$ ) used to calculate noise values. Blue dashed line, mean value. Data represents n=1 repeat.

### 8.2.5. Modelling sfGFP activity

The sfGFP\* reporter was selected for use in this thesis due to its quick folding/maturation rate and high folding fidelity. The stability of sfGFP\* was the only component of the system

known to fundamentally change the activity phenotype (Section 4.2.4, page 77). mRNA and protein experiments do not suggest a biochemical explanation for the difference in exponential and stationary phase fluorescent activity. Under inducing conditions there was high levels of transcript throughout the assay and the protein levels matched the activity profile (Figure 45 & Figure 55, page 155). sfGFP\* protein is known to be stable in *B. subtilis* for excess of 60 minutes during both exponential growth and in stationary phase (Syvertsson, 2013). These known parameters enable the use of a simple dynamic model to explain the effects of growth rate on gene expression (Figure 52). There is good fit between the model and the experimental data (Figure 34, page 96). At the production rate specified the fluorescent protein is diluted during exponential growth and accumulates when the growth rate is reduced. During rapid growth an equilibrium occurs between the change in cell volume and the rate of protein production. The model assumes a constant production rate of active protein and no degradation. Under these conditions growth rate becomes the limiting factor in the observed levels of fluorescence.

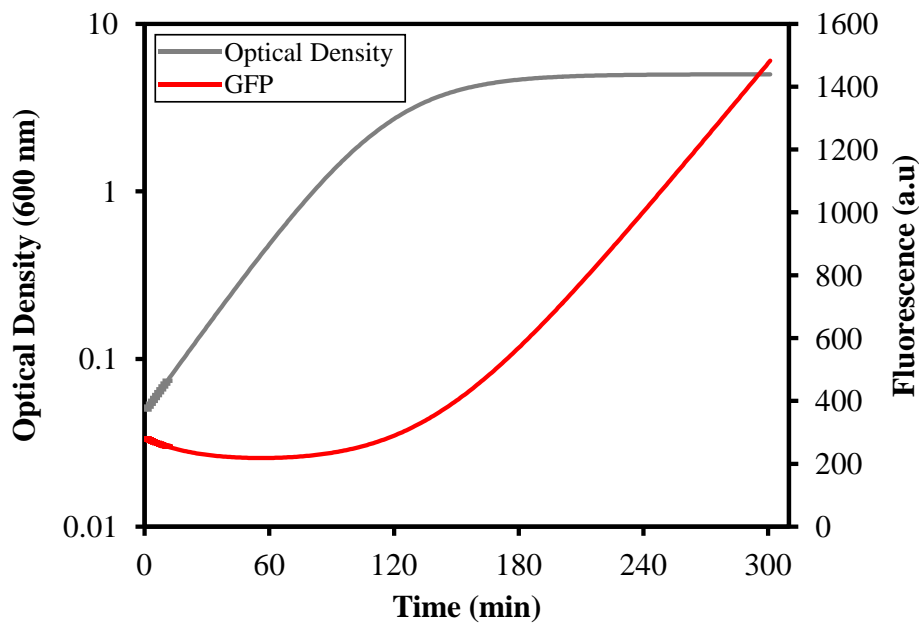


Figure 52. Deterministic dynamic model of growth, gene expression and dilution. The logistic growth equation calculates the optical density (OD) as  $OD = r \cdot OD \cdot (k - OD)$  with the rate constant  $r = 0.008$  and the carrying capacity  $k = 5$ . Total fluorescence ( $F_{total}$ ) at a given time ( $T_n$ ) is calculated as  $F_{total} = a \cdot OD + F_{total}(T_{n-1})$  with the rate constant  $a = 8$ . Fluorescence ( $F$ ) is defined as  $F = F_{total}/OD$ . Initial values at zero minutes are  $OD = 0.05$  and  $F_{total} = 14$ . The parameters are run through the StepODE function in the R based deSolve package (Soetaert *et al.*, 2010). R script produced by D. J. Wilkinson and adapted by T. P. Ewen.

### 8.3. Conclusions

Population levels of mRNA and protein were assayed together with fluorescent activity at single cell resolution. Under inducing conditions transcript levels were high throughout the assays in the single gene sfGFP\* strain and both dual reporter gene constructs while *mCherry2* mRNA levels in the single gene strain reduced over the course of the assay (Figure 45). A difference in *mCherry2* mRNA stability was observed between exponential growth and in stationary phase (Figure 46). This explains the reduction in *mCherry2* transcript levels. mRNA levels in the inducing conditions did not explain the difference in sfGFP\* and mCherry2 activity phenotypes. Without IPTG induction the noise phenotype observed with sfGFP\* reporters was transcription dependent (Figure 45). Under non-inducing conditions there was *mCherry2* transcript in the single gene mCherry2 strain and both dual reporter gene strains (Figure 45). This did not lead to an increase in mCherry2 noise, suggesting a translational mechanism for the difference in phenotype.

Immunoblotting was performed to compare protein levels with the fluorescent activity. In all assay conditions with all strains, sfGFP\* protein levels matched the fluorescent activity to a greater extent than mCherry2. The mCherry2 reporter produced varied protein levels depending on the strain. In the single gene strain mCherry protein levels were highest during exponential growth (Figure 48). The green-red strain produced most protein during stationary phase and the red-green strain had high levels of protein throughout the assay (Figure 49). Varying levels of mCherry2 protein produced relatively constant levels of fluorescent activity. These assays show higher levels of inactive mCherry2 protein, compared to sfGFP\*. These data are consistent with the extended maturation time observed with mCherry2 reporters. mCherry2 maturation times have been reported to take between 15 and 40 minutes (Shaner *et al.*, 2004; Merzlyak *et al.*, 2007). The presence of inactive protein suggest that the duration of standard assay conditions is insufficient to produce the growth phase dependent activity and noise phenotypes observed with sfGFP\*. The extended maturation time of the mCherry2 protein may shift the increase in fluorescence later into stationary phase. Extending the assay duration for 24 hours produced preliminary data in support of this hypothesis. mCherry2 activity and noise is dependent on the assay conditions. The extended maturation time of mCherry2 is likely to determine the absence of phenotype. Two simple alterations to the assay conditions could be introduced to test this hypothesis. Firstly, extending the duration of the assay would confirm if the increase in

fluorescence/noise is time-shifted due to a difference in maturation kinetics in the sfGFP\* and mCherry2 reporters. Secondly, reducing the growth rate in the assay conditions would offset the maturation time of mCherry2. An increase in stationary phase activity and noise is predicted to occur earlier with a reduced growth rate.

In the single gene sfGFP\* strain and the green-red strain protein levels and fluorescence increased through the assay. This suggests translation may be responsible for the sfGFP\* activity phenotype. However, in the red-green strain sfGFP\* protein levels were high throughout and the increase in fluorescence from exponential growth was still present (Figure 49). Data from the red-green strain suggests an alternative mechanism for the growth phase dependent sfGFP\* activity. This was confirmed through modelling the effects of growth rate on fluorescent activity (Figure 52). The model predicted an equilibrium of active protein production in exponential growth, followed by an accumulation of fluorescence during stationary phase. The model fits with the experimental data, explaining the growth phase dependent activity phenotype.

## Chapter 9. Discussion

### 9.1. Summary of findings

Caution must be advocated for synthetic and molecular biologists designing inducible systems. In this thesis the systems employed to investigate heterogeneity in gene expression produced unexpected phenotypes. *B. subtilis* exhibits growth phase dependent activity and noise with  $P_{\text{hyperspank}}$  driving expression of *sfGFP\**. The activity phenotype is due to the accumulation of a stable reporter protein. Placing gene expression within the context of the bacterial growth curve provided new understanding on the effects of dilution within rapidly growing cells. Dilution effects distorted the contribution of transcription and translation in this dynamic system. Transcription of *sfGFP\** was high throughout the growth phases and *sfGFP\** protein levels matched the fluorescent activity. Characterising the system with microscopy, Northern and Western blotting confirmed the *sfGFP\** activity phenotype was not dependent on gene expression. Use of multiple assay types provided sufficient data to model the dilution effects with good correlation to the experimental data.

High *sfGFP\** noise levels were only observed in stationary phase. Single time point assays during exponential growth would miss this phenotype entirely. This new observation was possible with the resolution of single cell microscopy with growing cultures, assayed over multiple growth phases. It provides detail to explain anecdotal reports of the “leaky” hyperspank promoter. This is a growth phase dependent de-repression of the promoter, occurring only within a subpopulation of cells. “Leakiness” not a generalised feature of the hyperspank promoter. It is a temporally dependent heterogeneous response, observed without induction and is due to an increase in stationary phase transcription.

The results of this project also show that reporters are not interchangeable. The mCherry2 protein being a clear example.  $P_{\text{hyperspank}}\text{-mCherry2}$  strains produced neither the growth phase dependent activity phenotype nor the stationary phase noise. Inactive mCherry2 protein suggests growth in the standard assay conditions was unsuitable for the folding and maturation kinetics of the mCherry2 reporter. The disparity between mCherry2 protein levels and activity suggest a delay in the maturation of correctly folded protein. This is supported by the literature and likely to result in mCherry2 levels reporting historic gene expression events, rather than a close correlation to real-time events (Shaner *et al.*, 2004; Merzlyak *et al.*, 2007). Dilution within rapidly growing cells will further distort transcription and translation

events within the system. It would be very difficult to observe dilution effects with the isolated use of *mCherry2* reporter strains.

A stated aim in this thesis was to characterise and evaluate regulatory gene circuits used to quantify heterogeneity in gene expression. Permutations in the composition of these gene circuits offer yet another cautionary tale in design. The single gene reporter constructs were different only in the open reading frames, yet produced a difference in transcript levels and in mRNA stability. Use of bicistronic reporter constructs altered the expected output of the system. Gene order in the dual reporter constructs affected the fluorescent signal from the second gene in the operon. A difference in mRNA secondary structure is candidate for explaining these differences. Folding of the 5' untranslated region, together with the RBS and first 18 bases of the open reading frames has been modelled using the Mfold server (Figure 53). There is a predicted difference in the stability of the secondary structures, suggesting access to the RBS and start codon will be energetically more favourable in the *mCherry2* transcript. Translation of genes is initiated at the 5' end of the mRNA. Contact between the Ribosome and the first gene in the transcript is known to disrupt the secondary structure of the second gene. This can allow greater access to the RBS, resulting in higher expression levels (Kozak, 2005; Studer and Joseph, 2006; Wen *et al.*, 2008; Chen *et al.*, 2013; Osterman *et al.*, 2013). The increased sfGFP\* activity levels in the red-green strain fit with this explanation. mRNA secondary structure is speculated to cause the difference in activity levels. It is inadequate to assume that an open reading frame exchange will produce the same transcription and translation dynamics in *B. subtilis*.

## 9.2. Implications

This thesis provides data highlighting the importance of both system design and characterisation. Dilution effects were observable with the combination of mRNA, protein and activity assays. The importance of growth and dilution cannot be understated when considering quantitative assertions regarding the physiology of bacteria. The concept of promoter activity defined by the fluorescence of reporter is potentially misleading. The rate of transcription may be stable and high and yet have large differences in the fluorescent activity of the reporter, due to the dilution effect. Defining the promoter activity based on the output of a fluorescent reporter should acknowledge and account for this dilution effect. Contrasting the output of *sfGFP\** and *mCherry2* genes illustrates a further consideration. A difference in the maturation of the reporters can present as a difference in the activity of the

promoter. In this example the quantifiable output is correctly folded, mature fluorescent protein. Maturation rate is key to accurate reporting. *gfp* has been fused to promoters involved in sporulation to report on the activity of phosphorelay genes (de Jong *et al.*, 2010b). A balance between maturation of the reporter and phosphorelay gene products is, perhaps unintentionally, implied. In reality the interplay between dilution and protein maturation may result in a growth phase dependent misrepresentation in the sporulation gene products. To overcome this molecular biologists may intuitively design conditions where dilution/maturation effects are minimised. Anecdotal suggestions that “mCherry works best at 30 °C” are one such example. Assay conditions with long generation times are likely to remove the dilution effect, producing comparable results with the hypothesised mechanism.

The stationary phase noise associated with  $P_{\text{hyperspank}}$  is an important observation in this thesis. Single time-point analysis during exponential growth do not capture the growth phase dependent difference in noise dynamics. The noise model Ozbudak *et al.* (2002) used to explain their results identifies rationale for translation producing the greater source of noise, in comparison to transcription. Once the single time-point measurements are incorporated into the bacterial growth curve the model no longer explains the behaviour. Noise in this thesis was found to be a transcription dependent feature of stationary phase cells. This finding has implications for the use of  $P_{\text{hyperspank}}$  in alternative applications. IPTG inducible promoters have been used in the design of synthetic gene networks (Moon *et al.*, 2012). They have been used to activate genetic toggle switches (Gardner *et al.*, 2000). Biosensors have utilised IPTG inducible promoters in regulatory gene networks (Kobayashi *et al.*, 2004). These synthetic regulatory systems need to be switched ON and OFF in response to sensor components. A synthetic regulatory network involving  $P_{\text{hyperspank}}$  may function correctly during exponential growth, then during stationary phase aberrant signalling will compromise the signal fidelity. High levels of stationary phase noise equate to growth phase dependent faults in the architecture of a synthetic gene circuit. Aberrant signalling is likely to be a common feature of IPTG inducible promoters during stationary phase. Gene expression noise is now a characterised feature of both the native *lac* promoter (Choi *et al.*, 2008) and the synthetic  $P_{\text{hyperspank}}$ .

Findings in this thesis raise questions for those using inducible systems with fluorescent reporters. The design of constructs will depend entirely on their intended purpose. Functionality of the system is paramount and will inform where tight control of the system is required and can be achieved. The promoters used in this thesis have offered a trade-off

between transcription levels and repression. Data from the *sfGFP\** and *mCherry2* open reading frame exchange would advocate consideration of mRNA secondary structure on both expression levels and stability. It should also be recognised that mRNA secondary structure in bicistronic operons may play a significant role in expression levels. Permutations in the constructs may be necessary to elucidate the specific mRNA effects potentially confounding interpretation of the data.

The phenotypic variation observed between fluorescent reporters supports the notion that translational control is critical when designing a regulatory gene circuit. The fundamental question is one of validity. Will the reporter accurately represent the physiology it is designed to work with? In *B. subtilis* grown in LB at 37 °C the *sfGFP\** reporter is superior to the *mCherry2* reporter as a quantitative tool to monitor gene expression. It is important to recognise the limitations of each reporter, supported by data in this thesis. Folding and maturation time should be assessed within the context of growth rate and dilution.

### **9.3. Future work**

Regulatory gene circuits have been characterised in their ability to monitor heterogeneous gene expression. There are still unanswered questions, interesting to those using inducible systems with fluorescent reporters. The rate of mature protein formation should be calculated during exponential growth and in stationary phase. This rate should be independent of the dilution effect. qPCR would provide numerical data relating mRNA levels during the different growth phases. Transcription rate could be established using the relative mRNA levels during standard assay conditions, together with mRNA from the degradation assays. The stability of both the *sfGFP\** and *mCherry2* protein should be defined during exponential growth and in stationary phase. Once established these data could be collated in a model to predict the rate of mature protein formation. Knowledge of reporter maturation time is a valuable reference for quantitative biology. The behaviour of the system could then be investigated at different temperatures. These data would provide a reference containing the detail necessary for assay design with quantitative validity.

### **9.4. Concluding remarks**

Advice must be given to fully investigate the system designed for a scientific purpose. This applies to molecular and synthetic biologists alike. The results presented in this work detail the level of attention required to produce a reporter system capable of functioning as



intended. Permutations in the system should be routinely applied to avoid unintended consequences. Constructs must be tested to confirm the validity of their function. The pursuit of knowledge demands careful application of the molecular tools available to the scientist.

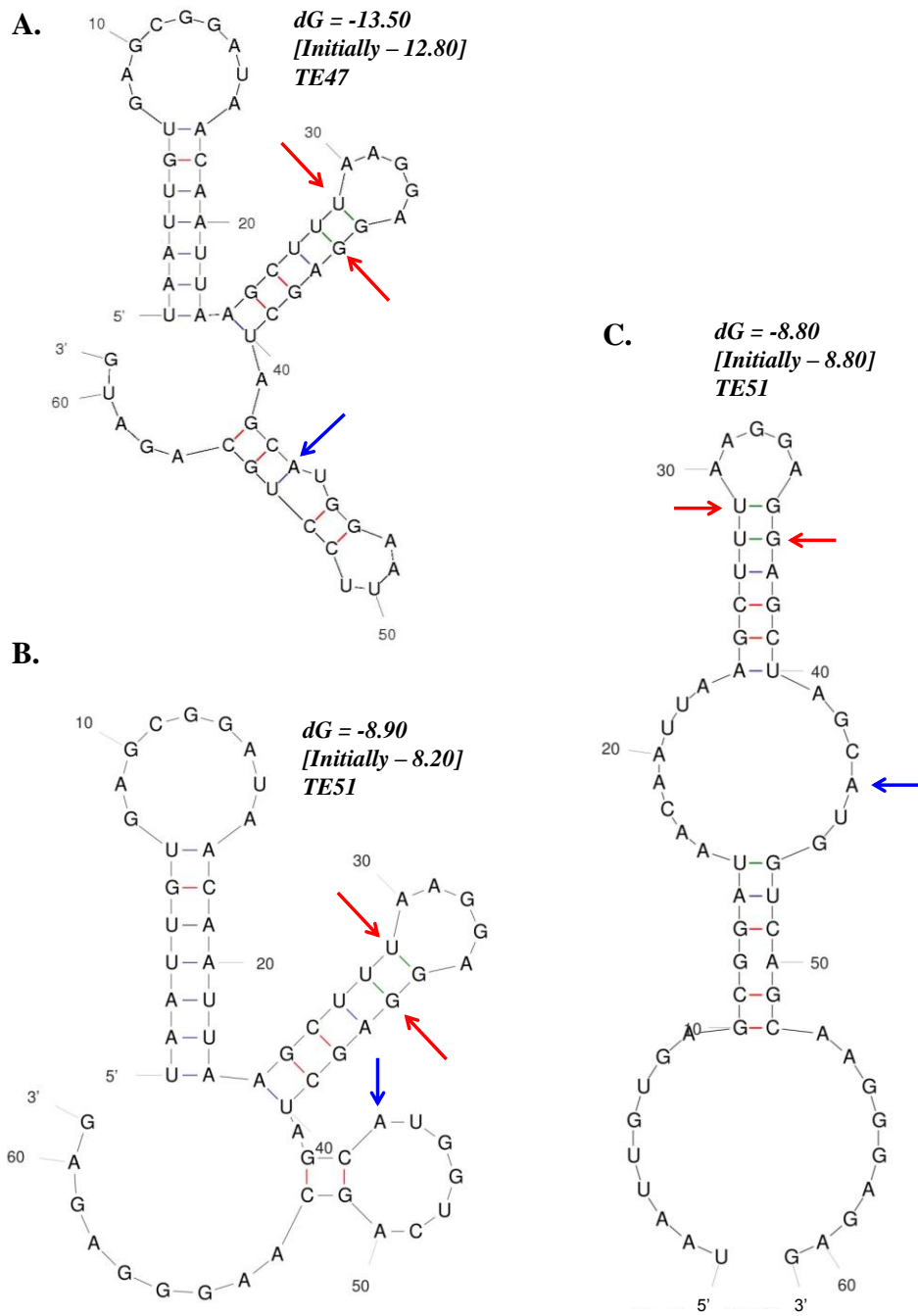


Figure 53. Predicted secondary structures in *sfGFP\** and *mCherry* transcripts. The first 70 bases transcribed have been entered into the Mfold web server (Markham, 1995-2016). **A**, the single predicted mRNA structure of strain TE47 (*P<sub>hyperspank</sub>-sfGFP\* lacI<sub>cis</sub>*). **B**, one of two predicted mRNA structures for strain TE51 (*P<sub>hyperspank</sub>-mCherry lacI<sub>cis</sub>*). **C**, the second predicted mRNA structure for strain TE51. Red arrows indicate the first and final base of the RBS. Blue arrows point to the first base in the start codon.

## Chapter 10. Appendix

### 10.1. Growth Media, Buffers and Solutions

Unless otherwise stated all buffers, solutions and growth media were prepared in deionised H<sub>2</sub>O (dH<sub>2</sub>O). Growth media was sterilised by autoclaving at 120 °C for 15 minutes, after which antibiotics were added where necessary.

#### 10.1.1. *Lysogeny broth (LB)*

The standard growth media used in this work was Lysogeny broth (LB). Composition:

Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	5 g/L

#### 10.1.2. *LB agar*

LB was prepared with 15 g/L agar and poured onto petri dishes for use as a solid growth media.

#### 10.1.3. *Spizizen minimal media (SMM)*

Spizizen minimal media (SMM) was prepared for use in *B. subtilis* transformations.

Composition:

Ammonium sulphate	0.2% w/v
Dipotassium phosphate	1.4% w/v
Potassium dihydrogen phosphate	0.6% w/v
Sodium citrate dihydrate	0.1% w/v
Magnesium sulphate	0.02% w/v

#### 10.1.4. *Minimal competence media (MM)*

Minimal competence media (MM) was prepared for *B. subtilis* transformations.

Composition:

SMM	10 ml
Glucose solution (40% w/v)	0.125 ml
Tryptophan solution (2 mg/ml)	0.1 ml

Magnesium sulphate (1 M)	0.06 ml
Casamino acids (20% w/v)	0.01 ml
Ferric ammonium citrate (0.22% w/v)	0.005 ml

#### **10.1.5. Starvation media**

Composition:

SMM	10 ml
Glucose solution (40% w/v)	0.125 ml
Ferric ammonium citrate (0.22% w/v)	0.06 ml

#### **10.1.6. RNA loading dye**

RNA loading dye was prepared for use in agarose gel electrophoresis. Composition:

Formamide	6 ml
Formaldehyde	1.2 ml
10 X MOPS (page 147)	2 ml
Glycerol	50%
Bromophenol blue	20 mg
Xylencyanol	20 mg

#### **10.1.7. 10 X MOPS running buffer**

MOPS running buffer was prepared for use in Northern blotting. The buffer was adjusted to pH 7.0 using 10 M NaOH and filter sterilised. Composition:

MOPS	200 mM
NaAc	50 mM
EDTA	10 mM

#### **10.1.8. 20 X Saline-sodium citrate buffer (SSC)**

Saline-sodium citrate buffer was prepared for use in Northern blotting. The buffer was adjusted to pH 7.0 using 1 M HCl and autoclaved. Composition:

Sodium chloride	3 M
Sodium Citrate	0.3 M

#### **10.1.9. Denaturation solution**

Denaturation solution was prepared for use in Northern blotting. The buffer was sterilised in the autoclave. Composition:

Sodium hydroxide	50 mM
Sodium chloride	10 mM

#### **10.1.10. Neutralisation solution**

Neutralisation solution was prepared for use in Northern blotting. The buffer was adjusted to pH 7.4 using HCl and autoclaved. Composition:

Trisma Base	10 mM
-------------	-------

#### **10.1.11. Malic acid**

Malic acid was prepared for use in Northern blotting. The buffer was adjusted to pH 7.5 and filter sterilised. Composition:

Maleic acid	1 M
Sodium Chloride	1.5 M
Sodium Hydroxide	1.8 M

#### **10.1.12. Pre-hybridisation solution**

Pre-hybridisation solution was prepared for use during the hybridisation steps in Northern blotting. The solution was used on the day of preparation. Composition:

dH <sub>2</sub> O	1520 µl
Sodium dodecyl sulfate (50%)	80 µl
N- Lauroylsarcosine (10%)	400 µl
Blocking reagent (page 149)	8 ml
20 X SSC (page 147)	10 ml
Formamide	20 ml

### **10.1.13. Hybridisation solution**

Composition:

Pre-hybridisation solution 10 ml

*Anti-Digoxigenin*-AP, Fab fragments (Roche) 2 µl

### **10.1.14. Blocking reagent**

Blocking reagent was prepared for use in Northern blotting. The reagent was autoclaved and held at 4 °C for storage < 2 years. Composition:

Blocking reagent (Roche: 11 096 176 001) 10% w/v

1 X Malic acid (page 148)

### **10.1.15. Blocking buffer**

Blocking buffer was prepared on the day of use in Northern blotting. The composition of blocking buffer for a single membrane was:

Blocking reagent 5 ml

10 X Malic acid (page 148) 5 ml

Sterile dH<sub>2</sub>O 40 ml

### **10.1.16. Wash solution 1**

Wash solution 1 was prepared for use in Northern blotting. Composition:

2 X SSC (page 147)

Sodium dodecyl sulfate (SDS) 0.01% w/v

Sterile dH<sub>2</sub>O

### **10.1.17. Wash solution 2**

Wash solution 2 was prepared for use in Northern blotting. Composition:

1 X SSC (page 147)

Sodium dodecyl sulfate (SDS) 0.01% w/v

Sterile dH<sub>2</sub>O

### **10.1.18. Tween wash buffer**

Tween wash buffer was prepared for use in Northern blotting. The pH was adjusted to 7.5 using NaOH and autoclaved. Composition:

Tween 20 (Sigma-Aldrich)	0.3% v/v
1 X Malic acid (page 148)	
Sodium Chloride	0.15 M
dH <sub>2</sub> O	

### **10.1.19. Alkaline phosphatase buffer**

Alkaline phosphatase buffer was prepared for use in Northern blotting. Tris, sodium chloride and water were autoclaved before the addition of Diethanolanin. The buffer was then adjusted to pH 9.5 using concentrated HCl. Composition:

Tris	100 mM
Sodium Chloride	100 mM
dH <sub>2</sub> O	990.37 ml
Diethanolanin	9.63 ml

### **10.1.20. Phosphate buffered solution (PBS)**

PBS was prepared for Western blotting. The pH was adjusted to 7.4 and autoclaved. Composition:

Sodium chloride	137 mM
Potassium chloride	2.7 mM
Sodium phosphate dibasic	10 mM
Potassium phosphate monobasic	1.8 mM

### **10.1.21. Phosphate Milk Tween (PMT)**

PMT was prepared for Western blotting on the day of use. The pH was adjusted to 7.4. Composition:

PBS (page 150)	
Milk	5%
Tween	0.1%

**10.1.22. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) buffer**

SDS-PAGE buffer was prepared for use in Western blotting. Composition:

Tris.HCl (pH 7.0)	62.5 mM
Glycerol	10%
Sodium dodecyl sulfate (SDS)	2%
$\beta$ -Mercaptoethanol	5%
Bromophenol Blue	A few grains

**10.1.23. Tricine gel buffer**

Tricine gel buffer was prepared for use during protein gel electrophoresis in Western blotting. Composition:

Tris (pH 8.45)	3 M
Sodium dodecyl sulfate (SDS)	0.3%
dH <sub>2</sub> O (sterile)	

**10.1.24. 10 X Anode running buffer**

Anode running buffer was prepared for protein gel electrophoresis in Western blotting. Composition:

Tris (pH 8.9)	3 M
dH <sub>2</sub> O (sterile)	

**10.1.25. 10 X Cathode Running Buffer**

Cathode running buffer was prepared for protein gel electrophoresis in Western blotting. Composition:

Tris (pH 8.25)	3 M
Tricine	1 M
Sodium dodecyl sulfate (SDS)	1%
dH <sub>2</sub> O (sterile)	



**10.1.26. Transfer buffer**

Transfer buffer was prepared for Western blotting. Composition:

Methanol	10%
CAPS (pH 10.5)	0.5 M
dH <sub>2</sub> O (sterile)	

**10.1.27. DNA agarose gel (1%)**

Composition:

Agarose	1%
Ethidium Bromide	0.4 µg/ml
TAE running buffer (Sigma)	

**10.1.28. RNA agarose gel (1.2%)**

Composition:

Agarose	1.2 g	
10 X MOPS (page 147)		10 ml
Formaldehyde (12.3 M)	18 ml	
dH <sub>2</sub> O	72 ml	

**10.1.29. Stacking acrylamide gel (3.96%)**

Composition:

Stacking acrylamide (49.5% - 4% T 3% C)	0.5 ml
Tricine gel buffer	1.55 ml
H <sub>2</sub> O	4.2 ml
Ammonium persulfate (APS) (10%)	50 µl
Tetramethylethylenediamine (TEMED)	5 µl

**10.1.30. Separating acrylamide gel (12%)**

Composition:

Separating Acrylamide (49.5% - 16.5% T 3%C)	3.6 ml
Tricine Gel Buffer	5 ml
50% Glycerol	5 ml
H <sub>2</sub> O	1.4 ml
APS (10%)	7.5 µl
TEMED	7.5 µl

## 10.2. Supplementary figures

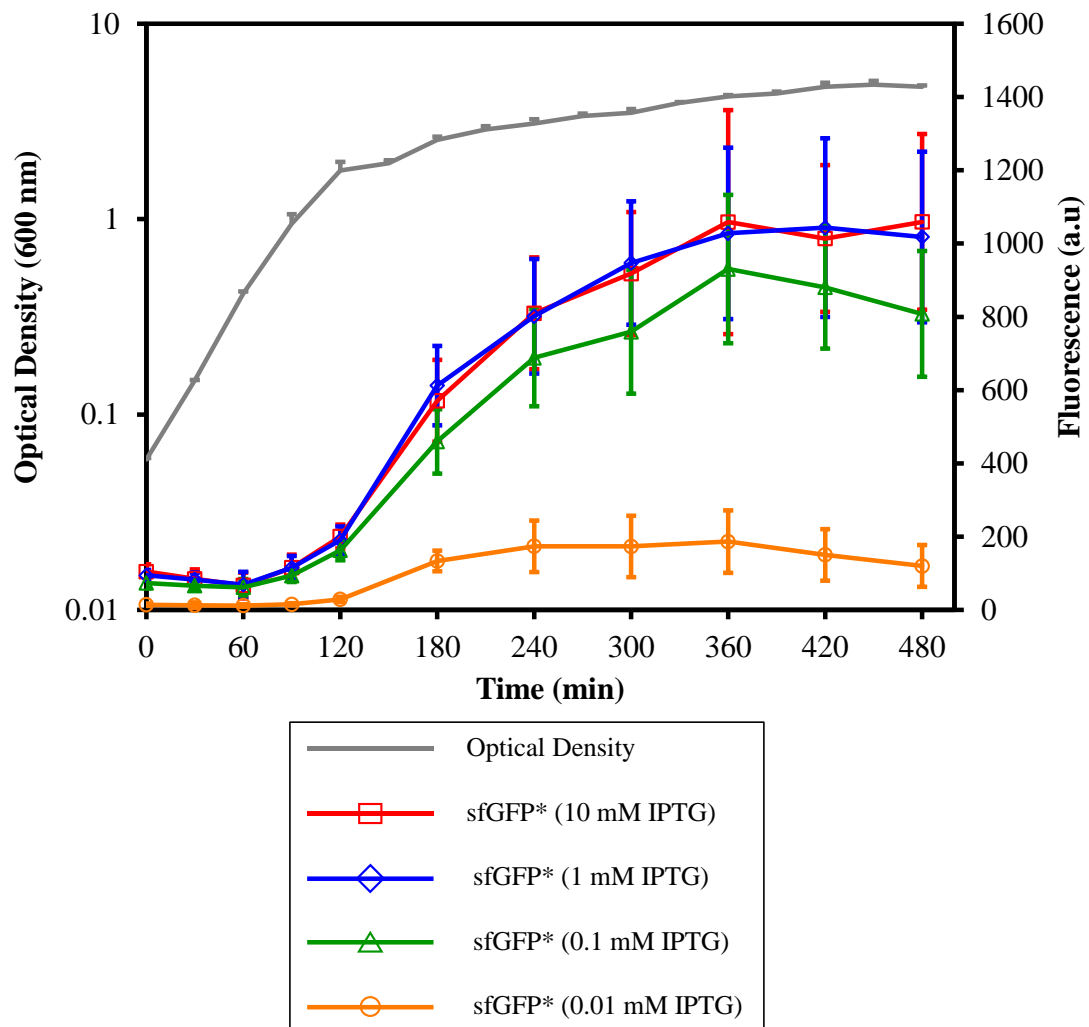


Figure 54. Preliminary data establishing IPTG induction parameters for the expression of *sfGFP\**. The *sfGFP\** reporter strain AH7 ( $P_{hyperspank}$ -*sfGFP\**-*lacI*) was induced with varied concentrations of IPTG to establish full induction conditions.

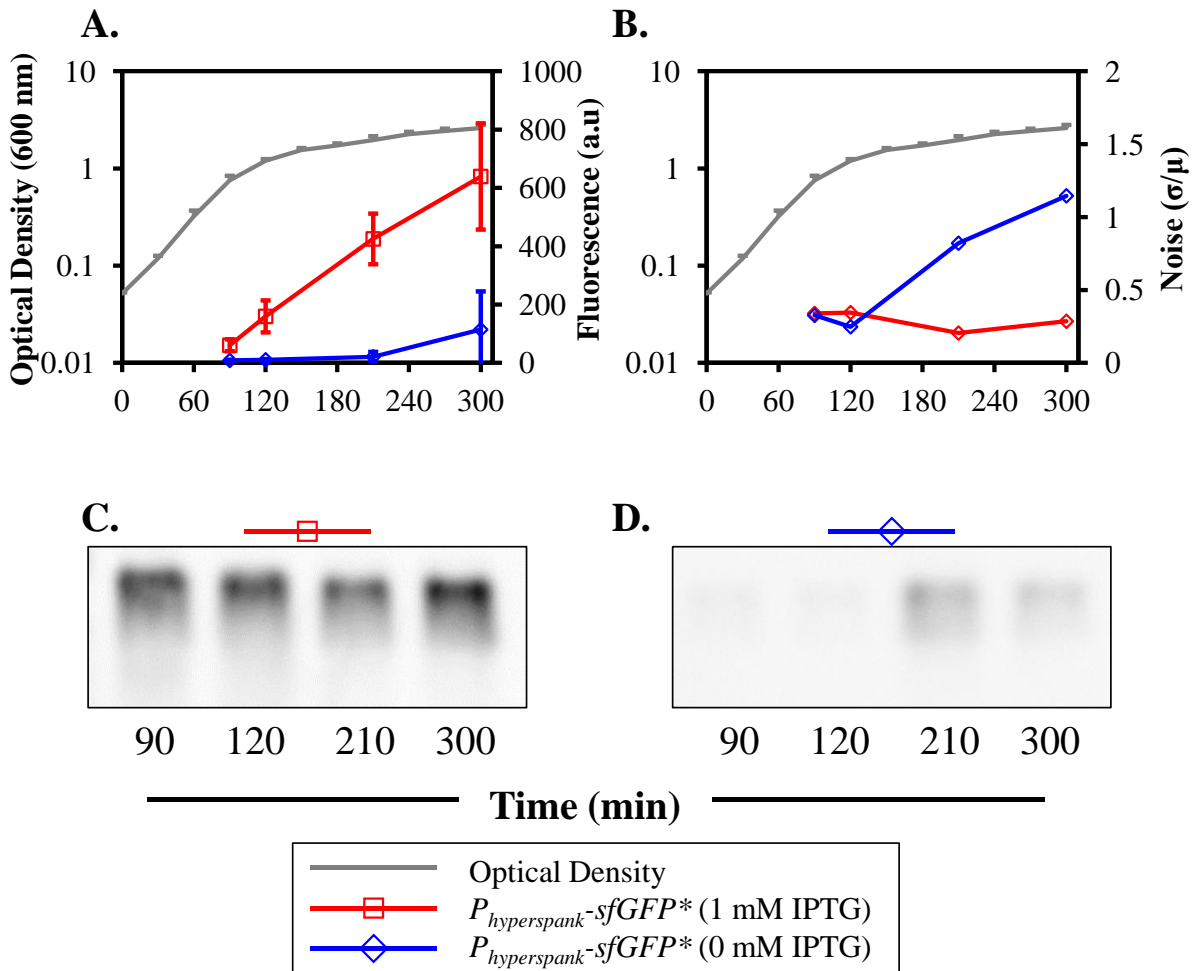


Figure 55. Transcript variation in the single gene  $sfGFP^*$  strain (TE47,  $P_{hyperspank}$ - $sfGFP^*$   $lacI_{cis}$ ). **A**, growth and fluorescence induced with 1 mM IPTG and 0 mM IPTG. **B**, noise levels calculated from data in A. **C**, Northern blot using an  $sfGFP$  probe under inducing conditions. **D**, Northern blot using an  $sfGFP$  probe without IPTG induction. Grey lines, optical density. Red lines, fluorescence/noise under inducing conditions. Blue lines fluorescence/noise without IPTG induction. Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 25, page 166. Northern blots are representative of  $n=3$  repeats.

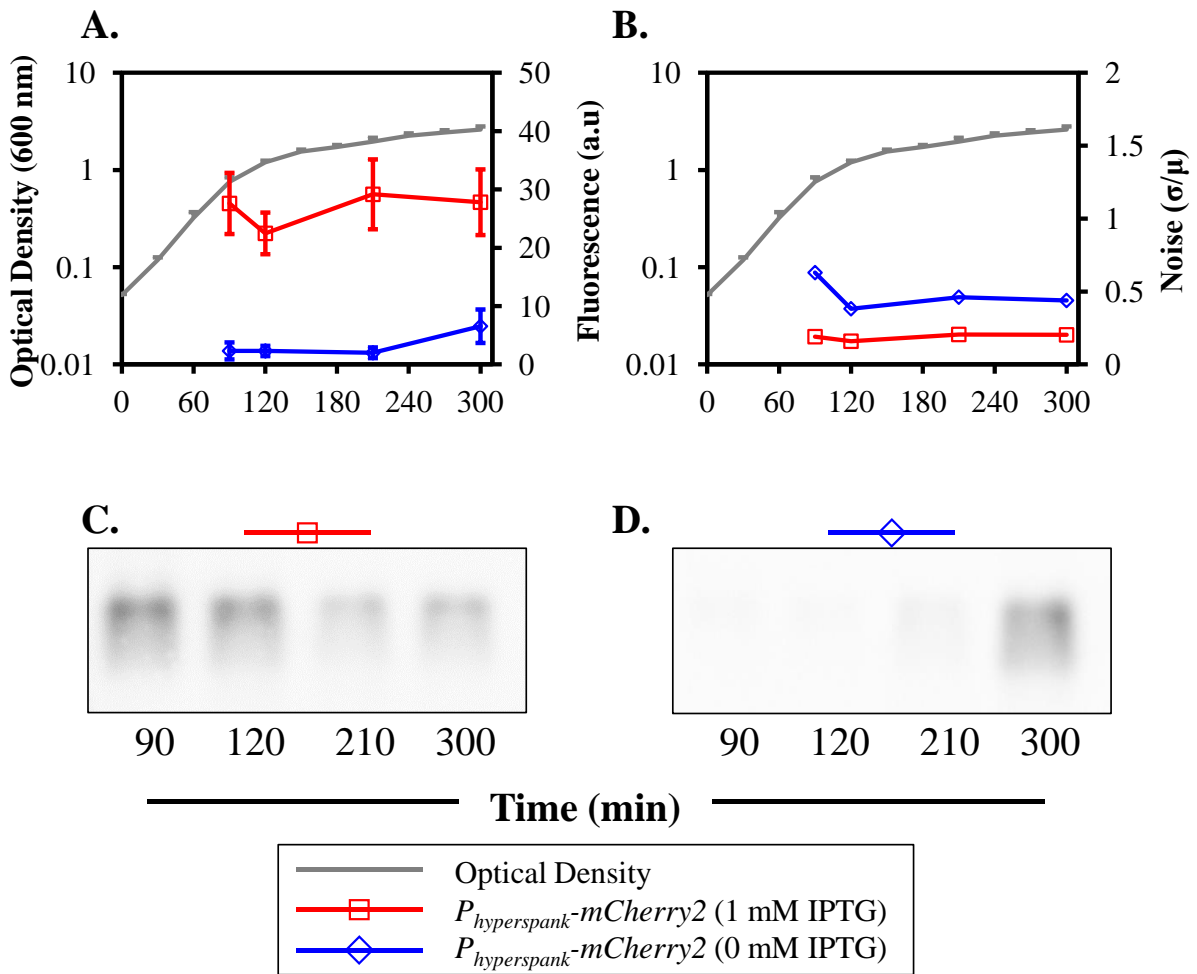


Figure 56. Transcript variation in the single gene *mCherry2* strain (TE51,  $P_{hyperspank}$ -*mCherry2 lacI\_cis*). **A**, growth and fluorescence induced with 1 mM IPTG and 0 mM IPTG. **B**, noise levels calculated from data in A. **C**, Northern blot using an *sfGFP* probe under inducing conditions. **D**, Northern blot using an *mCherry2* probe without IPTG induction. Grey lines, optical density. Red lines, fluorescence/noise under inducing conditions. Blue lines fluorescence/noise without IPTG induction. Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 26, page 166. Northern blots are representative of  $n=3$  repeats.

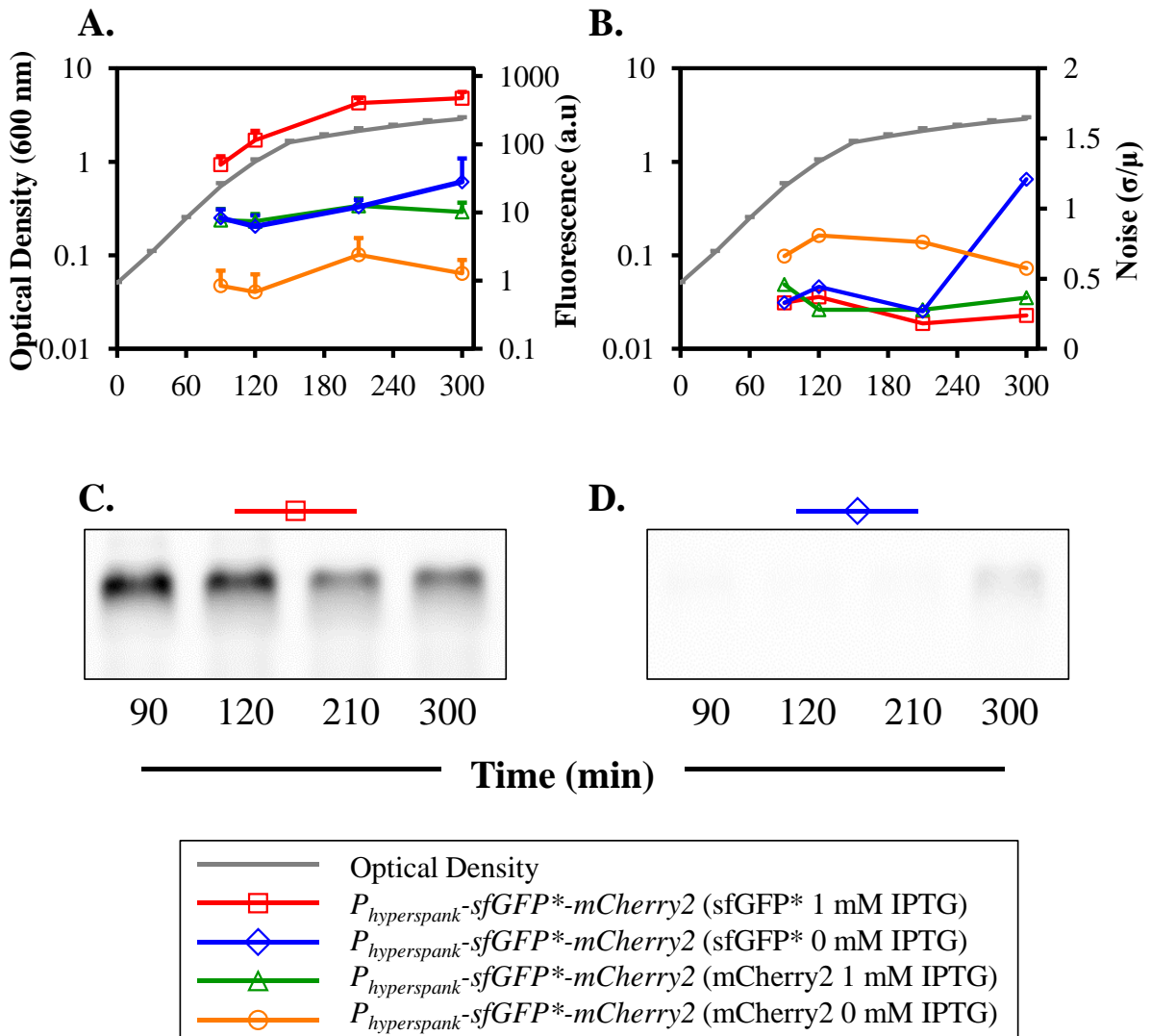


Figure 57. Transcript variation in the bicistronic  $sfGFP^*$ - $mCherry2$  strain. **A**, growth and fluorescence in the green-red strain (TE49,  $P_{hyperspank}$ - $sfGFP^*$ - $mCherry2$   $lacI_{cis}$ ) induced with 1 mM IPTG and 0 mM IPTG. **B**, noise levels calculated from data in A. **C**, Northern blot using an  $sfGFP$  probe under inducing conditions. **D**, Northern blot using an  $sfGFP$  probe without IPTG induction. Grey lines, optical density. Coloured lines, fluorescence and noise. Red lines,  $sfGFP^*$  under inducing conditions. Blue lines,  $sfGFP^*$  without induction. Green lines,  $mCherry2$  under inducing conditions. Orange lines,  $mCherry2$  without induction. Cell numbers corresponding to  $n=3$  independent repeats are recorded in Table 27, page 166. Northern blots are representative of  $n=2$  repeats.

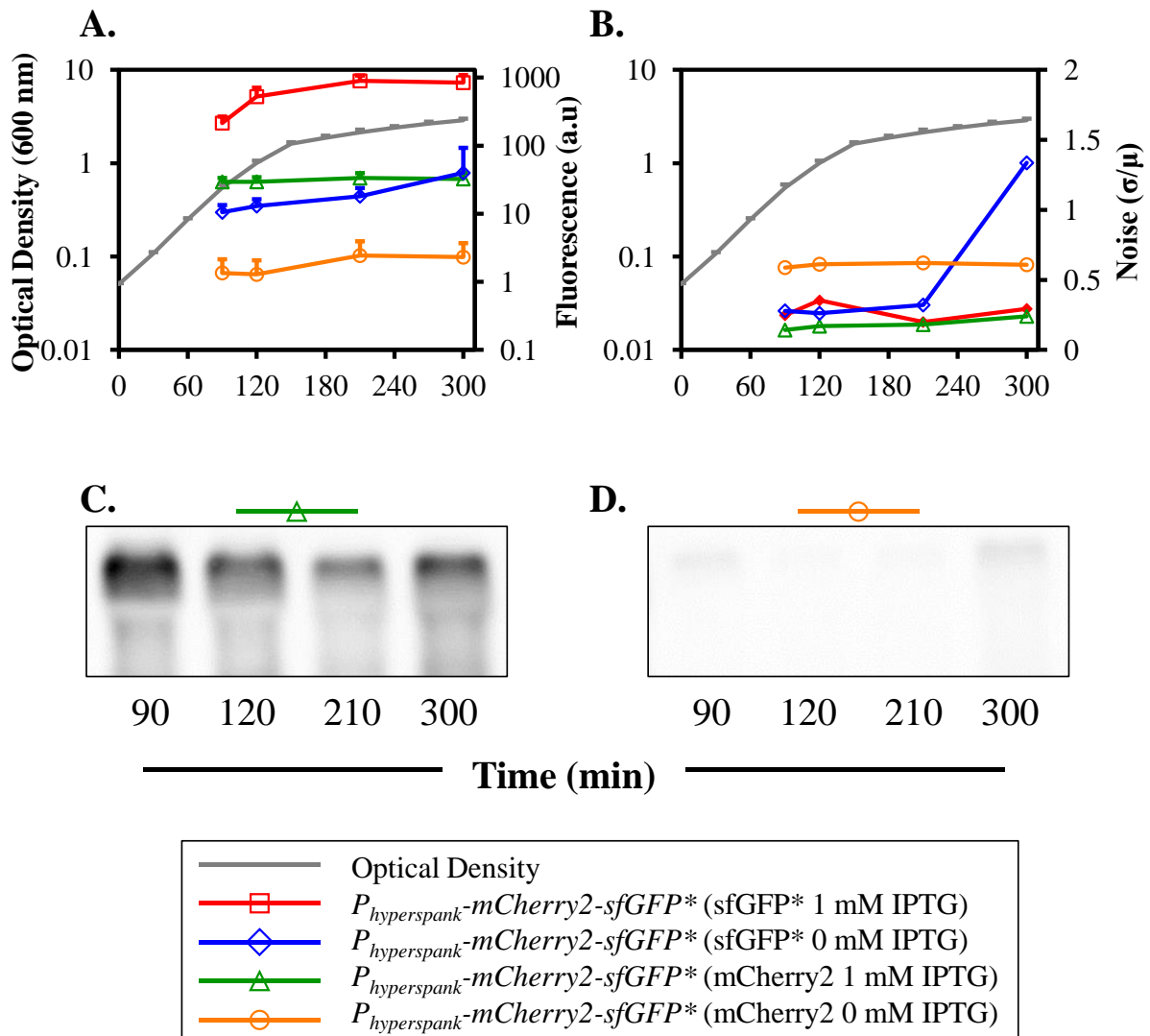


Figure 58. Transcript variation in the bicistronic *mCherry2-sfGFP\** strain. **A**, growth and fluorescence in the red-green strain (TE50,  $P_{hyperspank}$ -*mCherry2-sfGFP\** *lacI\_cis*) induced with 1 mM IPTG and 0 mM IPTG. **B**, noise levels calculated from data in A. **C**, Northern blot using an *sfGFP* probe under inducing conditions. **D**, Northern blot using an *sfGFP* probe without IPTG induction. Grey lines, optical density. Coloured lines, fluorescence and noise. Red lines, *sfGFP\** under inducing conditions. Blue lines, *sfGFP\** without induction. Green lines, *mCherry2* under inducing conditions. Orange lines, *mCherry2* without induction. Cell numbers corresponding to n=3 independent repeats are recorded in Table 28, page 167. Northern blots are representative of n=2 repeats.

### 10.3. Supplementary tables

<pre>% This file contains MicrobeTracker settings optimized for wildtype E. coli cells at 0.064 um/pixel resolution (using algorithm 4)  invertimage = 0  algorithm = 4</pre>		
<pre>% Pixel-based parameters areaMin = 120 areaMax = 2200 thresFactorM = 1 thresFactorF = 1 splitregions = 1 edgemode = logvalley edgeSigmaL = 3 edveSigmaV = 1 valleythresh1 = 0 valleythresh2 = 1 erodeNum = 1 opennum = 0 threshminlevel = 0.02</pre>	<pre>% Constraint parameters fmeshstep = 1 cellwidth = 13 fsmooth = 18 imageforce = 4 wspringconst = 0.3 rigidityRange = 2.5 rigidity = 1 rigidityRangeB = 8 rigidityB = 5 attrCoeff = 0.1 repCoeff = 0.3 attrRegion = 4 horalign = 0.2 eqaldist = 2.5</pre>	<pre>% Image force parameters fitqualitymax = 0.5 forceWeights = 0.25 0.5 0.25 dmapThres = 2 dmapPower = 2 gradSmoothArea = 0.5 repArea = 0.9 attrPower = 4 neighRep = 0.15</pre>
<pre>% Mesh creation parameters roiBorder = 20.5 noCellBorder = 5 maxmesh = 1000 maxCellNumber = 2000 maxRegNumber = 10000 meshStep = 1 meshTolerance = 0.01</pre>	<pre>% Fitting parameters fitConvLevel = 0.0001 fitMaxIter = 500 moveall = 0.1 fitStep = 0.2 fitStepM = 0.6</pre>	<pre>% Joining and splitting splitThreshold = 0.35 joindist = 5 joinangle = 0.8 joinWhenReuse = 0 split1 = 0</pre>
<pre>% Other bgrErodeNum = 5 sgnResize = 1 aligndepth = 1</pre>		

Table 5. Parameters used in MicrobeTracker image analysis.



IPTG concentration (mM)	Fluorescence relative to full induction (1 mM IPTG) at each assay time point (min)								Mean fold difference
	0	30	60	90	120	180	240	300	
<b>0.1</b>	0.63	0.84	1.07	0.56	0.82	0.92	0.82	0.75	0.80
<b>0.05</b>	0.49	0.64	0.70	0.56	0.61	0.74	0.70	0.60	0.63
<b>0.01</b>	0.19	0.36	0.34	0.18	0.13	0.12	0.11	0.11	0.19

Table 6. Relative sfGFP\* fluorescence in the auto-regulating lacI strain, induced with varied IPTG concentrations. Calculations for lacI (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*) are made against fluorescence with 1 mM IPTG induction.

IPTG concentration (mM)	Fluorescence relative to full induction (1 mM IPTG) at each assay time point (min)								Mean fold difference
	0	30	60	90	120	180	240	300	
<b>0.1</b>	1.13	1.01	0.91	0.91	1.00	0.93	0.97	0.94	0.98
<b>0.05</b>	0.87	0.78	0.95	0.73	0.74	0.81	0.83	0.80	0.81
<b>0.01</b>	0.31	0.31	0.35	0.28	0.24	0.29	0.36	0.30	0.31

Table 7. Relative sfGFP\* fluorescence in the negatively regulated lacI-cis strain, induced with varied IPTG concentrations. Calculations for lacI-cis (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI-cis*) are made against fluorescence with 1 mM IPTG induction.

IPTG concentration (mM)	Fluorescence relative to full induction (1 mM IPTG) at each assay time point (min)								Mean fold difference
	0	30	60	90	120	180	240	300	
<b>0.1</b>	0.95	1.04	1.03	0.89	1.06	0.98	1.01	1.02	1.00
<b>0.05</b>	0.79	0.98	0.87	0.82	1.01	0.92	0.97	0.93	0.92
<b>0.01</b>	0.31	0.31	0.34	0.35	0.43	0.42	0.46	0.43	0.38

Table 8. Relative sfGFP\* fluorescence in the negatively regulated lacI-trans strain, induced with varied IPTG concentrations. Calculations for lacI-trans (CJ3, *P<sub>hyperspank</sub>-sfGFP\* lacI-trans*) are made against fluorescence with 1 mM IPTG induction.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (1 mM)</b>	454	1170	1727	2641	1690	1891	1500	925
<b>AH7 (0 mM)</b>	540	845	1077	1381	1443	1536	2485	2078

Table 9. Cell numbers corresponding to the data presented in Figure 22, page 70.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE48 (1 mM)</b>	939	682	1198	2144	1032	1402	1517	1326
<b>TE48 (0 mM)</b>	542	578	1101	1746	1381	1359	1967	1546

Table 10. Cell numbers corresponding to the data presented in Figure 23, page 72.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (0 mM)</b>	540	845	1077	1381	1443	1536	2485	2078
<b>TE48</b>	2023	1838	3400	5636	3794	4118	5452	4418

Table 11. Cell numbers corresponding to the data presented in Figure 24, page 74.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (1 mM)</b>	698	879	1925	1897	1613	1419	1553	1163
<b>AH7 (0.1 mM)</b>	504	1060	1128	2701	1892	2870	1678	1463
<b>AH7 (0.05 mM)</b>	856	853	1543	1445	1939	1528	2054	1186
<b>AH7 (0.01 mM)</b>	556	755	1948	3034	1818	2469	2684	1267

Table 12. Cell numbers corresponding to the data presented in Figure 25, page 76.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (1 mM)</b>	454	1170	1727	2641	1690	1891	1500	925
<b>AH7 (0 mM)</b>	540	845	1077	1381	1443	1536	2485	2078
<b>AH6 (1 mM)</b>	518	1246	1253	2576	1178	2505	1902	2605
<b>AH6 (0 mM)</b>	595	1088	1801	2338	1895	1884	2762	2671
<b>AH5 (1 mM)</b>	507	1043	1356	1465	1299	1243	2638	1721
<b>AH5 (0 mM)</b>	512	557	1115	1664	1177	1513	1591	1890

Table 13. Cell numbers corresponding to the data presented in Figure 26, page 79.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (Before QC)</b>	177	671	712	1050	870	414	1323	1068
<b>AH7 (After QC)</b>	177	671	712	1050	870	414	166	221

Table 14. Cell numbers corresponding to the data presented in Figure 18, page 60.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (1 mM)</b>	498	1317	1409	1385	1828	1057	1223	1348
<b>AH7 (0 mM)</b>	637	1526	1156	1703	1357	1333	1170	1385
<b>TE47 (1 mM)</b>	721	1045	1580	2187	901	1455	1254	1578
<b>TE47 (0 mM)</b>	598	1071	1334	1261	1165	1257	1309	1491
<b>CJ3 (1 mM)</b>	479	681	1212	1497	1445	1517	1368	1471
<b>CJ3 (0 mM)</b>	418	832	1280	1361	1179	1393	1300	1889

Table 15. Cell numbers corresponding to the data presented in Figure 30, page 89 and Figure 31, page 91.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>AH7 (1 mM)</b>	698	879	1925	1897	1613	1419	1553	1163
<b>AH7 (0.1 mM)</b>	504	1060	1128	2701	1892	2870	1678	1463
<b>AH7 (0.05 mM)</b>	856	853	1543	1445	1939	1528	2054	1186
<b>AH7 (0.01 mM)</b>	556	755	1948	3034	1818	2469	2684	1267
<b>TE47 (1 mM)</b>	235	255	521	835	1264	953	1254	2090
<b>TE47 (0.1 mM)</b>	399	297	619	778	1012	1076	1001	1465
<b>TE47 (0.05 mM)</b>	429	268	495	758	1154	817	913	1390
<b>TE47 (0.01 mM)</b>	276	325	533	803	1187	721	987	1693
<b>CJ3 (1 mM)</b>	293	448	1096	1519	1001	1283	1057	990
<b>CJ3 (0.1 mM)</b>	339	427	878	1882	1062	1204	1171	1261
<b>CJ3 (0.05 mM)</b>	321	489	727	1898	956	999	1575	1773
<b>CJ3 (0.01 mM)</b>	347	774	1278	1552	842	1236	1626	1880

Table 16. Cell numbers corresponding to the data presented in Figure 32, page 93 and Figure 33, page 95.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	431	968	1472	1693	1403	1243	1119	1149
<b>CJ1 (0 mM)</b>	759	593	914	2552	1215	1407	1296	1068

Table 17. Cell numbers corresponding to the data presented in Figure 34, page 96.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	466	1515	1593	2000	1365	1230	2165	1193
<b>TE47 (0 mM)</b>	431	968	1472	1693	1403	1243	1119	1149
<b>TE52 (1%)</b>	501	834	1070	1692	1652	1599	2559	1402
<b>TE52 (0%)</b>	467	901	1062	1583	1509	1508	1235	1612
<b>TE53</b>	304	652	1107	1221	966	1465	940	1131

Table 18. Cell numbers corresponding to the data presented in Figure 35, page 101.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	317	369	758	1377	931	999	1373	1238
<b>TE58 (1 mM)</b>	369	450	843	1877	849	1193	742	1061

Table 19. Cell numbers corresponding to the data presented in Figure 36, page 104.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	367	369	984	1435	1878	1007	1340	1334
<b>TE58 (1 mM)</b>	369	450	843	1877	849	1193	742	1061
<b>TE78 (1 mM)</b>	283	344	719	1690	1712	1217	1259	1174

Table 20. Cell numbers corresponding to the data presented in Figure 38, page 107.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE51 (1 mM)</b>	420	326	804	1398	1257	1274	1547	2037
<b>TE51 (0 mM)</b>	446	425	660	1322	1948	1205	1307	1599
<b>TE47 (1 mM)</b>	1841	2433	4212	6451	6035	5313	5337	6855
<b>TE47 (0 mM)</b>	1038	1985	3678	4515	5576	4424	4188	4344

Table 21. Cell numbers corresponding to the data presented in Figure 39, page 110 and Figure 41, page 113.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	1841	2433	4212	6451	6035	5313	5337	6855
<b>TE47 (0 mM)</b>	1038	1985	3678	4515	5576	4424	4188	4344
<b>TE49 (1 mM)</b>	355	427	557	1115	1453	1113	1239	1215
<b>TE49 (0 mM)</b>	303	181	951	1358	1475	1211	1313	1390
<b>TE50 (1 mM)</b>	367	141	618	1019	1404	943	735	643
<b>TE50 (0 mM)</b>	318	456	536	1151	1082	835	1076	799
<b>TE51 (1 mM)</b>	516	443	1138	1847	3793	1814	1811	2752
<b>TE51 (0 mM)</b>	531	536	903	2108	2843	1548	1584	1948

Table 22. Cell numbers corresponding to the data presented in Figure 43, page 119 and Figure 44, page 121.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	N/a	N/a	N/a	1511	707	1284	N/a	1732
<b>TE51 (1 mM)</b>	N/a	N/a	N/a	1450	786	1169	N/a	1465

Table 23. Cell numbers corresponding to the data presented in Figure 48, page 131.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE49 (1 mM)</b>	N/a	N/a	N/a	1398	834	1017	N/a	1213
<b>TE50 (1 mM)</b>	N/a	N/a	N/a	1356	891	1086	N/a	1160

Table 24. Cell numbers corresponding to the data presented in Figure 49, page 133.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE47 (1 mM)</b>	N/a	N/a	N/a	1968	1416	528	N/a	1262
<b>TE47 (0 mM)</b>	N/a	N/a	N/a	1707	1473	858	N/a	1408

Table 25. Cell numbers corresponding to the data presented in Figure 55, page 155.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE51 (1 mM)</b>	N/a	N/a	N/a	1057	1471	379	N/a	906
<b>TE51 (0 mM)</b>	N/a	N/a	N/a	1909	785	560	N/a	968

Table 26. Cell numbers corresponding to the data presented in Figure 56, page 156.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE49 (1 mM)</b>	N/a	N/a	N/a	1780	2254	1825	N/a	1681
<b>TE49 (0 mM)</b>	N/a	N/a	N/a	1466	2369	2219	N/a	2384

Table 27. Cell numbers corresponding to the data presented in Figure 57, page 157.

Strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>TE50 (1 mM)</b>	N/a	N/a	N/a	1025	1594	1687	N/a	1895
<b>TE50 (0 mM)</b>	N/a	N/a	N/a	1723	1794	1558	N/a	2348

Table 28. Cell numbers corresponding to the data presented in Figure 58, page 158.

Time (minutes)	0	30	60	90	120	180	240	300
<b>lacI mean fluorescence</b>	56.7	35.7	39.4	55.6	93.8	227.1	351.8	410.0
<b>lacI-cis mean fluorescence</b>	67.3	34.0	51.2	64.9	122.7	322.7	481.8	596.1
<b>t</b>	-5.0	2.6	-19.6	-23.2	-26.4	-37.0	-44.5	-54.0
<b>Df</b>	1214	2176	2636	3567	2249	2446	2041	2657
<b>P-value</b>	$7.5 \times 10^{-7}$	$0.0108$ $26$	$5.7 \times 10^{-80}$	$6.6 \times 10^{-111}$	$2.3 \times 10^{-134}$	$1.2 \times 10^{-238}$	$4.1 \times 10^{-303}$	$0$
<b>lacI-cis mean fluorescence</b>	67.3	34.0	51.2	64.9	122.7	322.7	481.8	596.1
<b>lacI-trans mean fluorescence</b>	72.1	39.3	53.0	75.2	125.1	308.2	434.8	551.8
<b>t</b>	-1.9	-5.0	-2.3	-22.7	-2.2	5.3	13.4	10.3
<b>df</b>	1063	1089	2584	3447	2170	2870	2607	2986
<b>P-value</b>	0.0577	$7.1 \times 10^{-7}$	0.0203	$4.6 \times 10^{-106}$	0.0285	$1.4 \times 10^{-7}$	$7.6 \times 10^{-40}$	$2.3 \times 10^{-24}$

Table 29. Comparison of fluorescence in LacI regulated strains with full (1 mM) IPTG induction. Welch two-sample t-test comparing sfGFP\* fluorescence in strains: lacI (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*), lacI\_cis (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) and lacI-trans (CJ3, *P<sub>hyperspank</sub>-sfGFP\* lacI\_trans*). Significant differences in means are coloured red.



Time (minutes)	0	30	60	90	120	180	240	300
lacI Mean fluorescence	7.6	7.9	6.4	6.2	6.5	7.7	17.6	19.9
lacI-cis Mean fluorescence	10.1	9.2	7.3	7.8	7.9	12.5	35.0	57.9
t	-9.9	-12.1	-10.9	-18.8	-17.2	-31.6	-14.6	-19.3
df	429	2081	2446	2268	2272	1751	1989	1847
P-value	5.1 x 10 <sup>-21</sup>	1.3 x 10 <sup>-32</sup>	5.42 x 10 <sup>-27</sup>	1.1 x 10 <sup>-73</sup>	1.3 x 10 <sup>-62</sup>	3.6 x 10 <sup>-174</sup>	7.1 x 10 <sup>-46</sup>	2.0 x 10 <sup>-75</sup>
lacI-cis Mean fluorescence	10.1	9.2	7.3	7.8	7.9	12.5	35.0	57.9
lacI-trans Mean fluorescence	12.5	9.2	8.3	9.2	11.3	17.2	46.3	65.9
t	-5.0	-0.3	-12.2	-13.9	-31.5	-23.0	-6.8	-3.1
df	636	1450	2590	2620	2222	2648	2513	3245
P-value	8.7 x 10 <sup>-7</sup>	0.78	2.1 x 10 <sup>-33</sup>	3.7 x 10 <sup>-42</sup>	3.8 x 10 <sup>-180</sup>	3.8 x 10 <sup>-107</sup>	1.1 x 10 <sup>-11</sup>	0.0017

Table 30. Comparison of fluorescence in LacI regulated strains without IPTG induction. Welch two-sample t-test comparing sfGFP\* fluorescence in the lacI strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*), lacI\_cis (TE47, *P<sub>hyperspank</sub>-sfGFP\*-lacI\_cis*) and lacI-trans (CJ3, *P<sub>hyperspank</sub>-sfGFP\*-lacI\_trans*). Significant differences in means are coloured red.

lacI strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>Mean Fluorescence (1 mM IPTG)</b>	48.1	28.9	24.3	41.3	87.2	193.2	350.0	417.0
<b>Mean Fluorescence (0.1 mM IPTG)</b>	30.3	24.2	26.1	23.0	71.5	178.1	288.9	312.1
<b>t</b>	18	8	-5	40	19	10	33	43
<b>df</b>	1197	1351	2619	3849	3090	2330	3001	2374
<b>P Value</b>	2.0 x 10 <sup>-61</sup>	6.3 x 10 <sup>-16</sup>	3.7 x 10 <sup>-6</sup>	1.1 x 10 <sup>-293</sup>	3.5 x 10 <sup>-76</sup>	1.1 x 10 <sup>-24</sup>	8.2 x 10 <sup>-200</sup>	2.4 x 10 <sup>-300</sup>
<b>Mean Fluorescence (0.05 mM IPTG)</b>	23.4	18.4	16.9	23.0	52.9	142.2	243.4	248.7
<b>t</b>	10	15	27	0	30	30	31	31
<b>df</b>	719	1871	1900	3763	3545	3108	3572	2647
<b>P Value</b>	1.5 x 10 <sup>-21</sup>	7.6 x 10 <sup>-49</sup>	8.4 x 10 <sup>-138</sup>	0.929 839	3.2 x 10 <sup>-176</sup>	3.9 x 10 <sup>-171</sup>	3.9 x 10 <sup>-185</sup>	5.2 x 10 <sup>-177</sup>
<b>Mean Fluorescence (0.01 mM IPTG)</b>	9.3	10.4	8.2	7.3	11.1	22.8	39.0	46.8
<b>t</b>	45	26	48	108	110	120	205	124
<b>df</b>	1109	1235	1882	1825	2053	1671	2599	1989
<b>P Value</b>	5.6 x 10 <sup>-250</sup>	2.7 x 10 <sup>-118</sup>	0	0	0	0	0	0

Table 31. Welch two-sample t-test comparing sfGFP\* activity in the negative auto-regulating lacI strain under varied induction conditions. sfGFP\* fluorescence in the strain (AH7, *P<sub>hyperspank</sub>-sfGFP\*-lacI*) is compared in the 1 mM IPTG condition with a 10-fold reduction in inducer (0.1 mM IPTG). The 0.1 mM IPTG condition is compared with 0.05 mM IPTG and the 0.05 mM condition is compared to the 0.01 mM condition. Red colours indicate a statistically significant difference, accepted where P-values < 6.3 x 10<sup>-3</sup> (see 4.2.2).

lacI-cis strain	Time (min)							
	0	30	60	90	120	180	240	300
<b>Mean Fluorescence (1 mM IPTG)</b>	64.3	65.4	46.3	56.1	73.0	323.6	482.5	698.5
<b>Mean Fluorescence (0.1 mM IPTG)</b>	72.6	66.1	42.2	51.1	72.7	301.9	470.1	658.6
<b>t</b>	-3	0	5	7	0	7	3	10
<b>df</b>	480	538	888	1611	2125	1914	2251	3469
<b>P Value</b>	9.0 x 10 <sup>-4</sup>	0.809 087	1.6 x 10 <sup>-7</sup>	1.2 x 10 <sup>-13</sup>	0.671 009	5.4 x 10 <sup>-13</sup>	4.8 x 10 <sup>-3</sup>	1.7 x 10 <sup>-21</sup>
<b>Mean Fluorescence (0.05 mM IPTG)</b>	56.1	51.1	44.0	40.8	53.9	263.0	401.8	558.5
<b>t</b>	8	6	-2	15	33	11	15	23
<b>df</b>	800	559	809	1517	2154	1457	1829	2822
<b>P Value</b>	2.0 x 10 <sup>-16</sup>	1.1 x 10 <sup>-9</sup>	0.026 17	2.4 x 10 <sup>-47</sup>	7.8 x 10 <sup>-197</sup>	1.8 x 10 <sup>-27</sup>	1.2 x 10 <sup>-46</sup>	6.2 x 10 <sup>-111</sup>
<b>Mean Fluorescence (0.01 mM IPTG)</b>	20.1	20.1	16.4	15.9	17.6	92.3	173.8	207.8
<b>t</b>	27	19	37	47	88	54	59	93
<b>df</b>	507	290	572	866	1295	1038	1249	2486
<b>P Value</b>	1.5 x 10 <sup>-98</sup>	4.7 x 10 <sup>-53</sup>	4.5 x 10 <sup>-155</sup>	2.3 x 10 <sup>-243</sup>	0	3.2 x 10 <sup>-306</sup>	0	0

Table 32. Welch two-sample t-test comparing sfGFP\* activity in the negatively regulated lacI-cis strain under varied induction conditions. sfGFP\* fluorescence in the strain (TE47, *P<sub>hyperspank</sub>-sfGFP\* lacI\_cis*) is compared in the 1 mM IPTG condition with a 10-fold reduction in inducer (0.1 mM IPTG). The 0.1 mM IPTG condition is compared with 0.05 mM IPTG and the 0.05 mM condition is compared to the 0.01 mM condition. Red colours indicate a statistically significant difference, accepted where P-values < 6.3 x 10<sup>-3</sup> (see 4.2.2).

lacI-trans strain	Time (min)							
	0	30	60	90	120	180	240	300
Mean Fluorescence (1 mM IPTG)	62.4	60.6	53.1	69.9	91.0	299.9	478.9	479.8
Mean Fluorescence (0.1 mM IPTG)	59.5	63.2	54.5	62.1	96.5	293.4	485.2	487.0
t	1	-2	-2	13	-6	3	-2	-1
df	617	841	1864	3399	2022	2334	2128	1853
P Value	0.141 306	0.072 573	0.022 858	5.8 x 10 <sup>-40</sup>	3.0 x 10 <sup>-8</sup>	0.008 809	0.045 47	0.167 64
Mean Fluorescence (0.05 mM IPTG)	49.4	59.2	46.3	57.2	91.8	277.0	463.8	445.5
t	6	3	12	9	6	6	8	10
df	593	674	1601	3480	1940	2190	2603	2920
P Value	6.6 x 10 <sup>-10</sup>	1.8 x 10 <sup>-3</sup>	9.0 x 10 <sup>-34</sup>	3.1 x 10 <sup>-19</sup>	1.4 x 10 <sup>-8</sup>	1.2 x 10 <sup>-10</sup>	2.3 x 10 <sup>-14</sup>	2.8 x 10 <sup>-23</sup>
Mean Fluorescence (0.01 mM IPTG)	19.3	18.7	18.0	24.3	38.8	124.9	219.7	204.4
t	29	62	60	91	89	82	104	67
df	475	561	865	2744	1546	1559	2937	3192
P Value	1.0 x 10 <sup>-108</sup>	3.0 x 10 <sup>-252</sup>	0	0	0	0	0	0

Table 33. Welch two-sample t-test comparing sfGFP\* activity in the negatively regulated lacI-trans strain under varied induction conditions. sfGFP\* fluorescence in the strain (CJ3, *P<sub>hyperspank</sub>-sfGFP\* lacI<sub>cis</sub>*) is compared in the 1 mM IPTG condition with a 10-fold reduction in inducer (0.1 mM IPTG). The 0.1 mM IPTG condition is compared with 0.05 mM IPTG and the 0.05 mM condition is compared to the 0.01 mM condition. Red colours indicate a statistically significant difference, accepted where P-values < 6.3 x 10<sup>-3</sup> (see 4.2.2).

Time (minutes)	0	30	60	90	120	180	240	300
TE47 mean fluorescence (a.u)	47.4	56.7	41.4	63.3	105.3	246.8	494.7	499.6
TE58 mean fluorescence (a.u)	75.3	73.8	62.2	75.9	118.2	331.1	556.5	746.1
t	-13	-12	-25	-17	-10	-20	-12	-34
df	678	779	1541	3252	1562	2056	1033	1736
P-value	8.9 x 10 <sup>-33</sup>	2.5 x 10 <sup>-32</sup>	6.3 x 10 <sup>-106</sup>	7.5 x 10 <sup>-61</sup>	4.4 x 10 <sup>-23</sup>	3.1 x 10 <sup>-84</sup>	2.3 x 10 <sup>-31</sup>	1.1 x 10 <sup>-190</sup>

Table 34. Fluorescent output of the sfGFP\* reporter is reduced by the N-terminal translational linker coding for MEFLQ. Welch two-sample t-tests comparing strain TE47 (*P<sub>hyperspank</sub>-sfGFP\* lacI<sub>cis</sub>*) containing the linker and strain TE58 (*P<sub>hyperspank</sub>-sfGFP lacI<sub>cis</sub>*) without the linker. Fluorescent reporter genes are fully induced with 1 mM IPTG. Statistically significant differences have been coloured red.

Time (minutes)	0	30	60	90	120	180	240	300
TE58 mean fluorescence (a.u)	75.3	73.8	62.2	75.9	118.2	331.1	556.5	746.1
TE78 mean fluorescence (a.u)	103.6	99.5	84.5	112.9	185.7	420.6	584.5	798.3
t	-8	-16	-24	-55	-56	-28	-5	-7
df	437	766	1541	3084	1476	2098	1233	1739
P-value	3.1 x 10 <sup>-14</sup>	4.9 x 10 <sup>-49</sup>	6.1 x 10 <sup>-105</sup>	0	0	3.4 x 10 <sup>-145</sup>	2.8 x 10 <sup>-7</sup>	2.1 x 10 <sup>-12</sup>

Table 35. Fluorescent output of the sfGFP reporter is increased by codon optimisation. Mean fluorescence was compared using a Welch two-sample t-test for strain TE58 (*P<sub>hyperspank</sub>-sfGFP lacI<sub>cis</sub>*) containing the original *sfGFP* and strain TE76 (*P<sub>hyperspank</sub>-sfGFP\_RiboTempo lacI<sub>cis</sub>*) containing the codon optimised *sfGFP*. Statistically significant values have been coloured red.

Induction	1 mM IPTG			0 mM IPTG		
	90 (A) 120 (B)	90 (A) 210 (B)	90 (A) 300 (B)	300 (A) 90 (B)	300 (A) 120 (B)	300 (A) 210 (B)
Relative <i>mCherry</i> mRNA (A)	1	1	1	1	1	1
Relative <i>mCherry</i> mRNA (B)	0.5805	0.2749	0.2327	0.1132	0.0740	0.2860
T-value	4	11	8	35	51	2
Df	2	2	2	2	2	2
P-value	0.0483	0.0080	0.0173	0.0471	0.0258	0.1370

Table 36. Paired t-test comparing the relative mCherry mRNA levels during different growth phases. Strain TE51 (*P<sub>hyperspank</sub>-mCherry lacI<sub>cis</sub>*) was induced with 1 mM IPTG or without IPTG induction. The relative band intensities of n=3 Northern Blots are calculated using Image J. Statistically significant differences have been coloured red.

<b>Strain</b>	sfGFP*	mCherry2	green-red	green-red	red-green	red-green
<b>Probe</b>	<i>sfGFP</i>	<i>mCherry2</i>	<i>sfGFP</i>	<i>mCherry2</i>	<i>sfGFP</i>	<i>mCherry</i>
<b>Half-life Exponential Growth (min)</b>	8.3	17.4	10.0	5.6	16.6	19.8
<b>Standard deviation</b>	0.4	2.5	5.5	7.5	4.1	4.9
<b>Half-life Stationary Phase (min)</b>	4.8	3.4	2.5	5.5	7.5	4.1
<b>Standard deviation</b>	1.2	0.8	3.2	2.2	7.5	0.2
<b>T-value</b>	3	4	2	1	3	2
<b>Df</b>	2	2	2	1	2	1
<b>P-value</b>	0.073	0.052	0.150	0.436	0.122	0.336

Table 37. mRNA half-lives of strains in exponential growth and in stationary phase. Paired t-tests compare the: sfGFP\* strain (TE47,  $P_{hyperspank}$ -*sfGFP\** *lacI\_cis*), mCherry2 (TE51,  $P_{hyperspank}$ -*mCherry2* *lacI\_cis*), green-red (TE49,  $P_{hyperspank}$ -*sfGFP\**-*mCherry2* *lacI\_cis*) and red-green strain (TE50,  $P_{hyperspank}$ -*mCherry2*-*sfGFP\** *lacI\_cis*).

## Chapter 11. References

- Ababneh, Q.O. and Herman, J.K. (2015) 'RelA Inhibits *Bacillus subtilis* Motility and Chaining', *Journal of Bacteriology*, 197(1), pp. 128-137.
- Andrews, B.T., Schoenfish, A.R., Roy, M., Waldo, G. and Jennings, P.A. (2007) 'The rough energy landscape of superfolder GFP is linked to the chromophore', *Journal of Molecular Biology*, 373(2), pp. 476-490.
- Arrieta-Ortiz, M.L., Hafemeister, C., Bate, A.R., Chu, T., Greenfield, A., Shuster, B., Barry, S.N., Gallitto, M., Liu, B., Kacmarczyk, T., Santoriello, F., Chen, J., Rodrigues, C.D.A., Sato, T., Rudner, D.Z., Driks, A., Bonneau, R. and Eichenberger, P. (2015) 'An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network', *Molecular Systems Biology*, 11(11), p. 17.
- Bais, H.P., Fall, R. and Vivanco, J.M. (2004) 'Biocontrol of *Bacillus subtilis* against infection of *Arabidopsis* roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production', *Plant Physiology*, 134(1), pp. 307-319.
- Belas, R. (2014) 'Biofilms, flagella, and mechanosensing of surfaces by bacteria', *Trends in Microbiology*, 22(9), pp. 517-527.
- Blair, K.M., Turner, L., Winkelman, J.T., Berg, H.C. and Kearns, D.B. (2008) 'A molecular clutch disables flagella in the *Bacillus subtilis* biofilm', *Science*, 320(5883), pp. 1636-1638.
- Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R. and Kolter, R. (2001) 'Fruiting body formation by *Bacillus subtilis*', *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), pp. 11621-11626.
- Burbulys, D., Trach, K.A. and Hoch, J.A. (1991) 'Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay', *Cell*, 64(3), pp. 545-552.
- Butz, M., Neuenschwander, M., Kast, P. and Hilvert, D. (2011) 'An N-Terminal Protein Degradation Tag Enables Robust Selection of Highly Active Enzymes', *Biochemistry*, 50(40), pp. 8594-8602.

- Calvio, C., Osera, C., Amati, G. and Galizzi, A. (2008) 'Autoregulation of swrAA and motility in *Bacillus subtilis*', *Journal of Bacteriology*, 190(16), pp. 5720-5728.
- Cano, R.J. and Borucki, M.K. (1995) 'Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber', *Science*, 268(5213), pp. 1060-1064.
- Carabetta, V.J., Tanner, A.W., Greco, T.M., Defrancesco, M., Cristea, I.M. and Dubnau, D. (2013) 'A complex of YlbF, YmcA and YaaT regulates sporulation, competence and biofilm formation by accelerating the phosphorylation of Spo0A', *Molecular Microbiology*, 88(2), pp. 283-300.
- Caramori, T., Barilla, D., Nessi, C., Sacchi, L. and Galizzi, A. (1996) 'Role of FlgM in sigma(D)-dependent gene expression in *Bacillus subtilis*', *Journal of Bacteriology*, 178(11), pp. 3113-3118.
- Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P. and Sabatini, D.M. (2006) 'CellProfiler: image analysis software for identifying and quantifying cell phenotypes', *Genome Biology*, 7(10).
- Chai, Y., Beauregard, P.B., Vlamakis, H., Losick, R. and Kolter, R. (2012) 'Galactose Metabolism Plays a Crucial Role in Biofilm Formation by *Bacillus subtilis*', *Mbio*, 3(4).
- Chai, Y., Kolter, R. and Losick, R. (2009) 'Paralogous antirepressors acting on the master regulator for biofilm formation in *Bacillus subtilis*', *Molecular Microbiology*, 74(4), pp. 876-887.
- Chai, Y., Norman, T., Kolter, R. and Losick, R. (2010) 'An epigenetic switch governing daughter cell separation in *Bacillus subtilis*', *Genes & Development*, 24(8), pp. 754-765.
- Chen, C.L., Zhang, H.B., Broitman, S.L., Reiche, M., Farrell, I., Cooperman, B.S. and Goldman, Y.E. (2013) 'Dynamics of translation by single ribosomes through mRNA secondary structures', *Nature Structural & Molecular Biology*, 20(5), pp. 582-+.



Chen, Y., Ray, W.K., Helm, R.F., Melville, S.B. and Popham, D.L. (2014) 'Levels of Germination Proteins in *Bacillus subtilis* Dormant, Superdormant, and Germinating Spores', *Plos One*, 9(4), p. 10.

Choi, P.J., Cai, L., Frieda, K. and Xie, S. (2008) 'A stochastic single-molecule event triggers phenotype switching of a bacterial cell', *Science*, 322(5900), pp. 442-446.

Cozy, L.M. and Kearns, D.B. (2010) 'Gene position in a long operon governs motility development in *Bacillus subtilis*', *Molecular Microbiology*, 76(2), pp. 273-285.

Cozy, L.M., Phillips, A.M., Calvo, R.A., Bate, A.R., Hsueh, Y.-H., Bonneau, R., Eichenberger, P. and Kearns, D.B. (2012) 'SlrA/SinR/SlrR inhibits motility gene expression upstream of a hypersensitive and hysteretic switch at the level of sD in *Bacillus subtilis*', *Molecular Microbiology*, 83(6), pp. 1210-1228.

Cramer, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P.C. (1996) 'Improved green fluorescent protein by molecular evolution using DNA shuffling', *Nature Biotechnology*, 14(3), pp. 315-319.

Day, R.N. and Davidson, M.W. (2009) 'The fluorescent protein palette: tools for cellular imaging', *Chemical Society Reviews*, 38(10), pp. 2887-2921.

de Hoon, M.J.L., Makita, Y., Nakai, K. and Miyano, S. (2005) 'Prediction of transcriptional terminators in *Bacillus subtilis* and related species', *Plos Computational Biology*, 1(3), pp. 212-221.

de Jong, I.G., Veening, J.-W. and Kuipers, O.P. (2010a) 'Heterochronic Phosphorelay Gene Expression as a Source of Heterogeneity in *Bacillus subtilis* Spore Formation', *Journal of Bacteriology*, 192(8), pp. 2053-2067.

de Jong, I.G., Veening, J.W. and Kuipers, O.P. (2010b) 'Heterochronic phosphorelay gene expression as a source of heterogeneity in *Bacillus subtilis* spore formation', *J Bacteriol*, 192(8), pp. 2053-67.

De Lorimier, R.M., Smith, J.J., Dwyer, M.A., Looger, L.L., Sali, K.M., Paavola, C.D., Rizk, S.S., Sadigov, S., Conrad, D.W., Loew, L. and Hellinga, H.W. (2002) 'Construction of a fluorescent biosensor family', *Protein Science*, 11(11), pp. 2655-2675.

Doherty, G.P., Bailey, K. and Lewis, P.J. (2010) 'Stage-specific fluorescence intensity of GFP and mCherry during sporulation In *Bacillus Subtilis*', *BMC research notes*, 3, pp. 303-303.

Dong, H.J., Nilsson, L. and Kurland, C.G. (1996) 'Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates', *Journal of Molecular Biology*, 260(5), pp. 649-663.

Dsouza, C., Nakano, M.M., Corbell, N. and Zuber, P. (1993) 'Amino-acylation site mutations in amino acid-activating domains of surfactin synthetase: effects on surfactin production and competence development in *Bacillus subtilis*', *Journal of Bacteriology*, 175(11), pp. 3502-3510.

Dublanche, Y., Michalodimitrakis, K., Kuemmerer, N., Foglierini, M. and Serrano, L. (2006) 'Noise in transcription negative feedback loops: simulation and experimental analysis', *Molecular Systems Biology*, 2.

Dubnau, D. (1991) 'Genetic competence in *Bacillus subtilis*', *Microbiological Reviews*, 55(3), pp. 395-424.

Einstein, A.F., R (1956) 'On the Movement of Small Particles Suspended in a Stationary Liquid Demanded by the Molecular-Kinetic Theory of Heat', *Dover Publications, inc.*

Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) 'Stochastic gene expression in a single cell', *Science*, 297(5584), pp. 1183-1186.

Errington, J. (2003) 'Regulation of endospore formation in *Bacillus subtilis*', *Nature Reviews Microbiology*, 1(2), pp. 117-126.

Estacio, W., Santa Anna-Arriola, S., Adedipe, M. and Marquez-Magana, L.M. (1998) 'Dual promoters are responsible for transcription initiation of the *fla*/*che* operon in *Bacillus subtilis*', *Journal of Bacteriology*, 180(14), pp. 3548-3555.

Fajardo-Cavazos, P. and Nicholson, W.L. (2000) 'The TRAP-like SplA protein is a trans-acting negative regulator of spore photoprotein lyase synthesis during *Bacillus subtilis* sporulation', *Journal of Bacteriology*, 182(2), pp. 555-560.

Femino, A., Fay, F.S., Fogarty, K. and Singer, R.H. (1998) 'Visualization of single RNA transcripts in situ', *Science*, 280(5363), pp. 585-590.

Fengos, G. and Iber, D. (2013) 'Prediction stability in a data-based, mechanistic model of sigma(F) regulation during sporulation in *Bacillus subtilis*', *Scientific Reports*, 3.

Ferrell, J.E. (2002) 'Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability', *Current Opinion in Cell Biology*, 14(2), pp. 140-148.

Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G. and Collins, J.J. (2009) 'Synthetic Gene Networks That Count', *Science*, 324(5931), pp. 1199-1202.

Fujita, M., Gonzalez-Pastor, J.E. and Losick, R. (2005) 'High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*', *Journal of Bacteriology*, 187(4), pp. 1357-1368.

Fujita, M. and Losick, R. (2005) 'Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A', *Genes & Development*, 19(18), pp. 2236-2244.

Fukushima, T., Ishikawa, S., Yamamoto, H., Ogasawara, N. and Sekiguchi, J. (2003) 'Transcriptional, functional and cytochemical analyses of the *veg* gene in *Bacillus subtilis*', *Journal of Biochemistry*, 133(4), pp. 475-483.

Gamba, P., Jonker, M.J. and Hamoen, L.W. (2015) 'A Novel Feedback Loop That Controls Bimodal Expression of Genetic Competence', *Plos Genetics*, 11(6).

Gamba, P., Veening, J.-W., Saunders, N.J., Hamoen, L.W. and Daniel, R.A. (2009) 'Two-Step Assembly Dynamics of the Bacillus subtilis Divisome', *Journal of Bacteriology*, 191(13), pp. 4186-4194.

Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) 'Construction of a genetic toggle switch in Escherichia coli', *Nature*, 403(6767), pp. 339-342.

Gartner, D., Degenkolb, J., Ripperger, J.A.E., Allmansberger, R. and Hillen, W. (1992) 'Regulation of the Bacillus subtilis W23 xylose utilization operon: interaction of the Xyl repressor with the xyl operator and the inducer xylose', *Molecular & General Genetics*, 232(3), pp. 415-422.

Gartner, D., Geissendorfer, M. and Hillen, W. (1988) 'Expression of the Bacillus subtilis xyl operon is repressed at the level of transcription and is induced by xylose', *Journal of Bacteriology*, 170(7), pp. 3102-3109.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., III and Smith, H.O. (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods*, 6(5), pp. 343-U41.

Gillespie, D.T. (1977) 'Exact Stochastic Simulation of Coupled Chemical Reactions', *Journal of Physical Chemistry*, 81(25), pp. 2340-2361.

Golding, I., Paulsson, J., Zawilski, S.M. and Cox, E.C. (2005) 'Real-time kinetics of gene activity in individual bacteria', *Cell*, 123(6), pp. 1025-1036.

Gonzalez-Pastor, J.E., Hobbs, E.C. and Losick, R. (2003) 'Cannibalism by sporulating bacteria', *Science*, 301(5632), pp. 510-513.

Gottesman, S., Roche, E., Zhou, Y.N. and Sauer, R.T. (1998) 'The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system', *Genes & Development*, 12(9), pp. 1338-1347.

Grau, R.R., de Ona, P., Kunert, M., Lenini, C., Gallegos-Monterrosa, R., Mhatre, E., Vileta, D., Donato, V., Holscher, T., Boland, W., Kuipers, O.P. and Kovacs, A.T. (2015) 'A Duo of

Potassium-Responsive Histidine Kinases Govern the Multicellular Destiny of *Bacillus subtilis*', *Mbio*, 6(4), p. 16.

Haijema, B.J., Hahn, J., Haynes, J. and Dubnau, D. (2001) 'A ComGA-dependent checkpoint limits growth during the escape from competence', *Molecular Microbiology*, 40(1), pp. 52-64.

Hamoen, L.W., Kausche, D., Marahiel, M.A., van Sinderen, D., Venema, G. and Serror, P. (2003a) 'The *Bacillus subtilis* transition state regulator AbrB binds to the -35 promoter region of comK', *Fems Microbiology Letters*, 218(2), pp. 299-304.

Hamoen, L.W., Smits, W.K., de Jong, A., Holsappel, S. and Kuipers, O.P. (2002) 'Improving the predictive value of the competence transcription factor (ComK) binding site in *Bacillus subtilis* using a genomic approach', *Nucleic Acids Research*, 30(24), pp. 5517-5528.

Hamoen, L.W., Van Werkhoven, A.F., Venema, G. and Dubnau, D. (2000) 'The pleiotropic response regulator DegU functions as a priming protein in competence development in *Bacillus subtilis*', *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp. 9246-9251.

Hamoen, L.W., Venema, G. and Kuipers, O.P. (2003b) 'Controlling competence in *Bacillus subtilis*: shared use of regulators', *Microbiology-Sgm*, 149, pp. 9-17.

Henderson, A. (2012) *Heterogeneity and regulation of key cellular decisions in Bacillus subtilis*. Master of Research. Newcastle University.

Henkin, T.M., Grundy, F.J., Nicholson, W.L. and Chambliss, G.H. (1991) 'Catabolite repression of alpha-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli* lacI and galR repressors', *Molecular Microbiology*, 5(3), pp. 575-584.

Higgins, D. and Dworkin, J. (2012) 'Recent progress in *Bacillus subtilis* sporulation', *Fems Microbiology Reviews*, 36(1), pp. 131-148.

- Hillen, W. and Berens, C. (1994) 'Mechanisms underlying expression of Tn10 encoded tetracycline resistance', *Annual Review of Microbiology*, 48, pp. 345-369.
- Hoa, T.T., Tortosa, P., Albano, M. and Dubnau, D. (2002) 'Rok (YkuW) regulates genetic competence in *Bacillus subtilis* by directly repressing comK', *Molecular Microbiology*, 43(1), pp. 15-26.
- Hofler, C., Heckmann, J., Fritsch, A., Popp, P., Gebhard, S., Fritz, G. and Maschert, T. (2016) 'Cannibalism stress response in *Bacillus subtilis*', *Microbiology-Sgm*, 162, pp. 164-176.
- Hollenbeck, E.C., Douarhe, C., Allain, J.M., Roger, P., Regeard, C., Cegelski, L., Fuller, G.G. and Raspaud, E. (2016) 'Mechanical Behavior of a *Bacillus subtilis* Pellicle', *Journal of Physical Chemistry B*, 120(26), pp. 6080-6088.
- Howard, T.P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D.M., Taylor, G.N., Parker, D.A., Lee, R., Smirnov, N., Aves, S.J. and Love, J. (2013) 'Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp. 7636-7641.
- Hublitz, P., Albert, M. and Peters, A.H.F.M. (2009) 'Mechanisms of transcriptional repression by histone lysine methylation', *International Journal of Developmental Biology*, 53(2-3), pp. 335-354.
- Inacio, J.M. and de Sa-Nogueira, I. (2007) 'trans-acting factors and cis elements involved in glucose repression of arabinan degradation in *Bacillus subtilis*', *Journal of Bacteriology*, 189(22), pp. 8371-8376.
- Ito, M., Terahara, N., Fujinami, S. and Krulwich, T.A. (2005) 'Properties of motility in *Bacillus subtilis* powered by the H<sup>+</sup>-coupled MotAB flagellar stator, Na<sup>+</sup>-coupled MotPS or hybrid stators MotAS or MotPB', *Journal of Molecular Biology*, 352(2), pp. 396-408.

Jiang, M., Shao, W.L., Perego, M. and Hoch, J.A. (2000) 'Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*', *Molecular Microbiology*, 38(3), pp. 535-542.

Johansson, H.E., Dertinger, D., LeCuyer, K.A., Behlen, L.S., Greef, C.H. and Uhlenbeck, O.C. (1998) 'A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein', *Proceedings of the National Academy of Sciences of the United States of America*, 95(16), pp. 9244-9249.

Kampf, J. and Stuelke, J. (2015) 'Minor Cause-Major Effect: A Novel Mode of Control of Bistable Gene Expression', *Plos Genetics*, 11(6).

Kearns, D.B., Chu, F., Branda, S.S., Kolter, R. and Losick, R. (2005) 'A master regulator for biofilm formation by *Bacillus subtilis*', *Molecular Microbiology*, 55(3), pp. 739-749.

Khalil, A.S. and Collins, J.J. (2010) 'Synthetic biology: applications come of age', *Nature Reviews Genetics*, 11(5), pp. 367-379.

Khmelniskii, A., Meurer, M., Ho, C.-T., Besenbeck, B., Fueller, J., Lemberg, M.K., Bukau, B., Mogk, A. and Knop, M. (2016) 'Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers', *Molecular Biology of the Cell*, 27(2), pp. 360-370.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R. and Collins, J.J. (2004) 'Programmable cells: Interfacing natural and engineered gene networks', *Proceedings of the National Academy of Sciences of the United States of America*, 101(22), pp. 8414-8419.

Koh, A., Steyn, J., Sheth, H., Hall, P., Boyd, R.M., Woodhouse, S., Essa, Y., Tsu, D., Hallinan, D.J. and Wipat, P.A. (2010) *BacillaFilla: Fixing Cracks in Concrete*. iGEM competition. Newcastle University.

Kolodkin-Gal, I., Romero, D., Cao, S.G., Clardy, J., Kolter, R. and Losick, R. (2010) 'D-Amino Acids Trigger Biofilm Disassembly', *Science*, 328(5978), pp. 627-629.

Kozak, M. (2005) 'Regulation of translation via mRNA structure in prokaryotes and eukaryotes', *Gene*, 361, pp. 13-37.

Kuchina, A., Espinar, L., Cagatay, T., Balbin, A.O., Zhang, F., Alvarado, A., Garcia-Ojalvo, J. and Sueel, G.M. (2011) 'Temporal competition between differentiation programs determines cell fate choice', *Molecular Systems Biology*, 7.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., KlaerrBlanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., Oreilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., et al. (1997) 'The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*', *Nature*, 390(6657), pp. 249-256.

Leisner, M., Sting, K., Frey, E. and Maier, B. (2008) 'Stochastic switching to competence', *Current Opinion in Microbiology*, 11(6), pp. 553-559.

Lopez, D., Fischbach, M.A., Chu, F., Losick, R. and Kolter, R. (2009a) 'Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*', *Proceedings of the National Academy of Sciences of the United States of America*, 106(1), pp. 280-285.

Lopez, D., Vlamakis, H., Losick, R. and Kolter, R. (2009b) 'Paracrine signaling in a bacterium', *Genes & Development*, 23(14), pp. 1631-1638.



- Maamar, H. and Dubnau, D. (2005) 'Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop', *Molecular Microbiology*, 56(3), pp. 615-624.
- Magnuson, R., Solomon, J. and Grossman, A.D. (1994) 'Biochemical and genetic characterization of a competence pheromone from B. subtilis', *Cell*, 77(2), pp. 207-216.
- Markham, N.R.W., Andrew.; Zuker, Leslie. S.; Zuker, Michael. (1995-2016) 'Predicted mRNA folding', *The mfold Web Server* (Accessed: 25th April 2016).
- Marvasi, M., Visscher, P.T. and Martinez, L.C. (2010) 'Exopolymeric substances (EPS) from Bacillus subtilis: polymers and genes encoding their synthesis', *Fems Microbiology Letters*, 313(1), pp. 1-9.
- McAdams, H.H. and Arkin, A. (1999) 'It's a noisy business! Genetic regulation at the nanomolar scale', *Trends in Genetics*, 15(2), pp. 65-69.
- McKenney, P.T., Driks, A. and Eichenberger, P. (2013) 'The Bacillus subtilis endospore: assembly and functions of the multilayered coat', *Nature Reviews Microbiology*, 11(1), pp. 33-44.
- Merzlyak, E.M., Goedhart, J., Shcherbo, D., Bulina, M.E., Shcheglov, A.S., Fradkov, A.F., Gaintzeva, A., Lukyanov, K.A., Lukyanov, S., Gadella, T.W.J. and Chudakov, D.M. (2007) 'Bright monomeric red fluorescent protein with an extended fluorescence lifetime', *Nature Methods*, 4(7), pp. 555-557.
- Micka, B. and Marahiel, M.A. (1992) 'The DNA-binding protein HBsu is essential for normal growth and development in Bacillus subtilis', *Biochimie*, 74(7-8), pp. 641-650.
- Mirouze, N., Desai, Y., Raj, A. and Dubnau, D. (2012) 'Spo0A similar to P Imposes a Temporal Gate for the Bimodal Expression of Competence in Bacillus subtilis', *Plos Genetics*, 8(3).
- Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C. and Voigt, C.A. (2012) 'Genetic programs constructed from layered logic gates in single cells', *Nature*, 491(7423), pp. 249-253.

- Mordini, S., Osera, C., Marini, S., Scavone, F., Bellazzi, R., Galizzi, A. and Calvio, C. (2013) 'The Role of SwrA, DegU and P-D3 in fla/che Expression in *B. subtilis*', *Plos One*, 8(12).
- Morikawa, M. (2006) 'Beneficial biofilm formation by industrial bacteria *Bacillus subtilis* and related species', *Journal of Bioscience and Bioengineering*, 101(1), pp. 1-8.
- Moriya, N., Minamino, T., Hughes, K.T., Macnab, R.M. and Namba, K. (2006) 'The type III flagellar export specificity switch is dependent on FliK ruler and a molecular clock', *Journal of Molecular Biology*, 359(2), pp. 466-477.
- Mukherjee, S. and Kearns, D.B. (2014) 'The Structure and Regulation of Flagella in *Bacillus subtilis*', *Annual Review of Genetics*, Vol 48, 48, pp. 319-340.
- Murray, E.J., Kiley, T.B. and Stanley-Wall, N.R. (2009) 'A pivotal role for the response regulator DegU in controlling multicellular behaviour', *Microbiology-Sgm*, 155, pp. 1-8.
- Narula, J., Devi, S.N., Fujita, M. and Igoshin, O.A. (2012) 'Ultrasensitivity of the *Bacillus subtilis* sporulation decision', *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp. E3513-E3522.
- Narula, J., Kuchina, A., Lee, D.Y.D., Fujita, M., Suel, G.M. and Igoshin, O.A. (2015) 'Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication', *Cell*, 162(2), pp. 328-337.
- Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J. and Setlow, P. (2000) 'Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments', *Microbiology and Molecular Biology Reviews*, 64(3), pp. 548-+.
- Nishihara, T. and Freese, E. (1975) 'Motility of *Bacillus subtilis* during growth and sporulation', *Journal of Bacteriology*, 123(1), pp. 366-371.
- Norman, T.M., Lord, N.D., Paulsson, J. and Losick, R. (2013) 'Memory and modularity in cell-fate decision making', *Nature*, 503(7477), pp. 481-+.

Ogura, M., Liu, L., LaCelle, M., Nakano, M.M. and Zuber, P. (1999) 'Mutational analysis of ComS: evidence for the interaction of ComS and MecA in the regulation of competence development in *Bacillus subtilis*', *Molecular Microbiology*, 32(4), pp. 799-812.

Osterman, I.A., Evfratov, S.A., Sergiev, P.V. and Dontsova, O.A. (2013) 'Comparison of mRNA features affecting translation initiation and reinitiation', *Nucleic Acids Research*, 41(1), pp. 474-486.

Otsu, N. (1979) 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), p. 5.

Oudenaarden, A.R.a.A.v. (2008) 'Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences', *Cell*, 135.

Overkamp, W., Beilharz, K., Weme, R.D.O., Solopova, A., Karsens, H., Kovacs, A.T., Kok, J., Kuipers, O.P. and Veening, J.W. (2013) 'Benchmarking Various Green Fluorescent Protein Variants in *Bacillus subtilis*, *Streptococcus pneumoniae*, and *Lactococcus lactis* for Live Cell Imaging', *Applied and Environmental Microbiology*, 79(20), pp. 6481-6490.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. and van Oudenaarden, A. (2002) 'Regulation of noise in the expression of a single gene', *Nature Genetics*, 31(1), pp. 69-73.

Palani, S. and Sarkar, C.A. (2011) 'Synthetic conversion of a graded receptor signal into a tunable, reversible switch', *Molecular Systems Biology*, 7.

Patrick, J.E. and Kearns, D.B. (2009) 'Laboratory Strains of *Bacillus subtilis* Do Not Exhibit Swarming Motility', *Journal of Bacteriology*, 191(22), pp. 7129-7133.

Patrick, J.E. and Kearns, D.B. (2012) 'Swarming motility and the control of master regulators of flagellar biosynthesis', *Molecular Microbiology*, 83(1), pp. 14-23.

Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006) 'Engineering and characterization of a superfolder green fluorescent protein', *Nature Biotechnology*, 24(1), pp. 79-88.

- Pedraza, J.M. and van Oudenaarden, A. (2005) 'Noise propagation in gene networks', *Science*, 307(5717), pp. 1965-1969.
- Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G. and Cormier, M.J. (1992) 'Primary structure of the *Aequorea victoria* green-fluorescent protein', *Gene*, 111(2), pp. 229-233.
- Puigbo, P., Guzman, E., Romeu, A. and Garcia-Vallve, S. (2007) 'OPTIMIZER: a web server for optimizing the codon usage of DNA sequences', *Nucleic Acids Research*, 35, pp. W126-W131.
- Rao, C.V., Kirby, J.R. and Arkin, A.P. (2004) 'Design and diversity in bacterial chemotaxis: A comparative study in *Escherichia coli* and *Bacillus subtilis*', *Plos Biology*, 2(2), pp. 239-252.
- Ratnayake-Lecamwasam, M., Serror, P., Wong, K.W. and Sonenshein, A.L. (2001) '*Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels', *Genes & Development*, 15(9), pp. 1093-1103.
- Rosenberg, A., Sinai, L., Smith, Y. and Ben-Yehuda, S. (2012) 'Dynamic Expression of the Translational Machinery during *Bacillus subtilis* Life Cycle at a Single Cell Level', *Plos One*, 7(7).
- Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. and Elowitz, M.B. (2005) 'Gene regulation at the single-cell level', *Science*, 307(5717), pp. 1962-1965.
- Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. and Elowitz, M.B. (2007) 'Accurate prediction of gene feedback circuit behavior from component properties', *Molecular Systems Biology*, 3.
- Rybakova, K.N., Bruggeman, F.J., Tomaszewska, A., Mone, M.J., Carlberg, C. and Westerhoff, H.V. (2015) 'Multiplex Eukaryotic Transcription (In) activation: Timing, Bursting and Cycling of a Ratchet Clock Mechanism', *Plos Computational Biology*, 11(4).

- Sauer, C., Syvertsson, S., Bohorquez, L.C., Cruz, R., Harwood, C.R., van Rij, T. and Hamoen, L.W. (2016) 'Effect of Genome Position on Heterologous Gene Expression in *Bacillus subtilis*: An Unbiased Analysis', *Acs Synthetic Biology*, 5(9), pp. 942-947.
- Scheller, F.W., Wollenberger, U., Warsinke, A. and Lisdat, F. (2001) 'Research and development in biosensors', *Current Opinion in Biotechnology*, 12(1), pp. 35-40.
- Schultz, D., Wolynes, P.G., Ben Jacob, E. and Onuchic, J.N. (2009) 'Deciding fate in adverse times: Sporulation and competence in *Bacillus subtilis*', *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), pp. 21027-21034.
- Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. and Tsien, R.Y. (2004) 'Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp red fluorescent protein', *Nature Biotechnology*, 22(12), pp. 1567-1572.
- Shioi, J., Matsuura, S. and Imae, Y. (1980) 'Quantitative measurements of proton motive force and motility in *Bacillus subtilis*', *Journal of Bacteriology*, 144(3), pp. 891-897.
- Siebring, J., Elema, M.J.H., Vega, F.D., Kovacs, A.T., Haccou, P. and Kuipers, O.P. (2014) 'Repeated triggering of sporulation in *Bacillus subtilis* selects against a protein that affects the timing of cell division', *ISME Journal*, 8(1), pp. 77-87.
- Singh, A. and Soltani, M. (2013) 'Quantifying Intrinsic and Extrinsic Variability in Stochastic Gene Expression Models', *Plos One*, 8(12).
- Siuti, P., Yazbek, J. and Lu, T.K. (2013) 'Synthetic circuits integrating logic and memory in living cells', *Nature Biotechnology*, 31(5), pp. 448-+.
- Skinner, S.O., Sepulveda, L.A., Xu, H. and Golding, I. (2013) 'Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization', *Nature Protocols*, 8(6), pp. 1100-1113.
- Smits, W.K., Eschevins, C.C., Susanna, K.A., Bron, S., Kuipers, O.P. and Hamoen, L.W. (2005) 'Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development', *Molecular Microbiology*, 56(3), pp. 604-614.

- Soetaert, K., Petzoldt, T. and Setzer, R.W. (2010) 'Solving Differential Equations in R: Package deSolve', *Journal of Statistical Software*, 33(9), pp. 1-25.
- Sprinzak, D. and Elowitz, M.B. (2005) 'Reconstruction of genetic circuits', *Nature*, 438(7067), pp. 443-448.
- Stanton, B.C., Nielsen, A.A.K., Tamsir, A., Clancy, K., Peterson, T. and Voigt, C.A. (2014) 'Genomic mining of prokaryotic repressors for orthogonal logic gates', *Nature Chemical Biology*, 10(2), pp. 99-105.
- Studer, S.M. and Joseph, S. (2006) 'Unfolding of mRNA secondary structure by the bacterial translation initiation complex', *Molecular Cell*, 22(1), pp. 105-115.
- Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M. and Elowitz, M.B. (2006) 'An excitable gene regulatory circuit induces transient cellular differentiation', *Nature*, 440(7083), pp. 545-550.
- Suel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J. and Elowitz, M.B. (2007) 'Tunability and noise dependence in differentiation dynamics', *Science*, 315(5819), pp. 1716-1719.
- Sun, Q. and Margolin, W. (1998) 'FtsZ dynamics during the division cycle of live Escherichia coli cells', *Journal of Bacteriology*, 180(8), pp. 2050-2056.
- Syvertsson, S. (2013) *Bistable Differentiation in an Isogenic Cell Population*. PhD thesis. Newcastle University.
- Tabor, J.J., Salis, H.M., Simpson, Z.B., Chevalier, A.A., Levskaya, A., Marcotte, E.M., Voigt, C.A. and Ellington, A.D. (2009) 'A Synthetic Genetic Edge Detection Program', *Cell*, 137(7), pp. 1272-1281.
- Thompson, A. and Gasson, M.J. (2001) 'Location effects of a reporter gene on expression levels and on native protein synthesis in Lactococcus lactis and Saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 67(8), pp. 3434-3439.

- Tsien, R.Y. (1998) 'The green fluorescent protein', *Annual Review of Biochemistry*, 67, pp. 509-544.
- Tumbar, T., Sudlow, G. and Belmont, A.S. (1999) 'Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain', *Journal of Cell Biology*, 145(7), pp. 1341-1354.
- Turgay, K., Hahn, J., Burghoorn, J. and Dubnau, D. (1998) 'Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor', *Embo Journal*, 17(22), pp. 6730-6738.
- Vansinderen, D. and Venema, G. (1994) 'comK acts as an autoregulatory control switch in the signal transduction route to competence in *Bacillus subtilis*', *Journal of Bacteriology*, 176(18), pp. 5762-5770.
- Veening, J.-W., Smits, W.K. and Kuipers, O.P. (2008) 'Bistability, Epigenetics, and Bet-Hedging in Bacteria', in *Annual Review of Microbiology*. pp. 193-210.
- Vlamakis, H., Chai, Y., Beaugregard, P., Losick, R. and Kolter, R. (2013) 'Sticking together: building a biofilm the *Bacillus subtilis* way', *Nature Reviews Microbiology*, 11(3), pp. 157-168.
- Vreeland, R.H., Rosenzweig, W.D. and Powers, D.W. (2000) 'Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal', *Nature*, 407(6806), pp. 897-900.
- Wang, B., Barahona, M. and Buck, M. (2013) 'A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals', *Biosensors & Bioelectronics*, 40(1), pp. 368-376.
- Webb, C.D., Decatur, A., Teleman, A. and Losick, R. (1995) 'Use of green fluorescent protein for visualization of cell-specific gene expression and subcellular protein localization during sporulation in *Bacillus subtilis*', *Journal of Bacteriology*, 177(20), pp. 5906-5911.

- Wen, J.D., Lancaster, L., Hodges, C., Zeri, A.C., Yoshimura, S.H., Noller, H.F., Bustamante, C. and Tinoco, I. (2008) 'Following translation by single ribosomes one codon at a time', *Nature*, 452(7187), pp. 598-603.
- Wiegert, T. and Schumann, W. (2001) 'SsrA-mediated tagging in *Bacillus subtilis*', *Journal of Bacteriology*, 183(13), pp. 3885-3889.
- Wilking, J.N., Angelini, T.E., Seminara, A., Brenner, M.P. and Weitz, D.A. (2011) 'Biofilms as complex fluids', *Mrs Bulletin*, 36(5), pp. 385-391.
- Wilkinson, D.J. (2009) 'Stochastic modelling for quantitative description of heterogeneous biological systems', *Nature Reviews Genetics*, 10(2), pp. 122-133.
- Winkelman, J.T., Bree, A.C., Bate, A.R., Eichenberger, P., Gourse, R.L. and Kearns, D.B. (2013) 'RemA is a DNA-binding protein that activates biofilm matrix gene expression in *Bacillus subtilis*', *Molecular Microbiology*, 88(5), pp. 984-997.
- Wittman, V. and Wong, H.C. (1988) 'Regulation of the penicillinase genes of *Bacillus licheniformis*: interaction of the pen repressor with its operators', *Journal of Bacteriology*, 170(7), pp. 3206-3212.
- Wu, J.J., Schuch, R. and Piggot, P.J. (1992) 'Characterization of a *Bacillus subtilis* sporulation operon that includes genes for an RNA polymerase sigma factor and for a putative DD-carboxypeptidase', *Journal of Bacteriology*, 174(15), pp. 4885-4892.
- Yang, F., Moss, L.G. and Phillips, G.N. (1996) 'The molecular structure of green fluorescent protein', *Nature Biotechnology*, 14(10), pp. 1246-1251.
- Yansura, D.G. and Henner, D.J. (1984) 'Use of the *Escherichia coli* lac repressor and operator to control gene expression in *Bacillus subtilis*', *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 81(2), pp. 439-443.
- Yu, J., Xiao, J., Ren, X.J., Lao, K.Q. and Xie, X.S. (2006) 'Probing gene expression in live cells, one protein molecule at a time', *Science*, 311(5767), pp. 1600-1603.



Zhang, G. and Ignatova, Z. (2009) 'Generic Algorithm to Predict the Speed of Translational Elongation: Implications for Protein Biogenesis', *Plos One*, 4(4).