Investigation of the *marA*, *soxS*, *rob*, and *ramA* regulons of *Salmonella enterica* serovar Typhimurium

by

Alistair Middlemiss

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection
School of Biosciences
College of life and Environmental Sciences
University of Birmingham
February 2022

Abstract

Antimicrobials have revolutionised modern medicine, but in our hubris, we have turned a blind eye to the potential ramifications of overusing these wonder drugs. As a consequence, the incidences of antimicrobial resistance have increased considerably and pose a great threat to modern medicine. Synonymous with food poisoning, *Salmonella* species accounted for 213,000 deaths in 2017. Whilst typically associated with a self-limiting gastroenteritis, the incidences of *Salmonella* infections resistant to one or more antimicrobials is also quickly increasing.

As a foodborne pathogen that is generally transmitted through the faecal-oral route, *Salmonella* must be able to adapt to a wide variety of environmental conditions in order to ensure survival and transmission. This can be achieved through the complex, multicellular, structures of biofilms. Biofilms in *Salmonella* have important roles and greatly increase tolerance to environmental stresses, including antimicrobials. Previous work has linked the inhibition of biofilm formation in *Salmonella* to the homologous transcription factors, MarA, SoxS, Rob, and RamA, which are responsible for the control of antimicrobial resistance in both *Salmonella* and *E. coli*.

In this work, the genome-wide binding profiles of these transcription factors is determined using ChIP-seq. Subsequent Cappable-seq analysis, to identify TSSs, allows the elucidation of the direct and indirect cellular effects of these transcription factors. One notable observation was the binding of SoxS upstream of the master biofilm regulator *csgD*, part of the *csgDEFG*

operon. SoxS, and MarA, are known to indirectly inhibit *csgD* expression, and therefore biofilm formation, in *E. coli* through the *ycgZ-ymgABC* pathway. However, this pathway is absent in *Salmonella*.

This work identifies SoxS as a direct inhibitor of *csgDEFG* expression in *Salmonella* via binding to the upstream region of *csgDEFG* and repressing transcription. This observation is also observed in the presence of the master activator of *csgD* expression, MlrA; further highlighting the importance of SoxS in the inhibition of biofilm formation in *Salmonella*.

The results presented in this work bridge the gap between understanding how *E. coli* and *Salmonella* utilise the global regulators of antimicrobial stress to repress biofilm formation. Furthering the notion that the repression of biofilms by these global stress response regulators provides a survival mechanism against antimicrobials. It is hypothesised that if planktonic bacteria are subjected to antimicrobial stress, biofilm formation is repressed by direct binding of SoxS to the *csgDEFG* intergenic region. This counterintuitive repression of an antimicrobial resistance mechanism by a regulator of antimicrobial resistance could benefit the bacteria, as formation of a biofilm at this time would be energetically costly and insufficient to aid survival. Therefore, repression of biofilm formation by SoxS could allow the bacteria escape the harmful environment.

Acknowledgments

There are many people I would like to thank for their help and support during this PhD. First and foremost, I would like to thank my supervisor, Professor David Grainger, for his guidance and support. Thank you for accepting me into your group, for encouraging me and being patient with me when things weren't going as smoothly as we both would have liked. The Grainger Lab has been the best place to work, not just because of the high-quality of science, but the people who give the lab its soul. The lab was so welcoming and has been a constant pillar of support to me throughout this process. Special thanks go to Dr James Haycocks for being an excellent mentor. I really appreciate your teachings and support in the lab throughout my entire PhD. The jazz was alright too. To Drs Prateek Sharma, Gemma Warren, Lisa Lamberte, Jainaba Roussel and Shivani Singh: thank you for welcoming me so kindly into the lab and helping me find my feet. I also wish to thank Dr Joe Wade, Dr Anne Stringer and Carol Smith at the Wadsworth Centre, Albany, New York for allowing me to visit and for helping with the ChIP-seq experiments; without which, I would not have a project.

Lucas Walker and Dr David Forrest have been great colleagues and I am privileged to call you friends. Thanks for all the support, laughs, and puns. To Dr Rachel Kettles, Dr Emily Warman, Dr Tom Guest, Ali Trigg, and Charles Cooper: thanks for making the lab such a fun place to work and helping me when experiments were not going well.

I would also like to thank my friends and family. Thank you to my partner, Anna Westley, for all your love and support. Thank you for believing in me and encouraging me to be the best person I can be, I would not be half the person I am today without you by my side. Thank you to my Mum, Dad, Jenny, and Dan for being such a loving and supportive family throughout my life. I am lucky to have you all.

Table of Contents

Abstract ..... **i**

Acknowledgments ..... **iii**

List of Figures ..... **xii**

List of Tables ..... **xvi**

List of abbreviations ..... **xvii**

1. Introduction ***1***

1.1 The central dogma of genetics **2**

1.2 Bacterial RNA polymerase and genetic regulation **2**

1.2.1 Bacterial promoters 2

1.2.2 RNA polymerase 6

1.2.3 σ factors – The housekeeping σ factor 10

1.2.4 σ-factors – The alternative σ factors 11

1.2.5 Transcription factors 12

1.3 The process of transcription **22**

1.3.1 Promoter recognition and initiation 22

1.3.2 Elongation and termination 27

1.4 Biofilms **28**

1.4.1 Structure and function 28

1.4.2 Regulation of biofilms in *Salmonella* by CsgD 31

1.5 Salmonella enterica **34**

1.5.1 Overview 34

1.5.2 Infection 35

1.5.3 Virulence factors 39

1.5.4 Antimicrobial resistance in *Salmonella* 45

1.6 The transcription factors MarA, SoxS, Rob, and RamA 50

1.6.1 MarA 50

1.6.2 SoxS 53

1.6.3 Rob 55

1.6.4 RamA 55

1.7 Objectives of this study 59

2. Materials and Methods 60

2.1 Buffers and reagents 61

2.2 Bacterial strains and plasmids used 61

2.2.1 Bacterial strains 61

2.2.2 Plasmids 61

2.3 Growth of bacterial cultures and antibiotics used 61

2.3.1 Growth conditions 61

2.3.2 Solid media 64

2.3.3 Liquid media 65

2.3.4 Antibiotics 65

2.4 PCR reactions and oligonucleotides used 66

2.4.1 Buffers and reagents required 66

2.4.2 PCR 67

2.4.3 Colony PCR 67

2.4.4 Megaprimer PCR 67

2.4.5 Oligonucleotides 68

2.5 Preparation of competent cells 68

2.5.1 Buffers and reagents required 68

2.5.2 Calcium competent cells 71

2.5.3 Electrocompetent cells 71

2.6 Bacterial transformation methods 72

2.6.1 Heat shock transformation 72

2.6.2 Electroporation 72

2.6.3 Conjugation 73

2.7 Isolation of bacterial genomic and plasmid DNA 73

2.7.1 Genomic DNA prep 73

2.7.2 Plasmid DNA prep 74

2.8 Agarose gel electrophoresis 74

2.8.1 Buffers and reagents required 74

2.8.2 Agarose gel electrophoresis 75

2.9 Polyacrylamide gel electrophoresis 75

2.9.1 Buffers and reagents required 75

2.9.2 Polyacrylamide gel electrophoresis (PAGE) 76

2.9.3 SDS-PAGE 76

2.10 Extraction, precipitation, and purification of DNA 77

2.10.1 Buffers and reagents required 77

2.10.2 Phenol-chloroform extraction and ethanol precipitation 77

2.10.3 Agarose gel extraction 78

2.10.4 Polyacrylamide gel electrophoresis (PAGE) extraction 78

2.10.5 Agencourt AMPure XP magnetic bead clean up (Beckman Coulter) 79

2.11 Restriction digests 79

2.11.1 Buffers and reagents required 79

2.11.2 Restriction digests 79

2.12 Ligation of DNA fragments 80

2.12.1 Buffers and reagents required 80

2.12.2 Ligation of DNA fragments 80

2.13 Site-directed mutagenesis 80

2.13.1 Buffers and reagents required 80

2.13.2 Site-directed Mutagenesis (SDM) 81

2.14 Sequencing of DNA or RNA 81

2.14.1 Plasmid DNA and DNA fragments 82

2.14.2 DNA libraries 82

2.14.3 RNA libraries 83

2.15 β-galactosidase assay 83

2.15.1 Buffers and reagents required 83

2.15.2 β-galactosidase assay 83

2.16 Crystal violet biofilm staining assay 84

2.16.1 Buffers and reagents required 84

2.16.2 Crystal violet biofilm staining assay 85

2.17 Congo Red assays 85

2.18 Purification of recombinant proteins 86

2.18.1 Buffers and reagents required 86

2.18.1 Purification of MarA/SoxS/Rob/RamA/MlrA from SL1344 87

2.18.1 Purification of RNA polymerase from *E. coli* 88

2.19 Bradford assay 89

2.19.1 Buffers and reagents required 89

2.19.2 Bradford Assay 89

2.20 End-labelling of DNA fragments 90

2.20.1 Buffers and reagents required 90

2.20.2 End-labelling of DNA fragments 90

2.21 Electrophoretic mobility shift assay 91

2.21.1 Buffers and reagents required 91

2.21.2 Electrophoretic mobility shift assay (EMSA) 91

2.22 *In vitro* transcription assays 92

2.22.1 Buffers and reagents required 92

2.22.2 *In vitro* transcription assays 93

2.23 G + A ladder generation 93

2.23.1 Buffers and reagents required 94

2.23.2 Preparation of G + A ladder 94

2.24 ChIP-seq in *Salmonella enterica* Typhimurium SL1344 95

2.24.1 Buffers and reagents required 95

2.24.2 ChIP-seq in *Salmonella enterica* Typhimurium SL1344 96

2.25 Bioinformatic analysis of ChIP-seq data 99

2.25.1 Bioinformatic processing of ChIP-seq data 99

2.25.2 Bioinformatic analysis of ChIP-seq peak conservation 100

2.26 Cappable-seq in *Salmonella enterica* Typhimurium SL1344 103

2.26.1 Buffers and reagents required 103

2.26.2 Cappable-seq in *Salmonella enterica* Typhimurium SL1344 103

2.27 Bioinformatics analysis of RNA-seq data 104

2.27.1 FastQ processing workflow 104

2.27.2 Analysis of *Salmonella* transcription start sites using custom made Python scripts 106

3. Identification of transcription factor binding sites in the Salmonella SL1344 genome 111

3.1 Introduction 112

3.2 ChIP-seq analysis of MarA, SoxS, Rob, and RamA binding in *Salmonella* SL1344 113

3.3 MarA, SoxS, Rob, and RamA binding loci in SL1344 113

3.3.1 Properties of DNA bound by MarA, SoxS, Rob, and RamA 119

3.4 Biological functions of MarA, SoxS, RamA, and Rob bound genes in SL1344 123

3.5 Conservation of ChIP-seq binding targets amongst the Enterobacteriaceae.....126

3.6 Discussion 128

4. Identification of transcription start sites in Salmonella SL1344 by Cappable-seq..... 132

4.1 Introduction.....133

4.2 The distribution of directional transcription start sites across the *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344 genome 134

4.3 The distribution of divergent transcription start site pairs across the *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344 genome 138

4.4 Analysis of directional and bidirectional promoters -10 element sequences.....140

4.4.1 The spacing of transcription start sites within chromosomal bidirectional promoters in *Salmonella* SL1344.....140

4.5 The distribution of transcription start sites on *Salmonella* SL1344 virulence plasmids 143

4.6 Differential expression analysis of transcription start site use in response to MarA, SoxS, or RamA.....150

4.7 Discussion 157

5. SoxS represses the expression of the biofilm regulator csgD 166

5.1 The *csgBAC-csgDEFG* intergenic region.....167

5.2 MarA, SoxS, Rob, and RamA all bind the *csgD* intergenic region 171

5.3 MarA, SoxS, Rob, and RamA reduce expression of the *csgD* promoter.....174

5.4 SoxS represses *csgD* transcription directly.....178

5.5 SoxS can bind to both binding sites present in the *csgD* intergenic region 181

5.6 SoxS does not outcompete the activators MlrA or IHF when binding to the *csgD* intergenic region 181

5.7 SoxS-dependent repression of the *csgD* intergenic region requires SoxS site 2 185

5.8 Repression of CsgD-mediated biofilm formation *in vivo* 188

5.9 Discussion 191

6. Final conclusions 193

References 205

7. Appendix 225

7.1 Python Code 226

7.1.1 RNA_seq_analysis_multiple_samples_final.py 226

7.1.2 RNA_seq_combine_replicates.py 229

7.1.3 RNA_seq_total_combined_TSS.py 232

7.1.4 generate_EdgeR_inputs.py 235

7.1.5 extract_ChIP-seq_coordinates.py 239

7.1.6 merge_controls.py 241

7.1.7 bidirectional_analysis_final.py 243

7.1.8 extract_bidirectional_promoter_sequences.py 246

List of Figures

Figure 1.1 The central dogma of genetics 3

Figure 1.2 The elements of bacterial σ70 promoters 5

Figure 1.3 Subunits of the RNA polymerase holoenzyme 7

Figure 1.4 Mechanisms of activation by transcription factors 14

Figure 1.5 Mechanisms of repression by transcription factors 18

Figure 1.6 The process of transcription in Escherichia coli 25

Figure 1.7 The stages of biofilm development on a solid substrate 30

Figure 1.8 Simplified transcription factor-mediated regulation of csgD 33

Figure 1.9 Schematic representation of SPI-1 and SPI-2 40

Figure 1.10 Simplified regulation of SPI-1 and SPI-2 41

Figure 1.11 Mechanisms of antimicrobial resistance employed by Salmonella 47

Figure 1.12 The marRAB locus 51

Figure 1.13 The soxRS locus 54

Figure 1.14 The rob locus 56

Figure 1.15 The ramRA locus 58

Figure 3.1 The binding sites of SoxS, MarA, RamA, and Rob across the SL1344 genome 118

Figure 3.2 The overlap of the binding peaks observed for each transcription factor 120

Figure 3.3 The binding motifs of each transcription factor studied here compared to previously known binding motifs 121

Figure 3.4 The position of each transcription factor binding peak relative to the nearest start codon *122*

Figure 3.5 The biological function of each gene identified by ChIP-seq *125*

Figure 3.6 The conservation of binding sites identified by ChIP-seq across members of the Enterobacteriaceae *127*

Figure 4.1 A comparison of the number of TSSs, and the distribution of chromosomally encoded directional TSSs, identified by two different RNA-seq techniques in Salmonella ... *137*

Figure 4.2 The distribution of chromosomally encoded divergent TSS pairs, identified by two different RNA-seq techniques in Salmonella *139*

Figure 4.3 The -10 promoter elements of chromosomal canonical and bidirectional promoters identified by Cappable-seq *141*

Figure 4.4 The -10 promoter elements of chromosomal canonical and bidirectional promoters identified by dRNA-seq by Kröger et al. (2013) *142*

Figure 4.5 The spacing of transcription start sites of chromosomal divergent TSS pairs in Salmonella SL1344 ..... *145*

Figure 4.6 The distribution of transcription start sites on the pSLT virulence plasmid *147*

Figure 4.7 The distribution of transcription start sites on the pCol1B9 virulence plasmid ... *149*

Figure 4.8 Examples of transcription start sites identified by Cappable-seq in Salmonella SL1344 ..... *151*

Figure 4.9 Differential expression analysis of transcription start sites identified by Cappable-seq *155*

Figure 5.1 The SoxS binding peak identified by ChIP-seq *168*

Figure 5.2 The intergenic region between the csgBAC-csgDEFG divergent operons *170*

Figure 5.3 Comparison of the DNA sequences of the intergenic region between the csgBAC-csgDEFG divergent operons of Salmonella enterica and E. coli *172*

Figure 5.4 EMSA showing binding of MarA, SoxS, Rob, and RamA to the csgD intergenic region *175*

Figure 5.5 The effect of MarA, SoxS, Rob, and RamA on the expression of csgD in the absence of the activator MlrA *176*

Figure 5.6 The effect of MarA, SoxS, Rob, and RamA on the expression of csgD in the presence of the activator MlrA *177*

Figure 5.7 The effect of SoxS on the transcription of the csgD promoter *179*

Figure 5.8 Schematic representation of the wildtype csgD intergenic region compared to the mutants generated *182*

Figure 5.9 The effect of mutating the SoxS binding sites within the csgD intergenic region on SoxS binding *183*

Figure 5.10 EMSA showing competition between SoxS and MlrA in binding the csgD intergenic region *184*

Figure 5.11 EMSA showing competition between SoxS and IHF in binding the csgD intergenic region *186*

Figure 5.12 The effect of SoxS on the level of csgD expression when the SoxS binding sites are mutated individually or together *187*

Figure 5.13 Regulation of biofilm formation by SoxS in Salmonella SL1344..... 189

Figure 5.14 Regulation of curli fibre production by SoxS in Salmonella SL1344 190

Figure 6.1 Hypothetical regulatory network in response to ectopic production of SoxS, MarA, and RamA in Salmonella SL1344 199

List of Tables

Table 2.1 Bacterial strains used in this study 62

Table 2.2 Plasmids used in this study 63

Table 2.3 Oligonucleotides used in this study 69

Table 2.4 Strains submitted to a BLAST search for identifying conservation of ChIP-seq binding sites 102

Table 3.1 ChIP-seq binding targets of all four transcription factors on the SL1344 chromosome and virulence plasmids 114

Table 4.1 SoxS regulated TSSs that coincide with the binding of MarA, SoxS, and RamA.... 158

Table 4.2 SoxS regulated TSSs that coincide with the binding of SoxS..... 158

Table 4.3 MarA regulated TSSs that coincide with the binding of MarA, SoxS, and RamA .. 158

Table 4.4 MarA regulated TSSs that coincide with the binding of MarA..... 159

Table 4.5 RamA regulated TSSs that coincide with the binding of MarA, SoxS, and RamA . 159

Table 4.6 RamA regulated TSSs that coincide with the binding of RamA..... 160

List of abbreviations

A	Adenine
A (Ala)	Alanine
APS	Ammonium persulphate
AR	Activating region
APS	Ammonium persulfate
ATP	Adenosine triphosphate
Å	Angstrom
bp	Base pair
BSA	Bovine serum albumin
B. subtilis	*Bacillus subtilis*
C	Cytosine
C (Cys)	Cysteine
ChIP-exo	Chromatin immunoprecipitation coupled with exonuclease digestion
ChIP-seq	Chromatin immunoprecipitation coupled with next generation sequencing
Ci	Curie
CIAP	Calf intestinal alkaline phosphatase
CRP	Cyclic-AMP receptor protein
cAMP	3’-5’-cyclic adenosine monophosphate
CTD	Carboxy-terminal domain
°C	Degrees Celcius
D (Asp)	Aspartic acid
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate

ddH_2O	Distilled and de-ionised water
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxyribonucleoside 5'-triphosphate (N=A,C,G,T)
dRNA-seq	Differential RNA-sequencing
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
E	RNA polymerase core enzyme
E (Glu)	Glutamic acid
E. coli	*Escherichia coli*
EDTA	Diaminoethanetetra-acetic acid
EMSA	Electrophoretic mobility shift assay
ETEC	Enterotoxigenic *Escherichia coli*
F (Phe)	Phenylalanine
FNR	Fumarate and nitrate reductase
G	Guanosine
G (Gly)	Glycine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His (H)	Histidine
H-NS	Histone-like nucleoid structuring protein
IHF	Integration Host Factor
IPTG	Isopropyl β-D-1- thiogalactopyranoside
K (Lys)	Lysine
kDa	Kilodalton

kb	Kilobase
L (Leu)	Leucine
LB	Lennox broth
M (Met)	Methionine
Mbp	Mega base pairs
MarA	Activator of multiple antibiotic resistance
MarR	Repressor of multiple antibiotic resistance
MES	2- (*N*-morpholino)ethanesulphonic acid
Mg	Magnesium
mRNA	Messenger ribonucleic acid
N (Asn)	Asparagine
NGS	Next generation sequencing
nt	Nucleotide
NTD	Amino-terminal domain
NTP	Nucleoside triphosphate
O	Operator
OD	Optical Density
ONPG	*o*-nitrophenyl-β-D-galactopyranoside
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pppGpp/ ppGpp	Guanosine penta/tetraphosphate
Q (Gln)	Glutamine
R (Arg)	Arginine
RamA	Activator of multiple antibiotic resistance

RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
Rob	Right of origin binding protein
S (Ser)	Serine
S. typhimurium	*Salmonella enterica* subsp. *enterica* serovar Typhimurium
Salmonella 4/74	*Salmonella enterica* subsp. *enterica* serovar Typhimurium strain ST4/74
SL1344	*Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SoxS	Superoxide stress protein
T	Thymine
T (Thr)	Threonine
Taq	*Thermus aquaticus*
TEMED	N,N,N',N'-tetramethylethyene diamine
TF	Transcription factor
T_m	Melting temperature
Tris	Tris (hydroxymethyl) aminoethane
TSS	Transcriptoin start site
U	Uracil

UP element	Upstream promoter element
UTP	Uridine triphosphate
V (Val)	Valine
V	Volts
v/v	Volume per volume
w/v	Weight per volume
W (Trp)	Tryptophan
W	Watts
WT	Wildtype
Y (Tyr)	Tyrosine
Zn	Zinc

1. Introduction

1.1 The central dogma of genetics

The ability to respond to environmental signals has been essential for bacteria to colonise almost all environmental niches. In the field of genetics, one of the core concepts is the central dogma. This describes the process by which the genetic information in DNA is converted into a protein to provide a function for the organism (Figure 1.1). The information required to produce proteins is packaged in parcels called genes. When the organism requires access to this information it transcribes the gene, producing an mRNA transcript. This mRNA sequence is subsequently translated by ribosomes to make a protein (Figure 1.1). Hence, DNA is used to make RNA which is used to make proteins. To ensure tight control of these processes, and prevent waste, bacteria regulate each stage of the central dogma. The most effective regulation is at the transcriptional level, preventing unnecessary transcripts being produced.

1.2 Bacterial RNA polymerase and genetic regulation

1.2.1 Bacterial promoters

As mentioned above, transcription is regulated to maximise the fitness of bacteria. Bacterial promoters, the site at which transcription is initiated, are the gatekeepers of transcription (Browning and Busby, 2004). Transcription start sites (TSSs) are referred to as the "+1" position of transcription. The prefixes "-" and "+" are used to describe sequence features upstream or downstream of the TSS respectively. Binding of RNA polymerase (RNAP) to promoter sequences is facilitated by σ factors; In *Escherichia coli*, the most common σ-factor is σ^{70}.

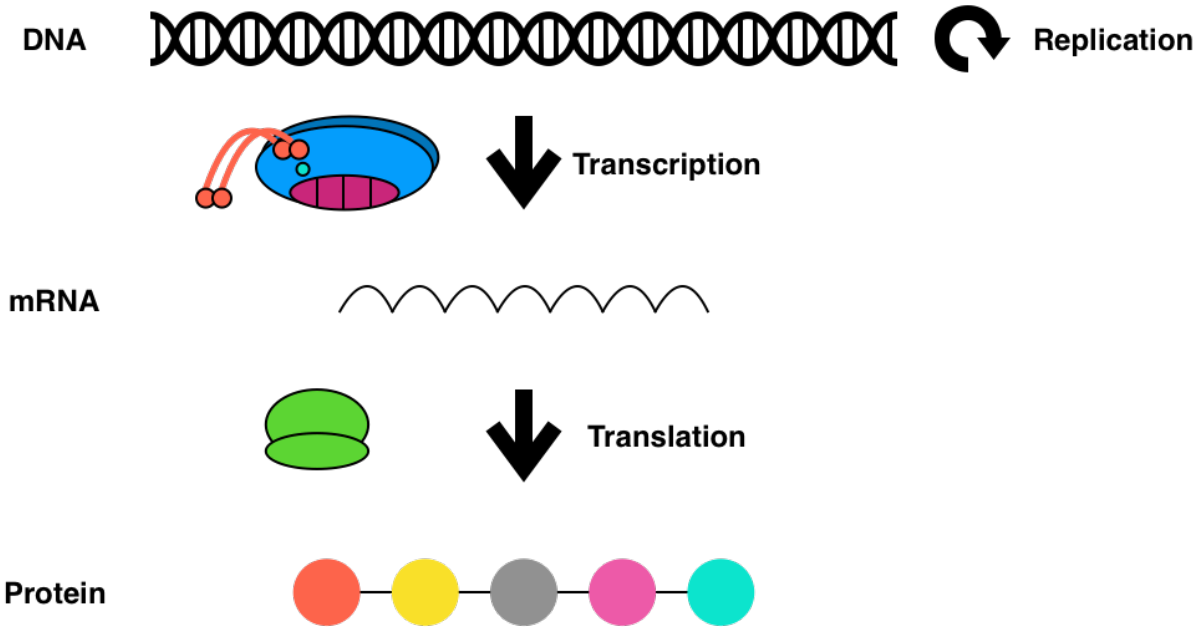


Figure 1.1 The central dogma of genetics. The flow of information within bacteria, termed the central dogma. A strand of DNA (oversimplified) is represented by a double helix, mRNA by a wavy black line and protein by coloured circles (representing individual amino acids) connected by peptide bonds (lines). The circular arrow represents DNA replication. Transcription is performed by RNAP holoenzyme shown as a blue, orange, and magenta complex. Translation is performed by ribosomes and is shown as green ovals.

As a dissociable subunit of RNAP, the σ factors associate with RNAP core enzyme (α2ββ'ω) (discussed below) to form the RNAP holoenzyme, which is competent for sequence specific transcription initiation (Borukhov and Nudler, 2008). The σ factors direct RNAP holoenzyme binding to promoter elements upstream of the target gene and facilitate the unwinding of the double helix, allowing transcription to initiate (Wosten, 1998). Housekeeping bacterial promoters are composed of up to 5 elements (Browning and Busby, 2016). The two most important are the -10 and -35 hexamers. The -10 hexamer, also known as the Pribnow box (Pribnow, 1975), has the consensus sequence 5'-TATAAT-3' and is located 10 base pairs (bp) upstream of the TSS; the -35 hexamer is located 35 bp upstream and has the consensus sequence 5'-TTGACA-3' (Busby and Ebright, 1994) (Figure 1.2). In addition to these sequences, σ factors can bind to two other promoter motifs, the extended -10 element (5'-TGTG-3') at positions -17 to -14 and the discriminator (5'-GGG-3') at positions -6 to -4 (Haugen *et al.*, 2008). These four components of the promoter structure form direct contacts with the σ-factors of the RNAP holoenzyme (Haugen *et al.*, 2008). The final element of the promoter structure does not contact the σ-factors but instead interacts with the α subunit of the RNAP; the UP element is a 20 base pair region at positions -37 to -58 and found in some of the strongest promoters (Browning and Busby, 2016, Haugen *et al.*, 2008) (Figure 1.2). The functions of these promoter elements are to align and orientate the σ-factor and RNAP complex to ensure transcription occurs in the right direction and to initiate the process of transcription. Alternative σ factors (i.e. those other than σ^{70} in *E. coli*), recognise different sequence elements and so redirect RNAP in response to specific stresses (Wosten, 1998, Browning and Busby, 2004).

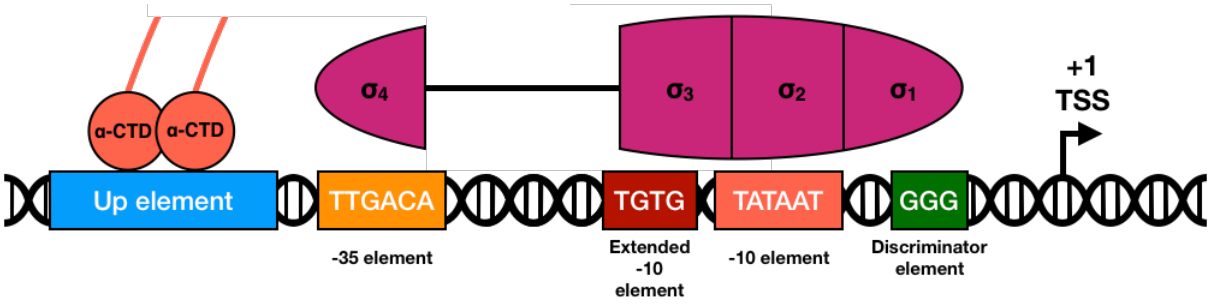


Figure 1.2 The elements of bacterial σ70 promoters. Bacterial σ70 promoters are comprised of five distinct elements, this figure provides a schematic representation of the elements and their interactions with the RNAP holoenzyme. The transcription start site (TSS) is denoted by the arrow. Consensus sequences shown are those of σ70 with the discriminator region at positions -4 to -6 shown as a green box, the -10 element at positions -7 to -12 shown as a light orange box, and the extended -10 element at positions -14 to -17 shown as a red box; further upstream, the -35 element at positions -30 to -35 is shown as a yellow box and the up element at positions -37 to -58 is shown as a blue box. The α-C-terminal domain (α-CTD) of the RNAP holoenzyme interacts with the up element and is shown as orange circles. The σ subunit of RNAP (magenta oval) provides the rest of the contacts with the bacterial promoter elements and the domains of the σ factor are shown above the promoter element with which they interact; domains σ3 and σ4 are connected by a flexible linker (shown as a black line). Reproduced and modified from Browning and Busby (2016).

1.2.2 RNA polymerase

As described above, all transcription is dependent on the large (400 kDa) multi-subunit RNA polymerase (RNAP) (Browning and Busby, 2004). The structure of RNAP is shown in Figure 1.3. The DNA-dependent core RNAP enzyme is capable of transcribing DNA but, without a σ factor, unable to initiate targeted transcription initiation (Murakami and Darst, 2003). The main body of the protein is composed of the β and β' subunits, which are assembled by the N-terminal domain (NTD) of two identical α subunits; the ω subunit aids the folding of the β subunits (Minakhin *et al.*, 2001b). The core RNAP forms a 'crab-claw' shape, with a 27 Å channel which can accommodate double stranded DNA and contains the Mg^{2+} cofactor (Zhang *et al.*, 1999).

1.2.2.1 RNAP β and β' subunits

Forming the catalytic machinery of RNAP, the β and β' subunits are 150 and 155 kDa respectively and comprise the majority of the enzyme's molecular weight (Zhang *et al.*, 1999, Sutherland and Murakami, 2018). This complex has multiple components which contribute to the efficient function of transcription. These can be categorised into stationary and mobile elements (Borukhov and Nudler, 2008). The stationary elements form the structural frame RNAP uses to interact with the σ factor and hold the DNA in place whilst transcription occurs. These elements also correctly position the catalytic aspartate residues for RNA synthesis and provide an entryway for nucleotide triphosphates (NTPs) to be used as substrates in this process (Bushnell *et al.*, 2002, Minakhin *et al.*, 2001a, Zhang *et al.*, 1999). β and β' each form the majority of one of the arms of the 'crab-claw' structure, leading to the formation of the

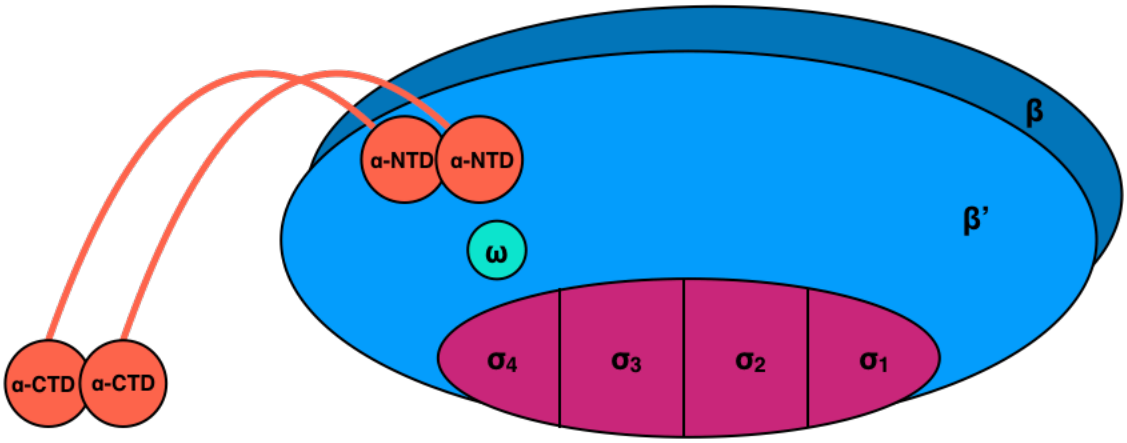


Figure 1.3 Subunits of the RNA polymerase holoenzyme. The composition of RNAP is shown as a cartoon schematic with the β and β' subunits shown as dark and light blue respectively. The two α subunits are shown as orange circles connected by flexible linkers. The σ subunit is shown as magenta and all domains are indicated. The ω subunit is shown as a green circle.

27 Å channel as mentioned above. This channel, highly conserved between prokaryotes and eukaryotes, contains the essential catalytic Mg^{2+} co-factor held in place by three aspartate residues (Borukhov and Nudler, 2008, Murakami and Darst, 2003, Zhang *et al.*, 1999, Ebright, 2000, Bushnell *et al.*, 2002). Interactions between β and β' at the base of the 'crab-claw' allow opening and closing of the structure. This allows DNA to enter and leave the complex when relaxed whilst holding the DNA in place during transcription (Borukhov and Nudler, 2008, Landick, 2001, Zhang *et al.*, 1999).

The mobile elements of RNAP aid the process of transcription by allowing RNAP to flex and translocate across the DNA template. In the acquisition of σ factors, RNAP utilises β lobes 1 and 2. These lobes contract and relax, allowing access to the β' main channel, which facilitates both binding and release of σ factors. Further to this, the β' lid and β flap interact with the σ factor, holding it in place, allowing it to recognise the -35 element and initiate transcription. Following σ factor acquisition and promotor binding, the DNA binding clamp binds to the +15-20 bp position and aids transcription initiation and the recognition of termination signals (Murakami *et al.*, 2002, Gusarov and Nudler, 1999). After transcription has initiated, the β lobes, flap, and β' lid relax, ejecting the σ factor. The β lobes, flap, and β' lid then re-contract, along with the β' main channel, zipper, and rudder, closing the RNA exit channel (Kuznedelov *et al.*, 2002a, Toulokhonov and Landick, 2003, Toulokhonov and Landick, 2006, Korzheva *et al.*, 2000). Then, during transcription elongation, the β lobes sequester the non-template DNA strand away from the template strand (Borukhov and Nudler, 2008, Korzheva *et al.*, 2000). The β flap also forms part of the RNA exit channel and interacts with the nascent RNA

transcript (Kuznedelov *et al.*, 2002b, Toulokhonov and Landick, 2003). As well as interacting with the σ factor, the β' lid, separates the DNA and RNA strands during transcription elongation (Toulokhonov and Landick, 2006). At the active site, the β' F(bridge)-helix and G(trigger)-loop ensure the correct NTP is incorporated into the nascent transcript; they also recognise pausing signals during transcription termination (Wang *et al.*, 2006).

1.2.2.2 RNAP α subunit

Whilst not involved with the transcription directly, the two α subunits have an important role in RNAP function. Each 329 amino acid, 37 kDa, subunit comprises two domains which interact with either the RNAP (NTD) or the UP element of the promoter (C terminal domain, CTD) (Gourse *et al.*, 2000). The NTD (residues 1-235) dimerises and is integral to the correct assembly of the β and β' subunits. The CTD (residues 249-329) binds to the minor grove of DNA at the UP elements using residue R265 (Igarashi *et al.*, 1991, Blatter *et al.*, 1994, Sutherland and Murakami, 2018). The N- and C-terminal domains are connected by a 13 amino acid flexible linker and can, independently, interact with transcription factors (TFs) (Jeon *et al.*, 1997).

1.2.2.3 RNAP ω subunit

The 91 amino acid ω subunit also aids assembly of RNAP by acting as a β' subunit chaperone and contacting the αNTD and αCTD (Minakhin *et al.*, 2001a, Zhang *et al.*, 1999). Interestingly, ω is the only non-essential part of RNAP as shown by deletion studies (Gentry and Burgess, 1989). Whilst non-essential, deletion of ω is not without consequence. A genome-wide

altered expression pattern has been seen in Δω strains, with RNAP interacting more frequently with alternative σ factors as well as changes to the supercoiling of DNA (Geertz *et al.*, 2011). It is thought that this genetic shift could be due to the interaction between ω and guanosine tetraphosphate (ppGpp) (Geertz *et al.*, 2011). The stress alarmone ppGpp is responsible for the stringent response (Irving *et al.*, 2021). The stringent response is a complex bacterial response to stresses such as nutrient depletion and heat shock. ppGpp, produced during the stringent response, binds to the interface between ω and β', and leads to a drastically altered transcription profile. The alarmone further modulates a wide set of cellular functions including virulence, nucleotide synthesis, transcription, and biofilms (Magnusson *et al.*, 2005, Ross *et al.*, 2013, Irving *et al.*, 2021).

1.2.3 σ factors – The housekeeping σ factor

As mentioned above, the RNAP core enzyme is catalytically competent but unable to initiate transcription without the binding of a σ factor (Borukhov and Nudler, 2008, Browning and Busby, 2004, Feklistov and Darst, 2011). σ factors are subdivided into families (the σ^{70} and σ^{54}) (Wosten, 1998). In *Escherichia coli*, σ^{70} is the housekeeping σ factor responsible for directing transcription of most genes involved with log-phase growth (Feklistov and Darst, 2011). Comprised of four domains connected by flexible linkers, σ^{70}, also called RpoD, uses each domain to recognise a specific element of the bacterial promoter (Figure 1.2). When bound to RNAP, the σ factor is positioned in a way that optimises contacts with the DNA (Borukhov and Nudler, 2008, Feklistov and Darst, 2011). Domain 1 recognises and binds to the discriminator region (-4 to -6) and is key in preventing DNA from entering the active site

of RNAP until the σ factor has recognised the promoter element, inducing a conformational shift (Mekler *et al.*, 2002). Domain 2 binds to the -10 element (-7 to -12) and unwinds DNA by chelating the A at position -11 and the T at position -7 when they are flipped out of the DNA base stack (Feklistov and Darst, 2011). Domain 2 also forms important contacts of the σ factor and RNAP by binding to the major docking site on β' (Borukhov and Nudler, 2008). Domain 3 binds to the extended -10 element (-14 to -17). Domain 4 binds to the -35 element (-30 to -35) using a helix-turn-helix (HTH) motif. Domains 3 and 4 are connected by a flexible linker, which interacts with the active site of the RNAP holoenzyme and aids the initiation of transcription before σ factor dissociation (Feklístov *et al.*, 2014).

1.2.4 σ-factors – The alternative σ factors

The alternative σ factors belong to the σ^{70} and σ^{54} families (Feklístov *et al.*, 2014, Wosten, 1998). Members of the σ^{70} family share structural characteristics, whilst those of the σ^{54} family share little to no sequence identity with the σ^{70} family (Wosten, 1998). The σ^{70} family is further subcategorised into groups 1 to 4 depending on the presence or absence of the domains they possess (Gruber and Gross, 2003, Paget and Helmann, 2003). Group 1 includes σ^{70} and any others that possess all four domains, including the 1.1 region of domain 1. Groups 2 and 3 are highly similar to group 1 and possess the same structures but are dispensable for growth under normal conditions. The group 4 σ factors contain only domains 2 and 4 and are termed the extracytoplasmic function family for their role in regulating the response to environmental conditions including infection of a host and metal toxicity (Österberg *et al.*, 2011, Gruber and Gross, 2003, Koo *et al.*, 2009). Unlike the σ^{70} family, the σ^{54} family recognise

promoter elements at positions -12 and -24 when forming a closed complex. Transcription initiation requires ATP-dependent activators to unwind the DNA (Yang *et al.*, 2015).

1.2.5 Transcription factors

Specific expression is often controlled by transcription factors (TFs), which bind the promoter region and modulate the interaction with RNAP (Griffith *et al.*, 2002, Browning and Busby, 2004). TFs have the ability to both positively and negatively regulate transcription (Pérez-Rueda and Collado-Vides, 2000). The process of activation or repression is complex and allows fine-tuning of the levels of transcription through the interaction of one or more TFs at promoters (Browning *et al.*, 2019). There are numerous ways a TF can influence transcription, with the mechanisms of activation discussed briefly in 1.2.5.1 and the mechanisms of repression in 1.2.5.2. The DNA sites bound by TFs are sometimes referred to as "boxes" or "operators" (Robison *et al.*, 1998, Pérez-Rueda and Collado-Vides, 2000).

Three quarters of TFs contain two domains, with one domain reserved for DNA binding whilst the other (termed regulatory domain) binds a ligand or is covalently modified (Madan Babu and Teichmann, 2003, Browning *et al.*, 2019). A good example is the modulation of the Lac repressor by binding of allolactose, leading to derepression of the *lac* operon (Müller-Hill, 2011). An alternative method of regulating TFs is as part of a two-component system (TCS). Here, the TF is activated by phosphorylation via a sensor kinase protein in response to a specific stimulus. This is typified by the EnvZ/OmpR system, in which the osmosensory sensor

kinase EnvZ phosphorylates OmpR to regulate porin expression in response to changing osmolarity (Cai and Inouye, 2002). TFs can also be controlled via expression level or proteolysis. An example is the oxidative stress response, controlled by SoxS. Normally repressed by SoxR and quickly degraded by Lon protease, SoxS levels are kept low in the absence of oxidative stress. When the cell encounters oxidative stress, SoxR flips into an activator and greatly increases transcription of *soxS* and, therefore, the intracellular concentration of SoxS (Duval and Lister, 2013). In a similar mechanism, TFs can be sequestered by a regulatory protein, preventing the TF from accessing the DNA and, therefore, controlling TF activity (Browning and Busby, 2004, Demple, 1996, Plumbridge, 2002, Stock *et al.*, 2000a). For example, in the phosphotransferase system, the presence of glucose causes derepression by Mlc, which is sequestered by binding to dephosphorylated Enzyme IIB (Plumbridge, 2002).

1.2.5.1 Mechanisms of transcriptional activation

Activation is used to increase levels of transcription at a given promoter, under specific conditions. Often, this is by enhancing recruitment of RNAP. Whilst the methods TFs employ to activate transcription can be complex, only simple cases are discussed here (Browning and Busby, 2004, Browning *et al.*, 2019). There are four methods of activation: class I activation, class II activation, activation by conformational change, and anti-repression. They are shown in Figure 1.4. Class I and class II activation is similar. In class I activation, the transcription factor binds upstream of the core promoter elements and recruits RNAP by interacting the α-CTD (Browning and Busby, 2016). In class II activation, the TF overlaps the -35 element and

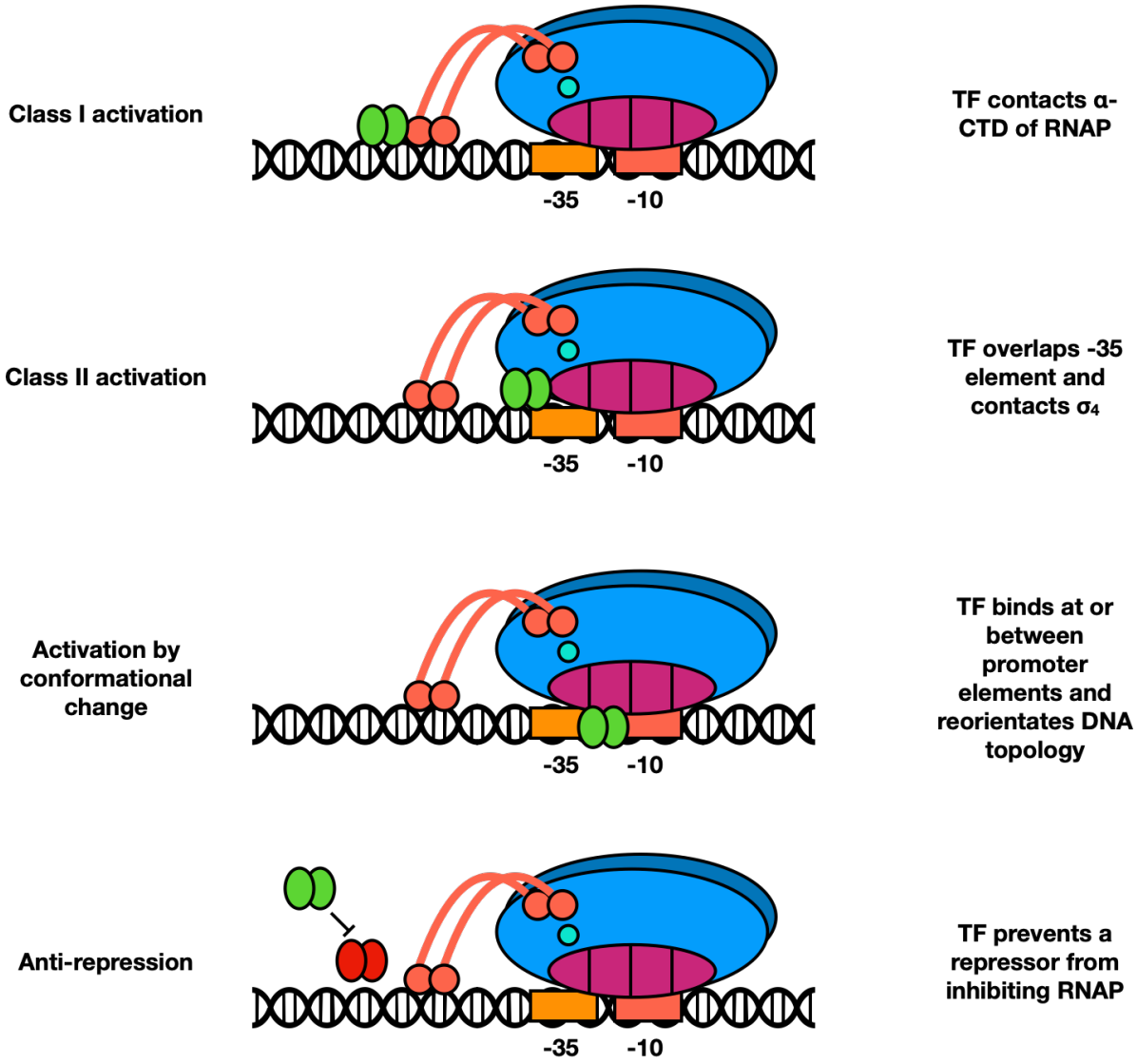


Figure 1.4 Mechanisms of activation by transcription factors. The mechanisms of activation by transcription factors. Green ovals indicate an activator and red ovals indicate a repressor. Inhibition is shown by a T shape. Promoter elements are indicated as boxes. Class I activation occurs when the TF binds to its cognate binding site and interacts with the α-CTD of RNAP, recruiting the RNAP to the promoter. Class II activation occurs in a similar mechanism to Class I activators but the TF binds to a site overlapping the -35 element and contacts σ^{70} domain 4. Activation by conformational change occurs when the TF binds to the DNA between the -10 and -35 promoter elements and reorientates the DNA to bring both promoter elements in line

with one another, allowing RNAP holoenzyme to bind. Anti-repression occurs when one TF prevents a repressor from inhibiting RNAP. Figure reproduced and modified from Browning and Busby (2004) and Browning *et al.* (2019).

contacts different combinations or domain 4 of the σ factor, α-NTD of RNAP and/or α-CTD. A class I activator may also be present allowing synergistic activation in response to multiple signals (Browning and Busby, 2016). In the cases where the -10 and -35 promoter elements are misaligned a TF, bound between the promoter elements, can induce a topological shift. This realigning of the promoter elements facilitates RNAP binding and transcription (Philips *et al.*, 2015, Browning and Busby, 2016). Activators can also counter the effects of a repressor. This is termed anti-repression and there are three mechanisms. First, the activator binds to the repressor, preventing it from binding DNA. Second, activators block DNA binding by the repressor. Third, the activator induces a conformational change that interrupts the repressor's mechanism of action (Smits *et al.*, 2007).

1.2.5.2 Mechanisms of transcriptional repression

Repression decreases transcription in response to specific stimuli. Mechanisms include steric hindrance, DNA looping, modulation of activators, and RNAP locking (Browning and Busby, 2004, Browning *et al.*, 2019). Steric hindrance occurs when the TF binds to an operator overlapping or close to the essential promoter elements. Thus, repressor binding stops RNAP:promoter interactions. Repression by looping also prevents the RNAP interacting with the promoter region. Looping occurs if TFs bind distal operators and interact with one another; a loop is formed by the intervening DNA (Browning and Busby, 2016, Browning *et al.*, 2019). Modulation of an activator involves a repressor that contacts an activator and disrupts normal function. These TFs are sometimes called anti-activators and, whilst they don't directly inhibit the RNAP-promoter binding, their method of

repression is similar to steric hinderance (Browning and Busby, 2016). The mechanisms above prevent RNAP from accessing the promoter region. The locking mechanism allows RNAP to be recruited and bind the promoter elements. However, RNAP is locked in place by the repressing TF and transcription cannot occur (Browning *et al.*, 2019). Mechanisms of repression are shown in Figure 1.5.

1.2.5.3 The major families of transcription factors

Based on structural similarities, TFs are grouped into the cyclic-AMP receptor protein, LysR, OmpR, LuxR, LacI, and AraC families. Of relevance to this study is the AraC family, which is discussed in 1.2.5.4. Other families are discussed in brief below.

The cyclic-AMP receptor protein (CRP) family

This family of TFs possess both activators and repressors and is exemplified by CRP. Generally, family members contain two domains, an N-terminal domain which responds to a stimulus and a C-terminal DNA-binding HTH domain (Pérez-Rueda and Collado-Vides, 2000). CRP, when activated, regulates cellular processes such as sugar and amino acid metabolism, protein folding, transport processes and the virulence traits of toxin production and pilus synthesis (Korner *et al.*, 2003). Other members of this family respond to stress signals such as temperature (PrfA), oxidative stress and nitrogen fixation (Fnr), and also regulate homeostasis during stationary phase (YeiL) (Korner *et al.*, 2003).

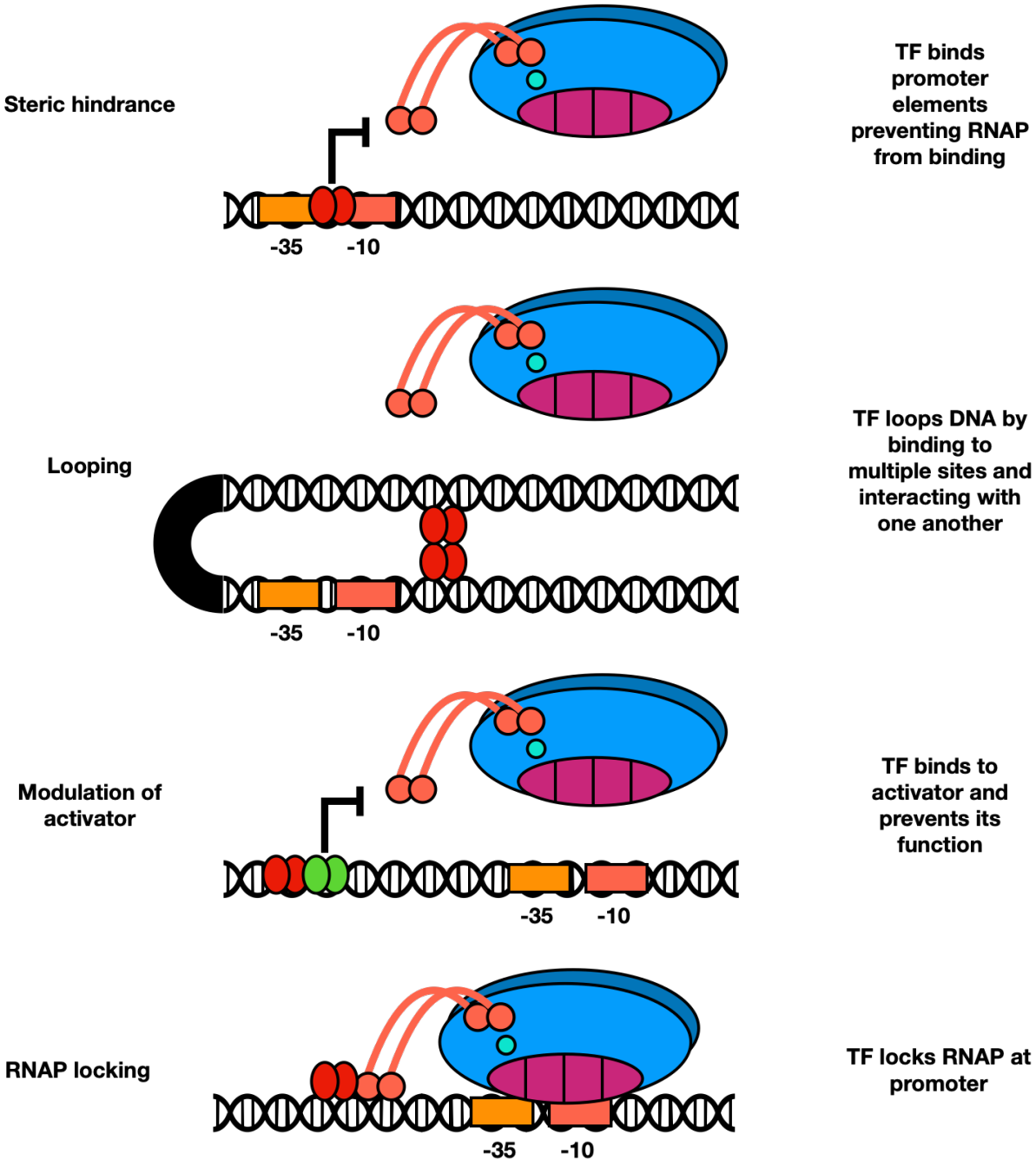


Figure 1.5 Mechanisms of repression by transcription factors. The mechanisms of repression by transcription factors. Red ovals indicate a repressor and green ovals indicate an activator. Inhibition is shown by a T shape. Promoter elements are indicated as boxes. Steric hindrance (top) occurs when the TF binds the DNA at the promoter elements, preventing RNAP from binding. Looping (middle) occurs when the TFs bind to distal sites, interact with each other,

and induce looping of the DNA, preventing RNAP from accessing the promoter elements. During repression by modulation of an activator (bottom), the TF binds to an activator and prevents recruitment of RNAP. RNAP locking occurs when RNAP binds to the promoter but is locked in place by the TF, preventing transcription. Reproduced and modified from Browning and Busby (2004) and Browning *et al.* (2019).

The LysR family

The LysR family is the most abundant family in prokaryotes (Schell, 1993, Maddocks and Oyston, 2008). Members bind DNA using an N-terminal HTH domain. The C-terminal domain binds respective ligands. Like the CRP family, members of the LysR family can be both activators and repressors. Whilst structurally similar, members of this family regulate a range of processes within the cell. These include oxidative stress (OxyR), biofilm formation (BsrA), antimicrobial resistance (MexT), quorum sensing and virulence (MvfR) (Maddocks and Oyston, 2008, Déziel *et al.*, 2005, Ochsner *et al.*, 2000, Yang *et al.*, 2019, Sobel *et al.*, 2005).

The OmpR family

The OmpR family of TFs are primarily the response regulators of TCSs. In these family members, the N-terminal domain of the response regulator is phosphorylated by the sensor kinase with the C-terminus containing the winged HTH DNA binding domain (Kenney, 2002). The prototypical example is the EnvZ/OmpR TCS, which regulates outer membrane porin expression (Martinez-Hackert and Stock, 1997). OmpR has also been shown to have a global regulatory effect and, in addition to outer membrane permeability, regulates fatty acid metabolism, motility, and biofilm formation (Brzóstkowska *et al.*, 2012).

The LuxR family

Another system in which TFs play a vital role is quorum sensing (QS). The LuxR family of TFs, themselves a subgroup of the TetR superfamily, have a similar structure to the CRP family and

are subdivided into two classes depending on activation mechanism (Zeng and Xie, 2011). LuxR family members involved in two component systems are activated by phosphorylation by their cognate response regulator. However, for LuxR family members involved in QS, their N-terminal domain responds to N-acyl-homoserine lactones, the effector molecules of QS, and the C-terminal domain contains the DNA binding HTH domain (Zeng and Xie, 2011). In *Vibrio cholerae*, the LuxR family member HapR is involved in a complex regulatory network, coupling the inhibition of biofilm formation and production of virulence factors to population density (Haycocks *et al.*, 2019).

The LacI family

Described in brief previously, the *lac* repressor is a member of the LacI family. These TFs bind inverted repeats of DNA as dimers, using HTH domains, and are modulated by ligand binding. The binding of a ligand to the regulatory domain induces a conformational change in shape, altering the DNA binding ability of the TF (Müller-Hill, 2011). Most of this family are involved in the sensing and regulation of sugar and carbohydrate utilisation genes (Ravcheev *et al.*, 2014).

1.2.5.4 AraC family of transcriptional regulators

The AraC family, so named after the regulator of the L-arabinose operon, share a 100 amino acid conserved region which forms two HTH motifs. This study focusses on 4 members of the AraC family, MarA, SoxS, Rob, and RamA, which are discussed individually in section 1.5. The

AraC family members generally regulate genes involved in carbon metabolism, pathogenesis, and stress response. This regulation is achieved through activation (both Class I and II) or repression (Martin and Rosner, 2001). For most members of the AraC family, each of the HTH motifs insert in adjacent major grooves of DNA, forming base-specific hydrogen bonds with specific nucleotides. This induces bending of the DNA by 35° (Martin and Rosner, 2001). This observation is true for MarA, SoxS, and RamA, but for Rob, only one HTH motif is used to bind the major groove. The other helix-turn-helix binds to the DNA backbone, decreasing the reliance on the MarA binding site consensus sequence and resulting in unbent DNA (Kwon *et al.*, 2000). The two HTH motifs do not share a conserved amino acid sequence and, therefore, these regulators bind to non-symmetrical sites (Gallegos *et al.*, 1997).

Generally, the N-terminal region of AraC members is used for oligomerisation or cofactor binding, but MarA, SoxS, and RamA lack an N-terminal domain (Gallegos *et al.*, 1997, Soisson *et al.*, 1997). In the example of AraC, a dimer binds to two sites ($araI_1$ and $araO_2$) separated by around 200 bps and represses transcription of the *araBAD* operon. The C-terminal HTH domains of each monomer bind the DNA sites, and the N-terminal domains interact with one another, leading to looping of the DNA. Each monomer of AraC can bind one molecule of arabinose in its N-terminal domain, causing a conformational change in shape. The AraC dimer now binds to the $araI_1$ and $araI_2$ half sites and promotes transcription (Gallegos *et al.*, 1997).

1.3 The process of transcription

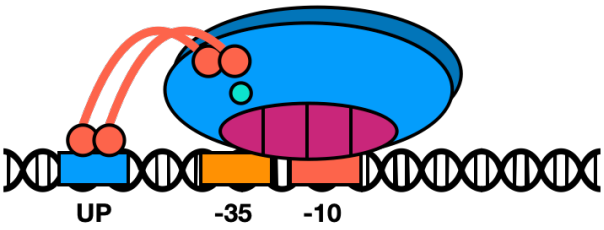
1.3.1 Promoter recognition and initiation

The RNAP holoenzyme locates the promoter region by sliding along the DNA and, once bound, the initial state is termed the closed complex (Figure 1.6A) (Sakata-Sogawa and Shimamoto, 2004). Whilst double stranded DNA is thermodynamically stable, base pairings are transient and have an average lifetime of milliseconds (Gueron and Leroy, 1995). This is in equilibrium with bases that are flipped out from the DNA (Feklistov and Darst, 2011). There is debate as to whether σ factors actively disrupt the base pairing at position -11 A or whether they simply catch the bases when they spontaneously flip out (Feklistov and Darst, 2011). In either instance, the transition into an open complex (Figure 1.6B) is mediated by σ factors and starts with a transcription bubble covering bases -11 to -7 before expanding downstream (Borukhov and Nudler, 2008, Feklistov and Darst, 2011, Gueron and Leroy, 1995, Murakami and Darst, 2003). Following the formation of the open complex, the +20 bp region enters into the DNA binding clamp of RNAP (Murakami and Darst, 2003). Now that the promoter DNA has been unwound to provide access to the template strand, and the downstream DNA clamped, the initiation of transcription can occur (Figure 1.6C).

The enzymatic process of transcription starts with entry of NTP substrates into the RNAP secondary channel. Upon correct binding of an NTP to the +1 and +2 positions, a phosphodiester bond forms between the first two nucleotides (Borukhov and Nudler, 2008). During this process, termed abortive initiation (Figure 1.6D), free energy is built up within the RNAP as DNA becomes scrunched up; this happens because the σ subunit is still bound the promoter preventing translocation. This free energy build up is usually not enough to relieve the contacts σ makes to the promoter and so the transcript is aborted, ejected, and the

A **Promoter recognition**

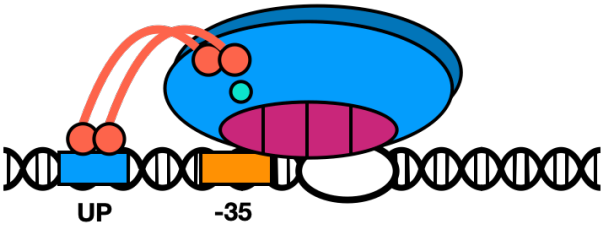
Closed complex



RNAP holoenzyme recognises and binds to the promoter

B

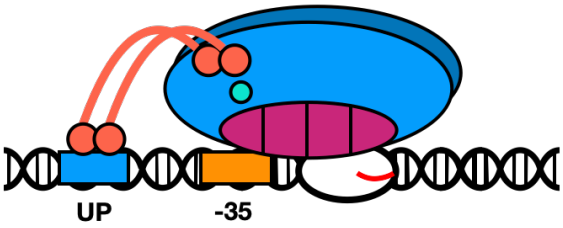
Formation of the open complex



DNA duplex unwinds with the formation of the open complex

C **Transcription initiation**

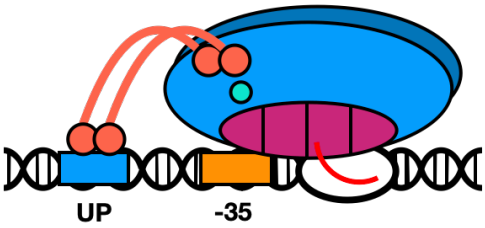
Transcription initiation



Small RNA transcripts are produced

D

Abortive initiation and DNA scrunching



Upstream DNA becomes scrunched as transcription occurs

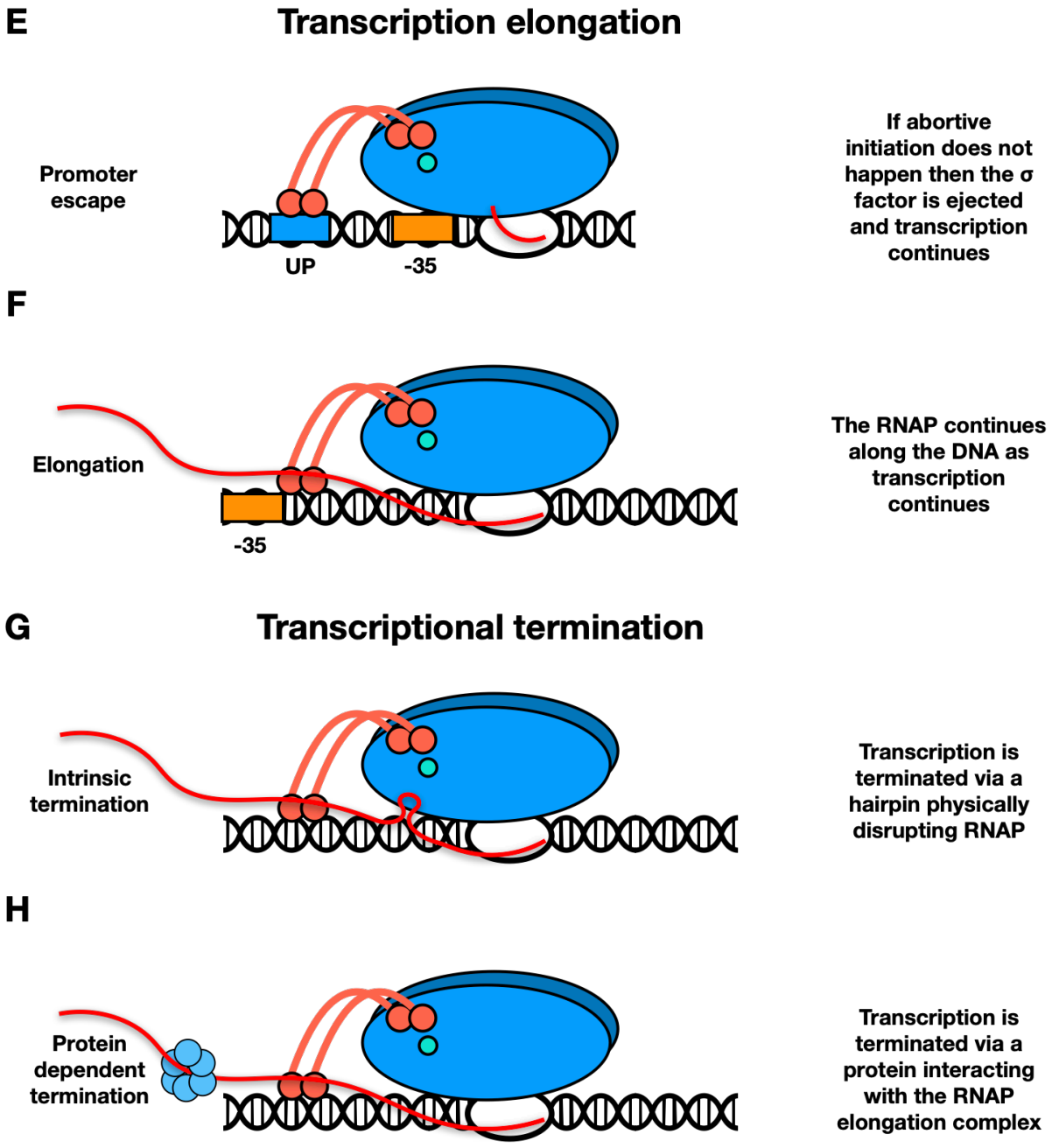


Figure 1.6 The process of transcription in *Escherichia coli*. The basic steps of bacterial transcription are presented above in a cartoon schematic, covering promoter recognition (A, B), transcription initiation (C, D), transcription elongation (E, F), and transcriptional termination (G, H). The structural composition of RNAP is shown with the β and β' subunits shown as dark and light blue respectively. The two α subunits are shown as orange circles

connected by flexible linkers. The ω subunit is shown as a green circle. DNA is shown as a double helix and the promoter elements shown by coloured boxes (Orange, yellow and blue for the -10, -35 and UP elements respectively). mRNA is shown as a red line and for protein dependent transcriptional termination, Rho is shown as a ring of light blue circles. Reproduced and modified from Browning and Busby (2004) and Browning and Busby (2016).

process restarts. But as the transcript reaches ~10 bp the free energy build can surpass the threshold set by the σ factor, leading to the dissociation of the RNAP from the promoter and ejection of the σ factor, forming the elongation complex (Borukhov and Nudler, 2008, Browning and Busby, 2016, Mooney *et al.*, 2005, Revyakin *et al.*, 2006) (Figure 1.6E).

1.3.2 Elongation and termination

During elongation (Figure 1.6F), NTP entry through the secondary channel is coordinated by the F(bridge)-helix and G(trigger)-loop (Vassylyev *et al.*, 2007). NTPs complementary to the template are added to the growing transcript with the formation of a phosphodiester bond between the 3' OH of the RNA strand and the α-PO_4 of the NTP substrate (Laptenko *et al.*, 2003). Once this reaction has taken place, the F(bridge)-helix and G(trigger)-loop stop interacting with the incorporated NTP and, instead, interact with the next NTP. This conformational change leads to the ratchet mechanism by which the RNAP translocates along the DNA (Bar-Nahum *et al.*, 2005, Korzheva *et al.*, 2000, Kuznedelov *et al.*, 2002a, Kuznedelov *et al.*, 2002b, Toulokhonov and Landick, 2003, Toulokhonov and Landick, 2006). In the next cycle of this ratchet mechanism, RNAP translocates forwards to provide access to the next base on the template strand. Factors including NusA and NusG control elongation and are also involved in pausing and terminating transcription (Schmidt and Chamberlin, 1987, Yakhnin *et al.*, 2016). NusA has multiple effects on transcription and enhances transcriptional pausing by stabilising interactions between RNAP and nascent hairpins. Conversely, NusG enhances the rate of transcription and interacts with the transcriptional termination factor Rho (Mooney *et al.*, 2009, Artsimovitch and Landick, 2000).

Uninterrupted, transcription elongation occurs until an intrinsic or factor dependent terminator is encountered (Ray-Soni *et al.*, 2016). Intrinsic termination occurs through a GC-rich RNA hairpin structure (Figure 1.6G). This forms in the RNA exit channel, leading to physical disruption of the elongation complex (Ray-Soni *et al.*, 2016, Larson *et al.*, 2008). Factor dependent termination involves Rho or Mfd (Roberts, 2019) (Figure 1.6H). The ATP-dependent RNA translocase Rho binds the nascent RNA transcript as a hexamer and translocates in the 3' to 5' direction until it reaches the elongation complex. The precise mechanism of termination by Rho is undetermined but it is thought that Rho either pulls the RNA transcript out of the RNAP or it pushes the RNAP destabilising the transcription bubble (Park and Roberts, 2006, Ray-Soni *et al.*, 2016, Richardson, 2002, Roberts, 2019). Mfd-dependent termination utilises the ATP-dependent DNA translocase Mfd, which binds to both DNA and RNAP simultaneously (Roberts and Park, 2004). Mfd uses ATP hydrolysis to push RNAP off the DNA by forward translocation (Ray-Soni *et al.*, 2016, Roberts and Park, 2004, Roberts, 2019).

1.4 Biofilms

1.4.1 Structure and function

Bacteria can exhibit motile planktonic lifestyles or exist as sessile biofilm forming cells (Steenackers *et al.*, 2012). Biofilms have a large impact on human civilisation and commonly cause contamination of infrastructure, medical devices, and foodstuffs, as well as being the origin of 80% of human infections (Srey *et al.*, 2013, Davies, 2003, Vishwakarma, 2020). Biofilms can form on both biotic and abiotic surfaces and are an important factor in the ability

of bacteria to colonise almost all niches on the planet (Berlanga and Guerrero, 2016, Baquero *et al.*, 2021). Biofilms can be formed by a single species or by multiple species, but both types show the same general macro structure (Davey and O'toole, 2000). Structural components of the biofilm matrix can vary depending on the substrate the biofilm is attached to.

Generally, there are five distinct steps in the lifecycle of a biofilm (Figure 1.7) (Steenackers *et al.*, 2012, Srey *et al.*, 2013, P. Stoodley *et al.*, 2002). The first step in biofilm formation is the initial, reversible, attachment of planktonic bacteria to a substrate and takes between 5 and 30 seconds (Mittelman, 1998). At this stage, the bacteria are held in place by physiochemical forces between cells and the surface and are still capable of movement, as there is not a large amount of exopolysaccharide (EPS) production (Chmielewski and Frank, 2003, Ferreira *et al.*, 2010, O'Toole and Kolter, 1998). This is depicted in section 1 of Figure 1.7. The transition to a state of irreversible attachment (Figure 1.7 section 2) occurs as EPS production increases and bacteria bond to both the surface and EPS; the time it taken can vary between 20 minutes and 4 hours (P. Stoodley *et al.*, 2002, Chmielewski and Frank, 2003). As biofilm bacteria divide, and continue to produce EPS, a microcolony forms (Figure 1.7 section 3). The growth of the microcolony is also supplemented by the recruitment of surrounding planktonic cells through quorum sensing (Srey *et al.*, 2013, Chmielewski and Frank, 2003, McLean *et al.*, 1997).

If the conditions are favourable, the biofilm can stabilise further to form a mature colony (Figure 1.7 section 4). The structure of mature biofilms can be either flat monolayers or

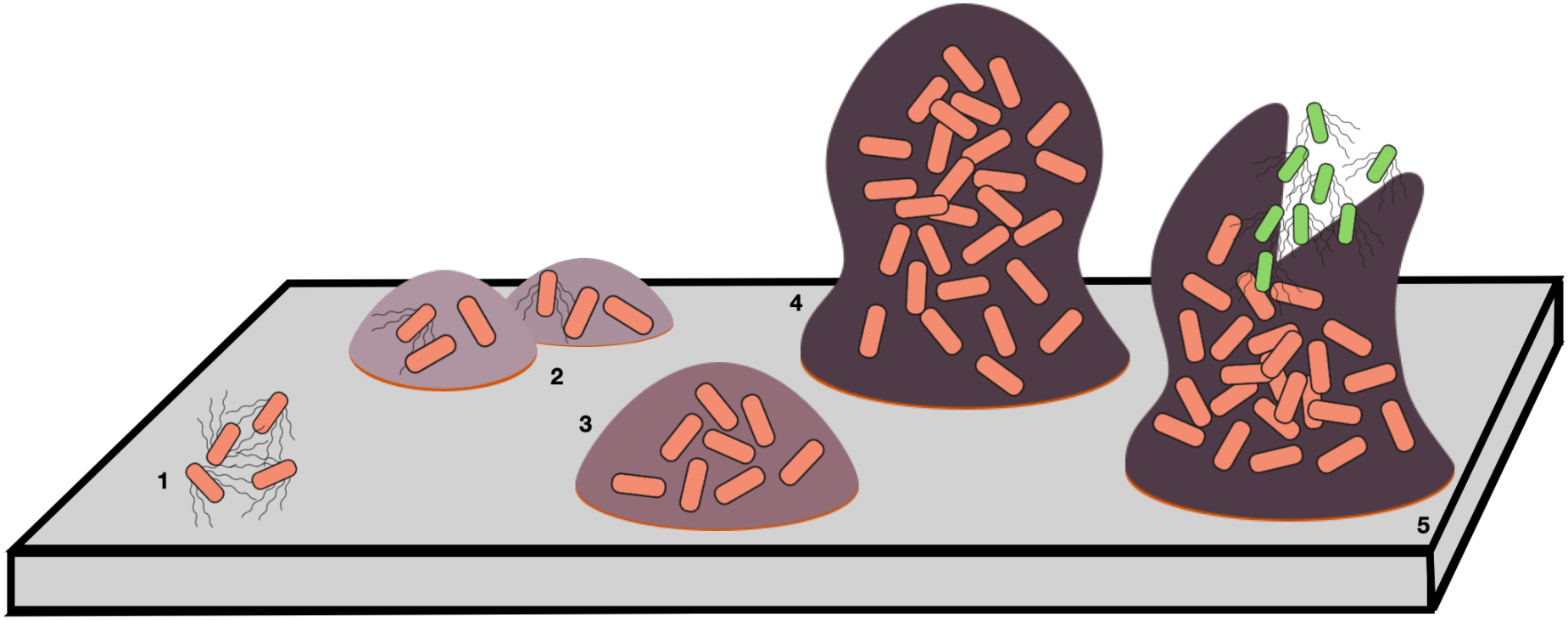


Figure 1.7 The stages of biofilm development on a solid substrate. The stages of biofilm development shown as a cartoon representation. 1 shows the initial reversible attachment to a substrate, 2 shows irreversible attachment and the production of EPS, 3 shows the development of microcolony formation, 4 shows a mature biofilm, and 5 shows the final step in the lifecycle of a biofilm - dispersion. Reproduced and modified from P. Stoodley *et al.* (2002).

balloon into a mushroom shape connected to the substrate via a stalk (Chmielewski and Frank, 2003). In both forms of mature biofilm, the EPS matrix contains water channels to allow nutrient flow. These structures take 10 days or more to fully develop (Davey and O'toole, 2000, P. Stoodley *et al.*, 2002). The final stage of biofilm formation is dispersion (Figure 1.7 into the environment (Sauer *et al.*, 2002, Srey *et al.*, 2013). Dispersion can be caused by internal and external factors. External factors include increased shear forces, that detach the biofilm, or fewer nutrients in the environment, prompting dispersal to colonise a new niche. Internal factors include the expression of EPS degrading enzymes (Stoodley *et al.*, 2002, Sauer *et al.*, 2002, O'Toole *et al.*, 2000, Kaplan *et al.*, 2004, Srey *et al.*, 2013).

1.4.2 Regulation of biofilms in *Salmonella* by CsgD

The best studied *Salmonella* biofilm is red, dry, and rough (rdar) in appearance. Such biofilms are visible on agar containing Congo Red (Römling *et al.*, 2007). This biofilm matrix is primarily formed of the EPS compound cellulose and various other EPSs and proteinaceous components including curli fibres, which make up 85 % of the extracellular matrix (Steenackers *et al.*, 2012, Barnhart and Chapman, 2006, Tursi *et al.*, 2020). The curli fimbriae, encoded by the *csgBAC-csgDEFG* divergent operons, are involved in the initial interactions between the cell and the substrate and also in anchoring the cell to the extracellular matrix of the biofilm and in cell-cell interactions (White *et al.*, 2003). Curli fibres, therefore, play a vital role in the formation and maintenance of biofilms in *Salmonella*.

Curli expression is controlled by the transcription factor *csgD*; a master regulator of biofilm formation (Gerstel and Römling, 2003, Fàbrega and Vila, 2013, Grantcharova *et al.*, 2010). A LuxR superfamily member, CsgD is a transcriptional regulator under the control of a complex regulatory network. There are 14 TFs known to regulate *csgD* expression (Figure 1.8), and a further 48 TFs are proposed to interact with the *csgD* promoter (Ogasawara *et al.*, 2020). The CsgD protein is usually expressed in the late exponential to early stationary phases (Ogasawara *et al.*, 2010a, Gerstel and Römling, 2003, Steenackers *et al.*, 2012). CsgD also positively regulates the production of the other main component of biofilms, cellulose, via the upregulation of *adrA*; which, in turn, upregulates the second messenger cyclic di-GMP (Zakikhany *et al.*, 2010). The N-terminal domain of CsgD forms the receiver for the response regulator and contains a conserved aspartate residue at amino acid 59 (D59), which is phosphorylated by CsgD's cognate histidine kinase (Zakikhany *et al.*, 2010). Once phosphorylated, CsgD undergoes a conformational change which facilitates binding to target DNA via a C-terminal, LuxR-like, helix-turn-helix domain (Stock *et al.*, 2000b, Zakikhany *et al.*, 2010).

The complex regulatory network controlling *csgD* (Figure 1.8) transcription allows the fine-tuning of *csgD* expression in response to environmental cues (Römling *et al.*, 1998, Gerstel and Römling, 2003, Steenackers *et al.*, 2012). Factors that influence *csgD* expression include nutrient concentration, temperature, pH, osmolarity, oxygen tension, and ethanol (Gerstel and Römling, 2003). Low nutrient availability, low temperatures (28 ^{o}C), alkaline pH, low osmolarity, low concentrations of ethanol, and microaerophilic conditions are associated with *csgD* expression (Gerstel and Römling, 2003). The intergenic region between the *csgBAC*-

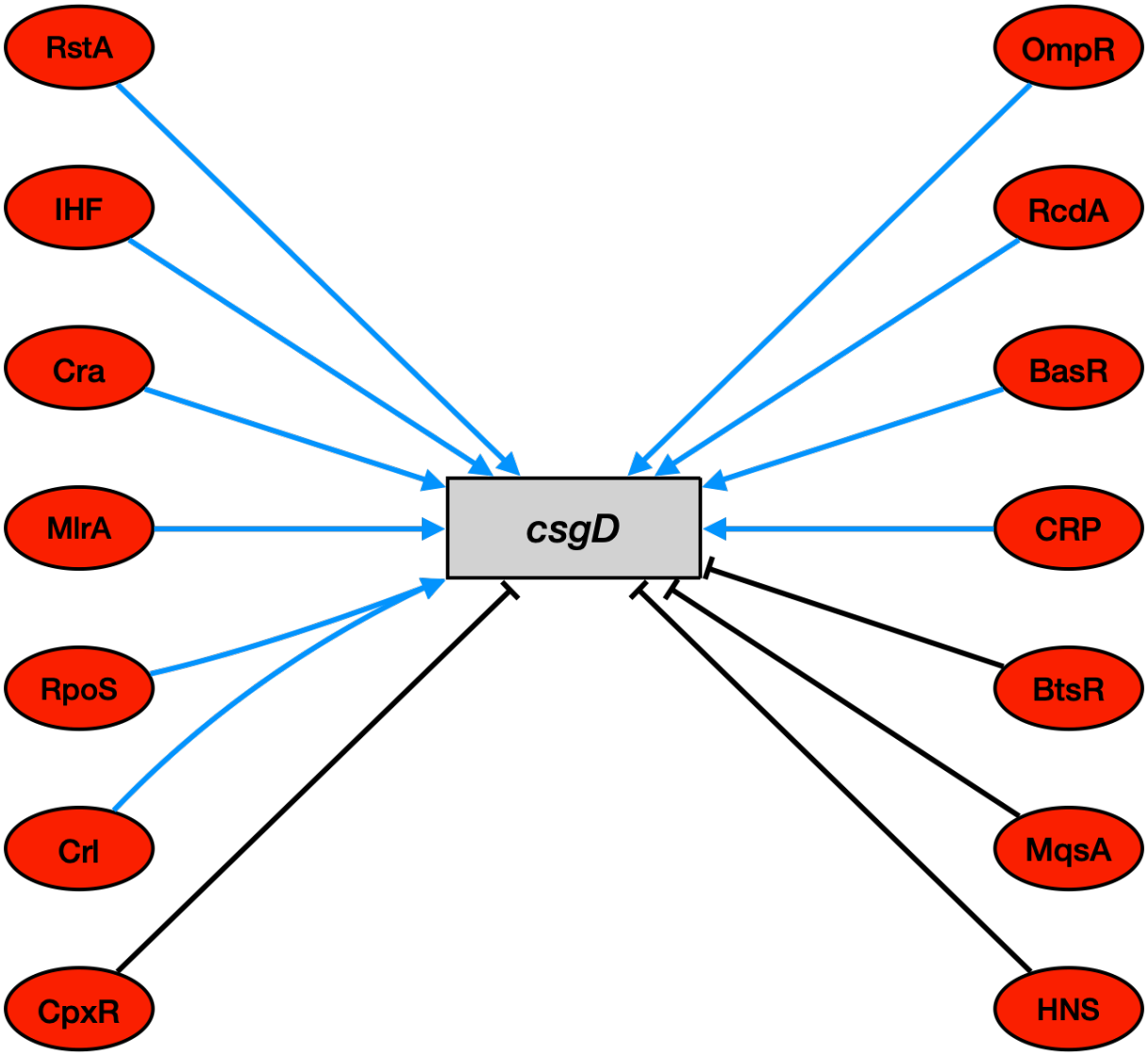


Figure 1.8 Simplified transcription factor-mediated regulation of *csgD*. A simplified schematic showing the TF-mediated regulation of *csgD*. Blue arrows represent activation and Black flat-headed arrows represent repression. Interactions between each of the regulators is not shown for clarity. Both RpoS and Crl contribute to *csgD* activation individually, but activation is strongest when they function together (shown by connecting arrows). Figure compiled using data from Ogasawara *et al.* (2020), Ogasawara *et al.* (2019), Ogasawara *et al.* (2012), Ogasawara *et al.* (2010a), Ogasawara *et al.* (2011), Shimada *et al.* (2012), Soo and Wood (2013), Steenackers *et al.* (2012).

csgDEFG divergent operons is more curved and less flexible than average (Prigent-Combaret *et al.*, 2001). Transcription is controlled by regulators including MlrA, H-NS, IHF, OmpR, RpoS, CpxR, and Crl in response to the conditions mentioned above (Steenackers *et al.*, 2012, Gerstel and Römling, 2003). Of note, MlrA, positively regulates *csgD* expression and is itself upregulated by RpoS (Brown *et al.*, 2001). MlrA binds to a 11 bp long inverted repeat with a 12 bp spacing between the repeats: AAAGTTGTACA(12N)TGCACAATTTT (Ogasawara *et al.*, 2010a). This sequence is found ~120 bp upstream from the *csgD* transcription start site (Ogasawara *et al.*, 2010a).

1.5 Salmonella enterica

1.5.1 Overview

Synonymous with food poisoning, the Gram-negative, facultatively anaerobic, bacillus shaped *Salmonella* species are members of the Enterobacteriaceae family and are closely related to *E. coli* (Rowley *et al.*, 2012, Su and Chiu, 2007). The first serovar identified, *Salmonella enterica* serovar Choleraesuis, was originally isolated from an outbreak of swine cholera and termed *Bacillus choleraesuis* by Daniel Salmon and Theobald Smith; however, the term *Salmonella* was proposed in 1900 to honour Daniel Salmon's achievements (Crump and Wain, 2017). The genus *Salmonella* is subdivided into two species, *bongori and enterica*, with the species *enterica* further divided into six subspecies: *enterica* (subspecies I), *salamae* (subspecies II), *arizonae* (subspecies IIIa), *diarizonae* (subspecies IIIb), *houtenae* (subspecies IV), and *indica* (subspecies VI) (Brenner *et al.*, 2000). *Salmonella* are further categorised into serovars, of which there are 2,600, based on the composition of their O and H antigens (Brenner *et al.*,

2000, Fookes *et al.*, 2011, Reeves *et al.*, 1989, Gal-Mor *et al.*, 2014). This study uses *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344, abbreviated to SL1344. SL1344 is a histidine auxotroph of the parental strain ST4/74 and is considered a good model for *Salmonella* research (Wray and Sojka, 1978, Kroger *et al.*, 2012). Following traditional nomenclature and notation, serovars of *Salmonella* species will be written as species and serovar, for example *Salmonella enterica* serovar Typhi will be written as *S.* Typhi.

1.5.2 Infection

Salmonella enterica cause typhoidal or non-typhoidal infection (Crump and Mintz, 2010, Fierer and Guiney, 2001). The 2017 Global Burden of Disease study estimated that non-typhoidal Salmonella infections caused 353,000 million human cases and 77,500 deaths; in the same period there were an estimated 14.3 million cases of typhoid and paratyphoid and 135,900 deaths (Stanaway *et al.*, 2019a, Stanaway *et al.*, 2019b). In England there were 80,000 laboratory reports of *Salmonella* infections between 2010 and 2019, with the serovars Enteritidis, Typhimurium, and Newport being the most common (UKHSA, 2021). In the case of typhoidal infections, about 3,900 cases were reported in England, Wales, and Northern Ireland between 2008 and 2017; 55% of which were caused by *S.* Typhi (PHE, 2018). Non-typhoidal Salmonellosis generally presents as a self-limiting gastroenteritis. Typhoidal infections cause severe systemic infection with escape of *Salmonella* from the gastrointestinal environment into the blood stream and organs (Crump *et al.*, 2004, Fierer and Guiney, 2001, Su and Chiu, 2007). Certain *Salmonella* serovars have evolved the ability to leave the intestine to cause systemic disease. This has led to the development of host-adapted serovars which

sacrifice the potential to infect multiple hosts for longevity within a single host (Tanner and Kingsley, 2018). Such host-adapted serovars, *S.* Typhi and *S.* Paratyphi, are now restricted to humans. Similarly, *S.* Choleraesuis and *S.* Dublin cause systemic infections in swine and cattle but a non-typhoidal infection in humans (Gal-Mor *et al.*, 2014). Recently, there has been an increase in invasive non-typhoidal salmonellosis (iNTS) caused by some strains of non-typhoidal serotypes. Of note is *S*. Typhimurium strain ST313, which has become the dominant invasive isolate in sub-Saharan Africa (Kurtz *et al.*, 2017, Balasubramanian *et al.*, 2019). These infections present similarly to typhoidal infection and are commonly associated with comorbidities such as HIV and malaria (Kurtz *et al.*, 2017).

Salmonella can modulate its metabolism to mirror the environment of its host, avoiding competition (Taylor and Winter, 2020). This ability allows *Salmonella* to survive in both animal and plant hosts, furthering its potential dissemination and transmissibility (Schikora *et al.*, 2012). The bacterium is usually acquired by a faecal-oral route, involving contaminated meat, dairy, or salad products (Hu *et al.*, 2018, Álvarez-Ordóñez *et al.*, 2012). Once *Salmonella* enters the intestinal lumen infection is established. The route to the intestine takes the *Salmonella* through the stomach. This triggers the bacterium's acid tolerance response. As a result, the intracellular pH of *Salmonella* is shifted, acid shock proteins are induced, and membrane composition is altered. This allows *Salmonella* to pass into the intestinal tract (Álvarez-Ordóñez *et al.*, 2012). Following adhesion to the intestinal mucosa, the path of infection splits depending on whether the strain is non-typhoidal or not (Manuela *et al.*, 2008).

During initial colonisation of the gut, *S.* Typhimurium and *S*. Typhi utilise a type III secretion system (T3SS) to invade the epithelium. This instigates an inflammatory response caused by the translocation of effector proteins and virulence factors (Tanner and Kingsley, 2018). In a non-typhoidal infection, this response is aggravated further by the long O antigens and the peritrichous flagella activating the complement system. Inflammation disrupts the commensal environment between fermenting gut microbiota and the intestinal epithelium. This disruption leads to recruitment of neutrophils, releasing reactive oxygen and reactive nitrogen species (ROS and RNS respectively). *Salmonella* can resist oxidative stress, allowing them to survive the neutrophil attacks and respire in this environment. Hence *Salmonella* outcompete other gut flora and proliferate (Taylor and Winter, 2020, Fàbrega and Vila, 2013, Tanner and Kingsley, 2018).

Dissemination beyond the gut, and establishment of systemic infection by *S.* Typhi, is achieved by downregulation of T3SSs and flagella expression. Further to this, O antigen lengths are altered, exopolysaccharides shield the LPS layer from recognition by the host complement system, and the typhoid toxin is produced (Tanner and Kingsley, 2018). Systemic infection is mediated by anti-inflammatory macrophages, recruited to the site of inflammation within the gut, that carry the bacterium around the body (Manuela *et al.*, 2008, Johnson *et al.*, 2018, Kurtz *et al.*, 2017, García-Gil *et al.*, 2018). Compared to non-typhoidal serovars, *S*. Typhi-induced inflammation promotes a weaker pro-inflammatory response from the host; in part because the Vi exopolysaccharide capsule prevents detection of the *Salmonella* pathogen-associated molecular patterns (PAMPs), the expression of the typhoid toxin, and the

production of specific effector proteins from the *Salmonella* pathogenicity island (SPI) 2 (Johnson *et al.*, 2018, Fàbrega and Vila, 2013, Tanner and Kingsley, 2018). Resistance to phagocytosis and survival within macrophages in a *Salmonella* containing vesicle is required for effective dissemination to other organs (Johnson *et al.*, 2018, Fàbrega and Vila, 2013, Tanner and Kingsley, 2018, Gal-Mor *et al.*, 2014).

Salmonella are able to form biofilms on stainless steel, rubber, glass, plastics, plants, and epithelial cells. This aids *Salmonella*'s ability to persist in the environment and facilitates transmission between hosts in the case of non-typhoidal, non-host restricted, serovars (Steenackers *et al.*, 2012). The ability of *Salmonella* to survive on plants has been the cause of numerous outbreaks (Berger *et al.*, 2010). Within hosts, *Salmonella* can form biofilms on epithelial cells (Ledeboer *et al.*, 2006). This could facilitate carriage in the intestines of livestock and contaminate meat during processing as well as increase persistence in the host during infection (Althouse *et al.*, 2003, Steenackers *et al.*, 2012). As mentioned previously, the failure to clear a *S.* Typhi infection can lead to a chronic carrier state in which the bacteria migrate to the gall bladder and persist as biofilms on gallstones; increasing antimicrobial resistance and, occasionally, requiring surgical intervention (Kurtz *et al.*, 2017, Gunn *et al.*, 2014, Prouty *et al.*, 2002, Steenackers *et al.*, 2012).

1.5.3 Virulence factors

As described above, the infection process of both non-typhoidal and typhoidal *Salmonella* serovars is complex and gaps in our knowledge remain (Johnson *et al.*, 2018). Irrespective of type, infection by *Salmonella* is not possible without virulence factors and their delivery mechanisms. Virulence factors are often translocated to the host cell using T3SSs. The main structure of the T3SS is formed from the needle complex. The basal body of the needle complex spans the bacterial inner membrane, peptidoglycan layer, and the outer membrane. The needle extends into the extracellular environment to interact with the target cell (Park *et al.*, 2018, Lou *et al.*, 2019). The tip of the needle, the tip complex, regulates effector protein secretion into the cells and forms the structure of the translocon – the structure in the host plasma membrane through which effector proteins are delivered (Dey *et al.*, 2019).

Virulence factors can be defined as any molecule or structure that aids the colonisation and infection of a host (Sharma *et al.*, 2017a). They can include injected toxins and flagella for movement. Of note in *Salmonella* are the adhesins; type I and curli fimbriae utilised to adhere to, and persist in, the host. These factors also support biofilm formation and are encoded chromosomally and on plasmids (Fàbrega and Vila, 2013). On the chromosome, virulence factors are primarily contained within conserved regions termed *Salmonella* pathogenicity islands (SPIs). Some virulence genes are encoded by plasmids (Fàbrega and Vila, 2013). The two main SPIs are discussed below (Figure 1.9), and their regulation is summarised in Figure 1.10. However, seventeen have been identified and show varying levels of conservation between subspecies and serovars (Kombade and Kaur, 2021).

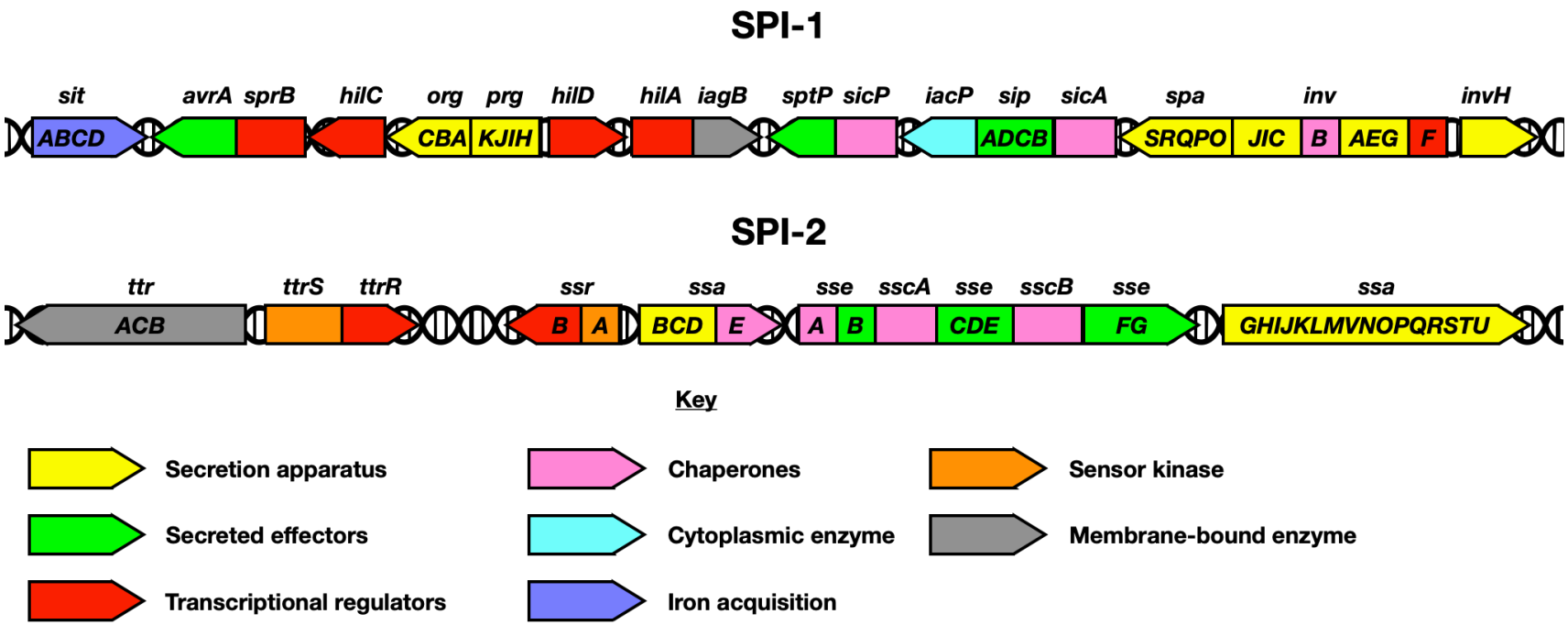


Figure 1.9 Schematic representation of SPI-1 and SPI-2. The genes of SPI-1 and SPI-2 are shown colour coordinated to their function, shown by the key. Not shown are other, chromosomally or plasmid encoded, genes expressed at the same time as the SPIs. Reproduced and modified from Fàbrega and Vila (2013) and Hensel *et al.* (1999).

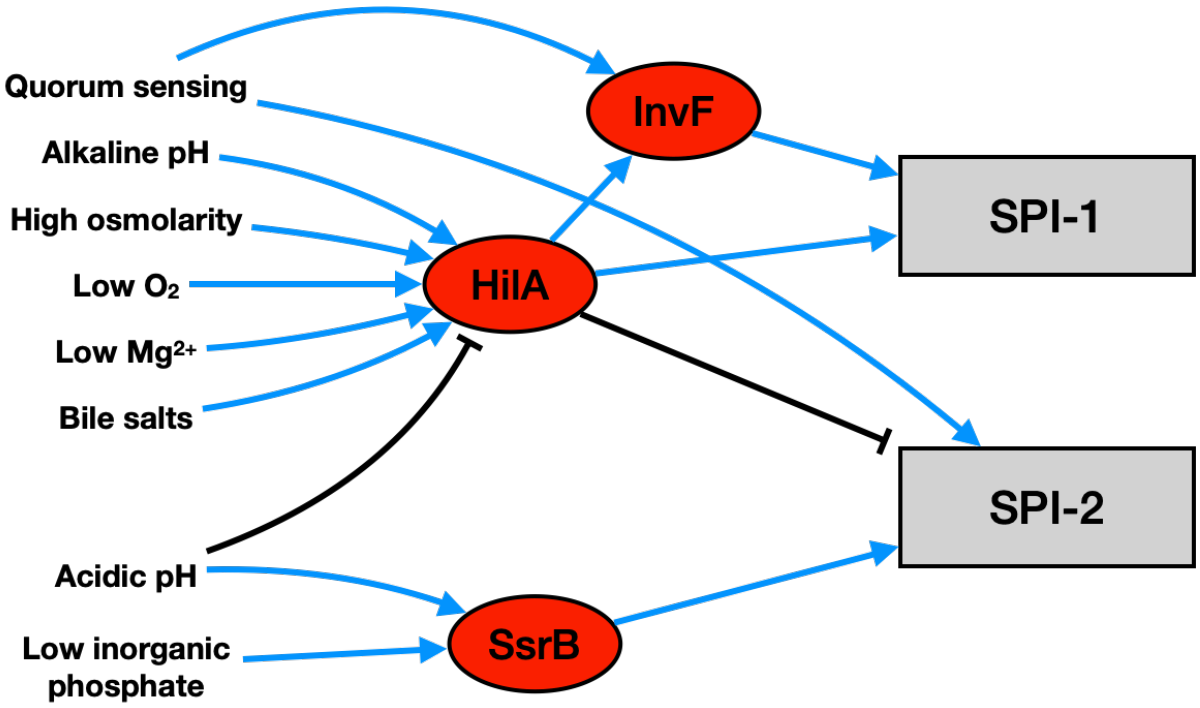


Figure 1.10 Simplified regulation of SPI-1 and SPI-2. The simplified regulation of SPI-1 and SPI-2 by their major regulators HilA/InvF and SsrA/SsrB respectively. Blue arrows represent activation, black flat-head arrows represent repression. The two-component system SsrA-SsrB is shown by the response regulator SsrB only. Reproduced and modified from Fàbrega and Vila (2013).

1.5.3.1 SPI-1

During initial colonisation of the gut, SPI-1 (Figure 1.9, Figure 1.10) is induced by the transcriptional activators *invF* and *hilA* (Ellermeier and Slauch, 2007). *HilA* is the major activator of SPI-1 and upregulates the encoded T3SS. This is achieved by upregulation of *sicA*. Further to this, *HilA* activates transcription of the AraC family member *invF*; which is dependent on *SicA* for the upregulation of effector proteins (Darwin and Miller, 2001, Lou *et al.*, 2019). The transcription of *hilA* is upregulated by a combination of a slightly alkaline pH, high osmolarity, low oxygenation, low Mg^{2+} concentration, and the presence of bile salts (Altier, 2005, Bajaj *et al.*, 1996, Lou *et al.*, 2019). The tip complex of the T3SS also binds bile salts to control the release of proteins through the T3SS translocon. Combined with the presence of bile, this may contribute to long-term colonisation of the gall bladder in chronic carriers where the *Salmonella* reside within the gall bladder (Kurtz *et al.*, 2017, Gunn *et al.*, 2014, Lou *et al.*, 2019). Interestingly, this is only seen in typhoidal serovars, in non-typhoidal serovars, bile salts repress the expression of the SPI-1 T3SS (Ellermeier and Slauch, 2007). SPI-1 encoded effector proteins are involved in actin cytoskeleton rearrangements needed to invade the gut M cells (Fàbrega and Vila, 2013, Lou *et al.*, 2019, Kurtz *et al.*, 2017). Of note are AvrA, SptP, proteins of the *sipABCD* operon, and factors chromosomally encoded by *sopABCD1D2E1E2* (Lou *et al.*, 2019, Fàbrega and Vila, 2013). AvrA has multiple roles in host cells: repression of the proinflammatory response, stimulation of proliferation at the gut epithelium, and control of tight junctions (Lin *et al.*, 2016, Ye *et al.*, 2007, Wu *et al.*, 2012). Host cell modulation is furthered by SptP, which disrupts the actin cytoskeleton, reduces membrane ruffling and suppresses the secretion of proinflammatory cytokines (Johnson *et al.*, 2017, Kaniga *et al.*, 1996, Button and Galán, 2011). The *Salmonella* invasion proteins (Sips)

are encoded in the *SipABCD* operon with SipB, SipC, and SipD involved in forming the translocon in the host membrane. SipA is involved in modulating and stabilising the actin cytoskeleton to facilitate bacterial uptake (Lou *et al.*, 2019, Zhou *et al.*, 1999, McGhie *et al.*, 2004). SipA also has an immunomodulatory effect; it induces the recruitment of polymorphonuclear leukocytes and the release of caspase-3 (McIntosh *et al.*, 2017, Srikanth *et al.*, 2010). The *Salmonella* outer proteins (Sops) have a wide range of functions. SopA, SopB, and SopE cause increased secretion of fluid into the intestinal lumen inducing diarrhoea and inflammation. SopA is also involved in regulating immune modulation and apoptosis, along with SopB. SopB, SopD, SopD2, SopE, and SopE2 cause cytoskeletal modifications and the subsequent invasion of *Salmonella* in to host cells. In addition to these functions, SopB, SopD, and SopE control replication and production of metabolites (Wood *et al.*, 2000, García-Gil *et al.*, 2018, Bertelsen *et al.*, 2004, Drecktrah *et al.*, 2005, Boonyom *et al.*, 2010, Bakowski *et al.*, 2007, Lim *et al.*, 2014, Vonaesch *et al.*, 2014, Bliska and van der Velden, 2012, Taylor and Winter, 2020, Lou *et al.*, 2019, Haraga *et al.*, 2008).

1.5.3.2 SPI-2

Following the adhesion and internalisation of *Salmonella* into the SCVs, SPI-2 is activated (Figure 1.9, Figure 1.10) (Kurtz *et al.*, 2017). The effectors encoded within SPI-2 are important for the development of systemic infection and intracellular replication within macrophages (Haraga *et al.*, 2008, Kurtz *et al.*, 2017). SPI-2 is under a complex regulatory network governed by the *SsrAB* TCS in response to low inorganic phosphate conditions and acidic pH (Fàbrega and Vila, 2013). SPI-2 is split into two sections. The smaller section contains the *ttrRSBCA*

operon responsible for tetrathionate reduction (Fàbrega and Vila, 2013, Löber *et al.*, 2006). Combined with the ability of *Salmonella* to resist host oxidative stresses this allows *Salmonella* to outcompete gut microbiota for nutrients and establish mucosal inflammation (Taylor and Winter, 2020, Winter *et al.*, 2010). One potential mechanism for outcompeting gut microbiota is *Salmonella*'s ability to resist reactive nitrogen stress via nitrate metabolism. *Salmonella* are known to produce higher levels of nitrous oxide compared to other members of the Enterobacteriaceae family (Torres *et al.*, 2016). Toxicity via nitrous oxide involves the inactivation of vitamin B_{12} and disruption of methionine synthesis (Deacon *et al.*, 1980). *Salmonella* can synthesise vitamin B_{12} *de novo* and also possess a vitamin B_{12} independent methionine synthase, in contrast to other organisms; this gives *Salmonella* the ability to withstand this toxic environment produced during nitrate metabolism, disrupt the gut microflora, and establish infection (Johnston, 2017).

1.5.3.3 Biofilms

Biofilms impact both the lifestyle and infection cycle of *Salmonella*. As for many bacteria, the base structure of such biofilms mainly comprises the extracellular polysaccharide cellulose and curli fimbriae. The cell surface protein BapA is also required for biofilm formation and has adhesive properties (Austin *et al.*, 1998, Solano *et al.*, 2002, Römling *et al.*, 1998, Fàbrega and Vila, 2013, Latasa *et al.*, 2005). Both typhoidal and non-typhoidal serovars form biofilms within the body, especially on gallstones. For *S.* Typhi, this can be part of chronic "carrier state" infections (Kurtz *et al.*, 2017, Gunn *et al.*, 2014, Prouty *et al.*, 2002). *S.* Typhimurium does not express curli fibres during mouse infection but does once it has left the host (White

et al., 2008, Barnhart and Chapman, 2006). The differential production of curli fibres is thought to be a response to the broader host range of non-typhoidal strains which are better able to persist in the environment and do not rely on host-to-host transfer. In the host, biofilm formation may help evasion of host defences and resistance to antimicrobial treatments administered. In the wider environment, biofilms allow the pathogen to withstand harsh conditions between hosts. Hence, biofilms aid transmissibility (White *et al.*, 2006, Gunn *et al.*, 2014, Prouty *et al.*, 2002, White *et al.*, 2008, Fàbrega and Vila, 2013).

1.5.4 Antimicrobial resistance in *Salmonella*

Incidence of antimicrobial resistance (AMR) amongst nontyphoidal *Salmonella* serovars increased 40% between 2004 and 2016. This caused 23,000 hospitalisations and 450 deaths per year in the US (Medalla *et al.*, 2021, Scallan *et al.*, 2011). Of all deaths caused by foodborne illnesses, nontyphoidal *Salmonella* species accounted for 28 % (Scallan *et al.*, 2011). Complications, hospitalisation, and deaths during infections have also increased with the prevalence of antimicrobial resistant *Salmonella*. This will continue unless novel therapeutic strategies are developed. The rise in AMR has led the World Health Organisation (WHO) to place fluoroquinolone resistant *Salmonella* serovars in the second highest priority group for which novel antibiotics are urgently required (WHO, 2017). *Salmonella* serovars strains resistant to 3 or more antibiotics account for 20,800 non-typhoidal infections each year in the US. There are 212,500 infections where the causative serovar is resistant to at least one essential antibiotic (CDC, 2019). Further, 74 % of all *S.* Typhi isolates are now resistant to ciprofloxacin (CDC, 2019).

Another problem associated with AMR in *Salmonella* is that there are many challenges of developing vaccines against *Salmonella enterica*. These include poor memory response, reactogenicity, narrow strain/serovar protection, and the need for multiple doses (MacLennan *et al.*, 2014). Further to this, previously established vaccines were not licensed for children, a demographic particularly susceptible to both typhoidal and non-typhoidal salmonellosis (MacLennan *et al.*, 2014, UKHSA, 2021, PHE, 2018, Qadri *et al.*, 2021). However, a new vaccine targeting *S.* Typhi has been licenced for use in children and shows good efficacy and safety (Qadri *et al.*, 2021, Crump and Oo, 2021). But, until there is a good catalogue of vaccines against multiple *Salmonella* serovars, the rising incidences of AMR within *Salmonella* infections will continue to increase in severity and lethality.

1.5.4.1 Mechanisms of antimicrobial resistance in Salmonella

The evolutionary pressures placed on an organism by antibiotics inevitably lead to resistance. *Salmonella* employ multiple mechanisms of resistance to withstand antimicrobial stress. In one mechanism, the intracellular concentration of the antimicrobial is reduced through efflux or cell envelope modification. *Salmonella* can also modify the intracellular target to disrupt the antimicrobials mechanism of action or modify and degrade the antimicrobial compound itself. At a multicellular level, community wide changes in behaviour to combat environmental stresses can also be used to increase resistance to antimicrobials (Martins *et al.*, 2011). Generally, resistance to a single antimicrobial compound involves a combination of these mechanisms (Cuypers *et al.*, 2018). These mechanisms are shown in Figure 1.11 and a brief example of each is given below.

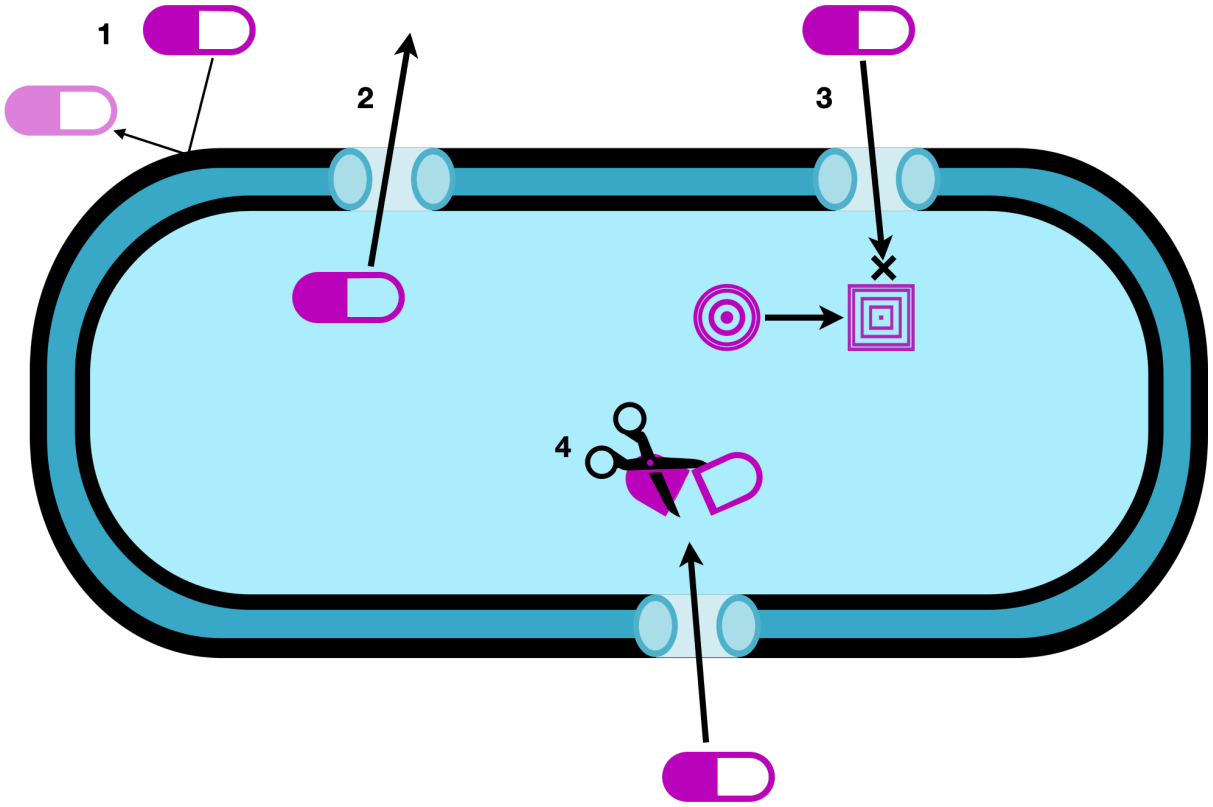


Figure 1.11 Mechanisms of antimicrobial resistance employed by *Salmonella*. An example of the resistance mechanisms employed by a bacterium in response to antimicrobial stress. 1. Modification of the cell envelop and downregulation of porin expression prevents the antimicrobial from entering the cell. 2. Upregulation of efflux pumps can actively pump out antimicrobial compounds, reducing intracellular concentrations. 3. Modification of the antimicrobial target inhibits the antimicrobial's mechanism of action. 4. Modification and degradation of the antimicrobial itself prevents it from functioning. Figure reproduced and modified from (CDC, 2019)

Efflux and modification of the cell envelope

The simplest mechanisms of resistance involve efflux of the antimicrobial or reduced uptake. Efflux is achieved by overexpression of efflux pumps and reduced uptake by modification of the cell envelope (Figure 1.11, section 1 and 2). The latter can involve decreased porin expression and increased levels of lipopolysaccharide in the outer membrane (Martins *et al.*, 2011). The prototypical example of an efflux pump is the tripartite AcrAB-TolC system, which is involved in antimicrobial resistance amongst almost all clinically relevant bacteria (Webber and Piddock, 2001, Whittle *et al.*, 2021, Piddock, 2006). These efflux pumps have a wide range of target molecules including quinolones, tetracyclines, oxazidiones, some macrolides, and chloramphenicol (Piddock, 2006). The benefit of overexpressing these efflux pumps is twofold. Firstly, this allows the intracellular concentration of the antimicrobial to be kept low, reducing access to its target. Secondly, by pumping the antimicrobial out of the cell, time for beneficial mutations to accumulate in the population increases, providing further levels of resistance (Piddock, 2006).

Modification of antimicrobial target

Mutations in target proteins, that prevent the antimicrobial from functioning, are a common response to fluroquinolones (Figure 1.11, sections 1, 2, 3) (Correia *et al.*, 2017). Such mutations are most commonly found within DNA gyrase (*gyrA*, *gyrB*) and topoisomerase IV (*parC*, *parE*) enzymes, which are essential for controlling DNA supercoiling and replication/transcription (Cuypers *et al.*, 2018). In *Salmonella* there are 12 residues of *gyrA* commonly mutated in strains of clinical importance. The mutations reduce the ability of

quinolones to bind DNA gyrase (Cuypers *et al.*, 2018). The most common mutation is of Ser83, and an acidic amino acid 4 residues downstream. These changes reduce quinolone binding and, therefore, reduce the disruption to normal functioning of DNA gyrase or topoisomerase IV (Aldred *et al.*, 2014). There are 9 common mutations observed in *gyrB* and *parC* and 10 mutations in *parE* (Correia *et al.*, 2017).

Modification of antimicrobial agent

Modification of the antimicrobial compound is a common response by bacteria to aminoglycosides (Figure 1.11, section 4). These antimicrobial compounds act on both the 16S and 30S subunit of the ribosome at the A site, inducing a conformational change in shape which locks the ribosome in a closed state (Frye and Jackson, 2013, Ramirez and Tolmasky, 2010). *Salmonella* use a suite of aminoglycoside modifying enzymes to inactive the aminoglycosides. These enzymes can be acetyltransferases, phosphotransferases, and nucleotidyltransferases. Modifications to the chemical structure of the antimicrobial compound disrupt its normal function, allowing *Salmonella* to resist the effect of the aminoglycoside (Ramirez and Tolmasky, 2010).

Population-wide responses to antimicrobials

The mechanisms of resistance discussed so far focus on the individual bacterium's response. However, AMR can also be achieved through population-wide collaboration. Quorum sensing (QS) is used to coordinate genetic responses to population density, fine-tuning gene

expression in response to both environmental conditions and other surrounding bacteria (Martins *et al.*, 2011). *Salmonella* use QS signalling molecules to coordinate virulence factor expression and colonisation of hosts (Martins *et al.*, 2011). QS is also linked to the formation of biofilms. Biofilms formed by *Salmonella* species are of great concern in food processing plants, as these biofilms show greatly enhanced resistance to environmental stresses and disinfectants (Steenackers *et al.*, 2012, Sereno *et al.*, 2017, Cadena *et al.*, 2019). Biofilms form protective environments which prevent, or severely reduce, the ability of antimicrobials to gain access to the individuals within the complex. This leads to cultures which are 10-1,000 fold more resistant to various classes of antimicrobials, and in the case of *Salmonella* this includes resistance to quinolones, sulphonamides, tetracyclines, cephalosporins, penicillins, aminoglycosides, sulphonamides, and monobactams (Davies, 2003, Sereno *et al.*, 2017).

1.6 The transcription factors MarA, SoxS, Rob, and RamA

1.6.1 MarA

The *marRAB* operon is widely conserved amongst enteric bacteria and important for the acquisition of multiple drug resistance (MDR) phenotype. In *E. coli* for example, the operon is implicated in conferring cross-resistance to quinolones and tetracyclines (Cohen *et al.*, 1993, Sharma *et al.*, 2017b). Under normal conditions *marRAB* is repressed by MarR homodimers (Figure 1.12) (Martin and Rosner, 1995). Each MarR molecule consists of 6 α helices and 3 β sheets with 1 wing region (Alekshun *et al.*, 2001). The N- and C-terminals of the protein interact to form the dimerisation domain, with residues 55-100 forming the DNA-binding

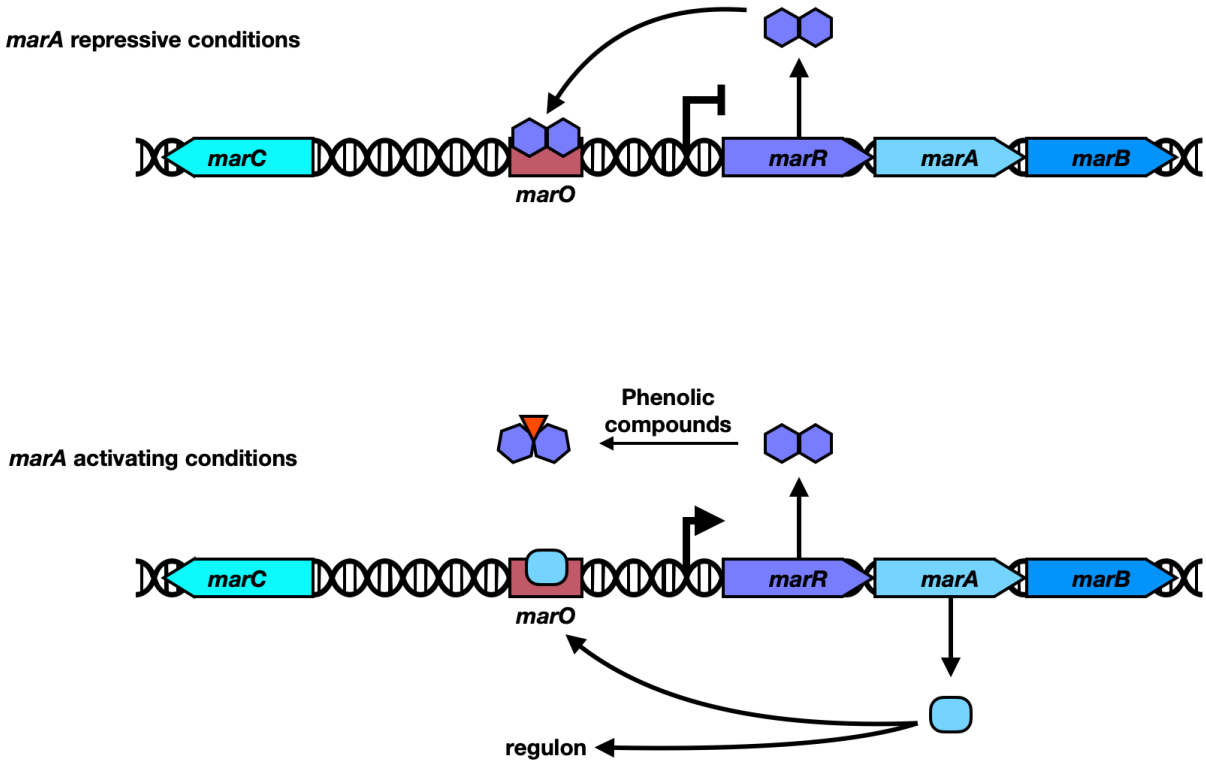


Figure 1.12 The *marRAB* locus. The *mar* locus under *marA* repressive conditions (top) and *marA* activating conditions (bottom). Flat head arrows indicate transcriptional repression, right-angled arrows indicate transcription. Under normal conditions, MarR homodimers, shown as twin hexagons, bind to the operator *marO* and repress transcription of the *marRAB* operon and *marC*. In response to phenolic compounds (inverted red triangles) the MarR homodimer undergoes a conformational change in shape and derepresses transcription. This allows MarA to be produced via transcription of *marA*. MarA positively regulates transcription at this locus and also regulates other genes. Figure reproduced and modified from Duval and Lister (2013).

winged helix-turn-helix motif (Alekshun *et al.*, 2001). MarR binds to two 21 base pair *marO* sites, upstream of *marR*, termed site 1 and site 2. Site 1 overlaps the -10 and -35 promoter elements (Martin and Rosner, 1995). MarR can bind multiple phenolic compounds (including salicylic acid and some napthaquinalones). Upon ligand binding a conformational shift occurs and the repression of the operon is lifted (Duval and Lister, 2013, Grove, 2013). Alleviation of MarR binding leads to transcription of the *marRAB* operon. MarA activates transcription of the operon and genes elsewhere on the chromosome (Martin *et al.*, 1996, Sharma *et al.*, 2017b). Mutations in either *marO* or *marR* lead to constitutive activation of *marRAB* and clinically relevant MDR (Sharma *et al.*, 2017b).

In *E.* coli, MarA directly regulates 33 genes (Sharma *et al.*, 2017b). MarA binds a 15-base pair, highly degenerate, sequence termed the marbox. The method by which MarA recruits RNAP to initiate transcription is different to the traditionally accepted model whereby the TF binds to the DNA then interacts with RNAP (Browning and Busby, 2004). MarA, has been shown to bind RNAP prior to interacting with DNA, which forms a promoter-scanning complex, in a method termed prerecruitment (Martin *et al.*, 2002). This promoter-scanning complex is also favoured for SoxS but not Rob (Duval and Lister, 2013). MarA is considered to be ambidextrous with regards to interaction with the RNA polymerase as it can act as both a Class I and a class II activator (Martin *et al.*, 2002, Sharma *et al.*, 2017b). Further to this, MarA also represses certain targets via steric hindrance (McMurry and Levy, 2010). Whilst this mechanism of repression is in contradiction to the prerecruitment theory, there is not enough

data to conclusively say whether prerecruitment is involved with repression (McMurry and Levy, 2010).

1.6.2 SoxS

SoxS, also a member of the AraC family, is a regulator of the oxidative stress response and is located in the divergent *soxRS* locus (Figure 1.13). SoxR, a MerR family member, homodimerises and functions as both an activator and repressor depending on whether its 2Fe-2S cluster has been oxidised by a redox cycling drug such as paraquat (Duval and Lister, 2013, Wu and Weiss, 1992, Gu and Imlay, 2011). SoxR shows similarities to MarR in that it homodimerizes and binds to the *soxRS* promoter region to prevent transcription via steric hindrance (Hidalgo *et al.*, 1998). However, unlike MarR, SoxR does not simply dissociate from the promoter upon signal recognition. Instead, SoxR switches to an active form which drives transcription of *soxS*. Whilst the direct mechanism of activation is not known, evidence suggests SoxR activates via conformational change (Hidalgo *et al.*, 1998, Watanabe *et al.*, 2008). SoxS binds the same consensus sequence as MarA and has an overlapping target regulon (Martin *et al.*, 2000, Li and Demple, 1994). SoxS (107 amino acids) is smaller than MarA, with which it shares a 41% identity and 67% similarity (Duval and Lister, 2013).

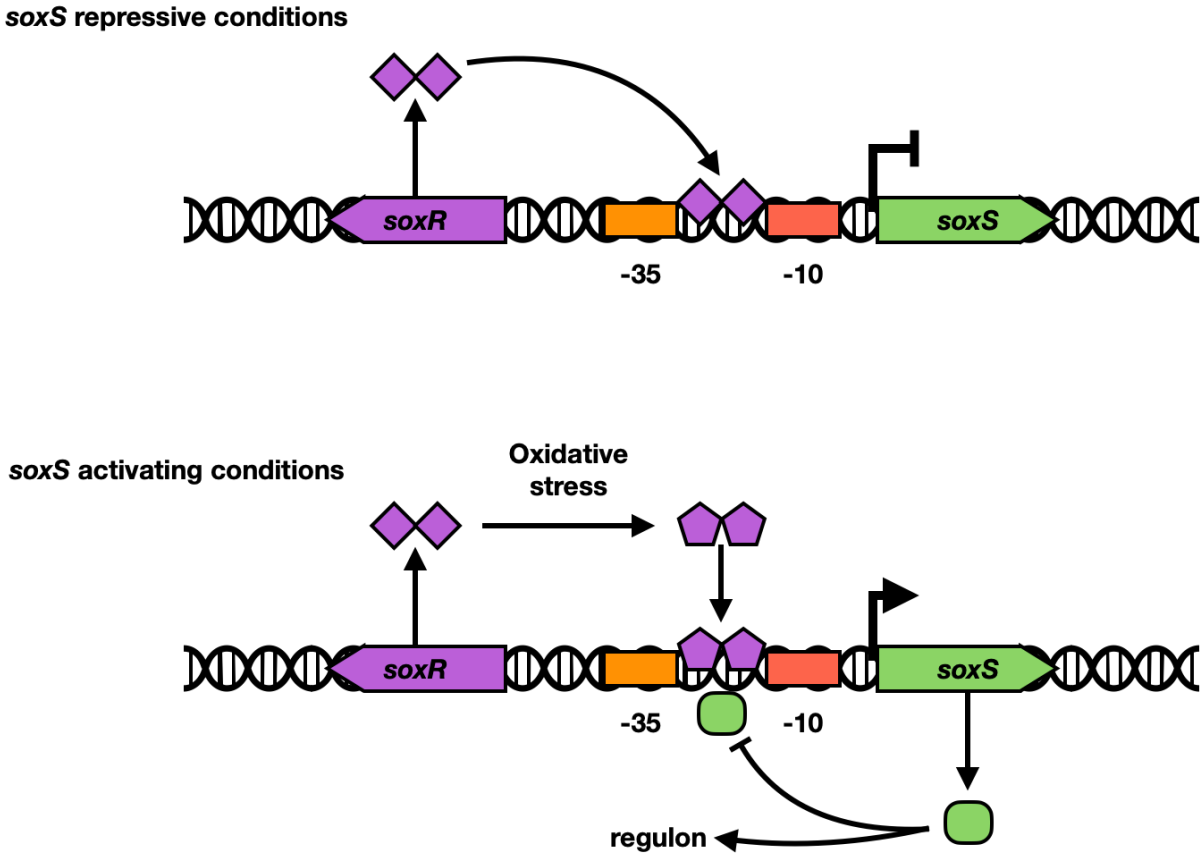


Figure 1.13 The *soxRS* locus. The *soxRS* locus under *soxS* repressive conditions (top) and *soxS* activating conditions (bottom). Flat head arrows indicate transcriptional repression, right-angled arrows indicate transcription. Under normal conditions, inactive SoxR homodimers (shown as twin diamonds) bind between the *soxS* promoter elements to repress transcription. Upon activation by oxidative stress, SoxR flips to an active state (shown as twin pentagons) and recruits RNAP to facilitate transcription of *soxS*. SoxS represses its own transcription and also regulates other genes. Figure reproduced and modified from Duval and Lister (2013).

1.6.3 Rob

Rob (289 amino acids) is larger than MarA, SoxS, and RamA and shows 51 % identity and 71 % similarity to MarA (Martin and Rosner, 2001). Discovered as a factor bound to the right border of *oriC*, Rob has no known role in the control of DNA replication (Skarstad *et al.*, 1993). Instead, Rob binds to the same DNA sequence as MarA, SoxS, and RamA (Jair *et al.*, 1996). Rob consists of two domains; the N-terminal domain resembles MarA and SoxS whilst the C-terminal domain prevents degradation by Lon protease and sequesters inactive Rob in intracellular foci (Li and Demple, 1994, Martin and Rosner, 2002, Griffith *et al.*, 2009). Rob is constitutively expressed at 5,000 to 10,000 copies per cell but is sequestered in inactive foci (Jair *et al.*, 1996, Skarstad *et al.*, 1993, Griffith *et al.*, 2002). In response to bile salts and fatty acids, the C-terminal domain of Rob releases the N-terminal domain, leading to the active form of Rob and the release from intracellular foci (Figure 1.14) (Griffith *et al.*, 2009, Duval and Lister, 2013).

1.6.4 RamA

RamA is found in *Salmonella* and some other members of the Enterobacteriaceae but not *E. coli*. RamR regulates *ramA* expression in a similar fashion to MarA are the *marRAB* locus (Figure 1.15). RamA also recognises the same DNA sequence as MarA, SoxS, and Rob (Weston *et al.*, 2018). As mentioned above, MDR in *E. coli* can be mediated by *marRAB* (Sharma *et al.*, 2017b). However, in *Salmonella enterica* species, the equivalent phenotype tends to result from mutations in *ramR* (Nikaido *et al.*, 2008). Hence, constitutive production of RamA is

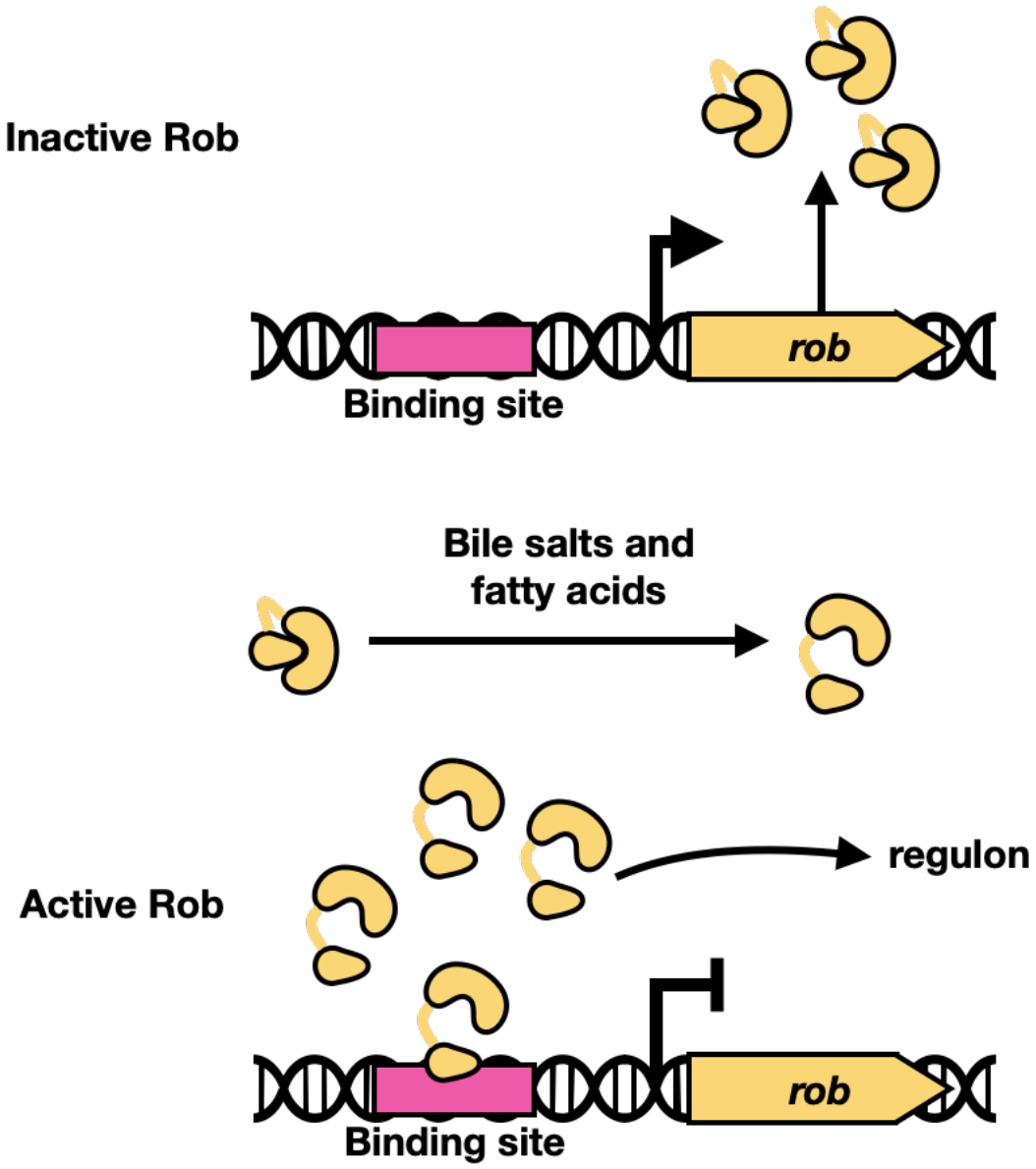


Figure 1.14 The *rob* locus. The *rob* locus shown under normal, inactive, conditions (top) and active conditions (bottom). Flat head arrows indicate transcriptional repression, right-angled arrows indicate transcription. Under normal conditions, constitutively expressed, inactive, Rob is sequestered in intracellular foci with the N-terminal domain sequestered by the C-terminal domain. In response to Bile salts and fatty acids, the C-terminal domain releases the N-terminal domain. Rob dissociates from the intracellular foci and binds to its binding site at

the *rob* locus repressing transcription as well as regulating other genes. Reproduced and modified from Duval and Lister (2013).

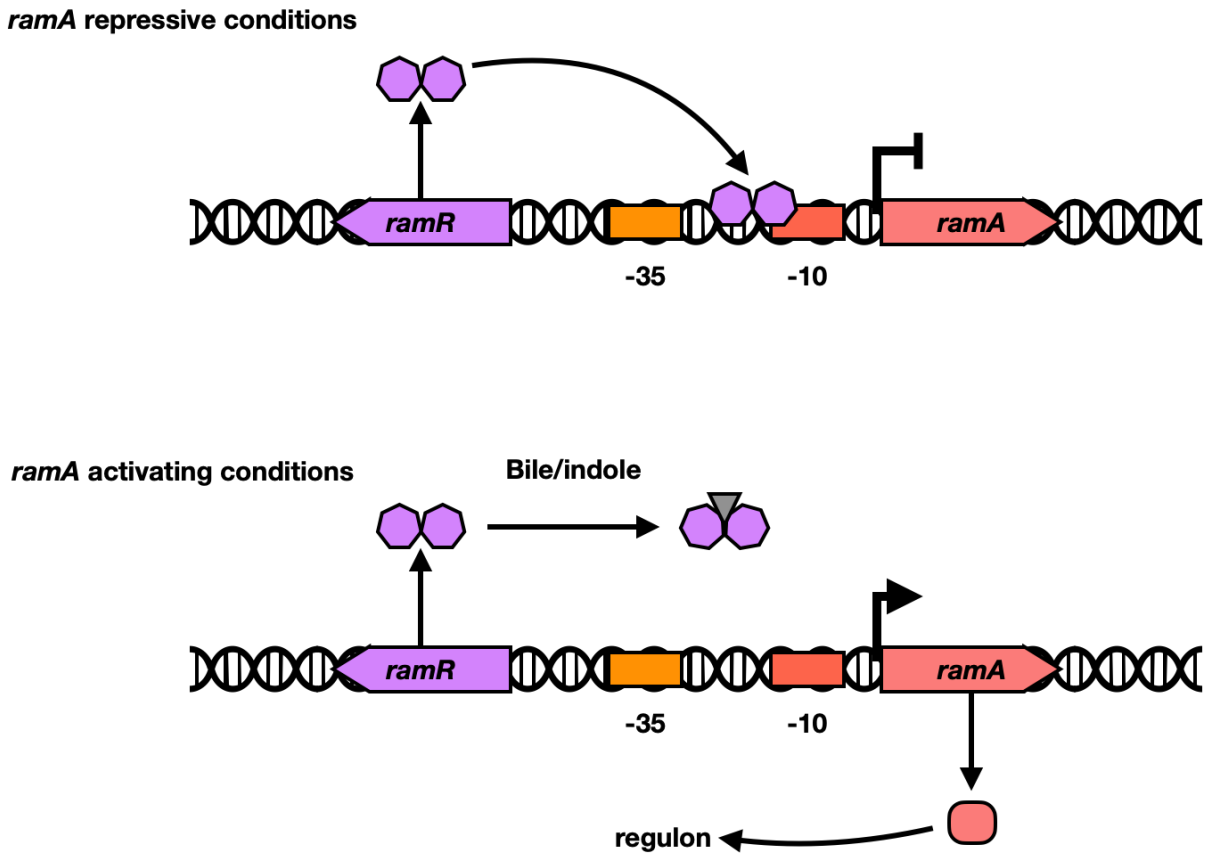


Figure 1.15 The *ramRA* locus. The *ramRA* locus shown under normal, *ramA* repressive conditions (top) and *ramA* activating conditions (bottom). Flat head arrows indicate transcriptional repression, right-angled arrows indicate transcription. Under normal conditions, RamR homodimers, shown as twin heptagons, and represses *ramA* transcription by binding to the core promoter elements. In response to bile, indole, and other drugs (inverted grey triangles), the binding affinity of the RamR homodimer is reduced and repression is relieved; *ramA* is subsequently transcribed and regulates its target genes. Reproduced and modified from Duval and Lister (2013).

required for the MDR phenotype and the apparent redundancy of MarA is not understood (Ricci *et al.*, 2006). As RamA assumes the functions of MarA in *Salmonella enterica*, it could be assumed that RamA will regulate transcription with the same promoter-scanning complex that MarA and SoxS do. RamA is upregulated by the intercellular signalling molecule indole and post translationally regulated by bile (Nikaido *et al.*, 2008). The DNA binding abilities of RamR are also reduced in response to certain drugs, including crystal violet and ethidium bromide (Yamasaki *et al.*, 2013).

1.7 Objectives of this study

There has been much study of the homologous TFs MarA, SoxS, and Rob in *E. coli (Duval and Lister, 2013)*. But this research has mainly focussed on individual proteins at specific targets. Even so, these TFs bind to the same degenerate sequence and are known to regulate each other. The level of overlap between each regulon is unknown. This study aimed to understand the DNA binding profiles of MarA, SoxS, Rob, and RamA in the clinically relevant species *Salmonella*. The implications of this work are important as these proteins are often associated with clinical resistance to antimicrobials; therefore, further study of them and their regulons is important in order to combat this looming threat.

2. Materials and Methods

2.1 Buffers and reagents

The buffers and reagents used in this work are presented in a sub-heading alongside each technique. All buffers were made using ddH_2O and autoclaved unless otherwise stated. All compounds were bought from Sigma or Life Technologies unless otherwise stated.

2.2 Bacterial strains and plasmids used

2.2.1 Bacterial strains

The bacterial strains used in this study are presented in Table 2.1.

2.2.2 Plasmids

The plasmids used in this study are presented in Table 2.2.

2.3 Growth of bacterial cultures and antibiotics used

2.3.1 Growth conditions

Unless otherwise stated, bacterial cultures were grown at 37 °C with shaking at 200 rpm. Usually, cultures were incubated overnight for 16 hours (5 pm - 9 am). Typically, LB was used as both the liquid and solid media for *E. coli*. For *Salmonella*, LB was used generally for both liquid and solid media, but Brilliant Green and Bismuth sulphite medium was used for *Salmonella* specific growth or selection. Following selective growth on these media,

Table 2.1 Bacterial strains used in this study

Strain	Genotype	Source
Escherichia coli JCB387	Δ*nirB* Δ*lac*	Grainger *et al.* (2007)
Salmonella enterica subsp. *enterica* serovar Typhimurium SL1344	Histidine auxotroph of parental strain ST4/74	Hoiseth and Stocker (1981), (Wray and Sojka, 1978)
Escherichia coli T7 Express	fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr- 73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 D(mcrC- mrr)114::IS10	New England Biolabs
Escherichia coli DH5α	fhuA2 D(argF-lacZ)U169 phoA glnV44 f80D(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolabs
Escherichia coli S17-1	TpR SmR recA, thi, pro, hsdR-M+RP4: 2-Tc:Mu: Km Tn7 λpir	Simon *et al.* (1983)

Table 2.2 Plasmids used in this study

Plasmid name	Features	Source
pAMNF	A pJ201 derivative with an N-terminal 3x FLAG tag upstream of *Hin*dIII and *Kpn*I restriction sites. Contains an Ori_pUC origin and a kanamycin resistance cassette.	This study, ordered from ATUM
pAMCF	A pJ201 derivative with a C-terminal 3x FLAG tag downstream of *Hin*dIII and *Kpn*I restriction sites. Contains an Ori_pUC origin and a kanamycin resistance cassette.	This study, ordered from ATUM
pAMNM	A pJ241 derivative with an N-terminal 8x Myc tag upstream of *Hin*dIII and *Kpn*I restriction sites. Contains an Ori_pUC origin and a kanamycin resistance cassette.	This study, ordered from ATUM
pAMCM	A pJ241 derivative with a C-terminal 8x Myc tag downstream of *Hin*dIII and *Kpn*I restriction sites. Contains an Ori_pUC origin and a kanamycin resistance cassette.	This study, ordered from ATUM
pSR	4 kb plasmid used for *in vitro* transcription. Encodes ampicillin resistance. Cloning site upstream of a λloop terminator. Derived from pBR322.	Kolb *et al.* (1995)
pRW50T	16 kb plasmid used for β-galactosidase assays. Encodes tetracycline resistance. Cloning site upstream of *LacZ* fusion.	Lodge *et al.* (1992)
pET28a	5.4 kb plasmid used for protein overexpression. Features an inducible T7*lac* promoter, both N- and C-terminal his tags, a thrombin cleavage site, and an internal T7 tag. Encodes kanamycin resistance.	Novagen
pUC57-MlrA	2.5 kb plasmid containing *mlrA* cloned downstream of the same promoter found in the pAM set of plasmids. Encodes ampicillin resistance.	Genewiz

All sequences cloned into the multiple cloning site of the pAM plasmids are under the control of a modified *lacUV5* promoter containing a -35, extended -10, and -10 element giving constitutive expression.

potential *Salmonella* colonies were confirmed by colony PCR of a *Salmonella* specific gene (*SseF*) alongside primers specific to any plasmid present in the strain of *Salmonella* grown.

2.3.2 Solid media

LB agar:

LB agar was purchased from Sigma as a powder (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 15 g/L agar) and dissolved in deionised and distilled water (ddH_20) at 35 g/L before autoclaving. Unless stated otherwise, all media and buffers/solutions were autoclaved at 121°C for 21 minutes. All agar plates were poured and dried in a flow hood before use or stored at 5 °C until required.

Congo Red agar:

500 mL of LB agar (as above) without salt is prepared and autoclaved. When the autoclaved medium has cooled to 50 °C, 2 mL of Congo Red solution (10 mg/mL) is added to give a final concentration of 40 μg/mL before any required antibiotics are added and plates poured.

Brilliant Green agar:

Brilliant Green agar was purchased from Oxoid as a powder. Formula is as follows: 10 g/L protease peptone, 3 g/L yeast extract, 10 g/L lactose, 10 g/L sucrose, 5 g/L sodium chloride, 0.08 g/L phenol red, 0.0125 g/L brilliant green, 12 g/L agar, pH 6.9 at 25 °C.

Bismuth Sulphite agar:

Bismuth Sulphite agar was purchased from Oxoid as a powder. Formula is as follows: 5 g/L peptone, 5 g/L 'Lab-Lemco' powder, 5 g/L glucose, 4 g/L disodium phosphate, 0.3 g/L bismuth sulphite indicator, 0.016 g/L brilliant green, 12.7 g/L agar, pH 7.2 at 25 °C.

2.3.3 Liquid media

LB broth:

LB broth was purchased from Sigma as a powder (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) and dissolved in ddH_2O at 20 g/L before autoclaving. If required, LB broth was supplemented with 1 % glucose before autoclaving.

2.3.4 Antibiotics

Stocks of antibiotics were made, and filter sterilised using a 0.45 µm pore filter before addition (at the stated concentrations) to media that had cooled below 50 °C. Antibiotics were added in a 1:1000 dilution (unless otherwise stated) to both solid and liquid media.

Ampicillin:

100 mg/mL was dissolved in ddH_2O and stored at 4 °C. A final concentration of 100 µg/mL was used.

Kanamycin:

50 mg/mL was dissolved in ddH_2O and stored at 4 °C. A final concentration of 50 µg/mL was used.

Tetracycline:

35 mg/mL was dissolved in methanol and stored at -20 °C. A final concentration of 35 µg/mL was used.

2.4 PCR reactions and oligonucleotides used

2.4.1 Buffers and reagents required

- MyTaq Red Mix (Bioline) – all-in-one mix, no dNTPs required
- 100 mM dNTP mix (Bioline) – 25 mM each dATP, dGTP, dCTP, dTTP. Diluted 1:10 with ddH_2O
- Velocity DNA polymerase (Bioline)
- 5x HiFi Buffer (Bioline) – PCR reaction buffer for use with Velocity DNA polymerase only. Mg^{2+} concentration is 10 mM
- Q5 High-Fidelity DNA Polymerase (New England Biolabs) – 5x PCR reaction buffer for use with Q5 DNA polymerase only. Mg^{2+} concentration is 2 mM at 1x

2.4.2 PCR

Throughout this study all PCR reactions were done in volumes of 25 or 50 µL using Velocity (Bioline), Q5 (New England Biolabs), or MyTaq Red mix (Bioline). For each reaction not using MyTaq Red Mix, 1 µL of 100 mM dNTPs (dGTP, dATP, dCTP, dTTP) were used. Usually, 100 – 200 ng of template DNA was used with 20 pmol of each primer and 5 or 10 µL of reaction buffer were used in a 25 or 50 µL reaction. A 98 °C denaturation step was used, and the annealing temperature was adjusted depending on the properties of the primer used – roughly 3-5 °C below the T_m of the primer. Generally, 30s elongation was used per 1kb of DNA amplified. Unless otherwise stated, PCR cycles were as follows: 98 °C 2 min, 35 cycles of 98 °C 30 s, 30s at specified annealing temperature and 72 °C 30s before a final extension step of 72 °C for 10 mins.

2.4.3 Colony PCR

Colony PCR was done in the same way as 2.4.2 but instead of the 1 µL template DNA being added, this was replaced with a colony picked from a plate and transferred directly to the PCR reaction tube.

2.4.4 Megaprimer PCR

For the preparation of large mutant DNA fragments (1 kb plus) in which the mutation required is not easily obtained through the use of traditional primers, megaprimer PCR reactions were done. This is a two-step PCR in which the first step amplifies a portion of the larger fragment with the mutation present in one of the primers; this fragment is then used as one of the primers in the second step. The first step involves a traditional forward primer at the 5' end of the fragment and a reverse primer (containing the desired mutation) that binds to the middle of the fragment. The amplified fragment is then sequenced to confirm the mutation before being used as the forward primer for the second step. In this step, the large fragment containing the mutation is used as the forward primer and the reverse primer anneals to the 3' end. Following this second PCR the fragment is again sequenced to confirm the mutation is present. All PCR reactions are done as above.

2.4.5 Oligonucleotides

The oligonucleotides used in this study are presented in Table 2.3.

2.5 Preparation of competent cells

2.5.1 Buffers and reagents required

- 100 mM Calcium Chloride.
- 50 % (v/v) glycerol.
- 15 % (v/v) glycerol.

Table 2.3 Oligonucleotides used in this study

Oligonucleotide (F – forward, R – reverse)	Sequence (5'→3')	Notes
Oligonucleotides used in ChIP-seq experiments		
MarA F	actgca**ggtacc**ATGTCCAGACGCAACACTGACGC	
MarA N-terminal R	tgcagt**aagctt**CTAGTAGTTGCCATGGTTCAGCGGC	Reverse primer containing stop codon for cloning into pAMNF or pAMNM
MarA C-terminal R	tgcagt**aagctt**GTAGTTGCCATGGTTCAGCGGC	Reverse primer lacking stop codon for cloning into pAMCF or pAMCM
SoxS F	actgca**ggtacc**ATGTCGCATCAGCAGATAATTCAGACCC	
SoxS N-terminal R	tgcagt**aagctt**CTACAGGCGGTGACGGTAATCGC	Reverse primer containing stop codon for cloning into pAMNF or pAMNM
SoxS C-terminal R	tgcagt**aagctt**CAGGCGGTGACGGTAATCGC	Reverse primer lacking stop codon for cloning into pAMCF or pAMCM
Rob F	actgca**ggtacc**ATGGATCAGGCTGGCATAATTCGCG	
Rob N-terminal R	tgcagt**aagctt**TTAACGGCGAATCGGGATCAGAAATTCGC	Reverse primer containing stop codon for cloning into pAMNF or pAMNM
Rob C-terminal R	tgcagt**aagctt**ACGGCGAATCGGGATCAGAAATTCGC	Reverse primer lacking stop codon for cloning into pAMCF or pAMCM
RamA F	actgca**ggtacc**ATGACCATTTCCGCTCAGGTTATCG	
RamA N-terminal R	tgcagt**aagctt**TCAATGCGTACGGCCATGCTTTTCTTTACG	Reverse primer containing stop codon for cloning into pAMNF or pAMNM
RamA C-terminal R	tgcagt**aagctt**ATGCGTACGGCCATGCTTTTCTTTACG	Reverse primer lacking stop codon for cloning into pAMCF or pAMCM
Oligonucleotides used in the amplification and mutation of the *csgDEFG* intergenic region		
AM0210	ggctgc**gaattc**GCTGTCACCCTGGACCTGGTCG	CsgD intergenic region long forward
AM0211	cgcccg**aagctt**CATGATGAAACTCCACTTTTTTTAATCGC	CsgD intergenic region long reverse
AM0212	GTATGATTTTTTAAATCTATGCAATCCCATAGCCCTGTACAACTTTACTATCAAATC	CsgD mutated SoxS binding site 1 mega primer forward
AM0213	ggctgc**gaattc**GGGGATGTTCTTATGCTTC	CsgD intergenic region short forward
AM0214	cgcccg**aagctt**GTGTTAAACTTGTAACCAAATG	CsgD intergenic region short reverse

AM0221	ggctgc**gaattc**GGGGATGTTCTTATGCTTCCCATGTGGGGCAATACGCACACCACTAGCCCCACTTCG	CsgD intergenic region short mutated SoxS binding site 2 megaprimer
AM0222	GGGGATGTTCTTATGCTTCCCATGTGGGGCAATACGCACACCACTAGCCCCACTTCG	Same as AM0221 but lacks restriction sites; used in the generation of a long fragment with SoxS binding site 2 mutated
AM0223	ggctgc**gaattc**GGGGATGTTCTTATGCTTCCCATGTGGGGCAATACGCACACCACTAGCCCCACTTCGTTTTTTTTGTCTTTGTGCTGTCCAGG	Extended version of AM0221 with 30 base clamp to ensure efficient binding to DNA template
AM0224	GGGGATGTTCTTATGCTTCCCATGTGGGGCAATACGCACACCACTAGCCCCACTTCGTTTTTTTTGTCTTTGTGCTGTCCAGG	Same as AM0223 but lacks restriction sites; used in the same way as AM0222
AM0225	CCTGGACAGCACAAAGACAAAAAAAACGAAGTGGGGCTAGTGGTGTGCGTATTGCCCCACATGGGAAGCATAAGAACATCCCC	Reverse complement of AM0224

Oligonucleotides used in the cloning of transcription factors for protein purification

MarA pET28a ovrexpr forward	gtagga**catatg**TCCAGACGCAACACTGACGC
MarA pET28a ovrexpr reverse	gccagt**ggatcc**CTAGTAGTTGCCATCCTTCAGCG
SoxS pET28a ovrexpr forward	gtagga**catatg**TCGCATCAGCAGATAAATTCAGACCC
SoxS pET28a ovrexpr reverse	gccagt**ggatcc**CTACAGGCGGTGACGGTAATCGCT
Rob pET28a ovrexpr forward	gtagga**catatg**GATCAGGCTGGCATAATTCGCG
Rob pET28a ovrexpr reverse	gccagt**ggatcc**TTAACGGCGAATCGGGATCAGAAATT
RamA pET28a ovrexpr forward	gtagga**catatg**ACCATTTCCGCTCAGGTTATCG
RamA pET28a ovrexpr reverse	gccagt**ggatcc**TCAATGCGTACGGCCATGCTTTTCTTTA

Bold text indicates restriction sites, lower case letters represent spacers to aid restriction enzyme function, underlined text represents mutated sequences.

2.5.2 Calcium competent cells

Calcium competent cells were made using 1 mL of overnight culture was inoculated into 50 mL fresh LB and incubated at 37 °C until OD_{600} = 0.6 (using a Jenway6300 spectrophotometer). The cells were then chilled on ice for 5-10 mins before centrifugation at 1,500 xg for 5 min and 4 °C. All subsequent steps were done on ice. The pellet was resuspended in 25 mL 100 mM ice cold $CaCl_2$ and incubated on ice for up to 30 mins before centrifuging as before. The resulting pellet was resuspended in 3.3 mL ice cold $CaCl_2$ and incubated overnight to maximise competency. The following day 1.1 mL of ice-cold sterile 50 % glycerol was added, and the cells were aliquoted into 200 µL volumes and stored at -80 °C until required. Cells were thawed on ice for 15-20 mins before use.

2.5.3 Electrocompetent cells

For electrocompetent cells, 3 mL of overnight culture was inoculated into 50 mL fresh LB and grown at 37 °C with shaking until OD_{600} = 0.6. The cultures were transferred to a 50 mL centrifuge tube and chilled on ice for 5-10 mins. All subsequent steps were done on ice. Cells were harvested by centrifugation at 1,500 xg for 10 mins at 4°C and the pellet was resuspended in 25 mL ice-cold sterile 15 % glycerol. This was repeated twice more. The final pellet was resuspended in 500 µL 15 % glycerol and aliquoted into 45 µL volumes for individual transformations and stored at -80 °C until needed. Cells were thawed on ice for 15-20 mins before use.

2.6 Bacterial transformation methods

2.6.1 Heat shock transformation

100 ng of plasmid DNA (or an entire ligation reaction) was mixed with 100 µL of calcium competent cells and incubated on ice for 90-120 mins. Following incubation on ice the cells were heat shocked for 90s at 42 °C before incubating on ice for 2-5 mins. 900 µL of fresh sterile medium (LB or SOC) broth was added and the cells were incubated at 37 °C for 20-40 mins for recovery. The cells were then harvested at 2,400 xg for 3 mins and 900 µL of the supernatant was removed. The cells were resuspended in the remaining 100 µL of medium and plated onto the appropriate plates containing the relevant antibiotics.

2.6.2 Electroporation

For the transformation of *Salmonella* SL1344, electroporation was used. In order to prevent arcing of the electrical current and the killing of the competent cells, salt concentrations in both the competent cells and the DNA must have a low salt concentration; therefore, a maximum of 10 % of the final volume can be DNA. To increase efficiency, the maximum amount of plasmid DNA (5 µL) was mixed with 45 µL electrocompetent cells and added to a pre-chilled 1 mm gap electroporation cuvette (supplied by Cell Projects). The cuvettes were incubated for up to 30 mins on ice before quickly transferring to an Eppendorf Eporator for electroporation at 2,500 V. Care was taken to ensure the electrodes on the cuvette were dry to ensure efficient electroporation. As soon as the electroporation had finished, 950 µL of sterile LB broth was added to recover cells and prevent loss of efficiency. The cells were then incubated at 37 °C for 20-40 mins. Following this, the cells were pelleted by centrifugation at

2,400 xg for 3 mins and 900 µL of supernatant was removed. The cells were resuspended in the remaining 100 µL of LB and plated onto the appropriate agar medium containing the relevant antibiotics.

2.6.3 Conjugation

For conjugations, pRW50T was electroporated into *E. coli* S17-1 to be used as the donor strain in the conjugations. Briefly, overnight colonies of both donor and reciepient strains were set up in LB. The following day, 1 mL of each overnight culture was centrifuged at 2,400 xg for 3 mins to pellet the cells. The pellet was then resuspended and washed with 500 µL 0.9 % NaCl, vortexed, and recentrifuged as before. This wash step was repeated. The pellet was then resuspended in 1 mL fresh LB and 200 µL of both donor and recipient cultures were added together, mixed with pipetting and 50 µL was spotted onto LB agar plates (no antibiotic) for incubation overnight at 30 °C. The following day, a sterile p100 tip was used to scrape the spotted clumps into 100 µL of 0.9 % NaCl before mixing via pipetting. The resuspended cultures were then plated onto selective LB plates and incubated overnight at 37 °C. Colonies were screened using colony PCR for the plasmids within each strain and a *Salmonella* specific gene as mentioned in 2.5.

2.7 Isolation of bacterial genomic and plasmid DNA

2.7.1 Genomic DNA prep

Genomic DNA preps were done using a Qiagen Blood and Tissue Kit following manufacturer's protocol. Briefly, 1 mL of overnight culture is pelleted and treated with lysozyme for 30 mins at 37 °C. The sample is then incubated with proteinase K for 90 mins at 56 °C before being RNase treated for 10 mins at room temperature. DNA is then bound to a silica membrane in a column using a high chaotropic salt buffer whilst contaminants are washed off before elution in water.

2.7.2 Plasmid DNA prep

Plasmids were extracted using Qiagen kits (Miniprep, Midiprep, and Maxiprep kits) and following the manufacturer's instructions. Qiagen kits use a high salt buffer to bind DNA to a silica membrane and a low salt buffer (or water as was used in this study) to elute the DNA. Miniprep and Midiprep kits were used to extract high copy number plasmids and maxiprep kits were used to extract low copy number plasmids or for any plasmids to be used in *in vitro* work.

2.8 Agarose gel electrophoresis

2.8.1 Buffers and reagents required

- Agarose powder – supplied by Bioline
- 5x TBE – 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA. Supplied by Fisher Scientific. Buffer was diluted to 1x with ddH_2O when used.

- 6x Loading dye – 15 % Ficoll-400, 60 mM EDTA, 19.8 mM Tris-HCl, 0.48 % SDS, 0.12 % Dye 1, 0.006 % Dye 2, pH 8. Supplied by NEB.

2.8.2 Agarose gel electrophoresis

Agarose gels were made to 1 % (w/v) by dissolving powdered agarose in 1x TBE. The powder was dissolved by microwaving for 1 min on high heat with regular shaking to mix. When the agarose solution had cooled below 50 °C, ethidium bromide or Sybr Safe (Thermo Fisher) was added to 1 % (v/v) before pouring. Gels were run at 120 V for 30-40 mins in 1x TBE. Gels were imaged using a UV transilluminator. All DNA was run with 6x loading dye to track gel progress.

2.9 Polyacrylamide gel electrophoresis

2.9.1 Buffers and reagents required

Polyacrylamide gel electrophoresis (PAGE):

- 5x TBE – 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA. Supplied by Fisher Scientific. Buffer was diluted to 1x with ddH_2O when used.
- Protogel – 30 % w/v Acrylamide/bisacrylamide (37.5:1 ratio). Supplied by Geneflow.
- 10 % (w/v) ammonium persulphate (APS) – 100 mg APS (Sigma) in 1 mL ddH_2O. Made fresh for each gel.
- TEMED (N,N,N',N'-Tetramethylethylenediamine).

SDS-PAGE:

- Protein ladder (New England Biolabs)
- 5x SDS loading dye – 10 mM Tris-HCl, e mM EDTA, 20 % glycerol, 0.025 % bromophenol blue, 0.025 % xylene cyanol, pH 7.5
- NuPAGE™ 4-12 % Bis-Tris protein gel (Invitrogen)
- 10x NuPAGE™ MES SDS Running Buffer (Invitrogen) – 500 mM MES, 500 mM Tris base, 10 % SDS, 10 mM EDTA, pH 7.3. Diluted to 1x before use.
- Staining buffer – 50 % (v/v) methanol, 10 % (v/v) acetic acid, 2 g Brilliant Blue R. Made to 1 litre.
- Fast destain buffer – 40 % (v/v) methanol, 10 % (v/v) acetic acid. Made to 1 litre.

2.9.2 Polyacrylamide gel electrophoresis (PAGE)

Acrylamide gels were made to 7.5 % acrylamide using 30 % (w/v) acrylamide, 5x TBE and ddH_2O. Polymerisation of the gels was done by adding 0.01 % volumes of 10 % (w/v) APS and 0.001 % volumes of TEMED. Gels were run at 30 mA for 20-30 mins or as required. Gels were stained in water containing 1 % ethidium bromide (v/v) for 10 mins. All DNA was run with 6x loading dye to track gel progress.

2.9.3 SDS-PAGE

SDS-PAGE gels were used to view the eluted fractions of purified proteins. SDS-PAGE gels were supplied by Invitrogen and run following manufacturer's recommendations. 2 µL Protein ladder was run alongside the proteins to determine the length of time the gel was run for. To

view proteins, staining buffer was added to the gel and incubated at room temperature for a minimum of 30 mins with gentle agitation. SDS-PAGE gels were destained using fast destain buffer in the same way as the staining buffer.

2.10 Extraction, precipitation, and purification of DNA

2.10.1 Buffers and reagents required

- Phenol/chloroform/isoamyl alcohol pH 8 (25:24:1)
- 100 % (v/v) Ethanol
- 70 % (v/v) Ethanol
- 3 M Sodium acetate pH5.2
- 20 mg/mL Glycogen

2.10.2 Phenol-chloroform extraction and ethanol precipitation

DNA extraction using phenol-chloroform was done by adding a 1:1 ratio of phenol-chloroform to the DNA to be extracted in a 2 mL 5PRIME Phase Lock Gel Heavy microfuge tube (Quantabio). The mixture was vortexed well for 20 seconds before centrifugation at 21,000 xg for 3 mins. The aqueous phase was then transferred into a fresh 1.5 mL microfuge tube containing 0.1 volumes of 3 M sodium acetate pH5.2 and 1µL 20 mg/mL glycogen and 3 volumes of ice cold 100% ethanol was added. Precipitation was done at -80 °C for 30-60 mins or -20 °C overnight and the pellet collected by centrifugation at 21,000 xg for 10 mins. The

pellet was washed with ice cold fresh 70 % (v/v) ethanol and re-centrifuged before drying in a speedvac and resuspended in the required volume of ddH_2O.

2.10.3 Agarose gel extraction

Agarose gel extractions were used in the purification of digested DNA plasmids or fragments and followed the directions of a Qiagen Gel Extraction Kit. Gel slices were excised using a fresh razor blade using a transilluminator.

2.10.4 Polyacrylamide gel electrophoresis (PAGE) extraction

PAGE gel extractions were done using an 8% acrylamide gel ran at 30 mA until the bromophenol blue band was 80 % of the way down the gel. DNA fragments were excised as described in 2.11.2; if using this method on a ChIP library the smear between 200 and 600 bp was taken. Excised fragments were placed into a 0.7 mL microfuge tube which had a small hole pierced through the bottom of the tube using a hot needle. This 0.7 mL microfuge tube was placed inside a 1.5 mL one and spun at 21,000 xg for 5 mins to shred the gel. The 0.7 mL microfuge tube was discarded, and 0.4 mL acrylamide extraction buffer (2.24.1) was added; the lid was sealed with Parafilm and rotated overnight at 4 °C. The following morning, the gel and liquid was transferred to a Corning Costar Spin-X column (Corning) and spun at 21,000 xg for 5 mins; the gel fragments were discarded, and the DNA was ethanol precipitated as above.

2.10.5 Agencourt AMPure XP magnetic bead clean up (Beckman Coulter)

Purification and size selection of ChIP-seq library DNA was done using an Agencourt AMPure XP magnetic bead clean up following the manufacturer's instructions. This technique is referred to as a bead clean up throughout this study.

2.11 Restriction digests

2.11.1 Buffers and reagents required

- 10x CutSmart buffer (New England Biolabs) – 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/mL BSA, pH 7.9
- Calf Intestinal Alkaline Phosphatase (New England Biolabs) – 25 mM Tris-HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 50 % glycerol, pH 7.5

2.11.2 Restriction digests

All restriction enzymes used were supplied by New England Biolabs. Unless otherwise stated restriction digestions were done at 37 °C for 1-3 hours. Single enzyme digests were done in a final volume of 50 µL, with 43.5 µL of DNA being digested by 1.5 µL of restriction enzyme and 5 µL of 10x CutSmart buffer. Double enzyme digests were done by setting up two single digests as described and mixing 45 µL of each single digest to give 90 µL final volume; the remaining single digests were run alongside the double digests to confirm that each restriction enzyme had cut correctly. For digestion of plasmid DNA, 4 µL of Calf Intestinal Alkaline

Phosphatase was added in the final 30-60 mins of the digestion to prevent re-ligation of the plasmid. Digests were check by running on an agarose gel as described above.

2.12 Ligation of DNA fragments

2.12.1 Buffers and reagents required

- T4 DNA ligase (New England Biolabs)
- 10x T4 DNA ligase buffer (New England Biolabs) – 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5

2.12.2 Ligation of DNA fragments

Ligations were done using T4 DNA ligase in 20 µL reactions. Reactions were typically made using a 3:1 insert:vector ratio with 2 µL of 10x T4 DNA ligase buffer, 1 µL T4 DNA ligase, and ddH_2O to give a final volume of 20 µL. Reactions were either left at room temperature for 2 hours or at 17 °C overnight. On occasion a specific Quick Ligase (New England Biolabs) was used, these ligations were left for 30-60 mins at room temperature.

2.13 Site-directed mutagenesis

2.13.1 Buffers and reagents required

- Q5 Site-Directed Mutagenesis Kit supplied by New England Biolabs

- Q5 Hot Start High-Fidelity DNA Polymerase Master Mix – Q5 DNA polymerase, reaction buffer, dNTPs, Mg^{2+}
- Enzyme Mix – 10x Kinase-Ligase-DpnI Mix (KLD)
- NEB 5-α Competent *E. coli*

2.13.2 Site-directed Mutagenesis (SDM)

SDM was done using a Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The primers used were designed using the New England Biolabs online primer design tool for SDM (www.NEBaseChanger.neb.com), in which you input your sequence and indicate the specific mutation required and the primer is recommended by the design tool. The technique used to introduce a single base substitution involves generating two primers whose 5' ends anneal back-to-back, with the desired mutation in one of the primers, note that this WT DNA fragment needs to be already in a plasmid. These primers are used in a 25 µL PCR reaction using the supplied Q5 DNA polymerase master mix in a standard PCR cycle using an elongation time of 30 seconds per kb. Following the amplification, 1 µL of the linearised plasmid PCR product is then phosphorylated and ligated together using the KLD mix, which also contains DpnI to digest any template DNA, for 5 mins at room temperature. Transformation into NEB 5-α is done using 5 µL of KLD reaction mix and follows a standard heat shock transformation protocol.

2.14 Sequencing of DNA or RNA

2.14.1 Plasmid DNA and DNA fragments

For general Sanger sequencing to confirm DNA fragments or plasmid sequences, either the Functional Genomics and Proteomics Facility at the University of Birmingham or Eurofins Genomics was used. The inhouse facility at the University of Birmingham required 10 µL volumes containing 3.2 pmol of primer and ~200-300 ng of DNA if possible. DNA was diluted accordingly if too concentrated; if too dilute, then the max volume was used if the fragment could not be remade. For Sequencing with Eurofins Genomics, DNA concentrations were required to be between 10-100 ng/µL, with varying volumes required depending on the number of sequencing reactions required. Primers could either be synthesised by Eurofins or be sent separately if required and 15 µL of 10 pmol/µL were typically sent.

2.14.2 DNA libraries

For sequencing DNA libraries following ChIP-seq, sequencing was done using an Illumina MiSeq (Illumina) and a MiSeq Reagent Kit V2. Following purification and amplification, pooled libraries (typically 2 nM) were denatured using 10 µL of 0.2 N NaOH for 5 mins at room temperature. 10 µL 200 mM Tris-HCl pH 7 was then added and the sample pH checked and adjusted to exactly pH 7 using concentrated HCl or NaOH. To this, 990 µL of HT1 buffer was added to give 1 ml of a denatured library of 10 pM. For the PhiX control, this process was repeated with 12.5 pM of PhiX control library. 6 µL of the control library was then added to 594 µL of the pooled library to give a final volume of 600 µL. The samples were then loaded onto the sample cartridge following the manufacturer's instructions and the sequencing run initiated.

2.14.3 RNA libraries

RNA sequencing was done by Vertis Biotechnologie AG using the Cappable-seq method described by Ettwiller *et al.* (2016). 5 µg of total RNA libraries extracted from each strain (2.26.2) was sent for processing and sequencing.

2.15 β-galactosidase assay

2.15.1 Buffers and reagents required

- 1 M sodium carbonate (Na_2CO_3)
- Z-buffer – 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_2O$, 0.75 g KCl, 0.25 g $MgSO_4$. Made to 1 litre.
- ONPG solution – 100 mg 2-nitrophenyl β-D-galactopyranoside (ONPG) supplied by Sigma and 677 µL β-mercaptoethanol (C_2H_6OS) dissolved in 250 mL Z-buffer to give a final ONPG concentration of 13 mM. ONPG solution was made fresh for each experiment.
- 1% (*w/v*) sodium deoxycholate ($C_{24}H_{39}O_4Na$)
- 100% (*w/v*) toluene (C_7H_8)

2.15.2 β-galactosidase assay

For β-galactosidase assays the promoter fragments were cloned upstream of *lacZYA* operon in pRW50T and the sequence confirmed by Sanger sequencing as described in 2.14.1. pRW50T is a derivative of the plasmid pRW50 but contains an origin of transfer for use in conjugation

experiments into bacteria that cannot be transformed by heat shock or in which the plasmid is too large for electroporation to work effectively. Following confirmation of sequencing, the pRW50T constructs were conjugated into the relevant strains of *Salmonella* for use in the β-galactosidase assays. Each strain was assayed three times using three individual overnight cultures from separate colonies. The following morning, subcultures were set up using 200 µL of overnight culture in 5 mL fresh LB and grown to an OD_{600} of 0.6. Toluene and 1 % sodium deoxycholate (two drops each) were used to lyse the cells. The lysates were incubated at 37 °C for 20-30 mins to allow toluene to evaporate. For the β-galactosidase assay, 100 µL of lysate was added to a clean test tube and 2.5 mL of ONPG solution was added at set intervals (usually 10 seconds) before incubating the samples for 30-40 mins at 37 °C until the samples had turned a straw-yellow colour. 1 M sodium carbonate was used to stop the reaction, 1 mL was added at timed intervals (usually 10 seconds) to ensure each sample had the same reaction time. The absorbance of each sample was measured at OD_{420}. The activity of the promoter is calculated in Miller units using the following equation:

$$Promoter\ activity\ = 1000\ \times \frac{Abs_{420} \times total\ reaction\ volume}{Abs_{650}\ \times volume\ lysate\ \times\ reaction\ time}$$

For each β-galactosidase assay done, a medium-only control was included for use as a blank. Another control was the use of empty pRW50T to show the background β-galactosidase activity.

2.16 Crystal violet biofilm staining assay

2.16.1 Buffers and reagents required

- 0.1 % (*w/v*) crystal violet ($C_{25}H_{30}N_3Cl$) solution. 50 mg crystal violet powder in 50 mL ddH_2O.
- 70 % (v/v) Ethanol

2.16.2 Crystal violet biofilm staining assay

Crystal violet biofilm staining assays were done as described in Baugh *et al.* (2014) and was used to quantify biofilm production in *Salmonella*. Each strain was repeated in duplicate and two individual overnight colonies were set up. The following day, the overnight cultures were diluted to an OD_{600} of 0.1. 200 µL of each was aliquoted into a well of a flat-bottomed 96-well polystyrene microtitre plate. Each sample was aliquoted into four wells per culture. The plate was incubated for 48 hours at 30 °C without shaking. Unattached cells were washed away using water before staining with 200 µL of 0.1 % crystal violet for 15 minutes. The wash was repeated to remove any free crystal violet before the remaining crystal violet was resolubilised using 200 µL of 70 % ethanol. A quantitative measure of biofilm formation was obtained by measuring the OD_{600} of the resolubilised crystal violet using a CLARIOstar (BMG LABTECH).

2.17 Congo Red assays

Congo Red assays were done by supplementing no salt LB agar with a final concentration of 40 µg/mL Congo Red before solidification. Bacterial cultures were grown overnight and then diluted 1:10,000 in sterile no salt LB broth before spotting 5 µL onto the Congo Red plate. The plates were incubated at 30 °C for 48 hours before photos taken of the colonies.

2.18 Purification of recombinant proteins

2.18.1 Buffers and reagents required

Purification of MarA, SoxS, Rob, and *RamA*:

- 1 M IPTG solution
- Lysis buffer – 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 M NaCl
- Wash buffer – 4 M Urea, 50 mM Tris-HCl (pH 8.5)
- Denaturing buffer – 6 M Guanidinium-HCl, 50 mM Tris-HCl (pH 8.5)
- Buffer A/B – 1 M NaCl, 50 mM Tris-HCl (pH 8.5), 1 M imidazole (Buffer B only)
- Dialysis buffer – 1 M NaCl, 50 mM HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesufonic acid, pH 8.5)

Purification of E. coli RNA polymerase:

- TGED buffer – 10 mM Tris-HCl, 5 %, glycerol, 0.1 mM EDTA, 0.1 mM DTT, pH 7.9
- Buffer A – TGED supplemented with 0.1 M NaCl
- Buffer B – TGED supplemented with 1 M NaCl
- Grinding buffer – 50 mM Tris-HCl, 2 mM $MgCl_2$, 2 mM EDTA 150 mM NaCl, 5 % glycerol, 1 mM β-mercaptoethanol, 0.1 mM DTT, 1 protease inhibitor cocktail per 10 mL (Roche cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail used here), 0.2 % triton X-100 or Tween 20, pH 7.9
- Storage buffer – 10 mM Tris-HCl, 50 % glycerol, 0.1 M NaCl, 0.1 mM EDTA, 0.1 mM DTT
- 10 % Polymin P (Polyethylenimine) – protected from light and adjusted pH to 7.9 using HCl before use, spun at 4,000 rpm for 5-10 mins and supernatant kept.

2.18.1 Purification of MarA/SoxS/Rob/RamA/MlrA from SL1344

Proteins were expressed using pET28a. The protein coding sequences were amplified by PCR using a genomic DNA prep and digested with *NdeI* and *Bam*HI before ligating downstream of an IPTG inducible T7*lac* promoter in pET28a. Ligations were transformed into *E. coli* T7 Express once the sequence had been confirmed. Protein purification was adapted from Jair *et al.* (1995). Proteins were overexpressed and purified from 2x 1 L LB + 1% glucose cultures. Each 1 L culture was subcultured with 40 mL overnight culture and grown to an OD_{600} of 0.8 before the addition of IPTG to a final concentration of 0.4 M. Cultures were incubated for a further 3 hours with vigorous shaking. Cells were then harvested by centrifugation at 1,600 xg for 15 mins at 4 °C and resuspended in 25 mL lysis buffer. The cells were harvested as before, and the pellet frozen at -80 °C until required.

The frozen pellet was thawed in 40 mL of lysis buffer and kept at 4 °C for the remainder of the purification. Cells were lysed using an EMULSIFLEX-C3 (Avestin) before centrifuging at 75,000 xg for 30 mins. The supernatant was discarded, and the pellet resuspended in 40 mL wash buffer before repeating the high-speed centrifugation. The supernatant was, again, discarded and the pellet resuspended in 40 mL of denaturation buffer. The centrifugation step was repeated, but the supernatant was kept and loaded onto a HisTrap 5 mL precharged Ni Sepharose High Performance column (supplied by Sigma). Unbound proteins were washed away using buffer A before elution of the bound protein with Buffer B. Buffer B was added in a linear gradient until the protein eluted (usually at between 0.2 and 0.4 M imidazole). Purified fractions were run on an SDS-PAGE gel to check for purity before dialysis overnight

into dialysis buffer. Purified, dialysed, protein was concentrated to 1 mg/mL using a Vivaspin 20 column (5,000 molecular weight cut off) before storage at -20 °C.

2.18.1 Purification of RNA polymerase from *E. coli*

For the purification of *E. coli* RNAP, a modified protocol from Burgess and Jendrisak (1975) was used. 2x 1 L cultures of *E. coli* JCB387 were grown to an OD_{600} of 0.8 before centrifugation at 1,600 xg for 15 mins to pellet the cells. Cells were resuspended in grinding buffer (10-20 mL of buffer for every g of cells) before lysozyme was added (0.25 mg/mL) and the cells incubated on ice for 30 mins. Unless specified otherwise, all remaining steps took place at 4 °C. The sample was split into 30 mL fractions and lysis was done using a tip sonicator for 4 x 30 second pulses at 20 % output for each 30 mL fraction. Lysates were pooled and centrifuged at 23,000 xg for 10 mins.

The supernatant was passed through a 0.45 µM filter and Polymin P was added to a final concentration of 0.35 % and stirred for 5 mins before centrifugation at 3,000 g for 15 mins. The pellet was resuspended in 70 mL buffer A and centrifuged as before. The pellet was then resuspended in 70 mL buffer B before centrifugation at 3,000 xg for 30 mins. The supernatant was kept and subjected to an ammonium sulphate cut by the gradual addition of 24.5 g finely ground ammonium sulphate to 50 % saturation whilst stirring for 5 mins. The sample was then centrifuged at 5,000 xg for 45 mins and the pellet resuspended in up to 50 mL buffer A. The sample was loaded onto a Heparin HiPrep FF 16/10 column equilibrated in buffer A. Elution of protein was done using a linear gradient of buffer B up to 100 % buffer B (1 M NaCl).

Fractions were checked using an SDS PAGE gel and fractions containing the polymerase were combined and subjected to another ammonium sulphate cut before centrifugation at 5,000 xg for 45 mins. The pellet was then resuspended in up to 50 mL buffer A and loaded onto a Mono Q HR 5/5 or HiTrap Q HP column equilibrated with buffer A. The protein was eluted with a linear gradient of buffer B as before and the fractions checked using an SDS PAGE gel. Fractions containing RNAP were collated, quantified by Bradford assays, and concentrated if required before dialysis overnight against storage buffer. Aliquots were stored at -80 °C long term or -20 °C short term.

2.19 Bradford assay

2.19.1 Buffers and reagents required

- Bradford Dye Reagent (Alfa Aesar)
- BSA (20 mg/ml) (New England Biolabs)

2.19.2 Bradford Assay

Bradford Dye Reagent and protein concentration calculations were done following manufacturers protocol. A BSA standard was used at concentrations of 0, 0.2, 0.6, 0.9, and 1.2 mg/mL. All standards were made to 100 µL and protein samples were used neat (or diluted in the relevant buffer if too concentrated), before 3 mL of room temperature Bradford Dye Reagent was added and the samples vortexed. Samples were left at room temperature for 5-

30 mins and all OD_{595} readings were taken within 5 mins of each other. The standards were used to generate a standard curve to which the samples were compared.

2.20 End-labelling of DNA fragments

2.20.1 Buffers and reagents required

- G-50 sephadex beads – resuspended and washed three times with Tris-EDTA (TE) before resuspension in TE to give a 12 % (w/v) slurry.
- TE – 10 mM Tris-HCl, 1 mM EDTA, pH 8.0
- T4 polynucleotide kinase (New England Biolabs)
- T4 polynucleotide kinase buffer (New England Biolabs) – 70 mM Tris-HCl, 10 mM $MgCl_2$, pH 7.6
- [γ-^{32}P]-ATP supplied by Perkin Elmer or Hartmann Analytic – 10 µCi/µL

2.20.2 End-labelling of DNA fragments

For radiolabelling DNA fragments for use in electrophoretic mobility shift assays, DNA fragments were amplified by PCR and cloned into pSR using *Eco*RI and *Hin*dIII. Whilst the pSR plasmid was primarily used for *in vitro* transcription assays, it was also used for the generation of DNA fragments for radiolabelling. Fragments were digested out of pSR using *Aat*II and *Hin*dIII and radiolabelled at both ends using T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions. 1 µL of [γ-^{32}P]-ATP was used in a 20 µL reaction

volume. Unincorporated radionucleotides were removed using two G-50 Sephadex columns (BioRad).

2.21 Electrophoretic mobility shift assay

2.21.1 Buffers and reagents required

Electrophoretic mobility shift assay:

- 10x TNSC buffer – 40 mM Tris acetate 10 mM $MgCl_2$, 1 M KCl, 10 mM DTT, pH 7.9
- 5x TBE – 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA. Supplied by Fisher Scientific.

7.5 % acrylamide gel electrophoresis:

- UreaGel Concentrate (National Diagnostics) – 237.5 g acrylamide, 12.5 g methylene bisacrylamide, and 7.5 M urea per 1 litre.
- UreaGel Diluent (National Diagnostics) – 7.5 M urea
- UreaGel Buffer (National Diagnostics) – 0.89 M Tris-Borate, 20 mM EDTA, 7.5 M urea, pH 7.8.3
- ProtoGel (National Diagnostics) – 37.5:1 acrylamide:bisacrylamide
- 10 % (w/v) APS solution – 100 mg APS (Sigma) in 1 mL ddH_2O. Made fresh for each gel
- TEMED

2.21.2 Electrophoretic mobility shift assay (EMSA)

DNA fragments labelled as in 2.20.2 were used in EMSA experiments. Reactions were set up as follows: labelled DNA fragment (~50 counts), purified protein (MarA, SoxS, Rob, RamA, or MlrA), herring sperm DNA (as a non-specific competitor) at a final concentration of 12.5 µg/mL and 1x TNSC buffer. Reactions were incubated at 37 °C for 20-30 mins before being run on a 7.5 % (w/v) polyacrylamide gel at 250 V in 0.5x TBE for ~2 hours. The gel was then vacuum dried and exposed to a phosphor screen overnight before imaging using a BioRad FX phosphoimager.

2.22 *In vitro* transcription assays

2.22.1 Buffers and reagents required

In vitro transcription assay:

- E. coli RNA polymerase core enzyme (purified in house)
- Sigma 70 (purified in house)
- 10x TNSC buffer – 40 mM Tris acetate 10 mM $MgCl_2$, 1 M KCl, 10 mM DTT, pH 7.9
- 1 mg/mL bovine serum albumin (BSA) – final concentration of 100 µg/mL
- [α-^{32}P]-UTP supplied by Perkin Elmer or Hartmann Analytic – 10 µCi/µL
- NTP mix – 1 mM ATP/CTP/GTP, 50 µM UTP. Final concentrations of 200 µM ATP/CTP/GTP and 10 µM UTP in reaction.
- STOP solution – 97.5 % (v/v) deionised formamide, 10 mM EDTA, 0.3 % (v/v) Bromophenol Blue/Xylene Cyanol.
- 5x TBE – 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA. Supplied by Fisher Scientific.

6 % denaturing sequencing gel:

- UreaGel Concentrate (National Diagnostics) – 237.5 g acrylamide, 12.5 g methylene bisacrylamide, and 7.5 M urea per 1 litre.
- UreaGel Diluent (National Diagnostics) – 7.5 M urea
- UreaGel Buffer (National Diagnostics) – 0.89 M Tris-Borate, 20 mM EDTA, 7.5 M urea, pH 7.8.3
- ProtoGel (National Diagnostics) – 37.5:1 acrylamide:bisacrylamide
- 10 % (w/v) APS solution – 100 mg APS (Sigma) in 1 mL ddH_2O. Made fresh for each gel
- TEMED

2.22.2 *In vitro* transcription assays

Promoter DNA fragments of interest were cloned into pSR using *Eco*RI and *Hin*dIII and used as a template for *in vitro* transcription experiments. For each reaction, 335 ng of pSR was combined with 100 µg/mL BSA, NTP mix, 1x TNSC buffer, and 4 µCi [α-^{32}P]-UTP to give a volume of 11 µL. Transcription was started by the addition of 5 µL of RNAP mix (RNAP core enzyme and σ-70 in a 10:1 ratio) and incubated at 37 °C for 10 mins. The reactions were stopped with 20 µL of STOP solution and 4 µL was run on a 6% (w/v) denaturing PAGE sequencing gel at 60-80 W in 1x TBE for 1.5-2 hours. The gel was vacuum dried and exposed to a phosphor screen as in 2.21.2.

2.23 G + A ladder generation

2.23.1 Buffers and reagents required

- 3 M sodium acetate
- 100 % ethanol (v/v)
- 70 % ethanol (v/v)
- 100 % (v/v) formic acid
- 1 M piperidine
- DNase I blue – 5 M urea, 20 mM NaOH, 1 mM EDTA, 0.025 % (v/v) bromophenol blue, 0.025 % (v/v) xylene cyanol.

2.23.2 Preparation of G + A ladder

A G + A ladder was generated by first cloning the required sequence into pSR before digestion with *Hin*dIII and treatment with Calf Intestinal Alkaline Phosphatase (New England Biolabs). The fragment is then cut from the linearised plasmid using *Aat*II and radiolabelled as in 2.20.2, leading to a fragment that is only labelled at the *Hin*dIII end. 12 µL of this fragment was mixed with 50 µL formic acid and incubated for 2.5 mins at room temperature. The reaction was stopped by the addition of 200 µL 0.3 M sodium acetate, 1 µL glycogen and 700 µL ice-cold ethanol (100 %) and DNA precipitated at -80 °C for 30 mins. DNA was centrifuged at 21,000 xg at 4 °C for at least 30 mins. The pellet was dried under vacuum and then resuspended in 100 µL 1 M piperidine (1 in 10 dilution from 10 M stock). The sample was then incubated at 90 °C for 30 mins. The DNA was then precipitated at -80 °C for 30 mins using 1 µL glycogen, 10 µL 3 M sodium acetate and 300 µL ice-cold 100 % ethanol. The sample was centrifuged as before and washed twice with 700 µL ice-cold 70 % ethanol (v/v) as described in 2.11.1. Once

the pellet had been vacuum dried, it was resuspended in 20 µL of DNase I blue and stored at -20 °C.

2.24 ChIP-seq in *Salmonella enterica* Typhimurium SL1344

2.24.1 Buffers and reagents required

- 10 mM Tris pH 8.0. Referred to as Tris 8 in this study.
- 10 mM Tris pH 7.5. Referred to as Tris 7.5 in this study.
- Immunoprecipitation buffer – 50 mM Hepes-KOH, pH7, 150/500 mM NaCl, 1 mM EDTA, 1 % (w/v) Triton X-100, 0.1 % (w/v) Sodium deoxycholate, 0.1 % (w/v) SDS. Two versions of immunoprecipitation (IP) buffer were made, one with 150 mM NaCl and one with 500 mM NaCl; which will be called IP 150 buffer or IP 500 buffer in this study.
- ChIP wash buffer – 10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5 % (w/v) Nonidet-P40, 0.5 % (w/v) Sodium Deoxycholate.
- ChIP elution buffer – 50 mM Tris-HCl, pH 7.5, 10 mM ETDA, 1 % (w/v) SDS.
- TE – 10 mM Tris-HCl, pH 8.0, 1 mM EDTA.
- 1x TBS – 20 mM Tris-HCl, pH7.4, 0.9 % (w/v) NaCl.
- Quick blunting and quick ligation kits supplied by NEB
- Klenow fragment (3'→5' exo-) supplied by NEB
- Protein A sepharose beads – Washed with ddH_2O and stored in a 50 % (v/v) slurry with 1x TBS. Supplied by GE healthcare.
- Agencourt AMPure magnetic beads XP DNA clean-up kit supplied by Beckman Coulter.
- NEXTflex ChIP-seq barcodes supplied by BioOscientific.

- Monoclonal anti-FLAG, anti-cMyc (Thermo Fisher), anti-σ^{70} antibody (Neoclone).
- Qubit flurometer/assay kit.
- NEBNext library quant kit for Illumina supplied by NEB.
- Acrylamide extraction buffer – TE as above supplemented with 0.4 M NaCl.

2.24.2 ChIP-seq in *Salmonella enterica* Typhimurium SL1344

All ChIP-seq experiments follow the procedure described in (Sharma *et al.*, 2017b) and were done in duplicate. For each ChIP-seq on a tagged protein (Table 2.2), a control was done using a strain of *Salmonella* SL1344 with an empty plasmid containing just the tag. An overnight culture of *Salmonella* SL1344 containing the tagged protein to be studied was sub-cultured into 40 mL of fresh LB + kanamycin and grown to mid-log phase (0.6 OD_{600}). Cells were then crosslinked with 1 mL formaldehyde (to give a final concentration of 1% (v/v) and incubated for 20 mins before quenching with 10 mL 2.5 M glycine. Centrifugation at 1,600 xg for 5 mins was used to recover the cells, which were then washed with 25 mL 1x TBS. The cells were re-harvested as before and re-suspended in 1 mL 1x TBS before transferring to a 1.5 mL microfuge tube and centrifuged at 21,000 xg for 1 min. The cell pellet was then re-suspended with 1 mL IP 150 buffer supplemented with either 2 or 4 mg/mL lysozyme and incubated at 37 °C for 30 mins to lyse the cells. Following this incubation step the cells were chilled on ice briefly and sonicated using a Bioruptor Plus sonicator (Diagenode) for 30 cycles of 30 s on 30 s off at 4 °C. The sonicated lysates were centrifuged at 21,000 xg for 5 mins to remove any cell debris and the supernatant divided into 4 separate microfuge tubes and diluted with 800 µL IP 150 buffer, one of which was used per immunoprecipitation (IP).

Prior to IP, Protein A sepharose beads were washed and made to a 50 % slurry (v/v) with 1x TBS. Due to their delicate nature blunt pipette tips were used in all steps containing Protein A beads. IPs were done using one 800 µL lysate, 25 µL of Protein A beads and 2 µL of anti-FLAG or anti-Myc antibodies; the IP cocktails were rotated for 90 minutes at room temperature.

Multiple sets of washes and enzymatic reactions were used to prepare the DNA fragments for Illumina sequencing. First, the Protein A bead-antibody-Protein-DNA complexes were collected by centrifugation at 1,600 xg for 1 min before resuspending in 700 µL fresh IP 150 buffer. The complexes were then transferred to a Spin-X column and rotated for 3 mins at room temperature, centrifuged at 1,600 xg for 1 min and the supernatant discarded. For the remaining wash steps of the ChIP-seq experiment (unless otherwise stated) all washes were done using 700 µL of the indicated buffer, rotated at room temperature for 3 min, centrifuged at 1,600 xg for 1 min and the supernatant discarded. The wash steps were separated into four steps, with an enzymatic reaction in between each step. Wash step 1 was as follows: 1 IP 150 buffer wash and 2 Tris 7.5 washes. The DNA fragments were blunted using a quick blunting kit (NEB) with 10 µL 10 of 10x blunting buffer, 10 µL of dNTP mix supplied with the kit and made to 100 µL with ddH_2O before the addition of 2 µL blunting enzyme. The reactions were then rotated at room temperature for 30 mins in a way that did not invert the Spin-X columns. Wash step 2 was then done as follows: 2 IP 150 buffer washes and 2 Tris 8 washes. The next enzymatic step added an 'A tail' to each DNA fragment using the Klenow fragment (3'→5' exo-), with each reaction consisting of 10 µL 10x NEB buffer 2 (supplied with the Klenow fragment), 2 µL 100 µM dATP and made up to 100 µL with ddH_2O before the addition of 2 µL Klenow

fragment. The reactions were incubated at 37 °C for 30 mins with rotation in such a way that did not invert the Spin-X columns. Wash set 3 was done as follows: 2 IP 150 buffer washes and 2 Tris 7.5 washes. The final enzymatic step ligated the NEXTflex barcoded adaptors (BioOscientific) to the DNA fragments using 100 µL 1x ligase buffer (NEB), 1 µL NEXTflex barcoded adaptors and 4 µL quick ligase (NEB). The reactions were rotated in such a way to prevent inversion for 15 mins at room temperature. Wash step 4 was done as follows: 2 IP 150 buffer washes, 1 IP 500 buffer wash, 1 ChIP wash buffer wash and 1 TE wash. DNA was eluted by transferring the Spin-X column basket to a fresh dolphin nosed tube and incubated at 65 °C for 10 mins in 100 µL ChIP elution buffer. Following the incubation, the reactions were quickly transferred to a centrifuge and spun for 1 min at 1,600 xg. The samples were then de-crosslinked by boiling for 10 mins.

Prior to library amplification, the samples were subjected to a 1.1x volume bead clean up and eluted in 13 µL ddH_2O. Following this bead clean up, 2 µL of library was used in either a PCR or a qPCR reaction to empirically determine the number of cycles to amplify each library for. Each PCR reaction was as follows: 2 µL ChIP-seq library, 1 µL of each NEXTflex primers 1 and 2 (BioOscientific), 1 µL 100 mM dNTPs, 10 µL 5x Velocity buffer (Bioline) and 1 µL Velocity (Bioline). A 65 °C annealing temperature was used, and the PCR reactions amplified for 33 cycles with 5 µL samples being taken every 3 cycles from cycle 18 onwards. PCR samples were run on an agarose gel and post stained with ethidium before imaging. The number of cycles for library amplification proper was chosen based on this gel, cycle number were chosen in such a way that maximised library amplification but minimised both NEXTflex barcode adapter

or PCR primer dimers and overamplified library concatemers with self-primed daisy chaining libraries. Libraries were then amplified as above for the required number of cycles. Amplified libraries were diluted to 200 µL with ddH_2O and subjected to a 0.7x bead clean up before imaging on an Agilent TapeStation 2200 (Agilent); should a peak be observed at 150 bp or a large shoulder (800-1,000 bp) seen on the library size distribution curve then an acrylamide gel extraction was done. If amplification was done using qPCR, then a threshold of 0.1 was set manually and each library was amplified for the required number of cycles using standard PCR.

Following library amplification, the library concentration was quantified using an NEBNext Library Quant Kit (NEB) using 1 µL of library following manufacturer's instructions. Libraries were diluted to a chosen concentration (0.5, 1 or 2 nM) before pooling to generate an equimolar mix for sequencing on an Illumina MiSeq.

2.25 Bioinformatic analysis of ChIP-seq data

2.25.1 Bioinformatic processing of ChIP-seq data

Bioinformatic analysis of ChIP-seq data was done as in (Sharma *et al.*, 2017b). Briefly, raw FASTQ files were extracted from the Illumina MiSeq and uploaded to usegalaxy.org for processing and analysis. FASTQGroomer was used to convert files to the FASTQ Sanger format, which were then aligned to the *Salmonella enterica* Typhimurium SL1344 genome (NC_016810.1), pCol1B9 (NC_017718.1), pRSF1010 (NC_017719.1), or pSLT (NC_017720.1)

using Bowtie 2 for Illumina. The resulting files were then converted to BAM format using SAM-to-BAM before determining the coverage per base using multiBamSummary. Further analysis was done using R. Normalising the data to the same average read depth and determining the mean coverage per base for each pair of replicates was done. The empty plasmid controls were subtracted from the IPs proper and viewed on Artemis (Carver *et al.*, 2012). Peaks were called from this coverage plot using a manual threshold determined by visual inspection of the data. Circular DNA plots were generated using DNAPlotter (Carver *et al.*, 2009).

Further bioinformatic analysis of the ChIP-seq binding targets was done by taking the centre of each peak with 100 bp either side and submitting this information to MEME (Bailey *et al.*, 2009) with the conditions that only 1 binding peak is expected per binding peak and that the minimum site for a motif was 15 bp. The P-value was generated by MEME and factors in the motif length and background DNA sequence to give the log likelihood ratio of the motif.

2.25.2 Bioinformatic analysis of ChIP-seq peak conservation

The conservation of the peaks identified in the ChIP-seq experiment was analysed as described in Sharma *et al.* (2017b). Briefly, 100 bases upstream and 100 bases downstream from the ChIP-seq peak centre were extracted from the *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344 reference genome (accession number: FQ312003.1 or NC_016810.1), giving a fragment 201 nucleotides in length with the centre of the ChIP-seq peak in the middle. These fragments were then submitted to BLASTn to search for similar

sequences in 26 strains (Table 2.4) known to contain a functional MarA protein (Sharma *et al.*, 2017b).

All ChIP-seq peaks were subjected to a BLASTn search on the taxid numbers of the strains in Table 2.4 and the output alignments saved. The search was limited to only the strains listed in Table 2.4; as they are known to encode functional MarA proteins without any mutations in the DNA binding helices (Sharma *et al.*, 2017b). Each ChIP-seq peak alignment text file was manually searched for the binding site identified by MEME (described above). The resulting binding sites identified in the alignments were manually entered into an excel sheet 1 cell per base and 1 sheet per ChIP-seq peak. If the binding site was only partially present, the aligning bases were input, and the rest left blank. Similarly, if multiple binding sites were present then the binding site with the closest match to the consensus was used. Each alignment binding site was compared against two reference binding sites, the binding site identified in the ChIP-seq experiment and shown in Table 3.1 (referred to here as the query binding site) and the consensus binding site (5'-gcactaattgctaaa-3') identified in *E. coli* by Sharma *et al.* (2017b) (referred to as the consensus binding site here). To determine if the regions were conserved, each base of the alignments was checked against both the query and consensus binding sites. A base was considered conserved if found in either reference binding sites and scored as a 1. The score for each alignment binding site was calculated by summing the base conservation scores for each alignment binding site. A score less than or equal to 12 was not considered conserved, a score of 13 was considered 'conserved with mismatches', a score of 14 or 15 was considered conserved.

Table 2.4 Strains submitted to a BLAST search for identifying conservation of ChIP-seq binding sites

Strain	NCBI Taxonomy ID number
Salmonella enterica subsp. *enterica* serovar Typhimurium LT2	taxid:99287
Salmonella enterica subsp. *enterica* serovar Typhimurium str. 14028S	taxid:588858
Salmonella bongori N268-08	taxid:1197719
Salmonella bongori NCTC 12419	taxid:218493
Salmonella bongori serovar 48:z41:-- str. RKS3044	taxid:1382510
Citrobacter rodentium ICC168	taxid:637910
Citrobacter koseri ATCC BAA-895	taxid:290338
Escherichia albertii KF1	taxid:1440052
Escherichia fergusonii ATCC 35469	taxid:585054
Escherichia coli MG1655	taxid:511145
Shigella flexneri 2a str. 301	taxid:198214
Kosakonia oryzae	taxid:497725
Kluyvera intermedia ATCC 33110	taxid:1218113
Lelliottia amnigena	taxid:61646
Leclercia adecarboxylata ATCC 23216	taxid:911008
Enterobacter cloacae complex sp. 20432	taxid:1812935
Cronobacter malonaticus LMG 23826	taxid:1159491
Raoultella ornithinolytica B6	taxid:1286170
Raoultella ornithinolytica ATCC 31898	taxid:1349784
Raoultella ornithinolytica 10-5246	taxid:883121
Klebsiella pneumoniae subsp. *pneumoniae* MGH 78578	taxid:272620
Pluralibacter gergoviae ATCC 33028	taxid:1354260
Shimwellia blattae DSM 4481	taxid:630626
Cedecea neteri ATCC 33855	taxid:1216978
Cedecea sp. M006	taxid:158822
Cedecea sp. ND02	taxid:158822

2.26 Cappable-seq in *Salmonella enterica* Typhimurium SL1344

2.26.1 Buffers and reagents required

- PBS buffer (10x) – 11.5 g N_a2HPO_4, 2 g KH_2PO_4, 80 g NaCl, 2 g KCl per 1 litre
- Trypsin-EDTA solution – 0.05 % trypsin (w/v), 0.53 mM EDTA, dissolve in 1xPBS
- RNA Lysis Buffer – 4 M guanidinium thiocyanate (GTC), 0.01 M Tris 0.97 % β-mercaptoethanol, pH 7.5
- DNase Stop Solution (concentrated) – 5 M GTC, 10 mM Tris-HCl, pH 7.5. Diluted with ethanol to give final concentration of 2 M guanidine isothiocyanate, 4 mM tris-HCl, 57 % ethanol (v/v), pH 7.5.
- RNA Wash Solution (concentrated) – 162.8 mM potassium acetate, 27.1 mM Tris-HCl, pH 7.5. Diluted with ethanol to give final concentration of 60 mM potassium acetate, 10 mM Tris-HCl, 60 % (v/v) ethanol, pH 7.5.
- Yellow Core Buffer – 0.0225 M Tris, 1.125 M NaCl, 0.0025 % yellow dye (w/v)

2.26.2 Cappable-seq in *Salmonella enterica* Typhimurium SL1344

Cappable-seq (Ettwiller *et al.*, 2016) was done by Vertis Biotechnologie AG on 5 µg of RNA extracted from each strain. All RNA extraction was done using the SV Total RNA Isolation System (Promega) following the manufacturer's instructions. In each strain, a transcription factor (MarA, SoxS, RamA, or an empty vector control) was overexpressed using the constitutive pAMNF or pAMNM plasmid as in the ChIP-seq experiments. Two replicates for each strain were used and grown to mid-log phase (OD_{600} of 0.6) before pelleting at 14,000 xg

for 2 mins and resuspending in 100 µL TE buffer containing 0.4 mg/mL lysozyme. The pellets were incubated for 3-5 mins at room temperature then 75 µL of RNA lysis buffer and 350 µL RNA dilution buffer was added, and the samples mixed by inversion. The lysate was then transferred to a fresh microcentrifuge tube and 200 µL 95% ethanol was added and mixed by pipetting. The samples were then centrifuged for one minute at 14,000 xg and the flow through discarded. 600 µL of RNA Wash Solution was added and the spin repeated. Following this wash the DNA within the samples were digested for 15 mins at room temperature using a 50 µL mixture comprising 40 µL Yellow Core Buffer, 5 µL 0.09 M $MnCl_2$ and 5 µL DNase I enzyme. To stop the reaction, 200 µL of DNase Stop Solution was added and the samples centrifuged for one minute at 14,000 xg. As before, 600 µL of RNA Wash Solution was added and the samples centrifuged for one minute at 14,000 g. The samples were washed further using 250 µL RNA Wash Solution and centrifuged at 14,000 xg for two minutes. Finally, the spin buckets were transferred to fresh microcentrifuge tubes and eluted using 100 µL of Nuclease-Free Water and centrifuging for one minute at 14,000 xg. The samples were stored at -80 ^{o}C until the quality was checked and concentration identified using an Agilent TapeStation 2200 (Agilent). RNA was shipped to Vertis Biotechnologies AG for Capable-seq (https://www.vertis-biotech.com).

2.27 Bioinformatics analysis of RNA-seq data

2.27.1 FastQ processing workflow

For the analysis of the RNA-seq data, following the Cappable-seq by Vertis Biotechnologie AG, the FASTQ files received were checked using `FastQC` (Version 0.11.9) to confirm that the

quality was sufficient for downstream processing, before aligning and mapping to the SL1344 reference genome and plasmids using `Bowtie2` and `SAMtools` (VERSION 1.3.1) in the following command (square brackets denote generic input or output files):

```
bowtie2 --local -x [genome-index] -U [sample.fastq.gz] | samtools view -bS - > [sample.bam]
```

Before identifying the transcription start sites (TSS) using the programs created by Ettwiller *et al* (Ettwiller *et al.*, 2016) for the analysis of Cappable-seq data sets, the BAM files were sorted using the following command from `SAMtools`:

```
samtools sort [sample.bam] -o [sample_sorted.bam]
```

The suite of Perl programs developed by Ettwiller *et al.* (2016) were employed to analyse the Cappable-seq files to identify the TSSs, generate .gtf files for further analysis, and the visualisation bam files for viewing the data; Artemis (Carver *et al.*, 2012) was used exclusively here and all references to visualisation refer to Artemis unless otherwise stated. The first Ettwiller *et al.* (2016) program `bam2firstbasegtf.pl` was used to generate the .gtf files for further analysis. Briefly, this program identifies the TSSs at single base resolution and generates a relative read score (RRS), which is calculated by taking the number of reads at each position (+ or -) and normalising to the total number of reads in the sample. The results are then filtered based on a cut off value, which Ettwiller *et al.* (2016) suggest as 1.5 as that is equivalent to 20 reads or more. This program was run using the following command:

```
perl bam2firstbasegtf.pl --bam [sample_sorted.bam] --cutoff 1.5 --out [sample_ettwiller.gtf]
```

Following the recommended workflow for samples without a control library (a sample without streptavidin enrichment), the `filter_tss.pl` program was not required and, therefore, `cluster_tss.pl` was used next. This program clusters nearby start sites based on a user specified distance and keeps the start site with the highest RRS. As the resulting file loses resolution, this program was used for the visualisation of the results only and not for further bioinformatic analysis (for example, in the analysis of bidirectional promoters). The program was run with a cut off value of 5 bases using the command:

```
perl cluster_tss.pl --tss [sample_ettwiller.gtf] --cutoff 5 --out [sample_ettwiller_cluster.gtf]
```

The output .gtf file from the above programs is not compatible with Artemis and a further program is required to view the start sites. This program was run using the following command:

```
perl bam2firstbasebam.pl --bam [sample.bam] --genome [reference_genome.fai] --out [sample_ettwiller_vis.bam]
```

Note that this program uses an index fasta (.fai) of the reference genome.

2.27.2 Analysis of *Salmonella* transcription start sites using custom made Python scripts

For further analysis, custom Python (Version 3.8) scripts were written for multiple functions. The code used is shown in full in the Appendix, but brief summaries follow here.

Processing of the Cappable-seq results

The first script used in the processing of the capable-seq results (`RNA_seq_analysis_multiple_samples_final.py`, 7.1.1) takes the non-clustered .gtf files (the output from `bam2firstbasegtf.pl`) and asks the user for the number of bases to be extracted upstream and downstream of the transcription start site (TSS); here, 100 bases upstream and 50 bases downstream were extracted. The flanking regions are then used to generate a set of coordinates for each TSS, which is saved as a .txt file for each strain. The script will then take the flanking regions and extract that sequence from the reference genome (supplied as a FASTA file), writing each TSS to a FASTA file with individual FASTA sequences for each TSS (151 bases in this analysis including the TSS). Further to this, the script will take 16 bases upstream of the TSS (inclusive), and 1 base downstream, and save these into a separate FASTA file for analysis of the -10 promoter element. Note that this script will extract the coordinates and sequences for each TSS in each replicate. A further two scripts were developed, similar to `RNA_seq_analysis_multiple_samples_final.py`, the first of which (`RNA_seq_combine_replicates.py`, 7.1.2) combines both replicates and removes any unique TSSs to give only the TSSs that are found in both replicates. The second script (`RNA_seq_total_combined_TSS.py`, 7.1.3) combines all .gtf files into one and removes any duplicated TSSs before extracting the coordinates and sequence fasta files as above.

Differential expression analysis

To determine the differential expression of each TSS in response to the transcription factors (MarA, SoxS, and RamA), EdgeR (McCarthy *et al.*, 2012, Robinson *et al.*, 2009) was used in a custom-made R script (written by Zuzana Palecková, not presented here). To generate a .csv input file for EdgeR, a Python script was made (`generate_EdgeR_inputs.py`, 7.1.4). This script takes four inputs, both replicates of a condition (MarA or SoxS or RamA) and both empty vector control replicates and combines them into one list containing only TSSs identified in all four replicates. The script then renames each TSS using the format 'TSS_[base]_[strand]'. EdgeR uses the level of coverage to determine if a gene is differentially expressed, which is not the same as the RRS score output by the Ettwiller *et al.* (2016) scripts. This coverage information is present in the output but contained within the name of each TSS, so a new column is generated using this value. From this, the script then saves a .csv file for each replicate which contains three columns, the TSS_ID, coverage for that replicate, and strand. These .csv files are then input into the EdgeR scripts and differential expression data is generated and plotted into a volcano plot using ggplot2 (Wickham, 2016).

In order to highlight the differentially expressed TSSs that fell within the ChIP-seq peaks identified, a custom-made python script was written. `Extract_ChIP-seq_coordinates.py` (7.1.5) takes the peak centre of each ChIP-seq binding target in Table 3.1 and creates a series of new columns, generating a column for both the start and end coordinates of 100 bases flanking or the start and end coordinates of 150 bases flanking. The ChIP-seq binding peaks are then extracted for each transcription factor to generate a separate list for each, with the start and end coordinates of each peak and the gene information being

saved as a .csv file. A complete list of all the ChIP-seq peaks was also saved separately. GenomicRanges (Lawrence *et al.*, 2013) was then used to find all instances where the TSSs identified in the RNA-seq data overlapped with the coordinates of the ChIP-seq peaks; these overlaps were then superimposed on the volcano plots.

Bidirectional promoter analysis

The number and distribution of bidirectional promoters was analysed using custom python scripts to identify bidirectional promoters and extract their -10 promoter element sequences. Then, the distribution of these promoters was assessed using spreadsheets to calculate the distribution of bidirectional promoters with respect to the nearest genes. Firstly, `merge_controls.py` (7.1.6) was used to combine both replicates of the empty vector control samples. This script takes both .gtf files output from the Ettwiller *et al.* (2016) scripts and merges them, keeping only the TSS found in both replicates. Only the TSS_ID, start, and strand columns are saved to a .csv file. Then, `bidirectional_analysis_final.py` (7.1.7) was used to identify both bidirectional and non-bidirectional promoters and save each into separate .csv files. For a pair of TSSs to be considered a bidirectional promoter, two conditions had to be met. Firstly, the upstream TSS must be on the minus strand and the downstream TSS on the plus strand and, secondly, that the distance between the start sites was between 7 and 25 bases as this allowed for overlap between the -10 promoter elements of each TSS. The script takes a given TSS and compares it to the downstream TSSs, until the distance between the two is greater than 25 bases. If the conditions are met by any of the pairs tested, then both TSSs are extracted, numbered as a bidirectional pair, and the

difference and coordinates of the pair are also saved. A list of non-bidirectional pairs is also saved, both complete and strand separated. The results are then compared to the reference genome, using the start and end coordinates of each gene. For the bidirectional promoter pairs, only the upstream TSS is used. Each TSS (bidirectional, non-bidirectional plus strand, and non-bidirectional minus strand) is then compared to the gene coordinates and orientation of each gene in the reference genome. The distribution of bidirectional or non-bidirectional promoters inside genes, between genes in the same orientation, between divergent genes, and between convergent genes is then calculated as a percentage.

To extract the -10 sequences from the bidirectional promoters identified above, `extract_bidirectonal_promoter_sequences.py` (7.1.8) was used. This script takes a reference genome as a FASTA file and asks for a user-specified number of flanking bases to be extracted. For this analysis, 20 bases upstream and downstream of the bidirectional pair was extracted. Briefly, the flanking bases are subtracted from the upstream TSS and added to the downstream TSS to give a set of coordinates, which are then extracted from the reference genome. As the most common distances between bidirectional TSSs identified in both *E. coli* (Warman *et al.*, 2021) and *Salmonella* (this study) was 7, 10, 12, 18, and 23 bases, the sequences of bidirectional promoters of these distances were each saved separately. Extracted sequences were submitted to Weblogo (Crooks *et al.*, 2004) to generate a sequence logo.

3. Identification of transcription factor binding sites in the *Salmonella* SL1344 genome

3.1 Introduction

This chapter presents chromatin immunoprecipitation and deep sequencing (ChIP-seq) analysis of binding targets for the transcription factors MarA, SoxS, Rob, and RamA in *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344. The roles of these TFs on the development of the MDR phenotype has been studied previously (Duval and Lister, 2013, Weston *et al.*, 2018, Demple, 1996, Pomposiello and Demple, 2000, Martin and Rosner, 2002, Baugh *et al.*, 2012, Piddock, 2006, Webber *et al.*, 2009, Webber and Piddock, 2001). However, prior studies into their respective regulons have relied on microarray technologies or molecular genetics approaches; and primarily focussed on *E. coli* (Martin and Rosner, 2002, Pomposiello and Demple, 2000, Duval and Lister, 2013, Seo *et al.*, 2015, Sharma *et al.*, 2017b).

The SL1344 genome has a slightly higher G+C content than *E. coli* K-12 MG1655 (Fookes *et al.*, 2011). Despite this, promoter elements are well conserved, with *S.* Typhimurium showing a more conserved extended -10 element but a slightly less conserved -35 element; suggesting that *Salmonella* places a stronger requirement on the -10 hexamer (Kroger *et al.*, 2012). As is expected for closely related species, their mechanisms of transcriptional initiation and regulation are similar (Lonetto *et al.*, 1992, Kroger *et al.*, 2012). In the following text, *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344 is abbreviated to SL1344 or *Salmonella* throughout. Both the binding motifs and sites identified by ChIP-seq are referred to as binding sites throughout this study.

3.2 ChIP-seq analysis of MarA, SoxS, Rob, and RamA binding in *Salmonella* SL1344

In the absence of specific inducers, MarA, SoxS, and Rob are repressed and present at low levels in the cell (Duval and Lister, 2013). Hence, to determine their genome-wide binding profiles, MarA, SoxS, Rob, and RamA were constitutively expressed at low levels from one of four plasmids: pAMNF, pAMCF, pAMNM, or pAMCM (Table 2.2). These permit tagging with either a 3x FLAG or 8x C-Myc tag, at the N- or C-terminus, respectively. Hence, four tagged variants of each TF were generated to facilitate ChIP-seq analysis. The resulting plasmids were used to transform SL1344. ChIP-seq was then done in duplicate for each variant and compared to control experiments with empty plasmid. The specific TF-tag combination that produced the best results was used for further analysis. For MarA, this was the 8x C-Myc tag at the N-terminus and for SoxS, Rob, and RamA this was the 3x FLAG tag at the N-terminus. The resulting genome-wide binding sites for MarA, SoxS, Rob, and RamA in SL1344 is shown in Table 3.1.

3.3 MarA, SoxS, Rob, and RamA binding loci in SL1344

ChIP-seq analysis revealed 100 binding loci for the four TFs in total. Of these 100 loci, 58, 38, 34, and 22 were bound by SoxS (green), MarA (blue), RamA (purple), and Rob (red) respectively (Table 3.1, Figure 3.1). Of the 58 identified SoxS binding targets, 38 are unique to SoxS (Table 3.1). Of all SoxS sites, 9 were previously identified in a ChIP-exo experiment in *E. coli* (Seo *et al.*, 2015). Further to this, 9 SoxS binding peaks were also seen to be bound by MarA in *E. coli* (Sharma *et al.*, 2017b).

Table 3.1 ChIP-seq binding targets of all four transcription factors on the SL1344 chromosome and virulence plasmids

Binding Protein	Peak Centre	Site Centre	Gene(s)	P-value[1]	Site Sequence (5' to 3')	*E. coli*[2] MarA[a] or SoxS[b]
MarA, RamA	52564	52620	*rpsT* < > *yaaY*	2.04E-04	CAAATCCATTGACAAAAGA	
MarA, Rob, SoxS	134070	134036	*leuL* < > *leuO*	1.38E-06	AGCACAATTAGCTAAAGTA	yes[a]
MarA, RamA, Rob, SoxS	156902	156874	*lpxC*	1.21E-04	GGCTCTTTGTGCTAAACTG	yes[b]
RamA	174940	174928	*aroP* < > *pdhR*	1.21E-04	TGCATTCGCGGCCACATAC	
MarA	202096	202029	*yadF* < > *yadG*	2.40E-04	TGCACTATGGTCAAAAATG	
SoxS	424246	n.d.	*hemB* < > *yaiU*	n.d.	n.d.	
SoxS	435010	435025	SL1344_0377	1.00E-04	GGAACCACCAGGAAAAAGA	
MarA	482680	482714	*phnS*	5.57E-04	GGCTTATATGACAAAACGA	
MarA, Rob, SoxS	497974	498029	*cyoA*	9.09E-06	ACCATCAATTGATAAAAAA	
SoxS	508034	507985	*cypD*	4.48E-04	AGCCTATTGTGACAAGAAA	
SoxS	515659	515659	SL1344_0452 < > *ybaO*	3.94E-07	GGCACAAAATGATAAATGG	yes[a,b]
MarA, RamA, SoxS	524017	524010	*ybaZ*	5.61E-05	GGCCCTGCCAGCTACATCC	
MarA, RamA, SoxS	533256	533255	*acrA* < > *aefA*	5.61E-05	GGCACGAAAAACCAAACAA	yes[b]
MarA, RamA	539646	539652	*priC* < > *apt*	4.16E-04	GGCGCAGGCGGTCAAAGAG	
SoxS	568176	568173	(ybbp)	2.23E-06	CGCACAATCGGATAAAACG	
Rob	598569	598459	*ppiB* < > *cysS*	4.57E-05	GGAACAGGATGCAAAAATG	
MarA	692417	n.d.	(*cspE*)	n.d.	n.d.	
MarA	711235	711324	*leuS* < > SL1344_0637	3.57E-04	GGCCCATAAAAATAAAGTC	
SoxS	757147	757191	*fldA*	2.38E-05	TGCACGCTCTGCTACACGA	yes[b]
RamA, SoxS	781807	781785	SL1344_0698	2.04E-04	GACAAAAATGGATACAGCA	
SoxS	792015	792022	SL1344_0709	2.59E-06	AGCATCGCGTGCTAAAAAA	
MarA, RamA, Rob, SoxS	844579	844497	*modE* < > *acrZ*	6.42E-04	ACCAGCTCCTGGTAAAAAG	yes[a,b]
RamA	898723	898691	*ybiF* < > *ompX*	3.06E-04	GAAACGTTCTGTTACATGA	
Rob	1014845	1014820	*pflB*	5.61E-05	TGCAGCAATGGCCAAAGTG	
MarA	1068935	1069003	SL1344_0962	2.04E-04	GGAATATACCACCAAAAAA	

SoxS	1186956	1186937	*csgD* < > *csgB*	1.87E-04	AGCACAAAGACAAAAAAAA	
RamA, SoxS	1292953	1292918	STnc1210	7.99E-06	GGCACAGATCGCTAAATAT	
MarA, RamA, Rob,	1416444	1416555	*lppB*	2.61E-04	TGCATTCCCATCAAAAAAA	
MarA, RamA	1465849	1465776	*purR* < > *ynhF*	6.88E-04	TGCCCGTTTCGCTACATCT	
MarA	1466921	1466965	*sodB*	3.86E-04	TAAACGACAGGATAAAATA	
RamA	1550776	n.d.	(STnc560)	n.d.	n.d.	
RamA, Rob, SoxS	1554865	1554877	*marR* < > *marC*	4.65E-06	GCCACGATTTGCTAAAAGG	yes[a]
SoxS	1603120	1603084	(*sfcA*)	8.27E-07	GGCACATTCTGCAAAATGT	
SoxS	1650151	1650167	*yncJ*	7.00E-06	CGCACTTATTGACAAACCG	
SoxS	1698924	n.d.	*nifJ*	n.d.	n.d.	
MarA, RamA, Rob	2064839	n.d.	(SL1344_1958)	n.d.	n.d.	
SoxS	2097359	2097370	(*cobU*)	1.32E-05	GGCACGTAGTGGTAAAAGC	
SoxS	2145247	2145230	(*yeeY*)	2.59E-06	AGCATTATTTGCTAAAATT	
MarA, RamA, SoxS	2364593	2364591	*ompC* < > *micF*	8.48E-08	AGCACTGAATGATAAAACA	yes[a,b]
SoxS	2520920	2520938	SL1344_2373 < > *ypeC*	3.23E-07	AGCATTTTTTGCTAAAACC	yes[a,b]
MarA	2594769	2594825	*ypfM* < > *yffB*	3.70E-05	AACCCAATTTGATAAAGTA	
MarA, RamA	2600650	2600628	*purC*	7.88E-04	CGAAATAGCGGTTAAATCG	
MarA, RamA, SoxS	2623708	2623719	*guaB* < > *xseA*	1.12E-08	AGCACTATTTGCAAAAAAA	yes[a]
MarA, RamA, SoxS	2759658	n.d.	(*isrJ*)		n.d.	
RamA, SoxS	2763581	2763538	SL1344_2584	2.23E-06	AGCACTTTTTGCAAAAGCT	
RamA	2767331	2767262	(*gpP*)	4.16E-04	AGCAGAAGTTGCTAACCAC	
Rob	2768467	2768394	*cIIa* < > SL1344_2594	3.86E-04	TGACTTGTTGGTAAAATGA	
RamA, Rob	2855168	n.d.	SL1344_2664 > < SL1344_2665	n.d.	n.d.	
SoxS	2891166	2891166	(SL1344_2712)	1.62E-06	AGCACATAGTGATAAAAAT	
SoxS	2984058	2984075	(*emrR*)	1.17E-06	AGCACTTCTTGCAAAAATG	
SoxS	2999313	2999307	*ygaD*	1.90E-06	AGCACAAACTGAAACAAAC	
RamA	3121825	3121821	(*pyrG*)	2.82E-04	GACCCCGCCGGTCACAAAA	
MarA, RamA, Rob, SoxS	3156896	3156839	(*gcvB*)	2.21E-04	ACAACCGTAAGCCAAAAGC	
MarA, SoxS	3219202	3219230	SL1344_3014 < > *idi*	7.36E-04	GAAAGGCATTACCAAAACA	

Rob	3242952	n.d.	*ygfA*	n.d.	n.d.	
MarA	3277271	3277264	(*yggJ*)	3.06E-04	TGAACGTCTGAACAAAAAG	
SoxS	3369166	3369096	*nudF* < > *tolC*	1.10E-04	AGCAATAATGATTAAATGA	yes[a]
MarA	3511816	3511923	*yhbL* < > *arcZ*	1.69E-05	GGCAAACGCGGAAAAAAAA	yes[a,b]
MarA	3515330	3515447	*yhcC* < > *gltB*	1.17E-05	AGCAAACGCTGAAAAAAGA	
SoxS	3550108	3550003	*yhcN*	3.23E-07	GGCATGATTTGCCAAATGA	
SoxS	3570791	3570803	(SL1344_3351)	1.50E-05	GGCATAGCTGGTTAAATGC	
SoxS	3581551	3581481	*acrE*	2.59E-06	GGCAATTAATGCCAAATGA	
SoxS	3602846	3602832	*sapG* > < SL1344_3378	4.03E-06	GACACCCACTGCCAAATCC	
MarA, RamA, Rob, SoxS	3618132	n.d.	*rpsJ* < > *hopD*	n.d.	n.d.	
MarA	3758162	3758200	*rpoH*	1.03E-05	ATCACTGTCTGATAAAAGA	
SoxS	3795560	3795552	(SL1344_3566)	7.99E-06	AGCATTTTTAGAAAAAGAA	
SoxS	3801691	3801721	*yhjB* < > *yhjC*	1.71E-07	AGCACATTTTGTTAAAAAA	
MarA	3829638	n.d.	STnc710	n.d.	n.d.	
MarA, RamA	3838439	n.d.	(SL1344_3597)	n.d.	n.d.	
MarA	3838473	n.d.	*dppA*	n.d.	n.d.	
MarA, RamA, Rob	3857661	n.d.	*cspA*	n.d.	n.d.	
SoxS	3867491	3867462	*yiaB*	5.34E-06	GGCATCGCCGGACAAATGC	
SoxS	3878460	3878455	*yiaM*	4.77E-07	AGCACAAAATGAAAAATAA	
SoxS	3879412	3879302	*yiaM*	5.19E-04	GGCATTGATTTCCAACAAT	
Rob	3926576	3926691	*kbl* < > *rfaD*	1.50E-05	GGCCCTGAATGATAAAGGT	
RamA	3962578	3962520	*gltS* < > *yicH*	2.21E-04	TGACCAGATGGTAAAAGCA	
SoxS	3974136	3974110	*rmbA*	2.04E-04	AACCCCACAAGCAAAATCA	
SoxS	3974754	3974833	*rmbA*	2.04E-04	GGCATTAAGTTACAAAATT	
SoxS	3975126	3975123	*rmbA*	1.57E-08	AGCACTATTTGCTAAATCA	
RamA, SoxS	4031825	4031845	*hslT* < > *yidQ*	4.03E-06	AGCACTGATTGTTAAAGTG	
RamA	4080491	4080583	*yieG* < > *yieH*	1.44E-04	TGCCGTCACAGTCAAAAAA	
MarA, RamA, Rob, SoxS	4130272	4130274	*comM* < > *ilvX*	6.20E-05	TGCAAGAATAGACAAAAAT	
Rob	4147006	4146922	*rho*	2.67E-05	GGAAGTGACGGATAAAACC	
SoxS	4167473	4167472	(*hemC*)	4.65E-06	GGCACATTATGTCAAAGAC	
MarA	4196382	4196381	*dlhH* < > *udp*	2.40E-04	TGCTTCTTCTGACAAACCC	
MarA	4198397	4198306	*ubiE*	1.00E-04	AGCCCGAACTGATAACCGA	
RamA, Rob	4230601	4230569	polA > < *engB*	2.04E-04	AAAATATTCAGCCAAATCC	

Rob	4231620	4231532	*engB* < > *csrC*	3.57E-04	GTAATTGTCTGAAAAAAAG	
SoxS	4248761	4248789	(*yihP*)	2.23E-06	AGCACGCAAGGATAAATAA	
SoxS	4287932	4287842	SL1344_4003 < > *sodA*	7.00E-06	GGCATCCGCTGAAAAAAAC	yes[b]
SoxS	4314627	4314649	*fpr*	7.54E-05	GGCTCTAACTAACAAATGC	yes[b]
MarA	4497504	4497448	*ssb*	2.38E-05	TGCATCTTCAGCTAAAGTA	
SoxS	4666487	4666601	*msrA* < > *ytfM*	3.86E-04	CCCACCCCTGGAAAAAATC	
SoxS	4673449	4673443	(SL1344_4345)	1.17E-05	AGCACCAGCCGACAAATCA	
Rob	4720090	4720013	*treR* < > *mgtA*	1.00E-04	CGCCATAATTGCCACAAAA	
Rob, SoxS	4844268	4844383	*deoB*	1.57E-04	GACACTCTGGGCCACATCG	yes[a]
MarA, RamA, SoxS	4851868	4851909	SL1344_4502	6.92E-07	AGCACAAATAGTTAAAACA	
RamA	4864134	4864193	*rob* < > *creA*	1.62E-06	AACACTGAATGCTAAAAGA	yes[b]
			Plasmid pSLT			
SoxS	74381	74501	SL1344_P1_0081	3.10E-09	AGCACAAATTGCTAAAGTG	
SoxS	78949	79068	*pefB*	2.80E-05	AGCACAAAAAATCAAAATA	

[1] P-value generated by MEME as a measure for false discovery rate (Bailey *et al.*, 2009).

[2] Comparison between *Salmonella* and *E. coli.* Genes that have been identified in an *E. coli* MarA ChIP-seq experiment (Sharma *et al.*, 2017b) are labelled a; those that have been identified in an *E. coli* SoxS ChIP-exo experiment (Seo *et al.*, 2015) are labelled b.

Note that brackets represent targets that were identified within genes, <> denotes divergent genes, and >< denotes convergent genes.

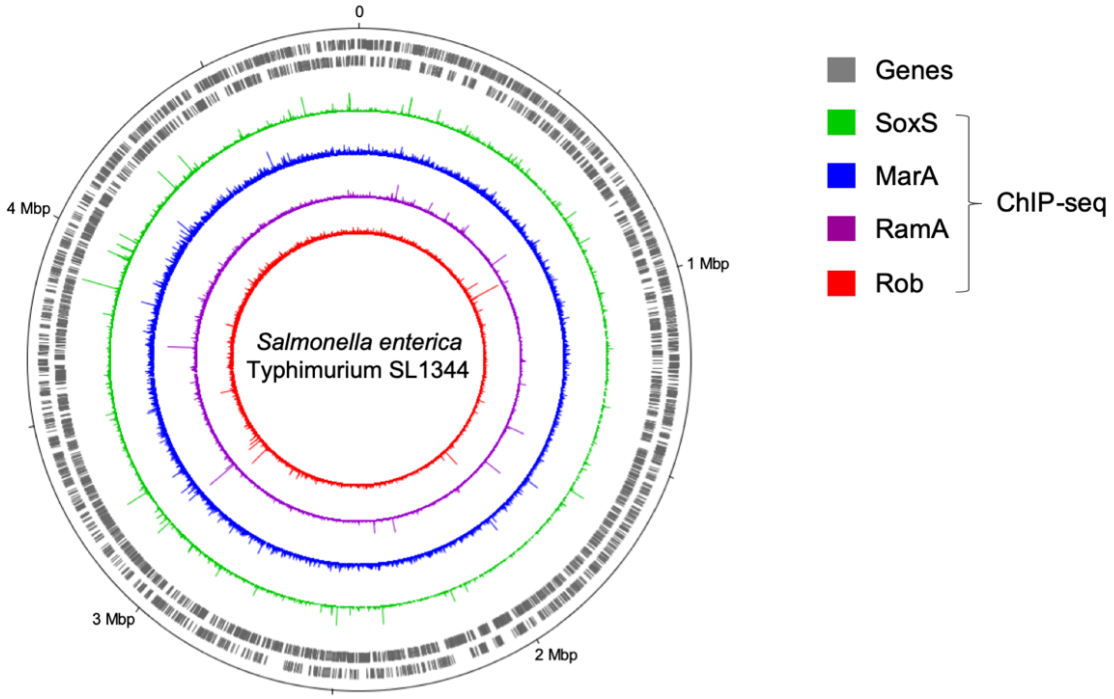


Figure 3.1 The binding sites of SoxS, MarA, RamA, and Rob across the SL1344 genome. The genes of the SL1344 chromosome are shown by the grey tracks on the outside of the circular plot, made using DNAPlotter software (Carver *et al.*, 2009). The ChIP-seq results for SoxS (green), MarA (blue), RamA (purple), and Rob (red) are shown.

The TF with the second largest number of observed peaks is MarA and 16 of the 38 targets are unique to MarA (Table 3.1). Sharma *et al.* (2017b) identified 33 binding targets for MarA in *E. coli* and 10 of these binding targets were also found in *Salmonella*. Of the 46 genes immediately adjacent to MarA binding targets in *E. coli* (Sharma *et al.*, 2017b), 13 of these genes were not identified in *Salmonella*. RamA, the dominant TF in the development of the MDR phenotype in *Salmonella* (Ricci *et al.*, 2006), has 34 binding peaks (Table 3.1), 8 of which are unique to RamA. Rob has 22 binding targets, with 8 being unique to Rob (Table 3.1). A Venn diagram in Figure 3.2 highlights overlap between the binding targets of each TF. Only 5 binding targets are bound by all 4 TFs. These were: *lpxC*, *modE* < > *acrZ*, (*gcvB*), *rpsJ* < > *hopD*, and *comM* < > *ilvX*.

3.3.1 Properties of DNA bound by MarA, SoxS, Rob, and RamA

Following identification of the binding peaks, the consensus DNA motif for each TF was determined as described in 2.25.1. For each regulator, 100 bp either side of ChIP-seq binding peaks was selected and the sequences were submitted to MEME for analysis (Bailey *et al.*, 2009). The minimum motif size was set to 15 bp and the distribution set to zero or one occurrence per sequence. The binding motifs are shown in Figure 3.3A and compared to binding motifs previously identified in *E. coli* for MarA (Figure 3.3B) (Sharma *et al.*, 2017b) and SoxS (Figure 3.3C) (Seo *et al.*, 2015). The positions of each TF binding site was also plotted relative to the nearest start codon (Figure 3.4). The majority of binding sites are between 0 and -300 bps upstream of a gene 5' end.

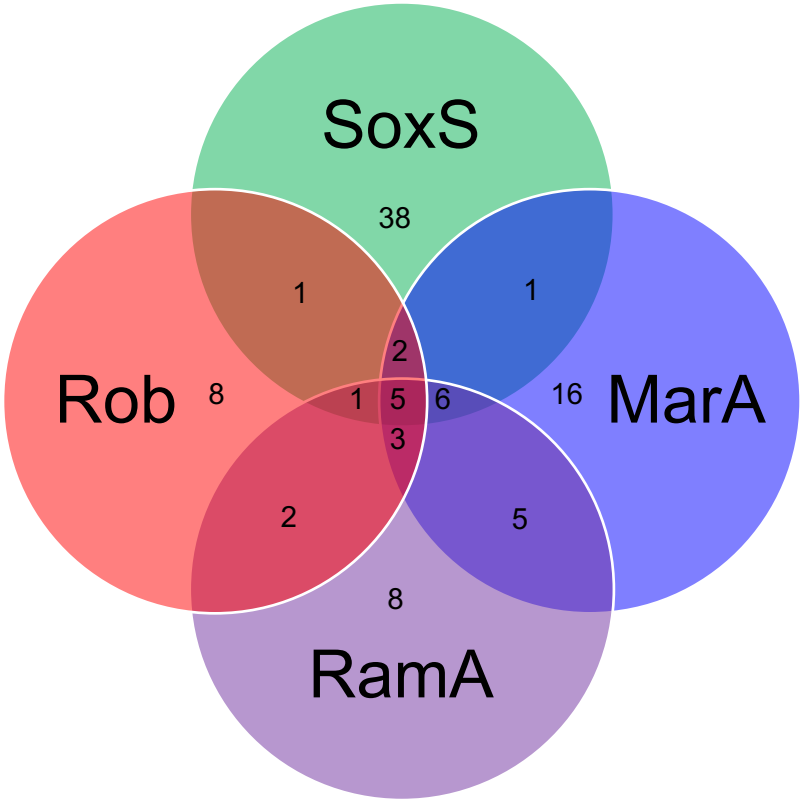


Figure 3.2 The overlap of the binding peaks observed for each transcription factor. A Venn diagram showing the overlap between each TF. The number of unique binding peaks is shown in each circle. The number of binding targets co-regulated by the TFs is shown in the overlapping regions.

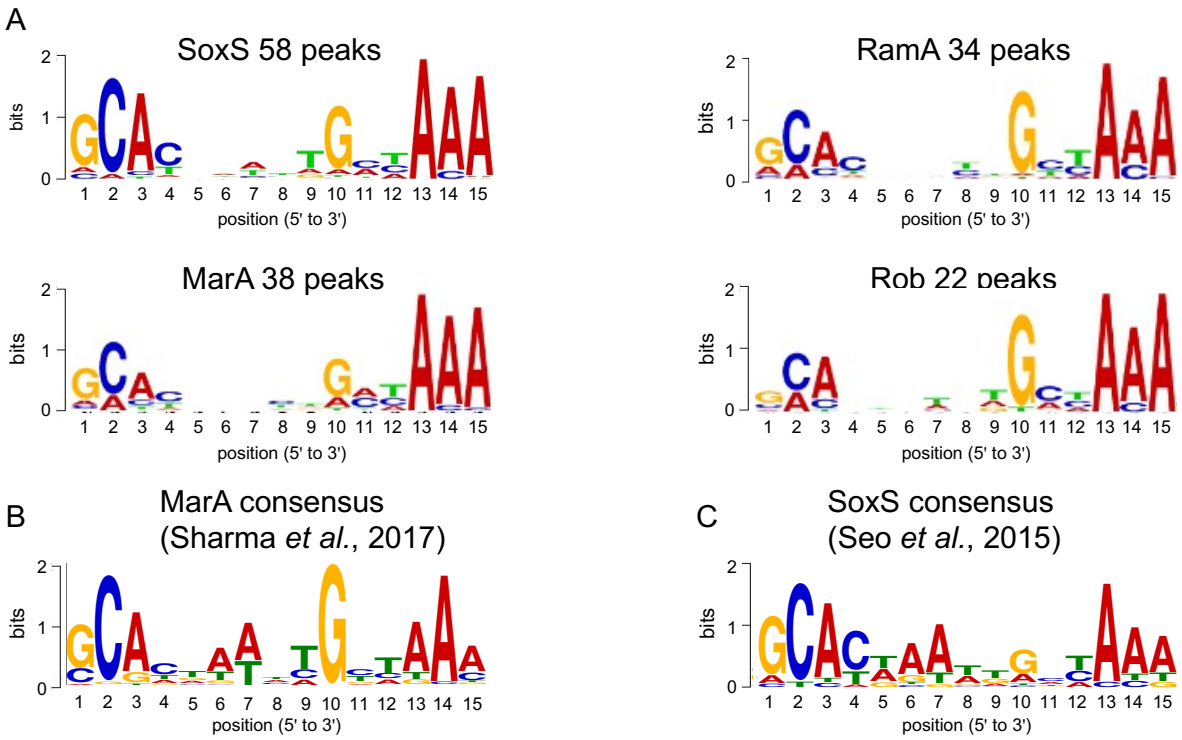


Figure 3.3 The binding motifs of each transcription factor studied here compared to previously known binding motifs. A) The binding motifs of each TF studied here generated using MEME (Bailey *et al.*, 2009) from the binding peaks identified by ChIP-seq for each TF. B) The MarA binding motif identified using a ChIP-seq experiment by Sharma *et al*. (Sharma *et al.*, 2017b) in *E. coli*. Motif generated using binding sites presented in Sharma *et al.* (2017b) using parameters defined by the authors. C) The SoxS binding motif identified using a ChIP-exo experiment by Seo *et al* (Seo *et al.*, 2015) in *E. coli*. Sequences were extracted from the *E. coli* MG1655 genome (NC_000913.2) using ChIP-exo binding coordinates presented by Seo *et al.* (2015) and motif was generated using parameters defined by the authors.

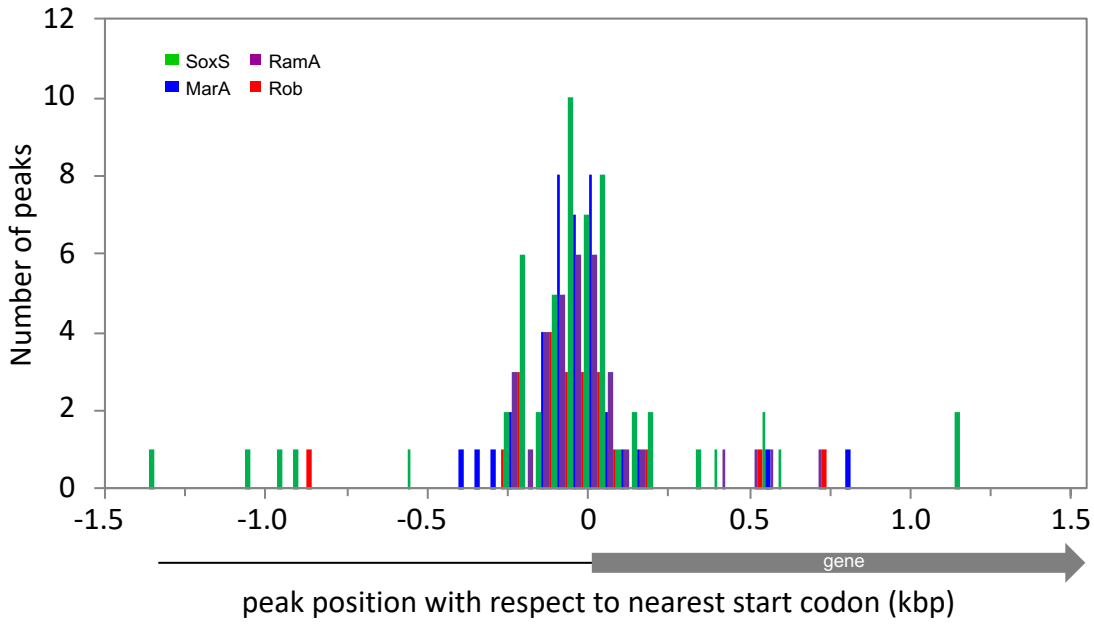


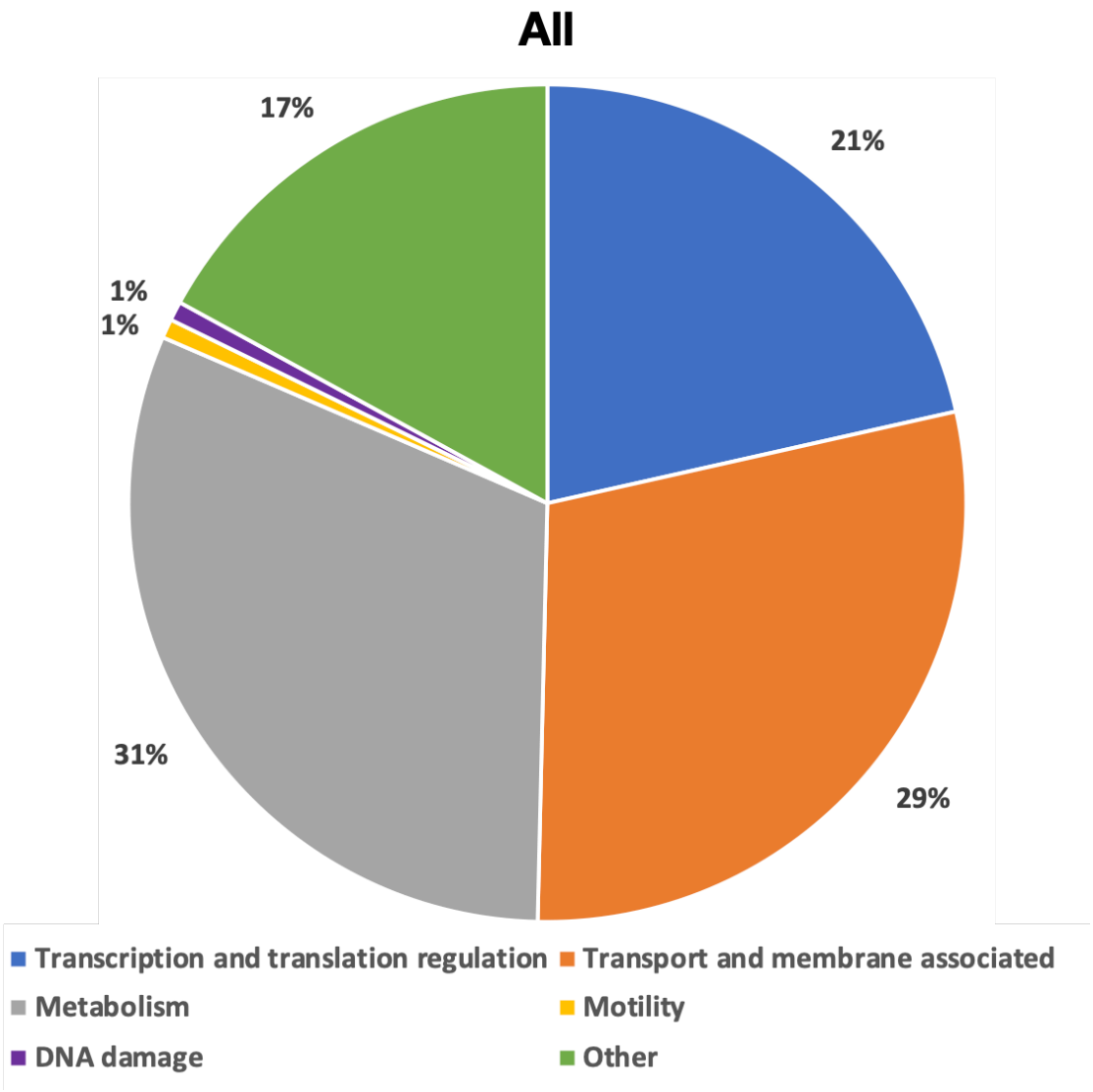
Figure 3.4 The position of each transcription factor binding peak relative to the nearest start codon. The binding sites of SoxS, MarA, RamA and Rob were plotted relative to the nearest start codon. The start site is represented by 0 and the binding sites are either upstream or downstream.

3.4 Biological functions of MarA, SoxS, RamA, and Rob bound genes in SL1344

The ChIP-seq binding peaks were associated with the genes shown in Table 3.1 and the biological function of these genes is shown in Figure 3.5. If a binding peak was observed between divergent or convergent genes, then both genes were listed. This led to a total of 135 genes immediately adjacent to the 100 binding peaks identified. The biological functions are grouped into the categories: transcription and translation regulation, transport and membrane associated, metabolism, motility, DNA damage, and other (Figure 3.5A). Of these categories, metabolism is the most numerous with 42 genes (32 %) being identified. Second most numerous is the transport and membrane associated group, with 38 genes (29 %) identified. Following this, transcription and translation regulation is the third most common group of genes with 28 genes (21 %) associated with the ChIP-seq peaks. Interestingly, only one gene identified is involved with DNA damage and one gene is involved with motility. Figure 3.5B shows the gene functions specific to targets for each regulator individually.

Unsurprisingly, given that SoxS bound the highest number of targets in ChIP-seq, more individual target genes were identified for this factor (73). The distribution of gene functions for SoxS (Figure 3.5B) was similar to the distributions of functions for genes targeted by the combined set of regulators (Figure 3.5A). Indeed, there is little difference in the distribution of the biological functions encoded by genes associated with peaks for binding of any of the 4 TFs (Figure 3.5B).

A


All
17%
21%
1%
1%
31%
29%
Transcription and translation regulation
Transport and membrane associated
Metabolism
Motility
DNA damage
Other

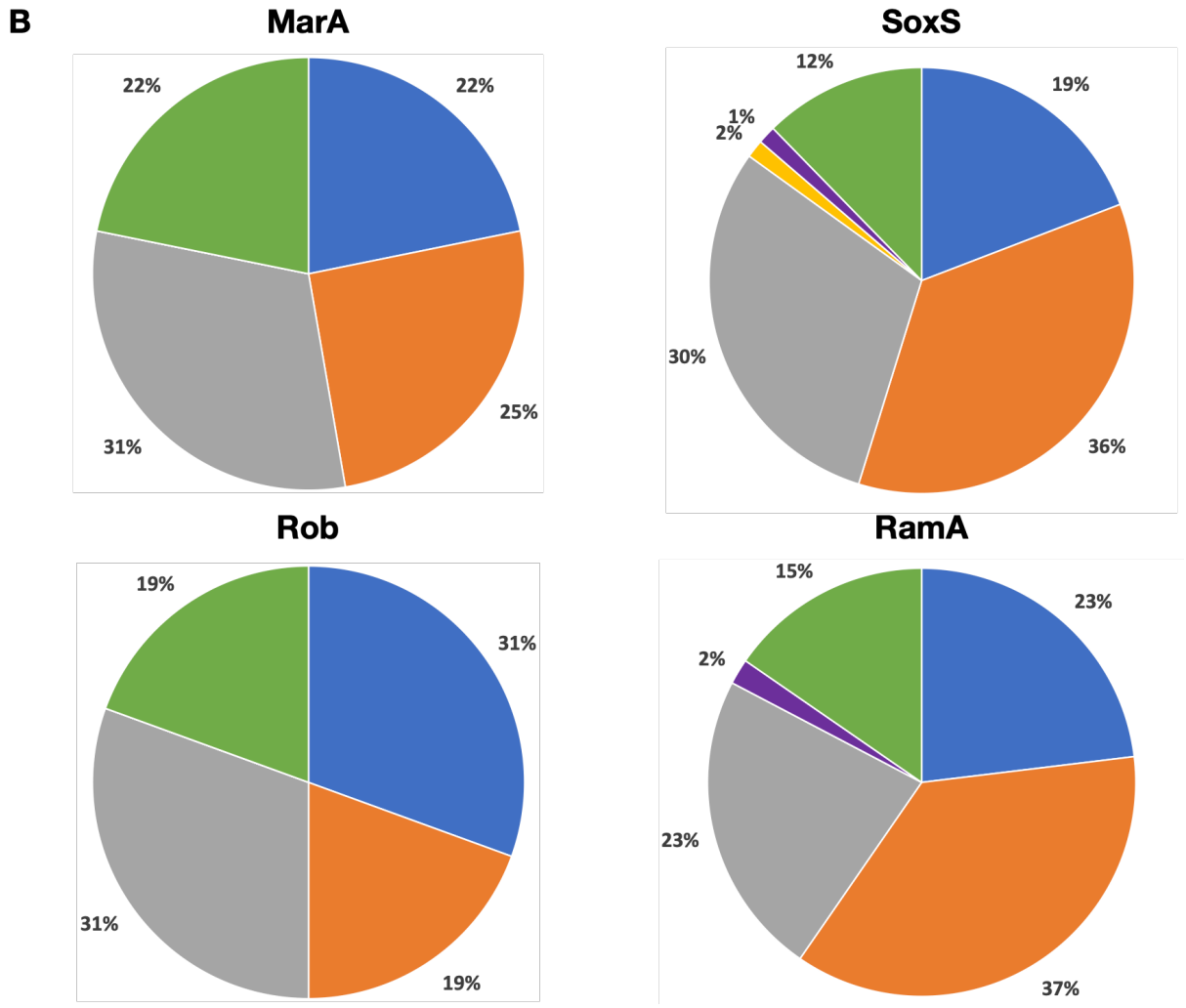


Figure 3.5 The biological function of each gene identified by ChIP-seq. Pie charts showing the general function of each gene identified in the ChIP-seq experiment. Each binding site identified by ChIP-seq was associated with one or more genes (Table 3.1). Each section presents the broad function, the number of genes that fall into that category, and the percentage of the total that function makes up. A. The distribution of gene function of all genes identified in the ChIP-seq experiment. B. The gene functions for all targets associated with each regulator.

3.5 Conservation of ChIP-seq binding targets amongst the Enterobacteriaceae

It is known that uncontrolled expression of MarA, SoxS, Rob, and RamA is associated with clinically relevant resistance to many antibiotics (Sharma *et al.*, 2017b, Weston *et al.*, 2018, Duval and Lister, 2013). As such, any gene under the control of one or more of these regulators might have clinical relevance or be a potential target for a novel therapeutic or adjuvant. The level of conservation of each ChIP-seq binding site within the Enterobacteriaceae was assessed. Briefly, the 150 bases upstream and downstream of each binding site identified by ChIP-seq (Table 3.1) was submitted to a BLASTn search. The search was limited to only the strains listed in Table 2.4; as they are known to encode functional MarA proteins without any mutations in the DNA binding helices (Sharma *et al.*, 2017b). The BLASTn search was unsuccessful for six strains: *Kluyvera intermedia* ATCC 33110, *Leclercia adecarboxylata* ATCC 23216, *Raoultella ornithinolytica* ATCC 31898, *Raoultella ornithinolytica* 10-5246, *Pluralibacter gergoviae* ATCC 33028, and *Cedacea neteri* ND02. None of the sequences submitted to a BLASTn search were identified in these strains and they have been omitted from Figure 3.6. The results are shown in Figure 3.6. In the figure, strains are ordered according to their evolutionary relationship determined by *polA* gene sequence as in Sharma *et al.* (2017b). If the binding site sequence submitted to BLASTn search was not widely conserved, it tended only to be found in closely related species. This is indicated by the general trend that the top half of the heatmap, which contains the *Salmonella* species and its closest relatives, shows the highest level of conservation. All but one target (*SL1344_2712*) was conserved within at least one of the *S. enterica* strains examined. Eleven binding sites are only conserved in the *Salmonella* species: *hemB*<>*yaiU*, *SL1344_0698*, *SL1344_0709*, *SL1344_0962*, *STnc1210*, *YncJ*, (*GpP*), *YgaD*, *YhcN*, (*SL1344_3566*), and *YhjB*<>*YhjC*. The

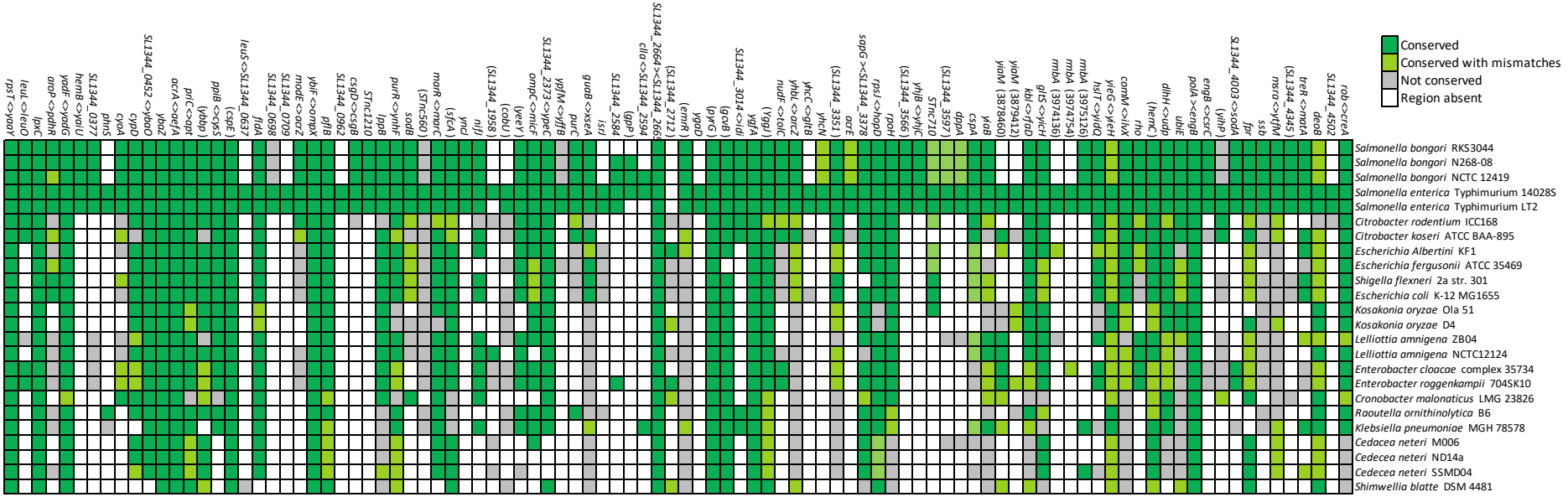


Figure 3.6 The conservation of binding sites identified by ChIP-seq across members of the Enterobacteriaceae. Conservation of each binding site identified by ChIP-seq in SL1344 (x-axis) amongst other Enterobacteriaceae family members (y-axis). Conservation is determined by the number of bases that match the MarA consensus sequence as determined by Sharma *et al.* (2017b). Dark green squares show conserved binding sites which contain a maximum of one mismatch, with light green squares showing a maximum of two mismatches. Grey squares indicate that the binding site was present (or partially) but not conserved. White squares indicates that the region was absent. Brackets represent targets that were identified within genes. <> denotes divergent genes and >< denotes convergent genes.

remaining binding sites show a variety of distributions amongst the Enterobacteriaceae family. Some binding sites are conserved in all species for which results were obtained, indicating that these genes are important for the Enterobacteriaceae stress response mediated by MarA, SoxS, Rob, and RamA.

3.6 Discussion

The purpose of experiments presented in this chapter was identification of targets for MarA, SoxS, RamA, and Rob in *Salmonella*. The results revealed 100 binding loci associated with at least 1 of the 4 regulators. SoxS had the most binding targets, followed by MarA, RamA, then Rob (Figure 3.1, Table 3.1). Of the binding sites identified, 98 were chromosomally encoded. Only two binding peaks were observed on the plasmid pSLT, both for SoxS (Table 3.1). No peaks were observed on plasmids pCol1B9 or pRSF1010. In fact, no sequence reads could be mapped to the pRSF1010 plasmid, suggesting that the strain of SL1344 used in this study has lost this genetic element. Whilst pRSF1010 is highly mobilizable and confers resistance to streptomycin and sulphonamides (Yau *et al.*, 2010); there is evidence to suggest that replication of this plasmid is repressed following conjugation alongside pSLT into a new host. It is proposed that this ensures maintenance of the pSLT plasmid whilst limiting the metabolic burden placed on the host cell by the newly transferred plasmids (Harrison and Brockhurst, 2012, El Mouali *et al.*, 2021). Therefore, the loss of the pRSF1010 plasmid in our strain of SL1344 does not significantly impact the validity of the results observed here with respect to AMR.

There is a large disparity between the number of potential binding sites for MarA, SoxS, Rob, and RamA (up to 65,000 in rapidly dividing cells) and the number of TF molecules expressed under optimal conditions (2,500 in the case of SoxS) (Griffith *et al.*, 2002). Furthermore, many sequences that resemble the binding site for these factors are inappropriately placed to control transcription. Therefore, it is not clear how these regulators recognise the correct sub-population of targets to control transcription. One possibility is the pre-recruitment model, whereby complexes formed between the transcription factor and RNA polymerase off the DNA scan the genome for sequences containing promoters and binding sites for the regulator. However, if such prerecruitment is the mechanism by which these regulators locate binding sites, this raises the question of how repression could occur. One possibility is RNAP locking, with the TF not releasing the RNAP from the promoter. This is discussed further in Chapter 5.

Only five binding sites were bound by all four TFs. This is interesting as, whilst all 4 regulators are involved in the global stress response and bind to the same sequence, the level of overlap between their binding targets is lower than expected. The consensus binding site for each regulator is nearly identical, as shown by Figure 3.3. This suggests that differences in the sequence bound by each TF could be influenced by the non-conserved amino acids within the HTH DNA-binding domain of each TF. These non-conserved amino acids could affect how specific DNA sequences interact with the specific amino acids of each TF by providing extra contacts with the DNA backbone, stabilising the interaction. Alternatively, binding site specificity could be influenced by the intracellular conditions in which each TF is expressed.

These altered conditions may affect the DNA-binding abilities of each TF (e.g. by changing intracellular pH or ionic strength of the cytoplasm). There is evidence to support this idea; previous work indicated that SoxS is less tolerant of deviations from the consensus site as SoxS lacks certain amino acids which facilitate non-specific DNA backbone contacts (Kettles, 2019). This led to differential binding affinities of SoxS or MarA to P*ycgZ* under differing salt conditions, in which key hydrogen bonds between the SoxS protein and DNA backbone were interrupted.

All but one binding site identified here was conserved within the *Salmonella enterica* species, and most were well conserved within *Salmonella bongori*. The level of conservation was generally good amongst closely related species including *Citrobacter*, *Escherichia*, *Shigella*, and *Klebsiella* (Figure 3.6). Whilst the binding sites identified in Table 3.1 were present in most species, the conservation was weaker the more distantly related the species were. The majority of genes associated with ChIP-seq binding peaks identified here belonged to the broad categories of metabolism or transport and membrane associated genes (Figure 3.5A). Similar to the results obtained by Sharma *et al.* (2017b), who identified the category of transport and membrane associated genes were the most numerous amongst MarA binding targets in *E. coli*. These observations fit with the roles of MarA, SoxS, Rob, and RamA in the development of clinically relevant AMR. As, in response to stressful conditions, including antimicrobial stress and nutrient starvation, efflux pumps are often upregulated in response to antimicrobials, and large alterations to the membrane occur (Martins *et al.*, 2011, Piddock, 2006). Almost 30 % of genes identified here are involved with transport or are membrane

associated, agreeing with this observation. In addition to this, 21 % of genes identified were involved in transcription and translation regulation. Indicating that there is a further subset of genes which are regulated indirectly, highlighting the power of these, high-level, global regulators to induce fundamental shifts in the cellular environment. The large amount of genes identified that were involved with metabolism also agree with evidence showing that these TFs can induce the formation of persister cells (Pu *et al.*, 2016). These phenotypic variants grant enhanced resistance to environmental stresses, including antimicrobials, furthering the complications of AMR (Wood *et al.*, 2013).

Whilst these results show many genes associated with each of the TFs, the data are unlikely to be complete. Some known targets were not detected by ChIP-seq. For example, with the MarA ChIP-seq data set, there was no binding peak observed at the *marRAB* operon, where MarA is known to bind and act as an autoactivator. This is not unexpected as ChIP-seq experiments often include false negative binding results. Furthermore, the various factors examined are likely to compete with each other to bind many targets. In summary, the data presented here show that MarA, SoxS, Rob, and RamA bind at a combined total of 100 unique locations in *Salmonella*. All TFs bind to the same sequence and these binding sites are generally well conserved within the Enterobacteriaceae. At least 135 genes are directly associated with MarA, SoxS, Rob, or RamA binding in *Salmonella*; however, the global effects are likely to be much larger, considering one third of genes associated with the binding peaks identified here are involved with transcriptional and translational regulation.

4. Identification of transcription start sites in *Salmonella* SL1344 by Cappable-seq

4.1 Introduction

Transcription start sites (TSSs) can be mapped accurately on a global scale using Cappable-seq. This tool enzymatically 'caps' the first nucleotide incorporated during transcription with a biotinylated GTP to enrich for unprocessed RNA 5' ends. As this step is very efficient, comparison with control (i.e. "uncapped") samples is not required. Hence, both the position and frequency of transcription initiation can be mapped at a single base resolution genome-wide. In *E. coli* and *B. subtilis*, Cappable-seq has identified unexpected patterns of transcription initiation (Warman *et al.*, 2021). Briefly, it was shown that the AT-rich promoter -10 element is sufficiently symmetrical to support bidirectional transcription initiation at many promoters (Warman *et al.*, 2021, Feklistov and Darst, 2011). Hence, divergent TSS pairs, with specific distances between them, are derived from bidirectional promoters. This is possible because transcription initiation by RNAP does not require a perfect match to the consensus -10 promoter element (Browning and Busby, 2004). For example, an adenine base following the 5'-TATAAT-3' -10 consensus generates 5'-TATAATA-3'. Consequently, a near consensus -10 hexamer simultaneously occurs on the opposite strand. This can support transcription initiation in the opposite direction with the forward and reverse initiation sites being 18 bp apart. Similar types of -10 element symmetry, with different distances between TSSs, can also give rise to "bidirectional" promoters. There is strong evidence that promoters within bacteria are frequently bidirectional (Singh *et al.*, 2014, Warman *et al.*, 2021).

In *Salmonella*, previous work has studied how the transcriptional landscape alters in response to conditions related to infection (Kröger *et al.*, 2013, Balkin *et al.*, 2021, Shah, 2014, Li *et al.*,

2017, Khajanchi *et al.*, 2019, Avital *et al.*, 2017). Many of these studies used standard RNA-seq, which does not provide good resolution at RNA 5' or 3' ends (i.e. TSSs) (Ozsolak and Milos, 2011, Ettwiller *et al.*, 2016). Differential RNA-seq (dRNA-seq) does provide information about TSS location. The approach uses an exonuclease to enrich for primary, unprocessed, RNA transcripts. However, control samples are needed so that 5' ends that result from processing can be excluded. Consequently, this method has a tendency to miss some genuine TSSs whilst identifying others erroneously. This approach has been surpassed by Cappable-seq (Ettwiller *et al.*, 2016).

The ChIP-seq results in the prior chapter provide information about binding targets but not associated regulatory effects or TSSs. The present chapter aimed to map global patterns of transcription initiation in *Salmonella*, the prevalence of bidirectional promoters, and the impact of MarA, SoxS, and RamA on the transcriptome. In combination with the prior ChIP-seq analysis, this should provide a comprehensive view of direct and indirect regulation by MarA, SoxS, and RamA.

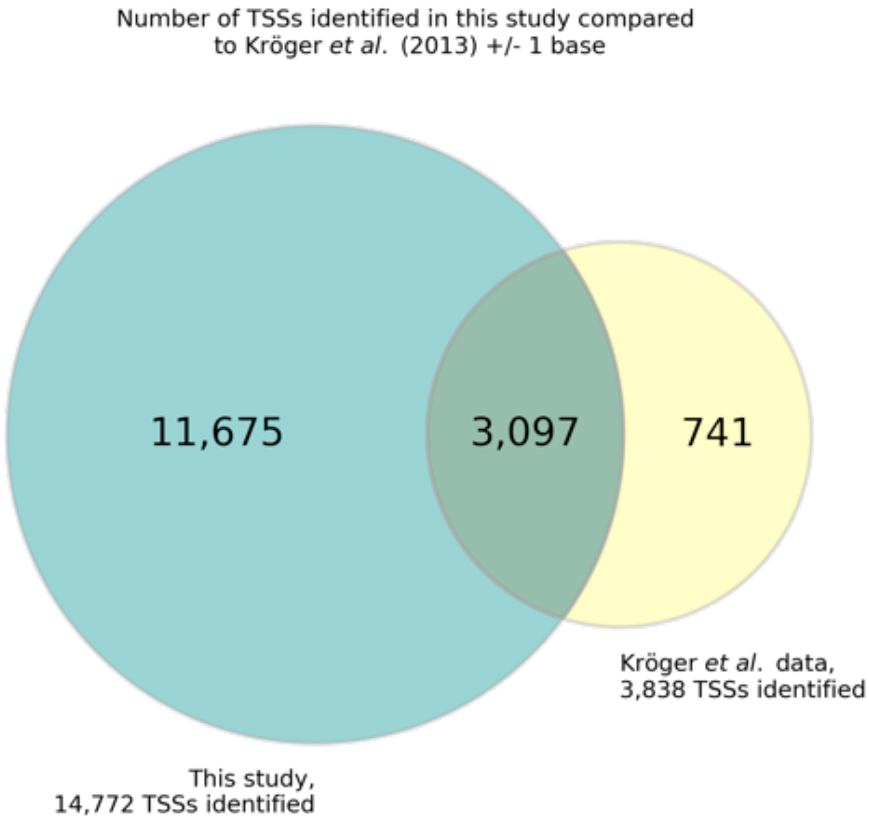
4.2 The distribution of directional transcription start sites across the *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344 genome

In an initial experiment, Cappable-seq was used to map TSSs in two individual replicates of wild type SL1344 cultures grown to mid-log phase in LB medium. The replicates identified 17,232 and 16,371 TSSs respectively and 14,772 TSSs were present in both replicates. These TSSs, shared by both replicates, were used in this analysis. First, the list of Cappable-seq TSSs

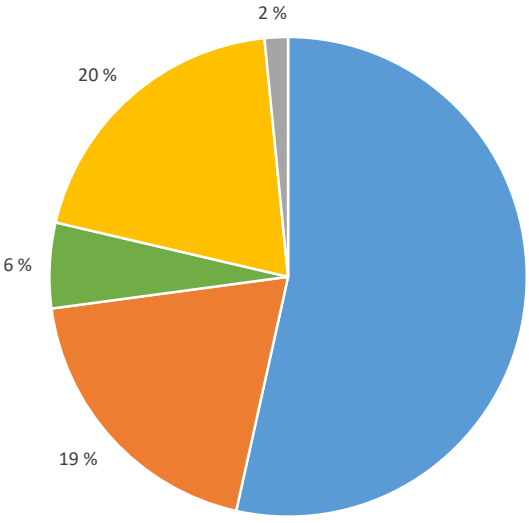
was compared to those TSSs identified using dRNA-seq by Kröger *et al.* (2013). To compare the two datasets, a window of +/- 1 base was used to account for differences in RNA sequencing technologies and TSS mapping strategies. Of the 3,838 TSSs identified by Kröger *et al.* (2013), 3,097 (81 %) were also identified by Cappable-seq (Figure 4.1A). A physiological reason may explain why Cappable-seq did not identify all TSSs found by Kröger *et al.* (2013) who studied 22 infection-relevant conditions.

Next, the location of TSSs with respect to coding features was determined. The majority of TSSs identified by Cappable-seq are within genes (Figure 4.1B, blue). Of the 25 % of TSSs located between co-orientated genes (i.e. genes in the same orientation) most are correctly orientated (orange) to drive mRNA transcription, with only a small number being positioned to drive antisense transcription (green). In total, 20 % of TSSs are between divergent genes (yellow). Finally, TSSs between convergent genes are rare (grey). Of the TSSs identified by Kröger *et al.* (2013), intragenic TSSs were also most numerous, although a smaller percentage of all TSSs (Figure 4.1C) Correspondingly, the proportion of TSSs located between divergent genes (yellow) or co-orientated genes (orange) was higher. The number of TSSs between convergent genes (grey) was again small.

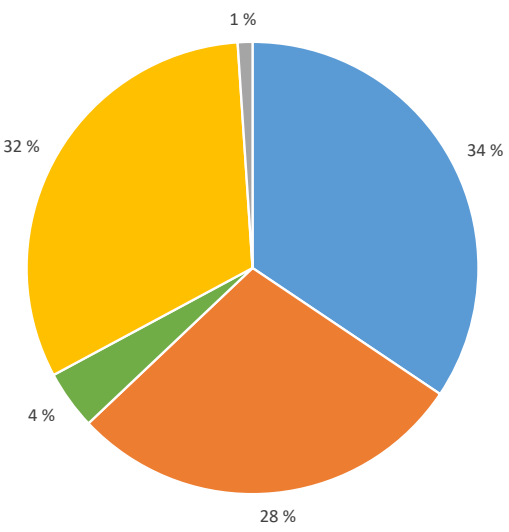
A


Number of TSSs identified in this study compared to Kröger et al. (2013) +/- 1 base
11,675
3,097
741
Kröger et al. data, 3,838 TSSs identified
This study, 14,772 TSSs identified

B


Distribution of directional TSSs identified here
2 %
20 %
6 %
53 %
19 %

C


Distribution of directional TSSs identified by Kröger et al. (2013)
1 %
32 %
34 %
4 %
28 %

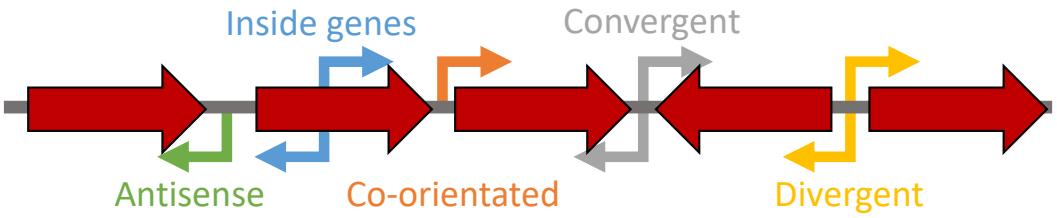

Inside genes
Convergent
Antisense
Co-orientated
Divergent

Figure 4.1 A comparison of the number of TSSs, and the distribution of chromosomally encoded directional TSSs, identified by two different RNA-seq techniques in *Salmonella*. A. The number of TSSs identified here by Cappable-seq in *Salmonella* compared to the results obtained by Kröger *et al.* (2013) using dRNA-seq. B. The distribution of directional TSSs on the *Salmonella* chromosome identified in each dataset. An example of the locations of these promoters is shown below the pie charts.

4.3 The distribution of divergent transcription start site pairs across the *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344 genome

Next, the number of TSSs corresponding to bidirectional promoters (i.e. oppositely orientated TSSs, separated by between 7 and 25 bp (Warman *et al.*, 2021)) was determined. Of the 14,772 total TSSs identified, a total of 13,190 TSSs were not part of a bidirectional pair. The remaining 1,582 TSSs corresponded to 2,594 divergent TSSs pairs (note that some TSSs can form more than one bidirectional pair, according to the selection criteria, if several TSSs occur in close proximity). Hence, 17.56 % of all TSSs identified in SL1344 are divergent pairs derived from a bidirectional promoter. Divergent TSS pairs were associated with 11 unique ChIP-seq binding peaks observed (Table 3.1): 5 were colocalised at MarA bound peaks, 4 at SoxS bound peaks, and 2 at RamA bound peaks. Of the 3,838 TSSs, in 22 infection-relevant conditions, identified by Kröger *et al.* (2013) 206 divergent TSSs pairs were identified, comprising 5.37 % of identified TSSs.

The distribution of bidirectional promoters with respect to coding sequences is shown in Figure 4.2A (Cappable-seq) and Figure 4.2B (dRNA-seq, Kröger *et al*., 2013). A similar distribution is observed in both datasets, with those identified between divergent genes being most common. However, the proportion of divergent TSSs identified between divergent genes identified by Kröger *et al.* (2013) (59 %, yellow) is greater than in this study (45 %). The number of divergent TSSs between co-orientated genes (27 %, orange) is also larger but fewer divergent TSSs were inside genes (14 %, blue). No divergent TSSs in the Kröger *et al.* (2013) were located between convergent genes (grey).

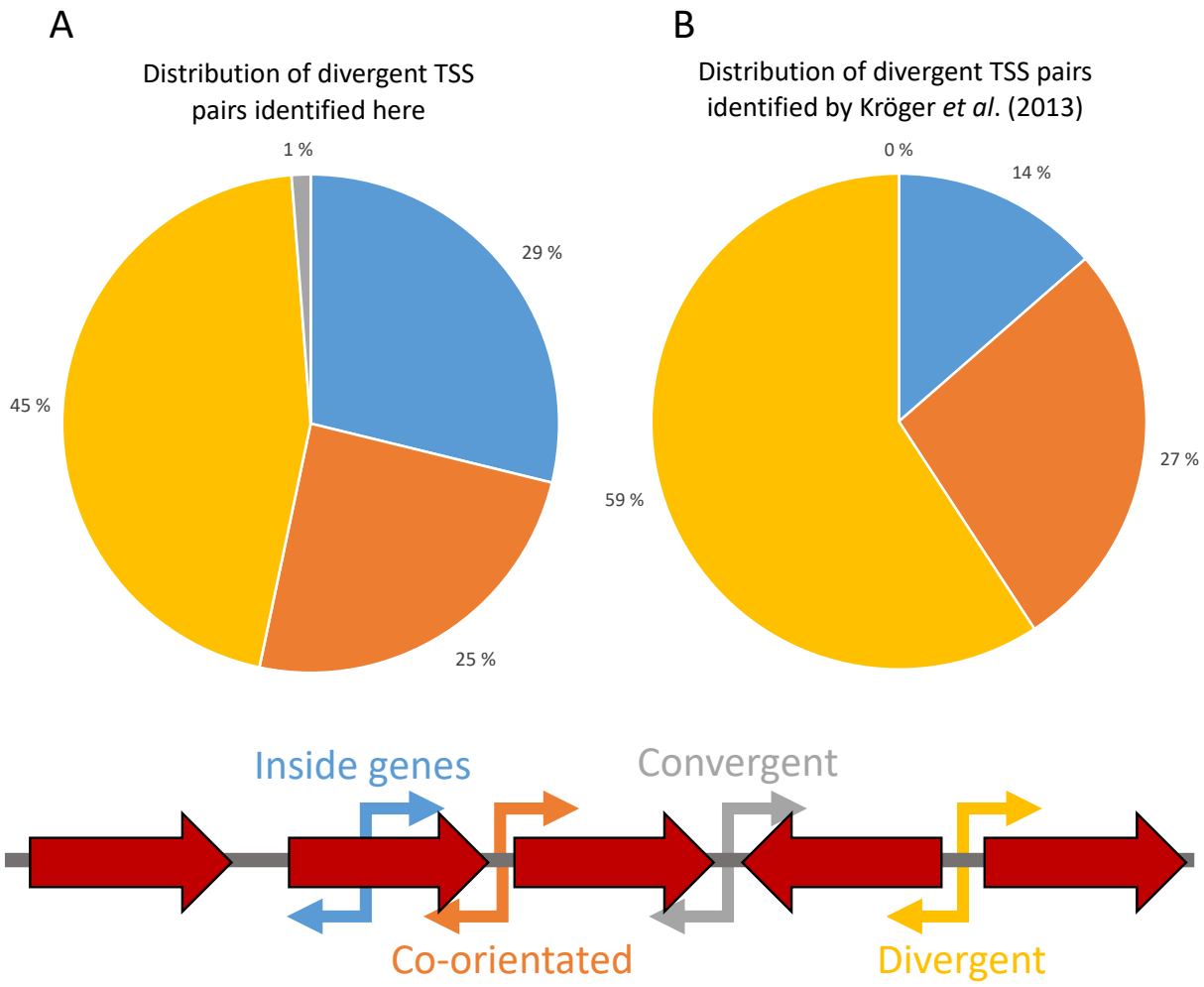


Figure 4.2 The distribution of chromosomally encoded divergent TSS pairs, identified by two different RNA-seq techniques in *Salmonella*. A. The distribution of bidirectional promoters identified here by Cappable-seq in *Salmonella*. B. The distribution of bidirectional promoters identified by Kröger *et al.* (2013) using dRNA-seq. An example of the locations of these promoters is shown below the pie charts.

4.4 Analysis of directional and bidirectional promoters -10 element sequences

Next, the sequences of promoter-10 elements identified by Cappable-seq and dRNA-seq were compared. To do this, DNA sequences from positions -15 to +2 and -20 to +4 were selected for directional TSSs, and bidirectional TSS pairs, respectively. Note that, for bidirectional TSS pairs, sequences were selected with respect to the top strand TSS. Sequences were aligned according to TSS position and visualised using Weblogo (Crooks *et al.*, 2004). The logos generated from Cappable-seq and dRNA-seq data are presented in Figures 4.3 and 4.4 respectively. Note that the top DNA strand TSS is labelled +1 for both directional and bidirectional promoters. As sequences are aligned by TSS and not by the -10 promoter element position, the DNA sequence logos generated do not show a consensus -10 element. This is because the distance between the TSS and -10 hexamer can vary by 1 or 2 bp between individual promoters (Busby and Ebright, 1994, Hawley and McClure, 1983).

4.4.1 The spacing of transcription start sites within chromosomal bidirectional promoters in *Salmonella* SL1344

In their study of divergent TSS pairs arising from bidirectional promoters in *E. coli* Warman *et al.* (2021) determined a series of preferred distances between the divergent TSSs. This analysis showed that separation of divergent TSS pairs by 18 bp was most frequent. However, there were additional preferred distances between the divergent TSSs corresponding to different symmetrical configurations of the promoter -10 element. For example, it was also common for divergent TSS pairs to be 23 bp apart so the key 5'-TA-3'

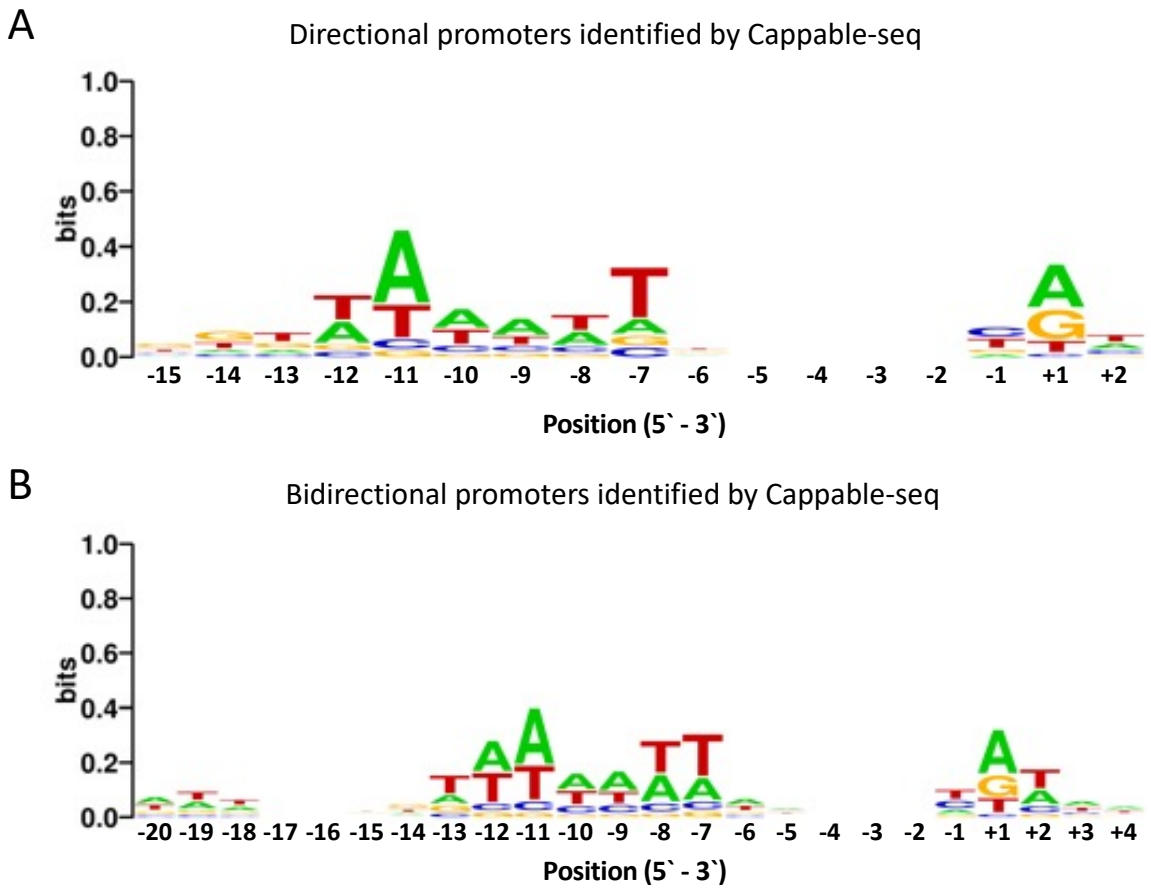


Figure 4.3 The -10 promoter elements of chromosomal canonical and bidirectional promoters identified by Cappable-seq. The -10 promoter element of directional (A) and bidirectional (B) promoters identified here by Cappable-seq on the *Salmonella* SL1344 chromosome, generated by Weblogo (Crooks *et al.*, 2004).

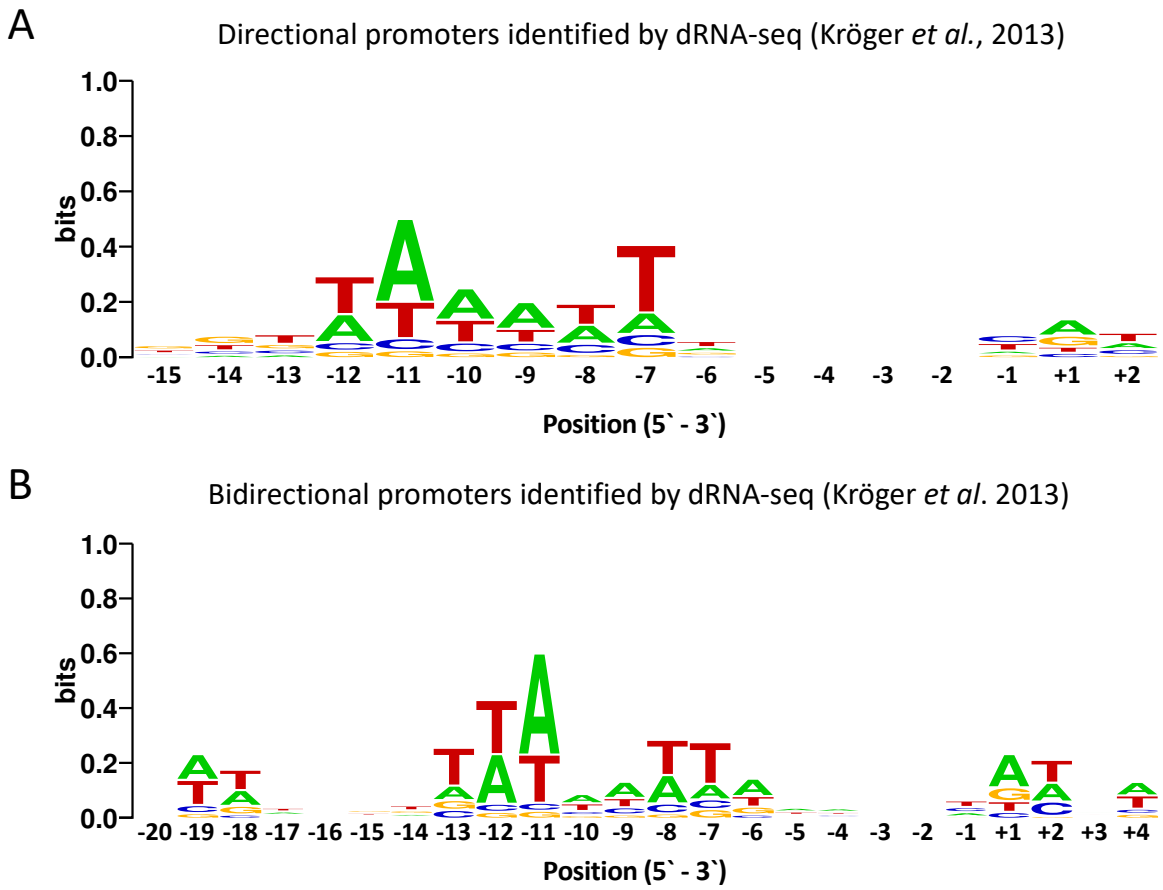


Figure 4.4 The -10 promoter elements of chromosomal canonical and bidirectional promoters identified by dRNA-seq by Kröger *et al.* (2013). The -10 promoter element of directional (A) and bidirectional (B) promoters identified by Kröger *et al.* (2013) using dRNA-seq on the *Salmonella* 4/74 chromosome, generated by Weblogo (Crooks *et al.*, 2004).

sequence at the start of the -10 hexamer precisely coincides on opposite DNA strands. For comparison, the number of divergent TSS pairs separated by different distances in *Salmonella* was studied here (Figure 4.5A). Consistent with prior observations for *E. coli*, divergent TSSs were most often separated by 18 bp. Other common spacings are 7 base (134 instances), 10 bases (110 instances), 12 bases (104 instances), and 23 bases (138 instances); this distribution is similar to that of *E. coli* (Warman *et al.*, 2021). For all such spacings, important bases for transcription initiation reciprocally coincide on opposite strands of the DNA and the resulting sequence logo is near symmetrical (Figure 4.5B). For instance, when TSSs are separated by 7 bp, the TSS (+1) on each DNA strand coincides with position -7 in the -10 element on the opposite DNA strand (Figure 4.5B).

4.5 The distribution of transcription start sites on *Salmonella* SL1344 virulence plasmids

Following the analysis of chromosomally encoded TSSs, the distribution of TSSs on the virulence plasmids pSLT and pCol1B9 was analysed. As with the ChIP-seq data, no RNA sequences were mapped to the pRSF1010 plasmid; further indicating that the strain of *Salmonella* SL1344 used here has lost this plasmid. A total of 1,134 TSSs were identified on pSLT, with 582 divergent TSS pairs identified. Similarly, 574 of the 1,283 TSSs mapping to pCol1B9 were parts of divergent pairs. Figures 4.6 (pSLT) and 4.7 (pCol1B9) show the distribution of TSSs, and distances between divergent TSS pairs, for each plasmid.

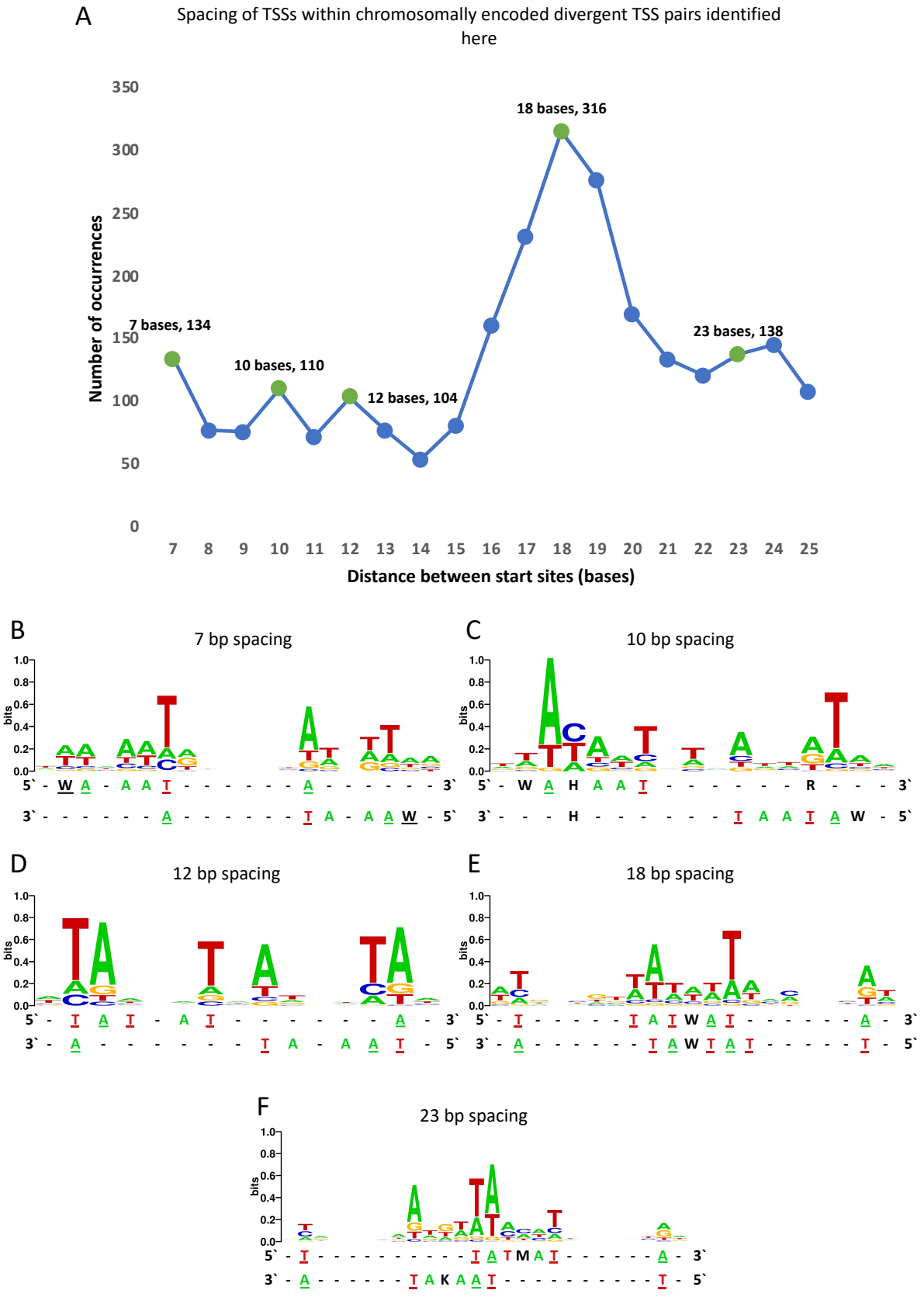
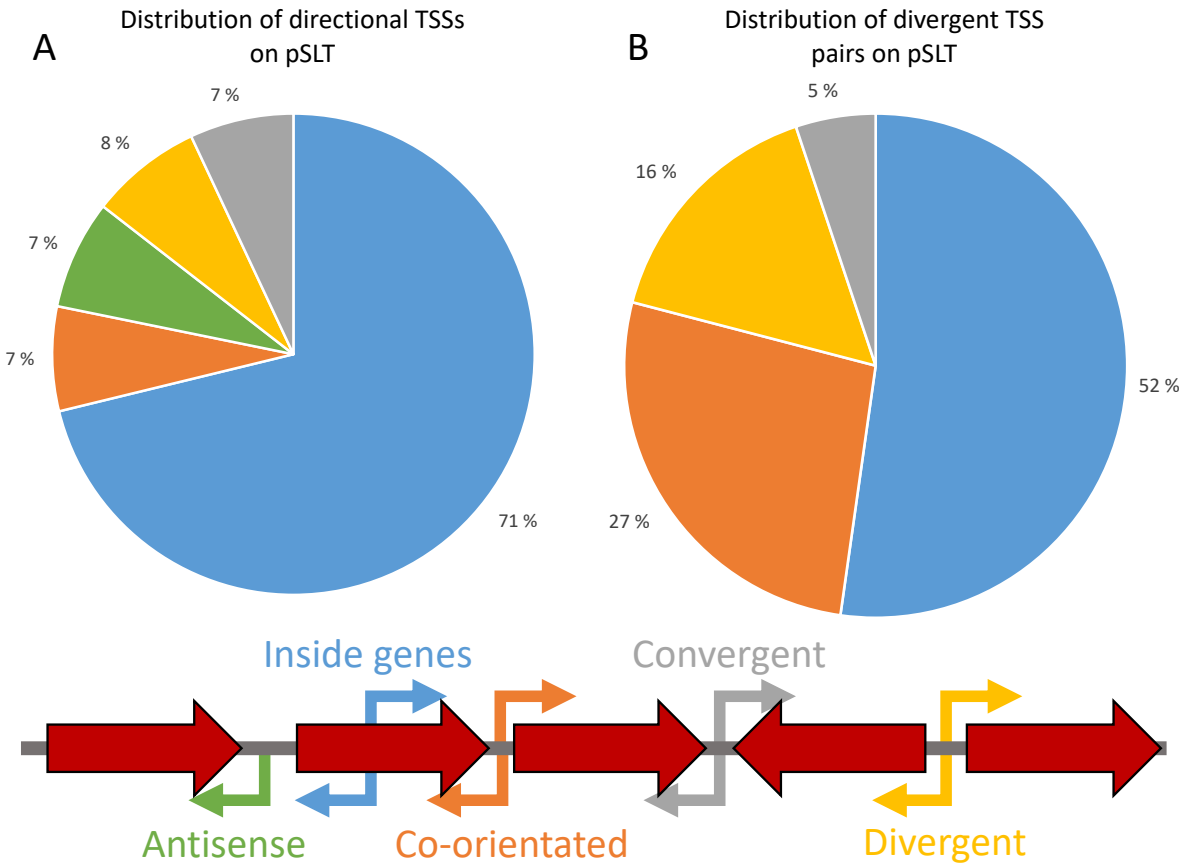

A
Spacing of TSSs within chromosomally encoded divergent TSS pairs identified here
Number of occurrences
Distance between start sites (bases)
7 bases, 134
10 bases, 110
12 bases, 104
18 bases, 316
23 bases, 138
B
7 bp spacing
C
10 bp spacing
D
12 bp spacing
E
18 bp spacing
F
23 bp spacing
bits

Figure 4.5 The spacing of transcription start sites of chromosomal divergent TSS pairs in *Salmonella* SL1344. The spacing of divergent TSS pairs identified as bidirectional promoters in SL1344. The distance between TSSs of divergent pairs plotted against the frequency of each spacing is shown in A. Notable spacings, which give rise to -10 promoter element overlaps are shown by green circles. The DNA-sequence motifs for bidirectional promoters separated by 7 bases (B), 10 bases (C), 12 bases (D), 18 bases (E), and 23 bases (F) are also shown.



C

Spacing of TSSs within pSLT encoded divergent TSS pairs identified here

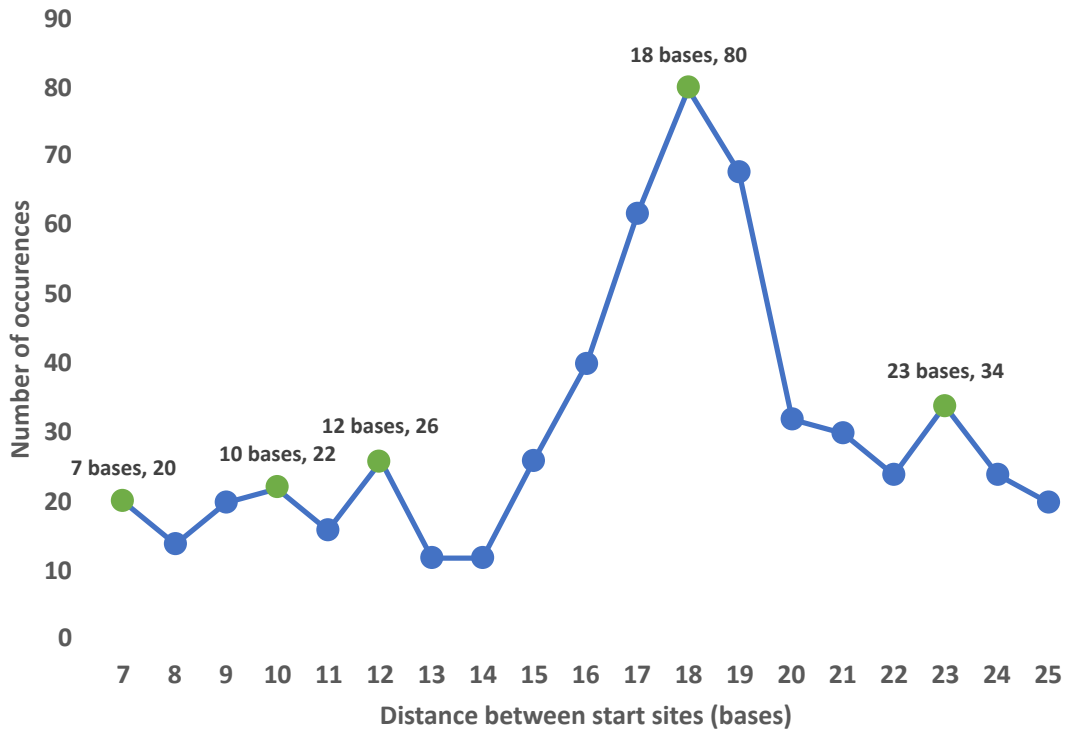
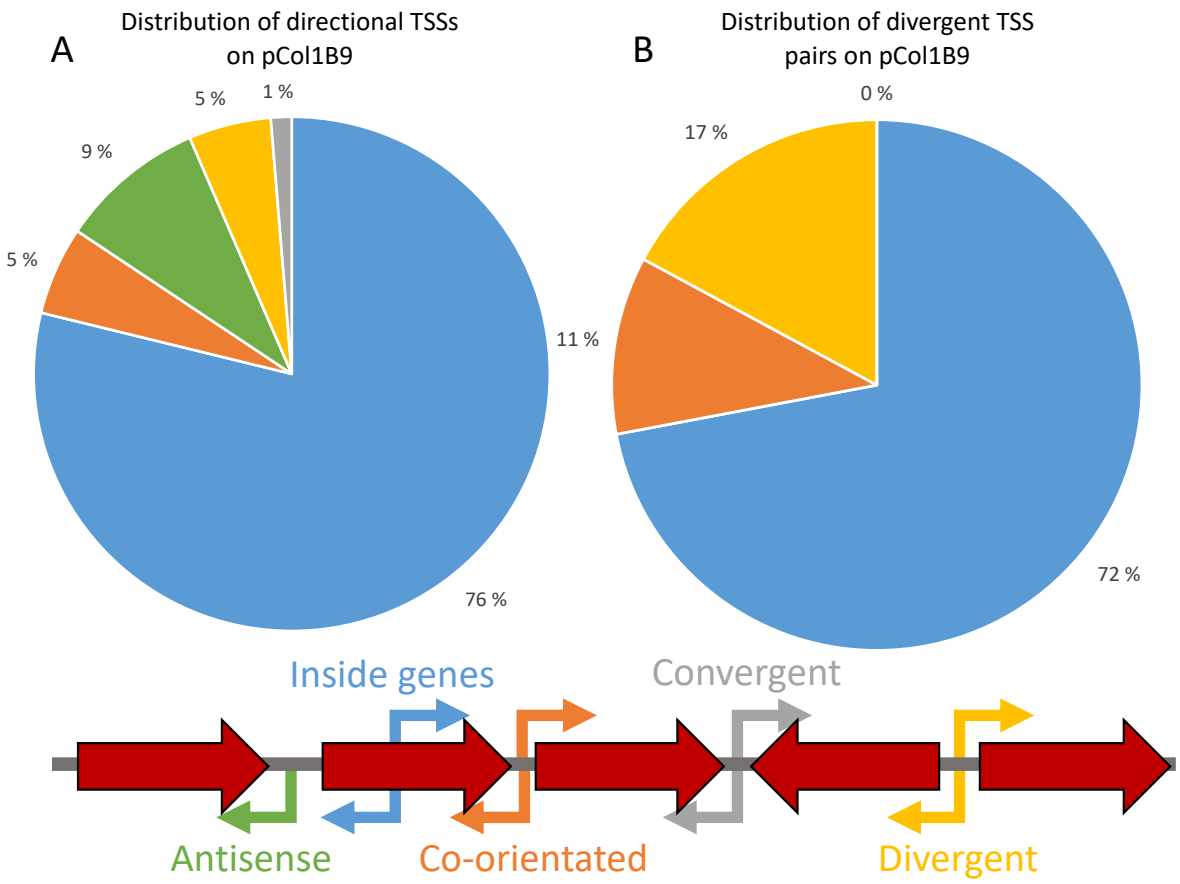


Figure 4.6 The distribution of transcription start sites on the pSLT virulence plasmid. The distribution of directional (A) and divergent (B) TSSs found within genes (blue), between co-orientated genes (orange), antisense TSSs (green), between divergent genes (yellow), and between convergent genes (grey). The distance between TSSs of divergent TSS pairs identified on pSLT plotted against the frequency of each spacing is shown in C.


A
Distribution of directional TSSs on pCol1B9
5 %
1 %
9 %
5 %
11 %
76 %
B
Distribution of divergent TSS pairs on pCol1B9
0 %
17 %
72 %
Inside genes
Convergent
Antisense
Co-orientated
Divergent

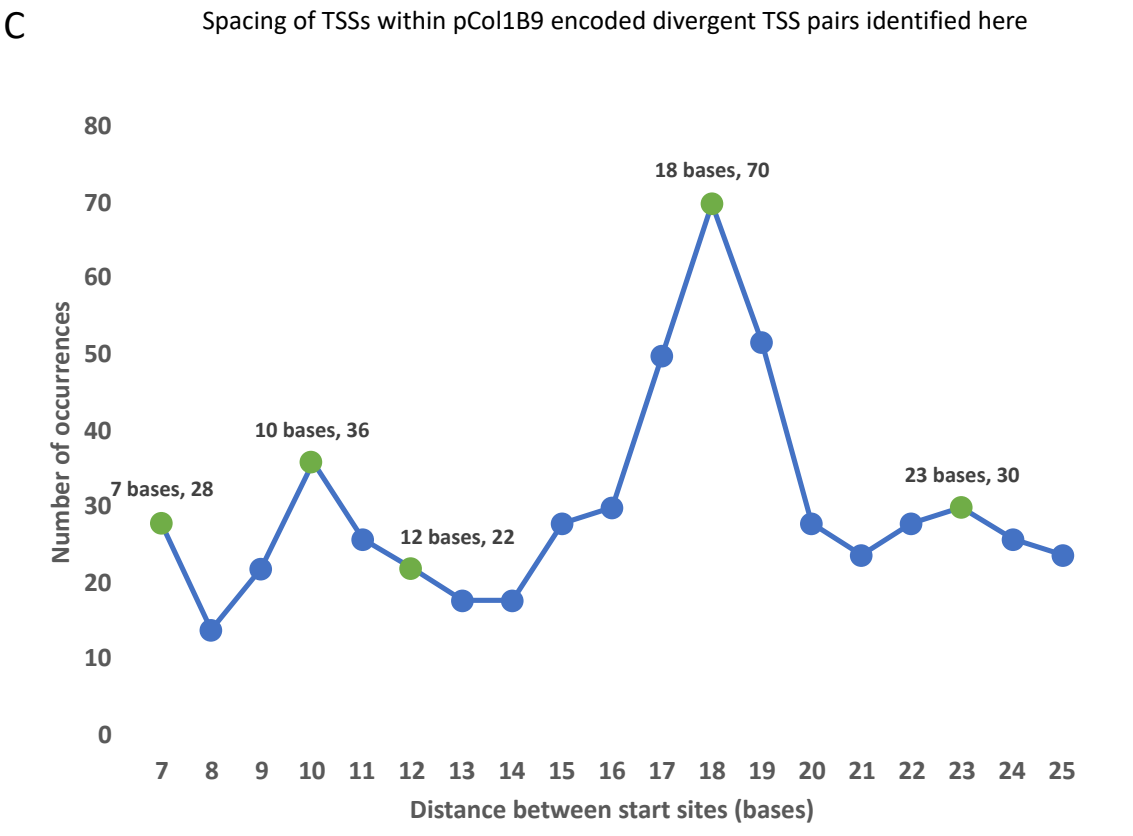

C
Spacing of TSSs within pCol1B9 encoded divergent TSS pairs identified here
Number of occurrences
Distance between start sites (bases)
7 bases, 28
10 bases, 36
12 bases, 22
18 bases, 70
23 bases, 30

Figure 4.7 The distribution of transcription start sites on the pCol1B9 virulence plasmid. The distribution of directional (A) and divergent (B) TSSs found within genes (blue), between co-orientated genes (orange), antisense TSSs (green), between divergent genes (yellow), and between convergent genes (grey). The distance between TSSs of divergent TSS pairs identified on pSLT plotted against the frequency of each spacing is shown in C.

As with chromosomally encoded promoters (Figure 4.1B), the majority of plasmid encoded TSSs are inside genes, although the proportion is much larger. Hence, on plasmids pSLT and pCol1B9, 71 % and 76 % of TSSs are intragenic (Figure 4.6A, Figure 4.7A). The remaining TSSs are evenly distributed for pSLT, with around 7 % each for co-oriented, divergent, and convergent genes. For pCol1B9 fewer TSSs were found between convergent genes (1 %).

4.6 Differential expression analysis of transcription start site use in response to MarA, SoxS, or RamA

Having assessed the position and strength of TSSs in WT SL1344 cells, the analysis was repeated in cells constitutively expressing either MarA, SoxS, or RamA. Rob was not included as the factor is sequestered in an inactive state by its C-terminal domain; as such, in order for Rob to be released, it would have to be activated by the presence of bile salts or other fatty acids, potentially altering the transcriptome in ways that were not solely due to Rob activation. Further to this, Rob had the smallest number of targets in the ChIP-seq analysis. Examples of observed TSS changes are shown in Figure 4.8 for targets of one or more of the regulators identified using ChIP-seq. As with the identification of RNA 5' ends in WT cells, TSSs were only called if they appeared in both biological replicates. Patterns of differential TSS use, due to ectopic production of each TF, are presented as volcano plots in Figure 4.9. In these volcano plots, each datapoint is a TSSs common to the control and TF-expressing cells. Differential TSS use was called if the log fold change in RNA 5' end abundance was greater than +/- 2 and had a P-value of less than 0.05 (presented as the -log10 (P-value)). TSSs identified as downregulated are coloured blue, those denoted upregulated are red. Figure

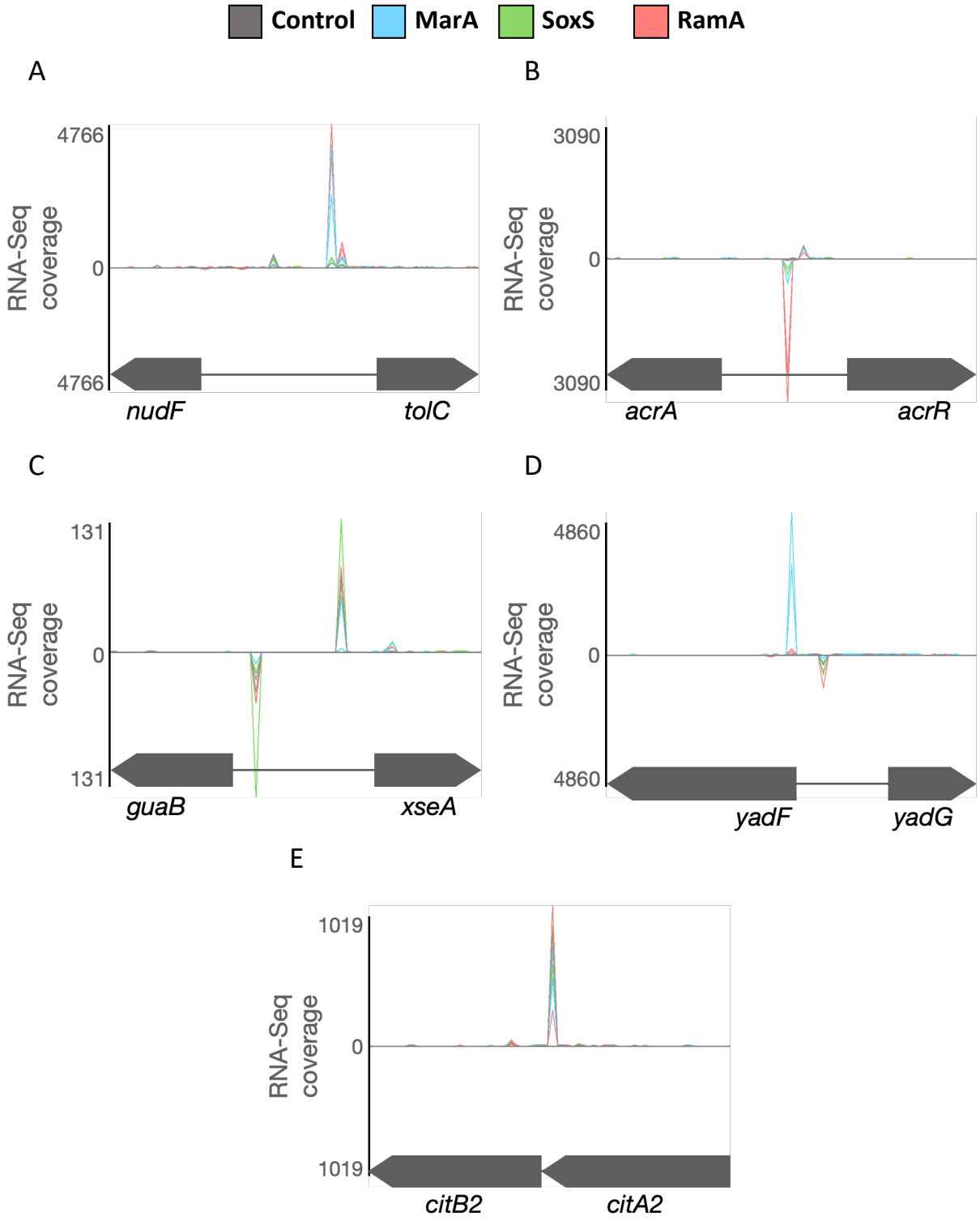
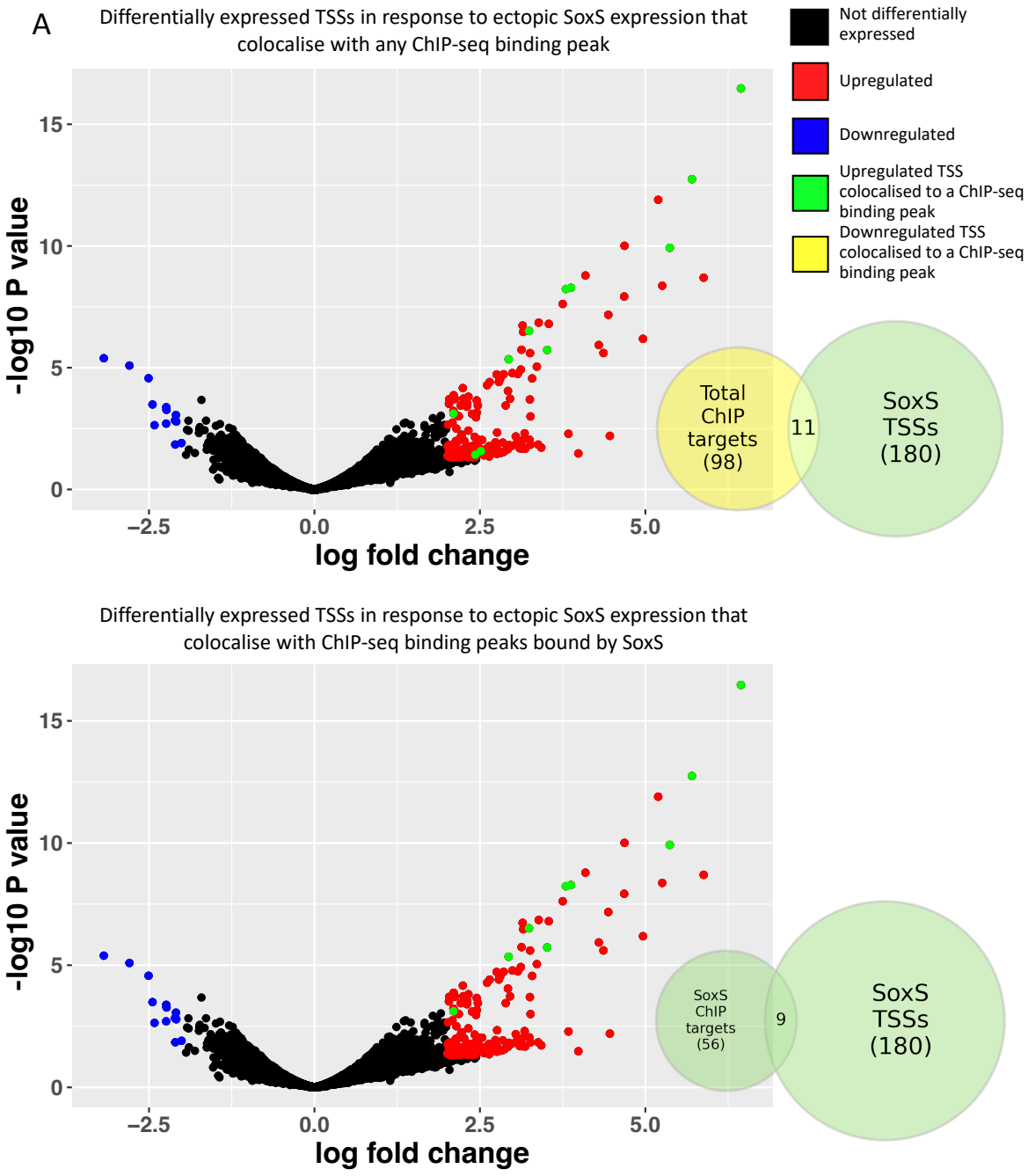


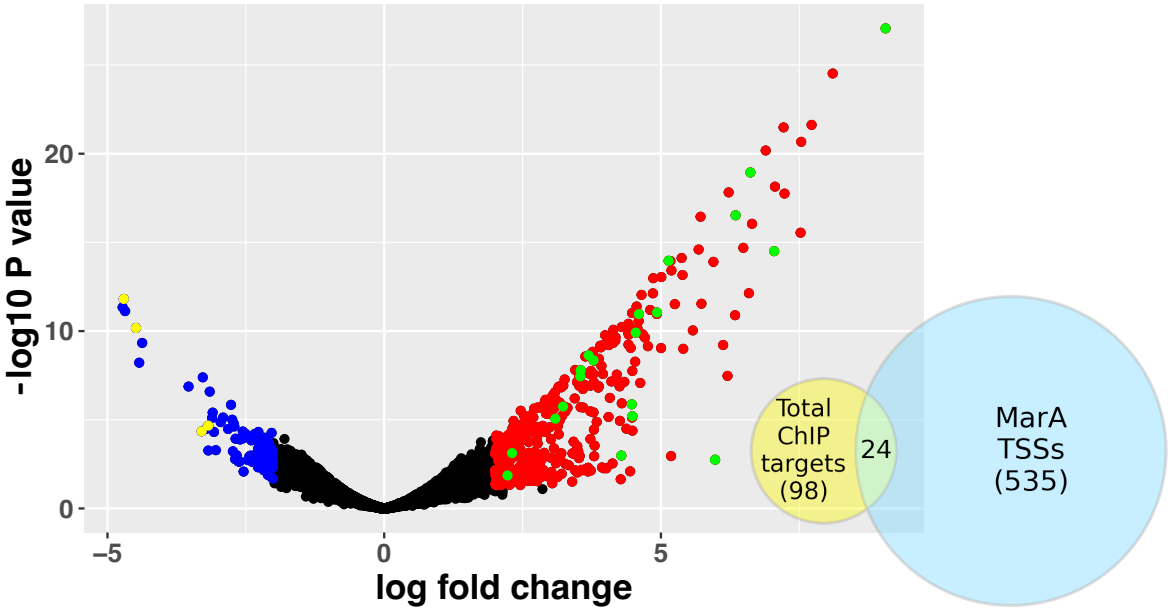
Figure 4.8 Examples of transcription start sites identified by Cappable-seq in *Salmonella* SL1344. Examples of transcription start sites identified by Cappable-seq are presented here.

MarA TSSs are shown in blue, SoxS in green, RamA in red, and the control (empty plasmid) is shown in grey. Various examples of TSSs identified are presented here: a differentially expressed TSS that is upregulated by all TFs when compared to the control (A), a bidirectional promoter (B), a SoxS binding site identified by ChIP-seq in Table 3.1 (C), the largest differentially expressed TSS (D), and an antisense promoter upregulated by all TFs (E).

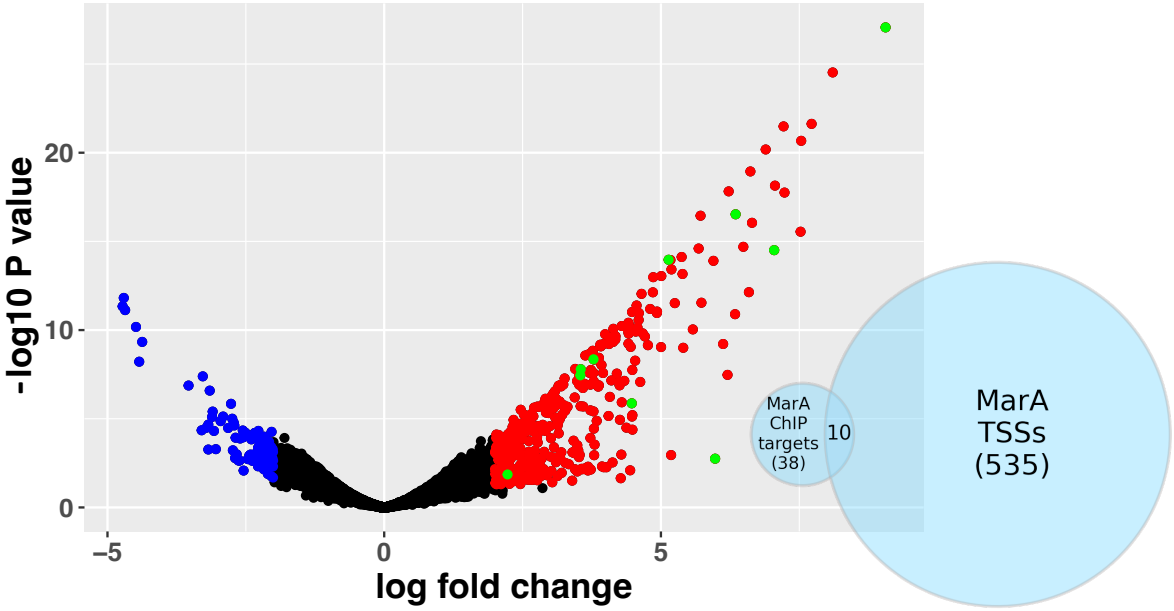

A
Differentially expressed TSSs in response to ectopic SoxS expression that colocalise with any ChIP-seq binding peak
Not differentially expressed
Upregulated
Downregulated
Upregulated TSS colocalised to a ChIP-seq binding peak
Downregulated TSS colocalised to a ChIP-seq binding peak
-log10 P value
log fold change
Total ChIP targets (98)
11
SoxS TSSs (180)
Differentially expressed TSSs in response to ectopic SoxS expression that colocalise with ChIP-seq binding peaks bound by SoxS
SoxS ChIP targets (56)
9
SoxS TSSs (180)

B

Differentially expressed TSSs in response to ectopic MarA expression that colocalise with any ChIP-seq binding peak



Differentially expressed TSSs in response to ectopic MarA expression that colocalise with ChIP-seq binding peaks bound by MarA



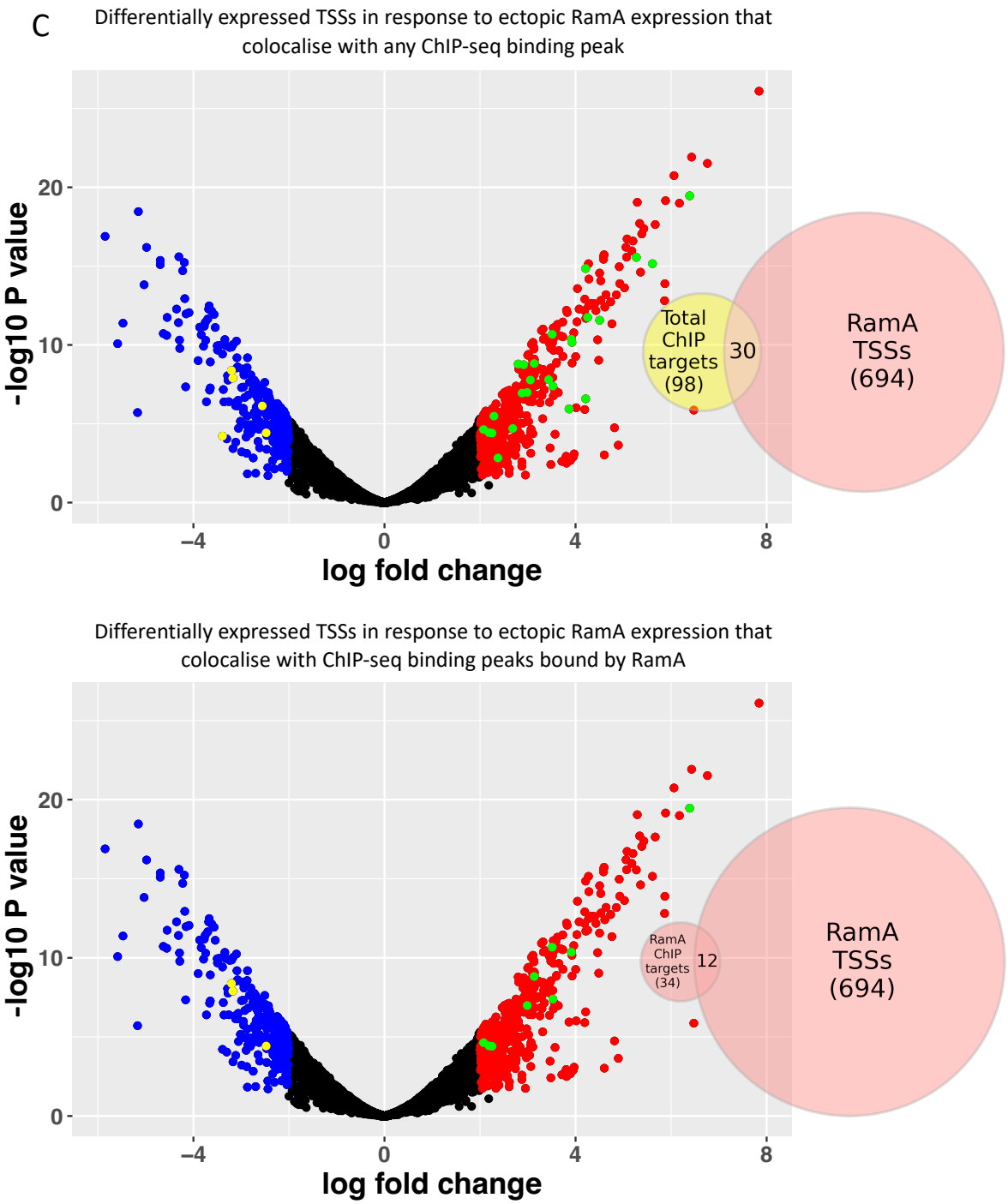


Figure 4.9 Differential expression analysis of transcription start sites identified by Cappable-seq. Volcano plots showing TSSs which are differentially expressed in the presence of SoxS

(A, green), MarA (B, blue), and RamA (C, red). TSSs that are located within 150 bases up or downstream of a binding peak identified ChIP-seq are highlighted in green for upregulated TSSs and yellow for downregulated TSSs. TSSs that overlap with any target identified by ChIP-seq are shown in the top panel, with TSSs that overlap with TF-specific targets shown in the bottom panel. Venn diagrams are shown to visualise the overlap between ChIP-seq and Cappable-seq data sets.

4.9A shows the changes resulting from SoxS expression. Of the 58 binding peaks for SoxS identified using ChIP-seq, 9 were associated with a change in TSS abundance. Of the total set of 100 ChIP-seq peaks (i.e. the combined set of peaks for MarA, SoxS, Rob, and RamA), 11 were associated with differential TSS use if SoxS was ectopically expressed. Note that, in the volcano plots, green and yellow data points are those TSSs associated with a ChIP-seq binding peak. For MarA, 437 TSSs were upregulated and 98 were downregulated (Figure 4.9B). The equivalent values for RamA were 424 and 270 respectively (Figure 4.9C). In both cases, a similar proportion (around one third) of the binding peaks from ChIP-seq experiments associated with a change in TSS use following ectopic expression of the regulator. The data for each factor are summarised in Tables 4.1-4.6.

4.7 Discussion

This chapter aimed to map TSSs in *Salmonella* and to understand the effects of MarA, SoxS, and RamA on TSS use. Of the 14,772 TSSs identified, 13,190 were associated with strictly directional promoters (Figure 4.1C). The distribution of these TSSs with respect to genes differed only slightly to previous reports for *E. coli* and *B. subtilis* (Warman *et al.*, 2021). In *E. coli*, 69 % of the 23,813 canonical TSSs were identified within coding regions; higher than the 53 % identified here (Figure 4.1C). Conversely, in *B. subtilis*, 44 % of the 5,282 canonical TSSs were identified within coding regions. These differences are likely due to the differing number of TSSs identified in each organism. Hence, as the number of TSSs identified increases, a greater percentage are found to be intragenic. In turn, this is likely because promoters within genes tend to be less active and generate unstable RNAs that are difficult to detect. As

Table 4.1 SoxS regulated TSSs that coincide with the binding of MarA, SoxS, and RamA

TSS_ID	ChIP-seq target	Log fold change	-log10 P value
TSS_156912_+	lpxC	3.79892242	8.230622674
TSS_435054_+	SL1344_0377	5.367998123	9.924453039
TSS_533179_-	acrA < > aefA	2.933565242	5.346787486
TSS_757124_-	fldA	2.100507873	3.114885968
TSS_1698914_+	nifJ	6.443940609	16.4698003
TSS_3570779_+	(SL1344_3351)	3.241708222	6.510041521
TSS_3857589_+	cspA	2.514413873	1.572209732
TSS_3962542_+	gltS < > yicH	2.43084701	1.429576344
TSS_4287925_+	SL1344_4003 < > sodA	3.515538478	5.728158393
TSS_4314586_-	fpr	5.702987278	12.74472749
TSS_4666474_-	msrA < > ytfM	3.875428793	8.281498311

Table 4.2 SoxS regulated TSSs that coincide with the binding of SoxS

TSS_ID	ChIP-seq target	Log fold change	-log10 P value
TSS_156912_+	lpxC	3.79892242	8.230622674
TSS_435054_+	SL1344_0377	5.367998123	9.924453039
TSS_533179_-	acrA < > aefA	2.933565242	5.346787486
TSS_757124_-	fldA	2.100507873	3.114885968
TSS_1698914_+	nifJ	6.443940609	16.4698003
TSS_3570779_+	(SL1344_3351)	3.241708222	6.510041521
TSS_4287925_+	SL1344_4003 < > sodA	3.515538478	5.728158393
TSS_4314586_-	fpr	5.702987278	12.74472749
TSS_4666474_-	msrA < > ytfM	3.875428793	8.281498311

Table 4.3 MarA regulated TSSs that coincide with the binding of MarA, SoxS, and RamA

TSS_ID	ChIP-seq target	Log fold change	-log10 P value
TSS_156912_+	lpxC	3.551561449	7.793174124
TSS_202067_+	yadF < > yadG	9.059428706	27.06651271
TSS_424164_-	hemB < > yaiU	3.092636716	5.057495894
TSS_435054_+	SL1344_0377	4.929817467	11.04575749
TSS_533179_-	acrA < > aefA	3.543951837	7.448550002
TSS_598505_-	ppiB < > cysS	2.310624636	3.119944799
TSS_692361_+	(cspE)	2.228770139	1.858541196

TSS_711381_+	leuS < > SL1344_0637	5.979392972	2.748259605
TSS_844593_+	modE < > acrZ	3.782318198	8.353596274
TSS_1466873_-	sodB	7.046464157	14.50584541
TSS_1698914_+	nifJ	3.693794463	8.632644079
TSS_2364629_+	ompC < > micF	4.473352185	5.869666232
TSS_3121974_-	(pyrG)	4.284775809	2.9794034
TSS_3219266_+	SL1344_3014 < > idi	5.139581416	13.9625735
TSS_3277254_+	(yggJ)	6.349900294	16.52870829
TSS_3369223_+	nudF < > tolC	4.601627888	10.95860731
TSS_3570779_+	(SL1344_3351)	6.616720415	18.94309515
TSS_4287925_+	SL1344_4003 < > sodA	3.232355849	5.747146969
TSS_4314586_-	fpr	4.54499831	9.913640169
TSS_4720050_+	treR < > mgtA	4.484524317	5.203425667
TSS_2763482_-	SL1344_2584	-4.485219807	10.17587417
TSS_4031831_+	hslT < > yidQ	-4.705416529	11.81247928
TSS_4167484_+	(hemC)	-3.178255588	4.669586227
TSS_4864158_-	rob < > creA	-3.298516113	4.349692477

Table 4.4 MarA regulated TSSs that coincide with the binding of MarA

TSS_ID	ChIP-seq target	Log fold change	-log10 P value
TSS_156912_+	lpxC	3.551561449	7.793174124
TSS_202067_+	yadF < > yadG	9.059428706	27.06651271
TSS_533179_-	acrA < > aefA	3.543951837	7.448550002
TSS_692361_+	(cspE)	2.228770139	1.858541196
TSS_711381_+	leuS < > SL1344_0637	5.979392972	2.748259605
TSS_844593_+	modE < > acrZ	3.782318198	8.353596274
TSS_1466873_-	sodB	7.046464157	14.50584541
TSS_2364629_+	ompC < > micF	4.473352185	5.869666232
TSS_3219266_+	SL1344_3014 < > idi	5.139581416	13.9625735
TSS_3277254_+	(yggJ)	6.349900294	16.52870829

Table 4.5 RamA regulated TSSs that coincide with the binding of MarA, SoxS, and RamA

TSS_ID	ChIP-seq target	Log fold change	-log10 P value
TSS_156912_+	lpxC	3.913139578	10.35359627
TSS_202067_+	yadF < > yadG	4.25233096	11.74958
TSS_435054_+	SL1344_0377	2.291175822	5.482804102

TSS_533179_-	acrA < > aefA	6.388804146	19.4609239
TSS_539629_+	priC < > apt	2.250377114	4.408935393
TSS_844593_+	modE < > acrZ	3.510695652	10.6716204
TSS_898764_+	ybiF < > ompX	2.18857325	4.436518915
TSS_1466873_-	sodB	5.610401172	15.15304467
TSS_1554915_+	marR < > marC	3.133759837	8.821023053
TSS_1650234_+	yncJ	4.210229521	14.85078089
TSS_1698914_+	nifJ	4.502968178	11.56383735
TSS_2364629_+	ompC < > micF	2.987983435	6.987162775
TSS_2521009_+	SL1344_2373 < > ypeC	3.436136612	7.798602876
TSS_3219265_+	SL1344_3014 < > idi	2.912610622	8.742321425
TSS_3277254_+	(yggJ)	2.878158297	6.943095149
TSS_3369224_+	nudF < > tolC	5.272277715	15.55752023
TSS_3550227_+	yhcN	2.685134834	4.701146924
TSS_3570779_+	(SL1344_3351)	2.798115949	8.787812396
TSS_3602899_+	sapG > < SL1344_3378	2.374347147	2.827872777
TSS_3962542_+	gltS < > yicH	2.074495709	4.632644079
TSS_4167543_+	(hemC)	3.053632834	7.774690718
TSS_4230504_+	polA > < engB	3.524263499	7.390405591
TSS_4287925_+	SL1344_4003 < > sodA	4.211052254	6.583359493
TSS_4314585_-	fpr	3.924338406	10.15989391
TSS_4719988_-	treR < > mgtA	3.863295639	5.93930216
TSS_781895_+	SL1344_0698	-3.175005381	7.896196279
TSS_2594735_-	ypfM < > yffB	-3.391297164	4.21395879
TSS_4031831_+	hslT < > yidQ	-3.203273526	8.378823718
TSS_4231598_+	engB < > csrC	-2.557727418	6.129011186
TSS_4864048_+	rob < > creA	-2.476855534	4.424812155

Table 4.6 RamA regulated TSSs that coincide with the binding of RamA

TSS_ID	ChIP-seq target	Log fold change	-log10 P value
TSS_156912_+	lpxC	3.913139578	10.35359627
TSS_533179_-	acrA < > aefA	6.388804146	19.4609239
TSS_539629_+	priC < > apt	2.250377114	4.408935393
TSS_844593_+	modE < > acrZ	3.510695652	10.6716204
TSS_898764_+	ybiF < > ompX	2.18857325	4.436518915
TSS_1554915_+	marR < > marC	3.133759837	8.821023053
TSS_2364629_+	ompC < > micF	2.987983435	6.987162775
TSS_3962542_+	gltS < > yicH	2.074495709	4.632644079
TSS_4230504_+	polA > < engB	3.524263499	7.390405591

TSS_781895_+	SL1344_0698	-3.175005381	7.896196279
TSS_4031831_+	hslT < > yidQ	-3.203273526	8.378823718
TSS_4864048_+	rob < > creA	-2.476855534	4.424812155

Cappable-seq is more accurate than dRNA-seq in the identification of RNA 5' ends, this is also likely to explain why the proportion of intragenic TSSs identified differ between Cappable-seq or dRNA-seq datasets.

Warman *et al.* (2021) noted that the number of divergent TSS pairs identified in a given TSS dataset correlates, but not linearly, with the total number of TSSs detected; hence, both TSSs in a divergent pair are exponentially more likely to be found as more total TSSs are found. According to the authors predictions, the 14,772 TSSs identified by Cappable-seq here should have identified ~1,500 divergent TSSs. However, the number of divergent TSSs observed here is 2,594 (17.56 %). Further to this, the distribution of divergent TSS pairs differed between *Salmonella* and *E. coli*. In *E. coli*, 51 % of divergent pairs were located within coding regions and 24 % were between divergent genes. In contrast, the majority of divergent TSS pairs in SL1344 (Figure 4.2D) located specifically between divergent genes (45 %), and those found within coding regions accounted for 29 % of TSSs.

The number, and distribution with respect to genes, of directional promoters observed on plasmids pSLT and pCol1B9, was markedly different. In a stark contrast to the chromosomally encoded TSSs, divergent TSS pairs account for 50 % of total TSSs on these plasmids. Additionally, for the plasmids, TSSs were observed more frequently within genes. These distributions may arise because the plasmids were relatively recently acquired by horizontal gene transfer. The differences do not seem a consequence of AT-richness, which is similar to

the SL1344 chromosome. Therefore, these results agree with the notion that divergent TSSs are a common site in newly acquired segments of DNA (Warman *et al.*, 2021).

Combining the genome-wide binding profiles of MarA, SoxS, and RamA, obtained by ChIP-seq, with TSS data from Capable-seq, allows us to conclude that each regulator controls their regulon by largely indirect means (i.e. there are many more apparent regulatory effects than there are ChIP-seq binding sites for the different factors). This is to be expected for global transcriptional regulators. Of the 100 total binding peaks identified in the ChIP-seq experiment (Table 3.1) 39 were associated with differential TSS use upon expression of one of MarA, SoxS, or RamA (Tables 4.1-6). In most cases, transcription was upregulated (Figure 4.9, green data points).

As noted above, dRNA-seq is also more prone to false positive signals and these may also account for some of the differences between the Cappable-seq analysis presented here and the work of Kröger *et al.* (2013). For instance, in this chapter, SL1344 cultures were grown to mid exponential phase (OD_{600} of 0.6) in LB, whereas Kröger *et al.* (2013) used 22 infection-relevant culture conditions, from which their final list of 3,838 TSSs was produced. It can be assumed that some of these conditions would lead to differing transcriptional profiles and could account for some of the 741 TSSs not identified here. Additionally, Kröger *et al.* (2013) used a more stringent cut-off value for TSSs selection in their study. Despite differences in overall numbers of TSS pairs, the spacing optima between these divergent TSSs is largely unchanged between Cappable-seq (Figure 4.5) and dRNA-seq (data not shown). The data are consistent with the observation of Warman *et al.* (2021) that the spacing between divergent

TSSs corresponds to reciprocal base pairing of the -10 element nucleotides and TSSs. Whether these divergent TSSs in *Salmonella* are subject to the same level of transcriptional silencing by nucleoid associated proteins like H-NS requires further study.

Another well studied *Salmonella* strain is *S.* Typhimurium 14028, which is similar to SL1344 (Clark *et al.*, 2011). A recent study utilised Cappable-seq to determine the transcriptional landscape of *S.* Typhimurium 14028s during *Acanthamoeba* infection (Balkin *et al.*, 2021). This identified 12,151 TSSs on the WT *S.* Typhimurium 14028 chromosome and 226 TSSs on the pSLT plasmid. Analysis of the TSSs presented by Balkin *et al.* (2021) identified 1,044 (8.59 %) divergent TSSs on the chromosome. Twenty-two (9.73 %) divergent TSSs were identified on the pSLT plasmid. Whilst a similar number of chromosomal TSSs were identified in both Cappable-seq studies, the number of divergent TSSs identified here was double. This apparent difference may arise because of the use of the `cluster_tss.pl` script by Balkin *et al.* (2021). This script clusters all TSSs within a user-defined distance, keeping only the TSS with the largest statistical score. Balkin *et al.* (2021) used a cut-off value of 5 bases, clustering TSSs within 5 base upstream and downstream. However, it was found here that the `cluster_tss.pl` script does not correctly cluster TSSs. For example, TSSs more than 15 bases apart were clustered together, even though a cut-off of 5 was used. Most likely, TSSs are compared using the specified cut-off value, clustered, and the distance check is erroneously repeated using the newly combined TSS as a reference. This creates a rolling window where TSSs originally separated by more than the defined cut-off value are clustered together. This will greatly impact the number of divergent TSS pairs identified in datasets

where this script is used. This is because the distance between divergent TSS pairs is 5-25 bases (Warman *et al.*, 2021). Hence, many divergent TSSs will be clustered together by this rolling window.

In summary, this chapter presents a more complete picture of the transcriptome of *Salmonella* SL1344 and the global transcriptional effects of MarA, SoxS, and RamA. Further to this, the work highlights the strength of Cappable-seq method for TSS identification and the combined use of ChIP-seq to identify direct gene regulatory effects.

5. SoxS represses the expression of the biofilm regulator csgD

5.1 The *csgBAC-csgDEFG* intergenic region

The ChIP-seq data in Chapter 3 revealed the gene encoding the biofilm regulator CsgD as a target for SoxS (Figure 5.1). CsgD is an activator of biofilm formation, controlling the production of *csgBAC*-encoded curli fibres (Tschowri *et al.*, 2012, Ogasawara *et al.*, 2010a). The *csgD* gene is located in the *csgDEFG* operon, adjacent to the divergent *csgBAC* transcription unit. The intergenic region between these genes, 755 bp in length, is amongst the largest in *E. coli* (Pedersen *et al.*, 2000). In addition, this intergenic region has one of the most complex regulatory landscapes. Currently, 14 transcription factors are known to regulate *csgDEFG* and a further 48 transcription factors may interact with the *csgD* intergenic region *in vitro* (Ogasawara *et al.*, 2020). In *E. coli*, *csgD* expression is indirectly regulated by MarA and SoxS. Briefly, *ycgZ-ymgABC* is activated by MarA or SoxS and this leads to production of *rprA* that represses *csgDEFG* (Kettles *et al.*, 2019). However, this regulatory pathway is not present in *Salmonella*. Instead, SoxS appears to target *csgDEFG* directly (Table 3.1). In this chapter the effect of SoxS on the expression of *csgDEFG* is explored.

Two potential SoxS binding sites were identified in the *csgBAC-csgDEFG* intergenic region. These are shown schematically in Figure 5.2A. One of these (pink) is 448 bp upstream of *csgD* and 300 bp upstream of the annotated *csgDEFG* TSS. This binding site aligns perfectly with the centre of the ChIP-seq binding peak. The second SoxS site (green), a better match to the consensus, is located 140 bp upstream of *csgD*, and overlaps a binding site for the key activator MlrA (yellow). The intergenic region sequence is shown in Figure 5.2B. The TSSs, as defined for *E. coli* (Hammar *et al.*, 1995, Ogasawara *et al.*, 2007) are shown by light blue bases

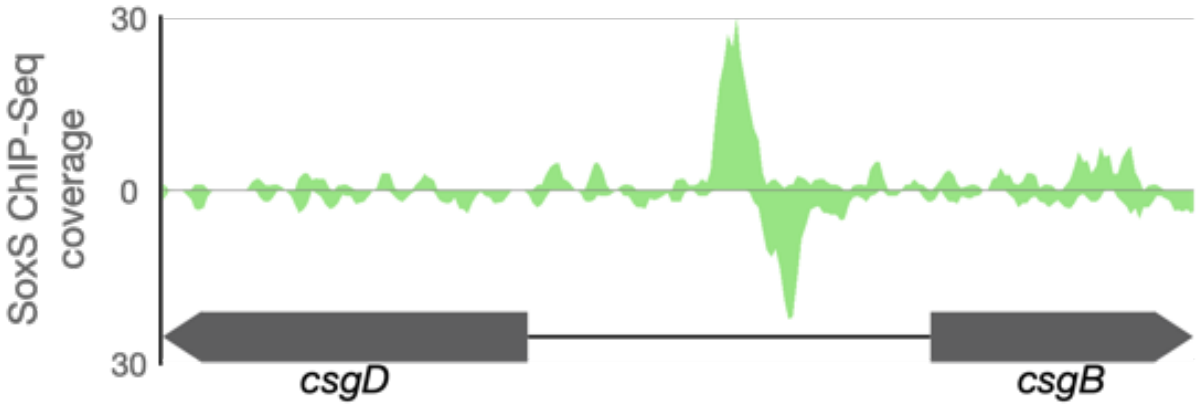
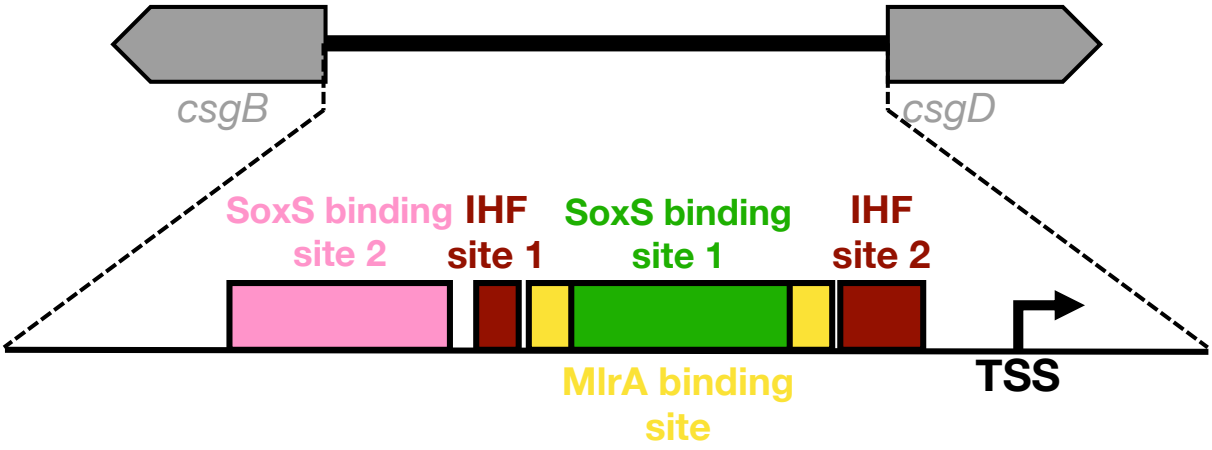


Figure 5.1 The SoxS binding peak identified by ChIP-seq. The binding peak of SoxS between the *csgBAC-csgDEFG* divergent operons identified by ChIP-seq. Coverage is shown on the Y axis. No binding peak was observed for MarA, Rob, or RamA, therefore coverage is not shown.



B

csgB

AAATAACAATTTGTTTTTCAT[GCTGTCACCCTGGACCTGGTCGTACATAGCGAAAATTATCTATTAC

CTTGTTAGCGACATGCGTTTTTTGTTAACGCGTCGTTACGATGAAGAGTATGTCCGTGGAAACATTT

TTAATAACTCACCCACGCGTGGTATTTTGTTATTTAAGCTCATACCAAAGTGCTAATAAAACGATAG

CCATGTGATTTTTTATAATTGATTTTTGGCCACAGAAGATAGTGTATCGCGCACCTAAAAAATGAAGT

GTTGGTGTGTGTTATGCCGCCAT[GGGGATGTTCTTATGCTTCCCATGTGGGGCAATACGCACAAG

ACGTTGACACACTTCGTTTTTTTTGTCTTTGTGCTGTCCAGGTTAATGCCACGTCTCAAATTTTAAGA

AAAAATAAAATCAAAACATAACATATAATAATTAAAATGATTAAAAATCAATGAATTATTATAATTTGTA

TGATTTTTTAAATCTATGCAATAACAGCGAAATGTACAACTTTACTATCAAATCTAAACTTCAAAAA

AACCCAAAAACAACATTTTAATATATATTTTTACATTTGGTTACAAGTTTAACAC[TTGCTTTAAGATTT

GTAATGGCTAGATTGAAAACAGTTAAAAGTATTTTCGTAAATATTTTTCTCTTTCTGGATAATGGGCT

ATTTCAACCCACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCCTTTATTTTATGGGGGCAGCTG

TCAGATGTGCGATTAAAAAAAGTGGAGTTTCATCATG[TTAATGAA

csgD

Figure 5.2 The intergenic region between the *csgBAC-csgDEFG* divergent operons. Panel A shows a schematic representation of the intergenic region showing only the relevant features analysed in this chapter. The 755 bp intergenic region between the *csgBAC-csgDEFG* divergent operons with relevant features highlighted is shown in B. Features are: the two SoxS binding sites (pink and green), the MlrA binding site (yellow), the IHF binding sites (dark red, underlined), the promoter elements and TSS (dark blue), the previously annotated TSSs in *E. coli* (light blue), the bases mutated to abolish binding site function (red), and the centre of the ChIP-seq binding peak (black box). The sequence contained within the green square brackets denotes the long fragment used for *in vivo* experiments and the sequence within the light blue square brackets denotes the short fragment used for *in vitro* experiments. Bases are numbered with respect to the TSS identified in this study.

with right angled arrows. The genes *csgD* and *csgB* are shown in grey text, and the centre of the ChIP-seq binding peak is shown by a black box. The intergenic region is numbered with respect to the TSS identified here (dark blue, underlined). Depending on the experiment, this region was used in two different forms. The first form, termed the long fragment (highlighted by green square brackets) was used for β-galactosidase assays, crystal violet and Congo Red biofilm staining assays, and *in vitro* transcription experiments. The second form, termed the short fragment (highlighted by light blue square brackets) was used for EMSA experiments. Finally, any base mutated to abolish binding site function is shown in red. Further to this, the DNA sequence of the intergenic region between the *csgBAC-csgDEFG* intergenic regions of SL1344 and *E. coli* MG1655 are compared in Figure 5.3. The relevant elements of this region are highlighted as in Figure 5.2. Interestingly only one SoxS binding site is conserved when using the scoring system described in 2.25.2 and Sharma *et al.* (2017b) (pink), the other (green) is considered not conserved. The MlrA binding site (yellow) is relatively well conserved with only 3 mismatches between *Salmonella* and *E. coli*. The IHF binding sites (red) are less conserved, with many mismatches. Throughout this chapter this intergenic region will be referred to as the *csgD* intergenic region, the binding site highlighted in green will be termed SoxS binding site 1 and the binding site highlighted in pink will be termed SoxS binding site 2.

5.2 MarA, SoxS, Rob, and RamA all bind the *csgD* intergenic region

Whilst SoxS was the only regulator identified as binding to the *csgD* intergenic region in the ChIP-seq experiment (Table 3.1), MarA, RamA, and Rob all bind the same DNA sequence motif (Martin *et al.*, 2000, Weston *et al.*, 2018). Hence, in an initial experiment, the binding of each

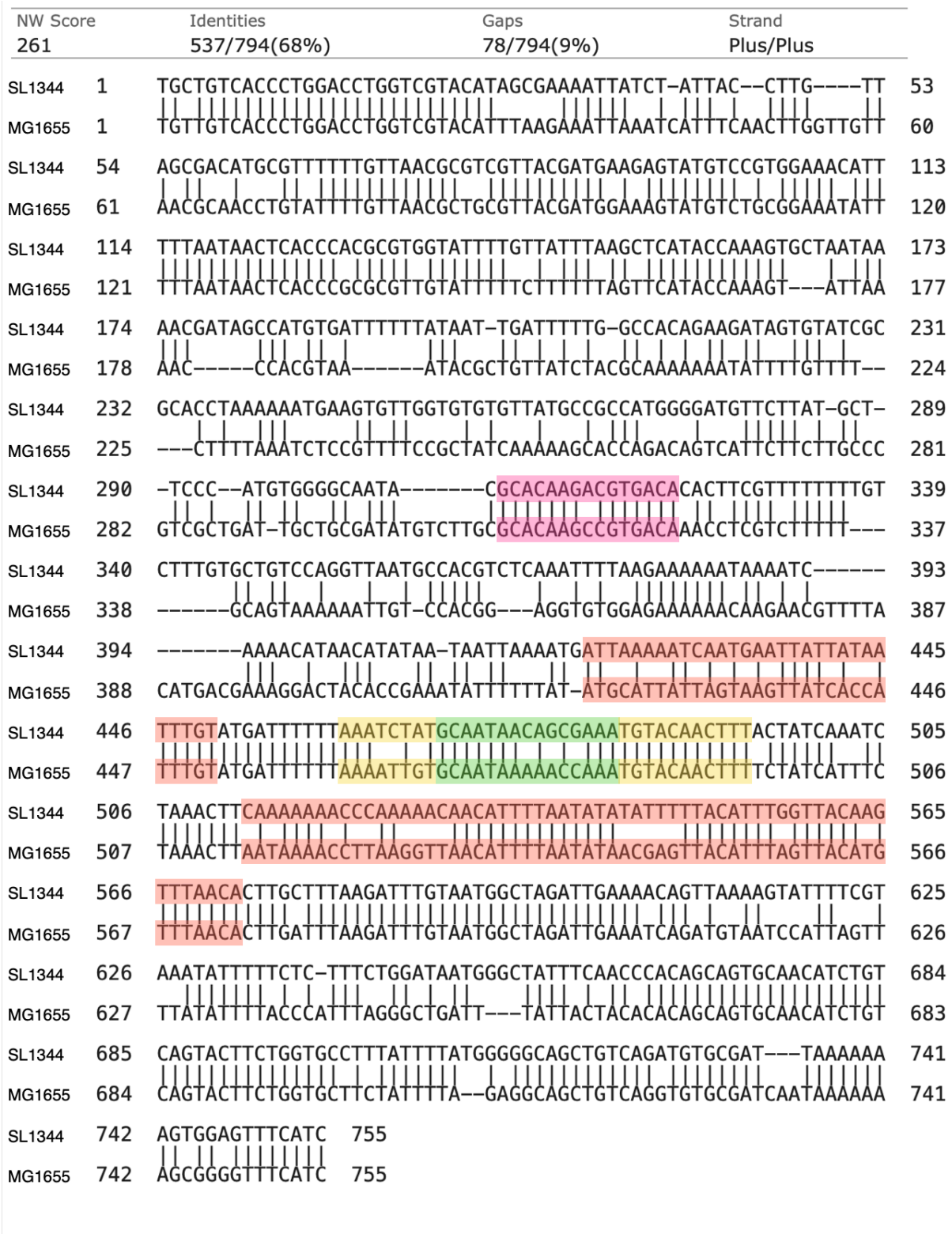


Figure 5.3 Comparison of the DNA sequences of the intergenic region between the *csgBAC*-*csgDEFG* divergent operons of *Salmonella enterica* and *E. coli*. The DNA sequences of the

intergenic region between the *csgBAC-csgDEFG* divergent operons of SL1344 and MG1655 are presented, with the relevant regions from Figure 5.2A highlighted. The two SoxS binding sites are highlighted in pink and green, the MlrA binding site is highlighted in yellow, and the IHF binding sites are highlighted in red.

regulatory protein to the short fragment using EMSAs. The results are shown in Figure 5.4. MarA, SoxS, and RamA bound when added at concentrations of 0.4 or 0.8 µM. Higher concentrations of MarA (0.8 µM and above) or SoxS (1.6 µM) caused non-specific binding, seen as a smear on the EMSA. Rob bound the DNA tightly but non-specifically at all concentrations. Note that this behaviour is typical for Rob (Sharma *et al.*, 2017b) and is caused by Rob's unusual mode of DNA recognition. It is notable that specific binding of MarA, SoxS, or RamA resulted in either one or two discrete bands. These likely indicate either one or two molecules of each regulator binding to the DNA. These bands are highlighted by red (single binding site occupancy) or pink (double binding site occupancy) arrows in Figure 5.4.

5.3 MarA, SoxS, Rob, and RamA reduce expression of the *csgD* promoter

The ability of each regulator to modulate *csgD* expression was studied using LacZ assays. The entire *csgDEFG* intergenic region was fused to *lacZ* in plasmid pRW50T. Resulting constructs were transferred into SL1344 by conjugation and low-level constitutive expression of MarA, SoxS, Rob, or RamA as provided using the plasmid system exploited for ChIP-seq analysis. Empty plasmid vector was used as a control. Assays were done in the absence or presence of pUC57-MlrA that constitutively expresses MlrA, the master activator of the *csgDEFG* operon. The results in the absence of ectopic MlrA expression are shown in Figure 5.5. SoxS, MarA, Rob, and RamA all repress promoter activity. The strongest repression was due to SoxS (green). As expected, in the presence of ectopic MlrA, overall promoter activity increased (Figure 5.6). Even so, SoxS (green) reduced expression by 60 %. Rob (yellow) also repressed promoter activity, but MarA (blue) and RamA (red) did not.

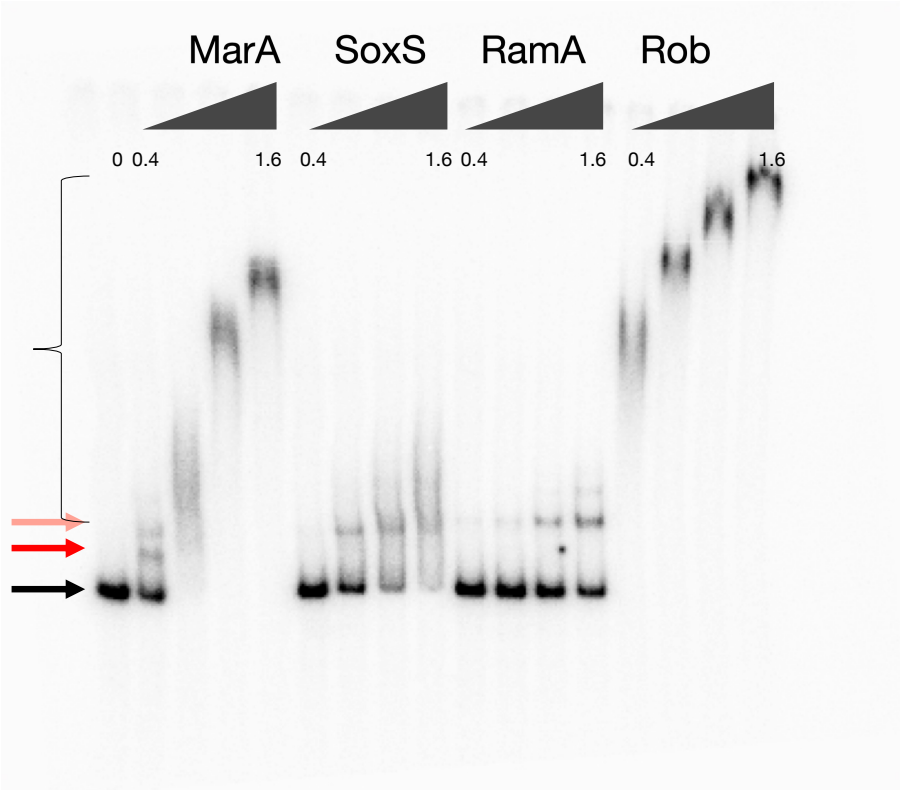


Figure 5.4 EMSA showing binding of MarA, SoxS, Rob, and RamA to the *csgD* intergenic region. EMSA experiment showing binding of MarA, SoxS, Rob, and RamA to the *csgD* WT intergenic region, run on a 6 % polyacrylamide gel. Concentrations used are 0.4, 0.8, 1.2, and 1.6 µM for each protein. An example of the features observed is given for MarA and shows unbound DNA marked by a black arrow, single binding site occupancy marked by a red arrow, double binding site occupancy marked by a pink arrow, non-specific binding indicated by a curly bracket.

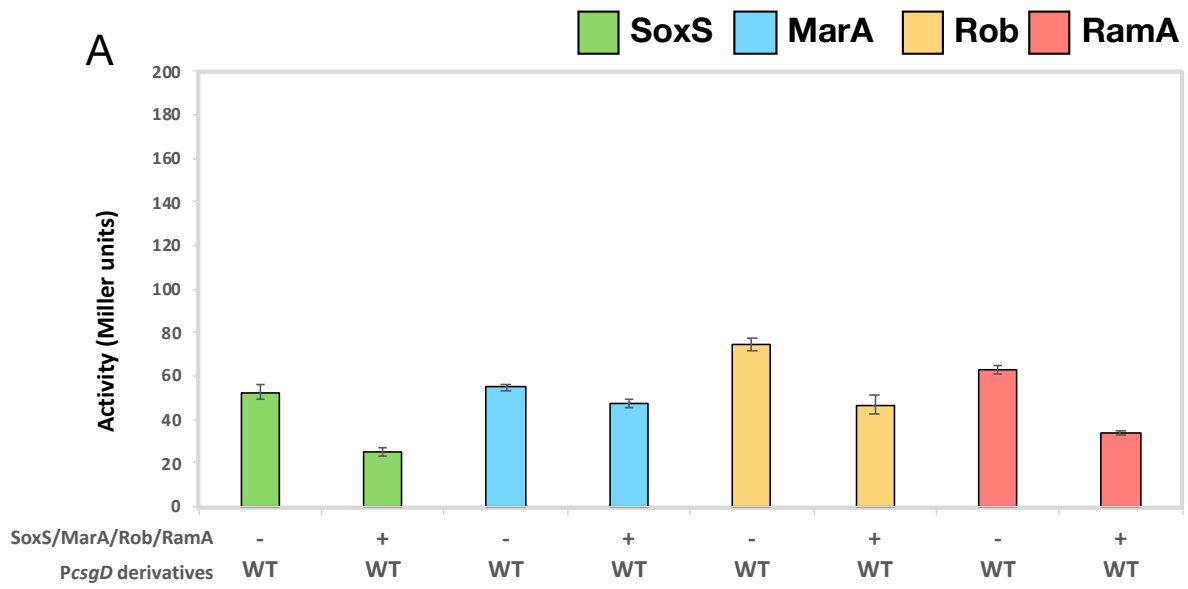


Figure 5.5 The effect of MarA, SoxS, Rob, and RamA on the expression of *csgD* in the absence of the activator MlrA. The effect of MarA (blue), SoxS (green), Rob (yellow), and RamA (red) on the level of transcription of the *csgD* promoter, in the absence of the activator MlrA, shown as the average promoter activity, ± standard deviation, of three biological replicates. Note that '-' denotes empty pAM plasmids without the transcription factor cloned in; in the case of SoxS, Rob, and RamA this was pAMNF and for MarA this was pAMNM.

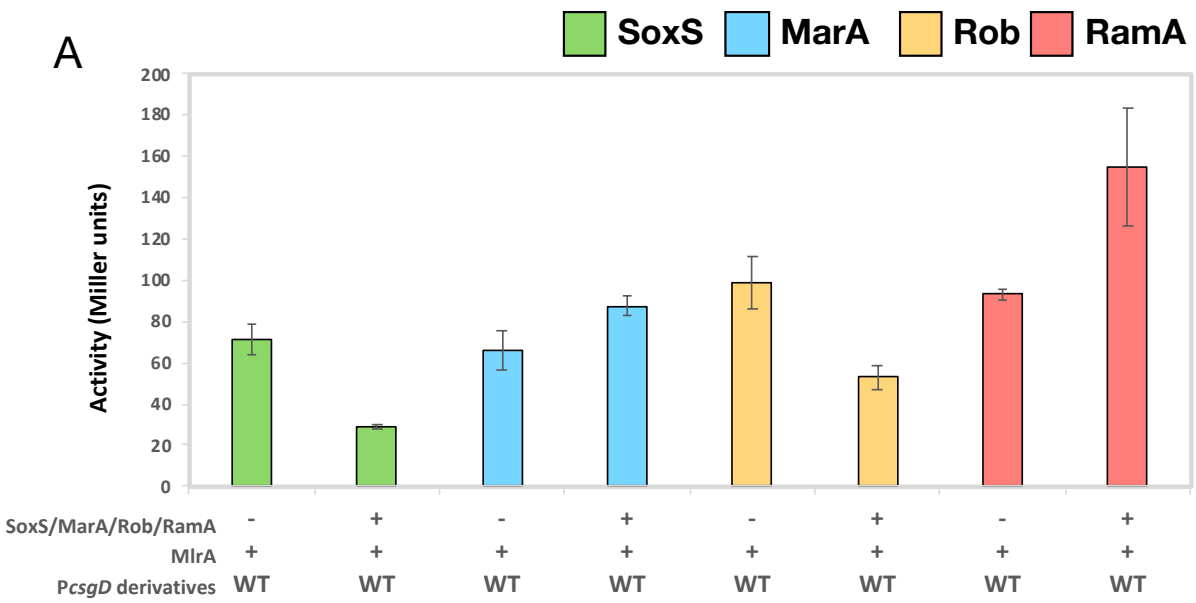


Figure 5.6 The effect of MarA, SoxS, Rob, and RamA on the expression of *csgD* in the presence of the activator MlrA. The effect of MarA (blue), SoxS (green), Rob (yellow), and RamA (red) on the level of transcription of the *csgD* promoter, in the presence of the activator MlrA, shown as the average promoter activity, ± standard deviation, of three biological replicates. Note that '-' denotes empty pAM plasmids without the transcription factor cloned in; in the case of SoxS, Rob, and RamA this was pAMNF and for MarA this was pAMNM.

5.4 SoxS represses *csgD* transcription directly

Together the data show that SoxS binds the *csgDEFG* regulatory region and represses promoter activity (Chapter 3, Figure 5.5 and 5.6). Conversely, MarA, Rob, and RamA binding was not detected using ChIP-seq and expression of these factors had both variable and smaller effects, compared to SoxS, *in vivo*. Hence, the interaction between SoxS and the *csgDEFG* intergenic region was studied further using *in vitro* transcription assays. The long DNA fragment was cloned in plasmid pSR upstream of the λ*oop* terminator. Hence, transcription initiating from the *csgDEFG* promoter IS terminated by λ*oop* and the resulting 244 nt RNA can be detected after electrophoresis. As noted above, the *csgD* intergenic region is one of the largest in *E. coli*, and another characteristic of this region is its high curvature and AT content (Pedersen *et al.*, 2000). As such, the nucleoid associated proteins H-NS and IHF are known to interact with this region. In particular IHF acts as a positive regulator and binds to multiple sites including one just upstream of the MlrA binding site (Figure 5.2) (Ogasawara *et al.*, 2010a). Hence, the effect of IHF in the *invitro* transcription assays was also examined. The results are shown in Figure 5.7. Addition of SoxS resulted in reduced transcription from the *csgDEFG* regulatory region, including of the expected 244 nt transcript (lanes 2-4). Addition of MlrA had little effect (lanes 6-9), whilst addition of SoxS to reactions also containing MlrA reduced transcription (lanes 10-12). Similar observations were made when SoxS was added to reactions in the presence of IHF with or without MlrA (lanes -14-28). Overall, the pattern of transcription was difficult to interpret because of the high-level background transcription. Most likely, this is a consequence of the DNA fragment's high AT-content. Such DNA sequences cause transcription to initiate non-specifically because they resemble promoter elements for RNA polymerase (Warman *et al.*, 2021, Singh and Grainger, 2013).

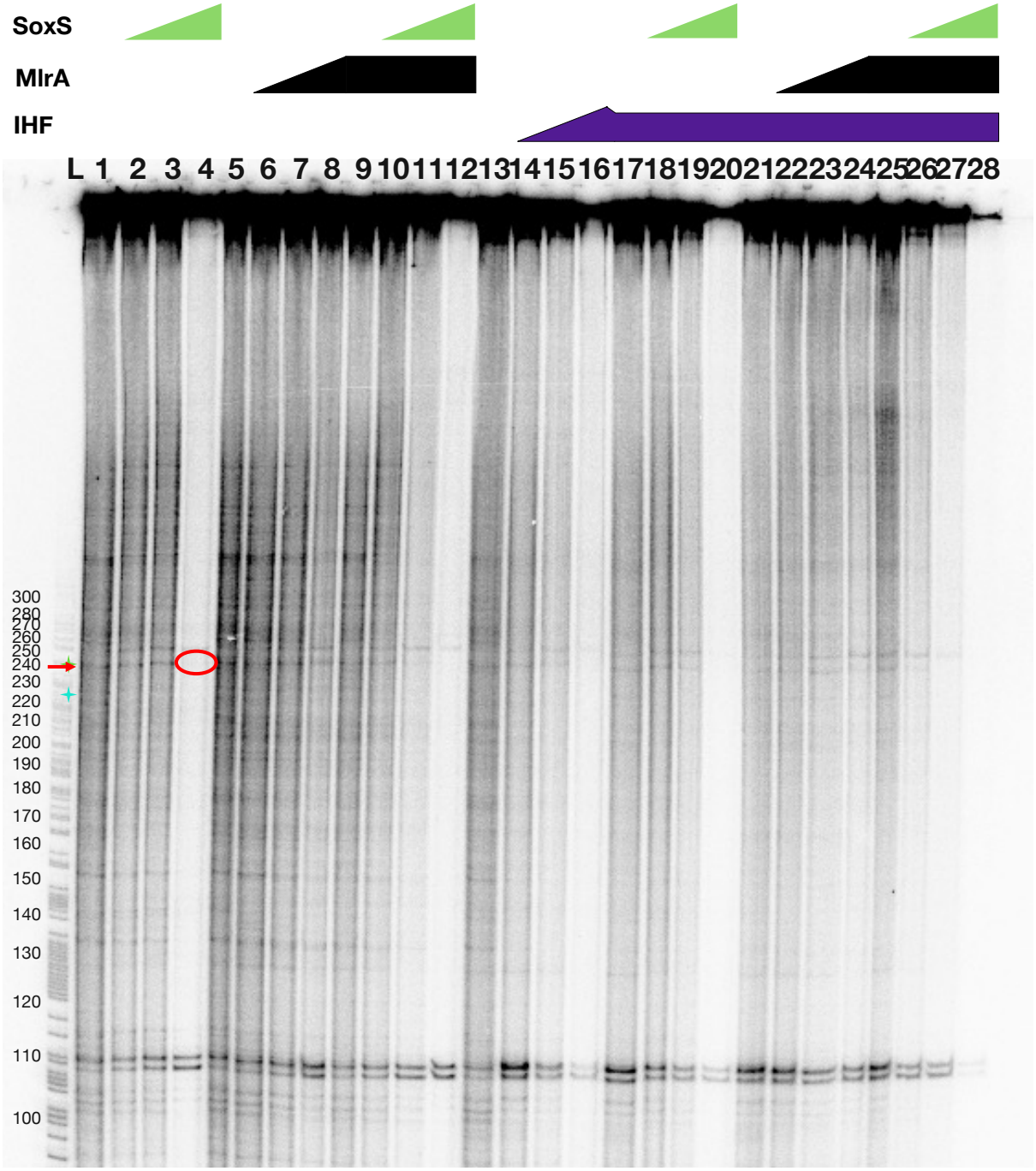


Figure 5.7 The effect of SoxS on the transcription of the *csgD* promoter. The effect of SoxS on the transcription from the *csgD* promoter assayed using *in vitro* transcription. Reactions were analysed using a 6 % denaturing acrylamide gel. Proteins were added at concentrations of 1, 2, and 4 µM as indicated. Transcription was done using *E. coli* RNAP. The two annotated

TSSs in *E. coli* are shown by blue and green stars. The TSS identified here is indicated by a red arrow, and the corresponding DNA band is highlighted in red for the first instance.

5.5 SoxS can bind to both binding sites present in the *csgD* intergenic region

Having confirmed direct repression by SoxS *in vitro* the interaction between SoxS and the intergenic region was probed further. As shown in Figure 5.2, there are two potential SoxS binding sites present within the *csgD* intergenic region: one between promoter positions -140 and -126 (site 1) and another at bases -301 to -287 (site 2). To determine which of these are bound by SoxS, the sites were mutated individually or together. A schematic representation of the mutant DNA fragments is shown in Figure 5.8. The specific mutations used to abolish binding site function are shown by red text in Figure 5.2 (red bases).

To test SoxS binding, an EMSA was done (Figure 5.9). Binding to the wildtype (WT) fragment was as in Figure 5.3, with a non-specific smear observed at concentrations of 2 µM SoxS. Both the Mut 1 (lacking site 1) and Mut 2 (lacking site 2) DNA fragments show reduced binding compared to WT. Binding was not abolished completely when both sites were mutated (Mut 1+2). These results suggest that SoxS interacts with both sites.

5.6 SoxS does not outcompete the activators MlrA or IHF when binding to the *csgD* intergenic region

Further EMSA experiments were done to understand the effect of SoxS on binding of the activators MlrA and IHF. First, competition between SoxS and MlrA was tested (Figure 5.10). SoxS and MlrA (both added at a final concentration of 0.5 µM) were incubated with the WT,

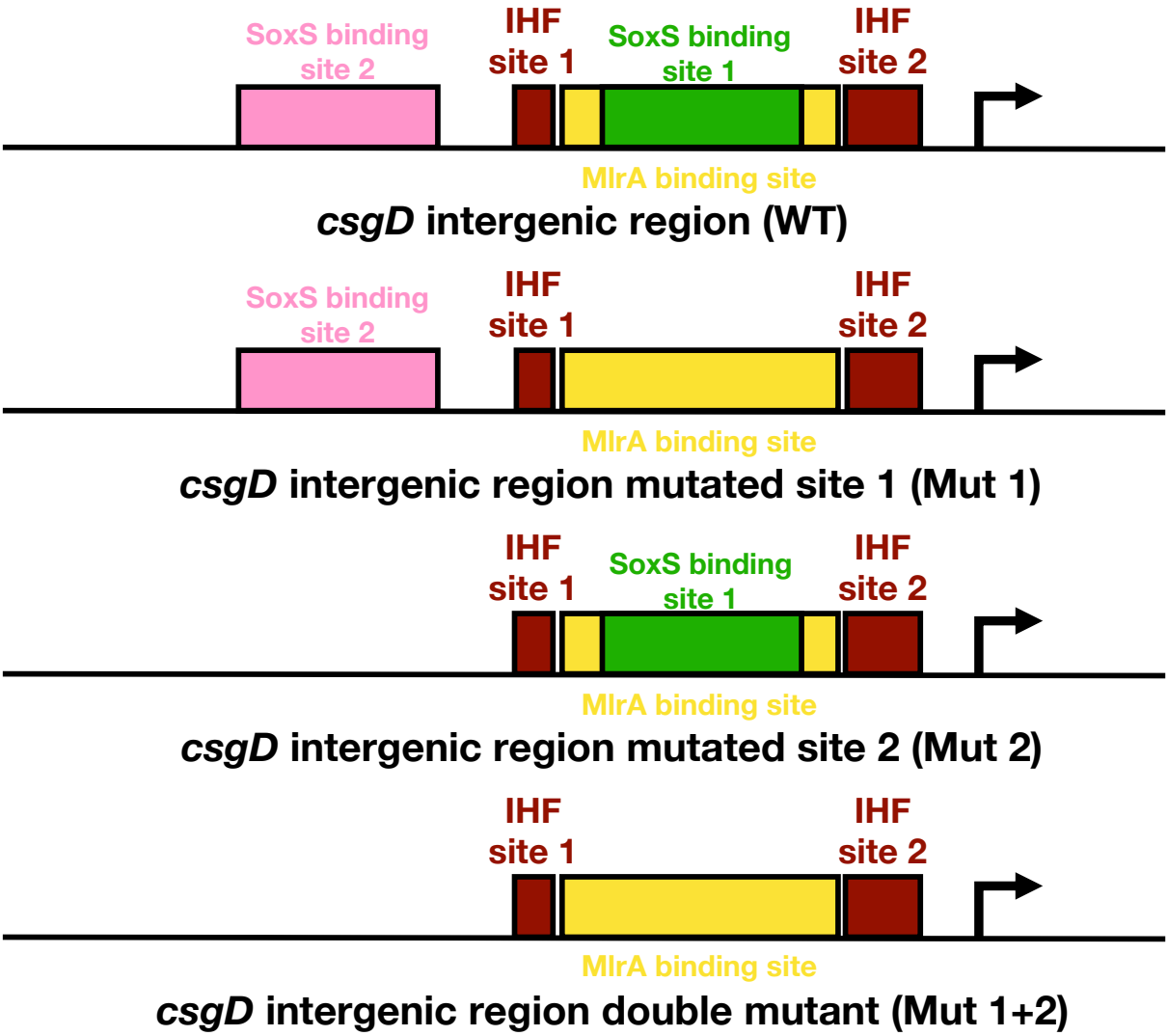


Figure 5.8 Schematic representation of the wildtype *csgD* intergenic region compared to the mutants generated. A schematic representation of each *csgD* intergenic region generated here. TF binding sites are shown as coloured boxes and TSSs are shown as right-angled arrows. In this study, each mutant is referred to by the name given to it in brackets.

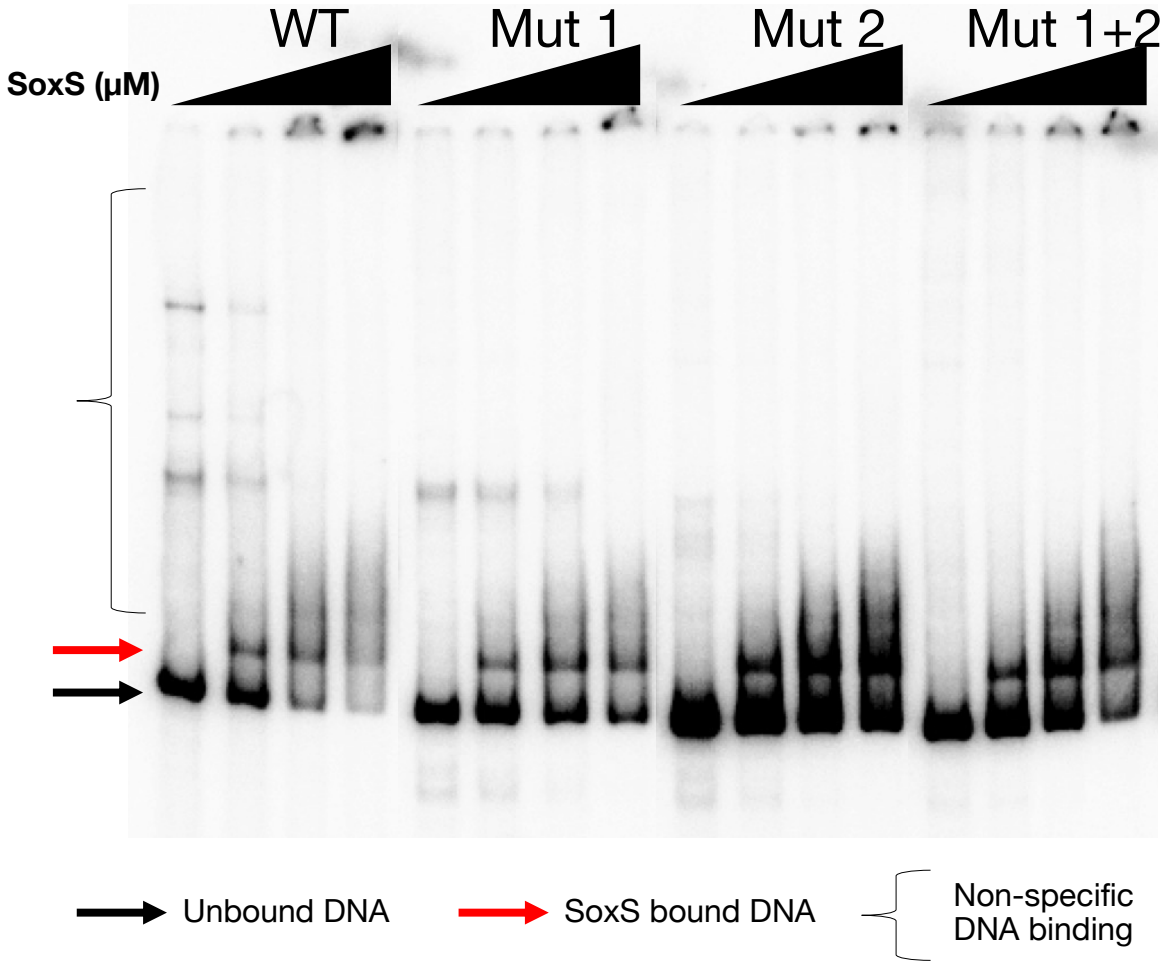


Figure 5.9 The effect of mutating the SoxS binding sites within the *csgD* intergenic region on SoxS binding. EMSA experiment showing binding of SoxS to the *csgD* intergenic region mutants, run on a 6 % polyacrylamide gel. Triangles indicate increasing concentrations of SoxS. Concentrations of SoxS used are 0, 0.5, 1, 1.5 and 2 µM for each mutant fragment. The wildtype intergenic region is shown as WT. Each individual mutant is named after the binding site mutated, with the Mut1+2 fragment having both sites mutated as described in Figure 5.1 and 5.7.

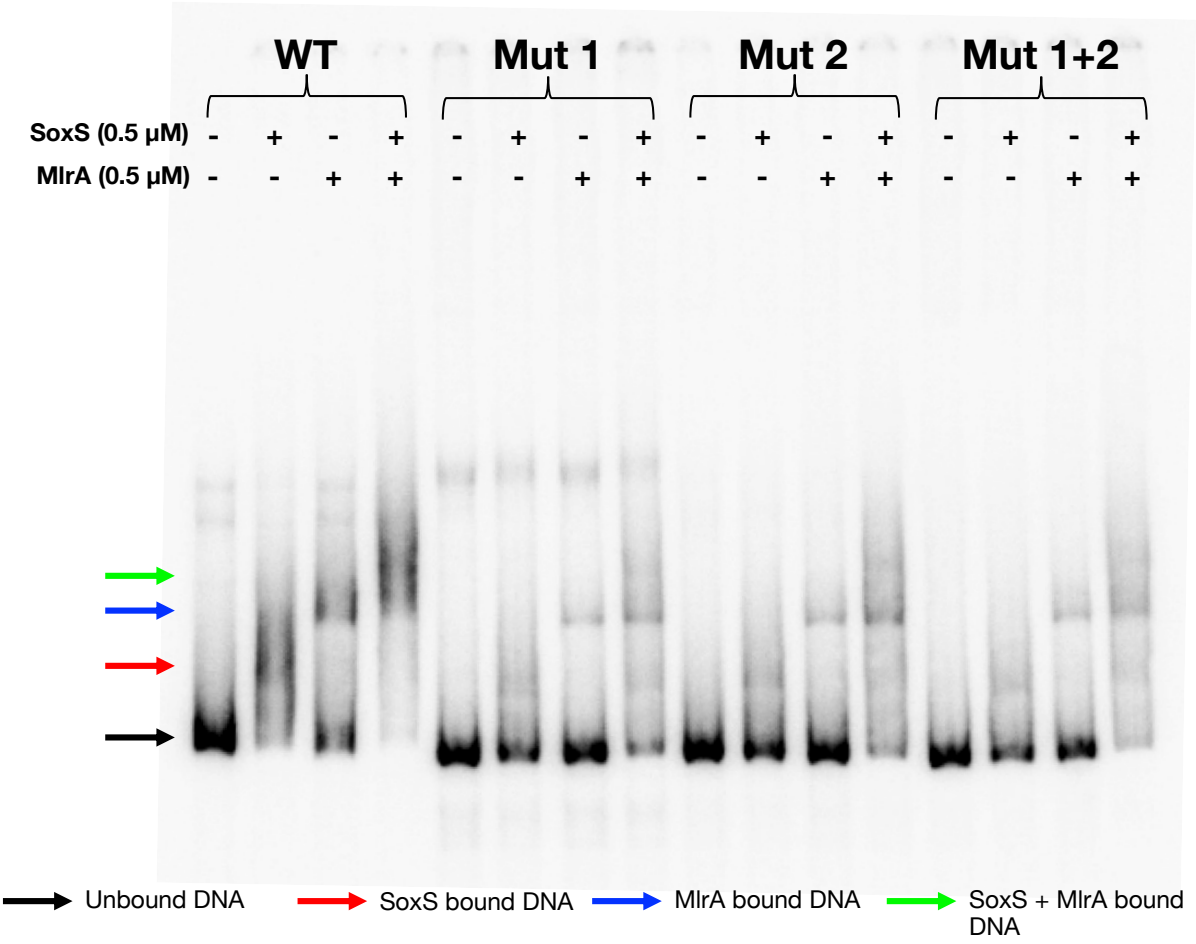


Figure 5.10 EMSA showing competition between SoxS and MlrA in binding the *csgD* intergenic region. EMSA experiment, run on a 6 % polyacrylamide gel, showing binding of SoxS or RamA to the wildtype (WT), Mut 1, Mut 2, or Double mutant fragments of the *csgD* intergenic region. SoxS and MlrA are supplied at 0.5 µM concentrations either individually, or together, the presence of each protein is indicated with a '+'.

Mut 1, Mut 2, or Mut 1+2 DNA fragments individually or together (Figure 5.10). SoxS (lane 2) and MlrA (lane 3) form different mobility complexes in conjunction with WT DNA. The complex observed when SoxS and MlrA are co-incubated (lanes 4) indicates binding of both SoxS and MlrA. Dual occupation of the fragment by both SoxS and MlrA is not observed with the mutants Mut 1, Mut 2, Mut 1+2. As expected, binding of SoxS alone is severely reduced with all mutant fragments. The ability of MlrA to bind mutated fragments was unaffected. It was concluded that SoxS represses *csgD* by some method other than steric hindrance of MlrA.

The binding sites of IHF are located just upstream and downstream of the MlrA binding sites, and SoxS binding site 1, at the *csgDEFG* promoter (Ogasawara *et al.*, 2010a). Hence EMSA experiments comparing the binding of SoxS and IHF were also done (Figure 5.11). IHF (lane 2) and SoxS (lane 8) formed complexes with different mobilities. IHF binding was similar in the presence and absence of SoxS (lanes 2-6 and 9-12). Therefore, it was concluded that IHF binding is unimpeded by SoxS.

5.7 SoxS-dependent repression of the *csgD* intergenic region requires SoxS site 2

To understand which of the SoxS sites were required for repression of *csgDEFG* promoter activity, *lacZ* fusion assays were done to test the different mutant DNA fragments (Figure 5.12). As expected, with the wildtype fragment (green) SoxS repressed *csgD* expression in both the presence and absence of MlrA. When site 1 is mutated (Mut 1) the level of repression is unchanged, indicating that this binding site is not required for repression of *csgD*. However, when site 2 is mutated (Mut 2), repression was reduced. Specifically, in the absence of ectopic

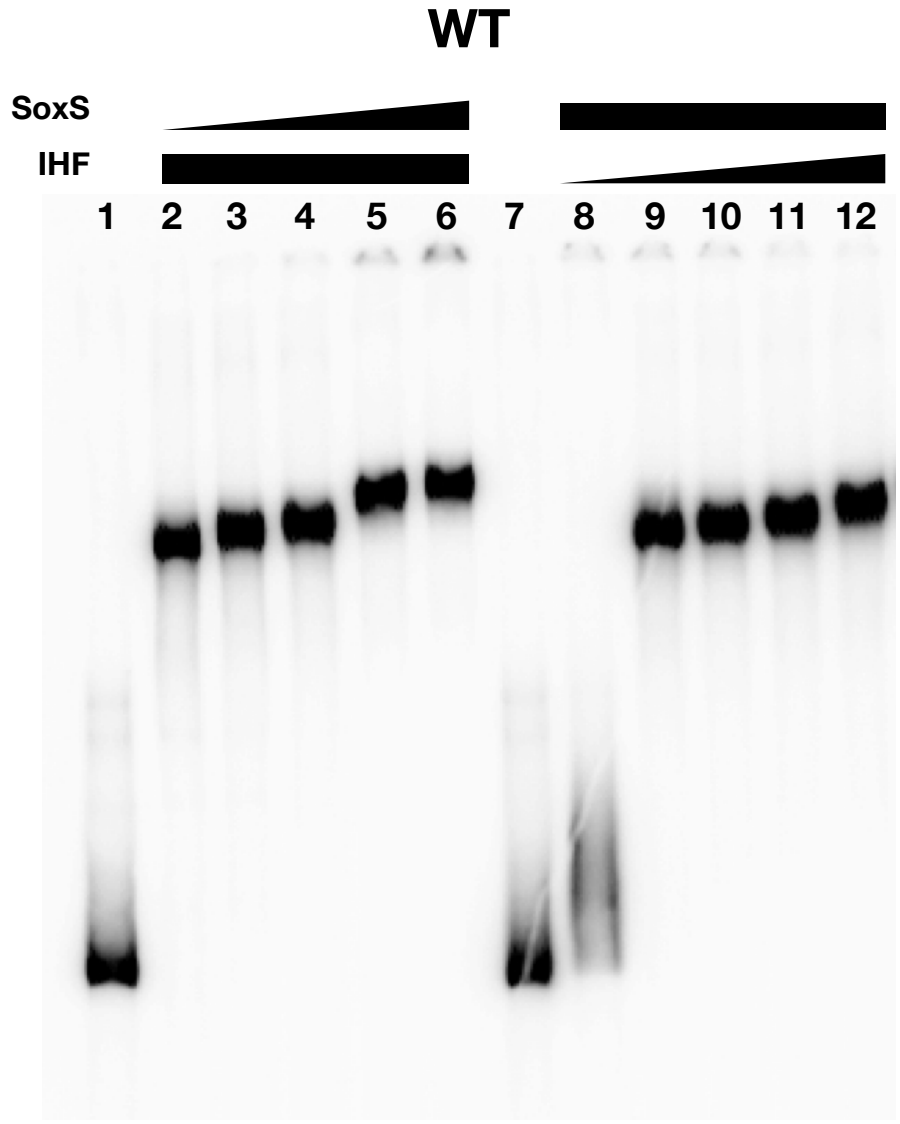


Figure 5.11 EMSA showing competition between SoxS and IHF in binding the *csgD* intergenic region. EMSA experiment, run on a 6 % polyacrylamide gel, showing binding of SoxS or IHF to the wildtype (WT) fragment of the *csgD* intergenic region. The static concentrations of SoxS (1 µM) and IHF (2 µM) are represented by a black bar. Titrations of SoxS and IHF are shown by triangles and the concentration was increase in 0.5 µM increments from 0-2 µM.

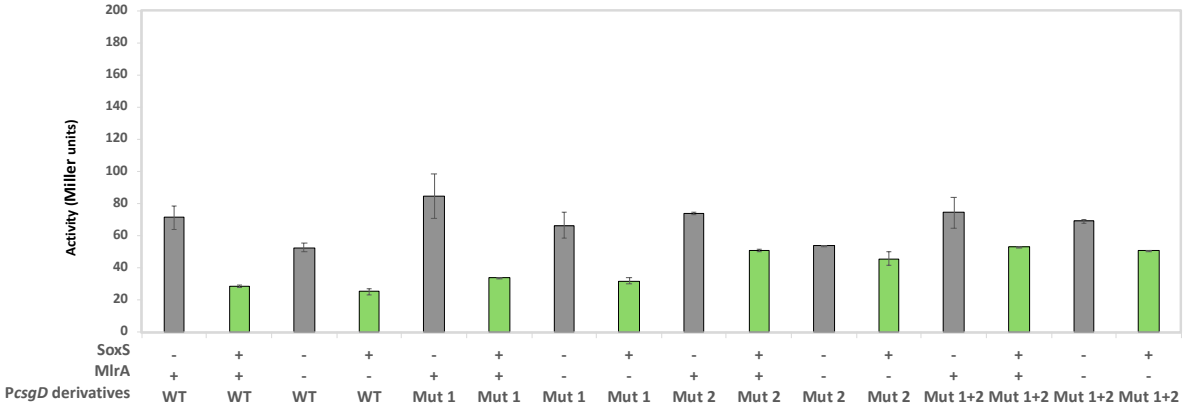


Figure 5.12 The effect of SoxS on the level of *csgD* expression when the SoxS binding sites are mutated individually or together. The effect of SoxS on the level of transcription of the *csgDEFG* promoter shown as the average promoter activity, ± standard deviation, of three biological replicates is shown. The wildtype *csgD* intergenic region with functional binding sites is shown as WT. The individual binding site mutants, Mut 1 and Mut 2, are named after the binding site which has been mutated. Both binding sites are mutated in the Mut 1+2 mutant. Note that '-' denotes empty pAM plasmids without the transcription factor cloned in; in the case of SoxS, Rob, and RamA this was pAMNF and for MarA this was pAMNM. For MlrA, this represents the absence of the pUC57-MlrA plasmid.

MlrA expression, the level of transcription was near identical in the presence and absence of SoxS. A small repressive effect of SoxS was seen when MlrA was expressed ectopically. The Mut 1+2 fragment behaved similarly to Mut 2. It was concluded that SoxS site 2 is key for repression.

5.8 Repression of CsgD-mediated biofilm formation *in vivo*

To assess how SoxS impacts biofilm formation, crystal violet assays were done in the presence or absence of ectopic MlrA or SoxS expression (Figure 5.13). Briefly, the strains were grown overnight, before dilution to an OD_{600} of 0.1, in 200 µL aliquots, in a flat-bottomed 96-well microtitre plate. The strains were then incubated at 30 °C for 48 hours. Wells were then washed, and biofilms stained with crystal violet. The innate level of biofilm formation, without expression of MlrA was low (Figure 5.13, white). When MlrA is expressed, biofilm formation increased considerably. Expression of SoxS reduced biofilm formation.

Crystal violet assays provide an indirect measure of biofilm forming capability through hydrostatic interactions between the dye and cell wall components, leading to retention of the dye (Choong *et al.*, 2016). Congo Red binds directly to the exopolysaccharide components (including curli fibres) of biofilms providing a more direct measure of biofilm formation (Choong *et al.*, 2016). Hence, the effects of SoxS and MlrA were tested using this assay (figure 5.14). Overnight cultures were diluted 1:10,000 before 5 µL was spotted onto no salt LB agar containing Congo Red (40 µg/mL), and the plates incubated for 48 hours at 30 °C. SoxS reduced Congo Red binding in the absence of ectopic MlrA expression.

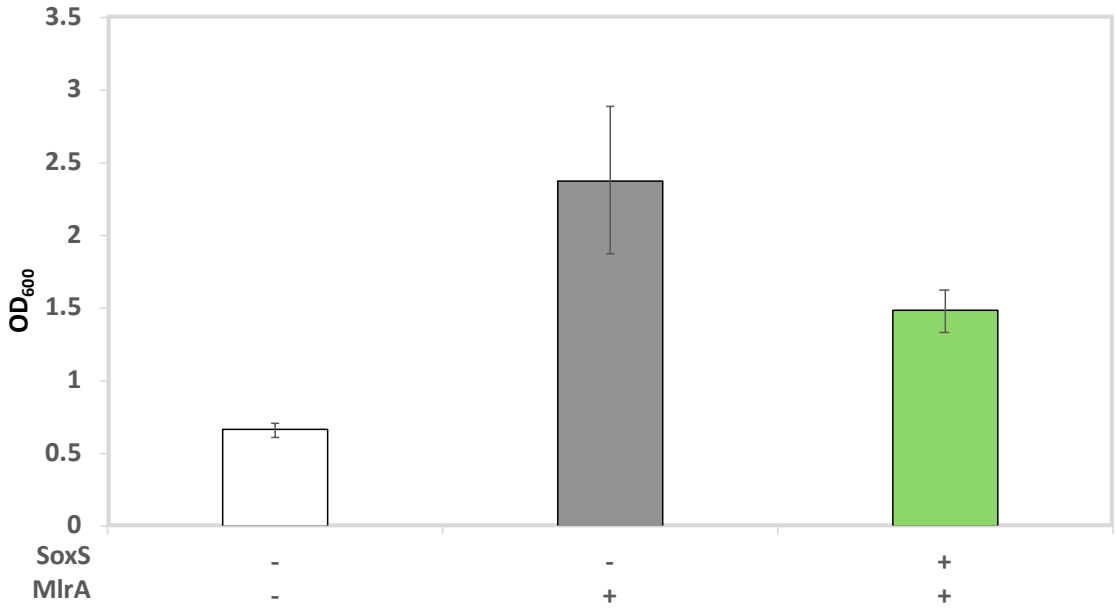


Figure 5.13 Regulation of biofilm formation by SoxS in *Salmonella* SL1344. Crystal violet biofilm staining assays showing the level of biofilm formation by SL1344 in response to the ectopic expression of either MlrA (grey) or SoxS and MlrA (green). An overnight culture of each strain was diluted to and OD_{600} of 0.1 in LB. 200 µL aliquots were incubated in a flat-bottomed 96-well plate for a further 48 hours at 30 °C without shaking. Biofilms were stained with 0.1 % crystal violet for 15 mins before solubilisation with 70 % ethanol and OD_{600} measurements taken. The average OD_{600} is shown, ± standard deviation, of 4 biological repeats. Note that '-' denotes either the empty pAMNF plasmid without *soxS* cloned in, or the absence of the pUC57-MlrA plasmid.

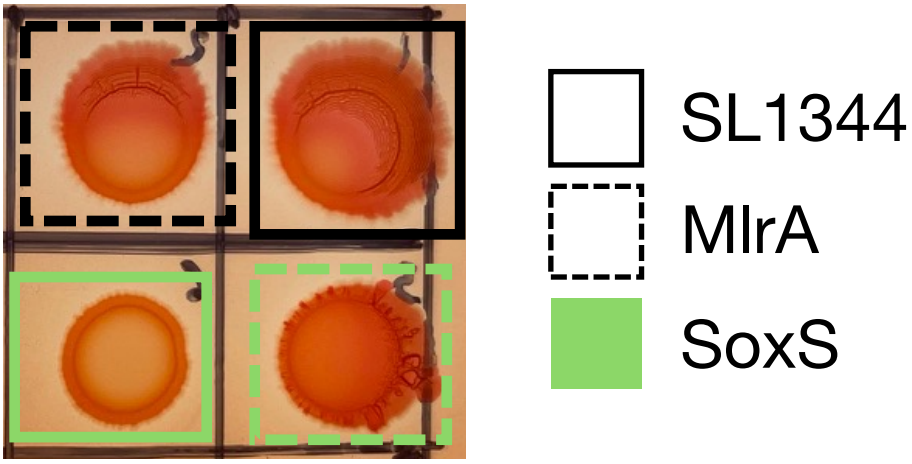


Figure 5.14 Regulation of curli fibre production by SoxS in *Salmonella* SL1344. Congo Red assay to assess curli fibre production by SL1344 in response to ectopic expression of SoxS (green) in the presence or absence of MlrA (dashed boxes). WT curli fibre production is shown in black. Overnight cultures were diluted 1:10,000 before 5 µL was spotted onto no salt LB agar containing Congo Red (40 µg/mL), and the plates incubated for 48 hours at 30 °C. Photos were taken with an iPhone.

5.9 Discussion

Previous work identified indirect repression of *csgD* expression, and biofilm formation, in *E. coli* by MarA (Kettles *et al.*, 2019). Briefly, MarA activates the *bluR-bluF-ycgZ-ymgAB* operon that triggers a cascade ultimately leading to *csgD* repression. However, the *ycgZ-ymgAB* operon is not present in *Salmonella*. Hence, the Mar, Sox, Rob, and Ram family of regulators must repress *csgD* directly. Despite the fact that *csgD* was only identified in the SoxS ChIP-seq experiment it was shown that all four homologous TFs interact with this region both *in vitro* and lead to differential *csgDEFG* expression *in vivo* (Figures 5.4-6). Both MarA and Rob bound the *csgD* intergenic region more tightly than SoxS and RamA (Figure 5.4). This was not expected for MarA as it was not identified as binding to the *csgD* region in the ChIP-seq data, but it was expected for Rob as it is known to bind DNA more tightly than the other TFs (Kettles, 2019). Unlike MarA and SoxS, rob recruits RNAP traditionally and is constitutively expressed but sequestered in an inactive state in intracellular foci (Martin and Rosner, 2002, Griffith *et al.*, 2009). A combination of these factors may explain why the tight binding of Rob to its targets does not translate to a constitutively active Rob phenotype.

This 754 bp intergenic region between the divergent *csgBAC-csgDEFG* operons has an AT content of 67 %, 20 % higher than the chromosomal AT percentage. The high level of non-specific activity observed in the *in vitro* transcription assays was likely due to this factor. The high AT content of this region presumably allowed spurious transcription initiation to occur, complicating the results (Lamberte *et al.*, 2017, Singh *et al.*, 2014, Singh and Grainger, 2013). This observation is likely to be irrelevant with respect to *in vivo* conditions, as H-NS is known

to bind this region (Ogasawara *et al.*, 2010a, Gerstel and Römling, 2003). H-NS binding is likely to ensure transcription specificity, as has been observed previously at the *ehxCABD* operon (Singh and Grainger, 2013).

Repression by SoxS of *csgDEFG* does not involve direct competition with two known activators of the operon, MlrA and IHF (Figures 5.10 and 11). However, the mechanism of repression must be caused by direct SoxS binding to this intergenic region, as *in vitro* transcription assays show reduction in transcription of P*csgD* when SoxS is present (Figure 5.7). Further to this, mutation of binding site 2 greatly reduces repression by SoxS.

The resulting effects on biofilm formation agree with previous work showing that the overexpression of MarA, SoxS, and RamA repress biofilm formation in *S*. Typhimurium 14028S (Baugh, 2014, Holden and Webber, 2020). Further to this, the results presented here bridge the gap between current observations that overexpression of MarA, SoxS, Rob, and RamA repress biofilm formation in *Salmonella* (Kettles *et al.*, 2019, Holden and Webber, 2020, Thota and Chubiz, 2019, Baugh *et al.*, 2012, Baugh, 2014). In the absence of the *bluR-bluF-ycgZ-ymgAB* pathway, *Salmonella* still repress *csgD* expression but do so directly by utilising SoxS.

6. Final conclusions

Antimicrobials have been used for thousands of years; with the ancient Greeks treating a variety of medical conditions, including burns, with treatments derived from moulds or plant extracts that we now know have antimicrobial properties (Lindblad, 2008). Whilst antimicrobials have revolutionised modern medicine, we disregarded the potential ramifications of overuse. With up to half of all antimicrobial use being unnecessary (CDC, 2017), the incidences of AMR have increased dramatically (O'Neill, 2016, Murray *et al.*, 2022). One further breeding ground for antimicrobial resistance is use in livestock, giving rise to drug resistant zoonotic pathogens (Silbergeld *et al.*, 2008). Whilst the focus of AMR is clinical relevance, the overuse of antimicrobials in agriculture is vastly underestimated, and is the largest use of antimicrobials worldwide (Silbergeld *et al.*, 2008).

The ability for AMR to develop is a consequence of evolution, which is born out of competition for space and nutrients (Butler and Buss, 2006). In a landmark report, it was estimated that, if left unchecked, AMR would cause 10 million deaths a year by 2050, more than cancer kills today (O'Neill, 2016). Unfortunately, recent evidence suggests that this report may have underestimated the global burden of AMR. In 2019 there were almost 5 million deaths associated with bacterial AMR (Murray *et al.*, 2022). If this trend increases, the global burden of AMR will be much greater than initially thought. It is, therefore, important that our stewardship of antimicrobials is improved in order to stem the spread and rate of AMR, whilst also extending the lifespan of current medicines. In addition to this, novel antimicrobial therapies are also desperately needed (WHO, 2020).

Whilst typically associated with a self-limiting gastroenteritis, *Salmonella* species accounted for 213,000 deaths in 2017 (Stanaway *et al.*, 2019a, Stanaway *et al.*, 2019b). *Salmonella* infections are frequently the result of contaminated food, leading to a lifestyle which alternates from hosts to the environment (Schikora *et al.*, 2012). As such, the majority of *Salmonella* species are able to form biofilms on a variety of biotic and abiotic surfaces, giving biofilms a role in virulence and transmission (White *et al.*, 2006, Gunn *et al.*, 2014, Prouty *et al.*, 2002, White *et al.*, 2008, Fàbrega and Vila, 2013). The shift from one lifestyle to another, for example from a planktonic state to sessile/biofilm state, requires a large reprogramming of cellular machinery (Steenackers *et al.*, 2012). As this shift utilises a lot of cellular resources, the switch from one lifestyle to another is tightly regulated by TFs (O'Toole *et al.*, 2000, Bjarnsholt, 2013).

TFs respond to environmental signals and modulate genetic expression levels, bridging the gap between the environment and the cellular response (Browning and Busby, 2004). The key regulators of AMR in *E. coli*, MarA, SoxS, and Rob have been implicated in clinically relevant AMR (Duval and Lister, 2013). In addition to MarA, SoxS, and Rob, *Salmonella* possess a fourth homologous TF, RamA, which is required for the MDR phenotype (Weston *et al.*, 2018). Previous work has identified genome-wide binding profiles for MarA and SoxS in *E. coli* (Sharma *et al.*, 2017b, Seo *et al.*, 2015). Further to this, there is also growing evidence that these TFs play a large role in the regulation of biofilm formation and motility within these pathogens (Kettles *et al.*, 2019, Holden and Webber, 2020, Thota and Chubiz, 2019, Baugh *et al.*, 2012, Baugh, 2014).

Whilst biofilms are involved in up to 80 % of all infections (Bjarnsholt *et al.*, 2018), they also have a large economic impact through their contamination of infrastructure, medical devices, and foodstuffs (Srey *et al.*, 2013, Davies, 2003, Vishwakarma, 2020). Developing our understanding of the complex interactions within the regulatory network of biofilm formation is essential in the development of treatment strategies to combat this phenomenon. Further to this, treatments which minimise the evolutionary pressure on the bacterium to develop resistance are of great desire. One common avenue of research is into anti-virulence drugs, which prevent virulence but are not bactericidal. Due to the involvement of biofilms in both infections and industrial applications, treatment of biofilms with an anti-virulence drug, biofilm dispersal agents, or adjuvant therapy could be important in the mitigation of these structures (Verderosa *et al.*, 2019).

To further our understanding of the effects of MarA, SoxS, Rob, and RamA on biofilm regulation, the aims of this study were threefold: firstly, to identify the genome-wide binding profile for four, clinically relevant TFs in *Salmonella*; secondly, to assess how these TFs alter the transcriptional landscape of *Salmonella*; and thirdly, to identify further links between these TFs and the regulation of biofilms and motility. In pursuit of the first aim, the genome-wide binding profiles of these regulators in *Salmonella* was identified for the first time. To achieve this, ChIP-seq was done in *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain SL1344. The binding profiles of MarA, SoxS, Rob, and RamA were associated with a total of 100 genes, 98 of which were chromosomally encoded. Only two binding peaks were observed on the virulence plasmid pSLT, and none were observed on plasmid pCol1B9. The

results obtained in this study also indicate that the SL1344 strain used here has lost the pRSF1010 plasmid. Considering these TFs recognise the same binding site, the level of overlap between the binding profiles observed for each TF was lower than expected. Unfortunately, the results obtained here did not provide new evidence as to how these TFs differentiate their binding targets when their consensus sites are highly similar. Future work could study this phenomenon by studying what is different between the regulators rather than what is similar, as small differences may play a larger role than initially considered. This could be achieved by analysing the surrounding sequences rather than the consensus binding site; or whether other parts of the protein may affect its DNA binding ability. Previous work has started to ask these questions and suggested that the level of salt present could be a factor in disrupting hydrogen bonds between the DNA-binding helices and the DNA backbone (Kettles, 2019). However, this avenue of research is still young and requires extensive further study.

Following ChIP-seq, the effect of MarA, SoxS, and RamA on the transcriptional landscape was studied. A comparison of bidirectionality, between *Salmonella* and *E. coli*, was also done. *Salmonella* has a similar number of bidirectional promoters compared to *E. coli* (Warman *et al.*, 2021), although their distributions in relation to coding sequences was different. Overexpression of MarA lead to the differential use of 535 TSSs. Only 24 of these TSSs were identified within 150 bases of any of the binding targets identified by ChIP-seq. The overexpression of SoxS showed a smaller transcriptional change, with only 180 differentially expressed TSSs; 11 of which correlated with the ChIP-seq data. In contrast, RamA showed the largest effect on the transcriptional landscape of *Salmonella*, with 694 TSSs differentially used

when RamA is overexpressed. This is to be expected as RamA is the dominant TF in the MDR phenotype of *Salmonella* (Weston *et al.*, 2018). Interestingly, only 30 of the RamA-induced, differentially used, TSSs correlate with ChIP-seq results. This could be because about 20 % of genes associated with these high level regulators are involved in transcription and translation regulation, leading to a large amount of indirect regulation. To illustrate this, a simplified hypothetical regulatory network generated using the data in Tables 3.1 and 4.1-6 is presented in Figure 6.1. Note that ovals show proteins, arrows show effects on expression of proteins mediated by transcriptional or translational regulation of associated genes. Blue arrows represent activation, red arrows represent repression. For clarity, self-regulation has been omitted and black arrows represent general regulation of the wider cellular processes (grey boxes) and include both activation and repression. Of the transcription factors studied here, RamA (red) showed the largest effect on the cellular landscape through both direct and indirect regulation of other regulatory proteins or small regulatory RNAs. Of note is the increase in expression of the small regulatory RNA *micF* by both RamA and MarA (blue), which has a large regulatory cascade itself and controls the expression of *cpxR* and *lrp* (Delihas and Forst, 2001, Holmqvist *et al.*, 2012, Weatherspoon-Griffin *et al.*, 2014). Furthermore, RamA repressed *csrC* which has been shown to be involved in the activation of SPI-1 via derepression of *csrA* (Fortune *et al.*, 2006, Liu *et al.*, 1997). SoxS (green) has been included even though none of the differentially expressed TSSs which co-localised with ChIP-seq binding peaks are transcription factors. Cappable-seq was not done in a strain of SL1344 ectopically producing Rob (yellow) so no regulatory network could be built hypothesised. Further work will need to be done in order to establish how SoxS and Rob contribute to this network, and time course experiments could be done to build a more accurate picture of how this network forms.

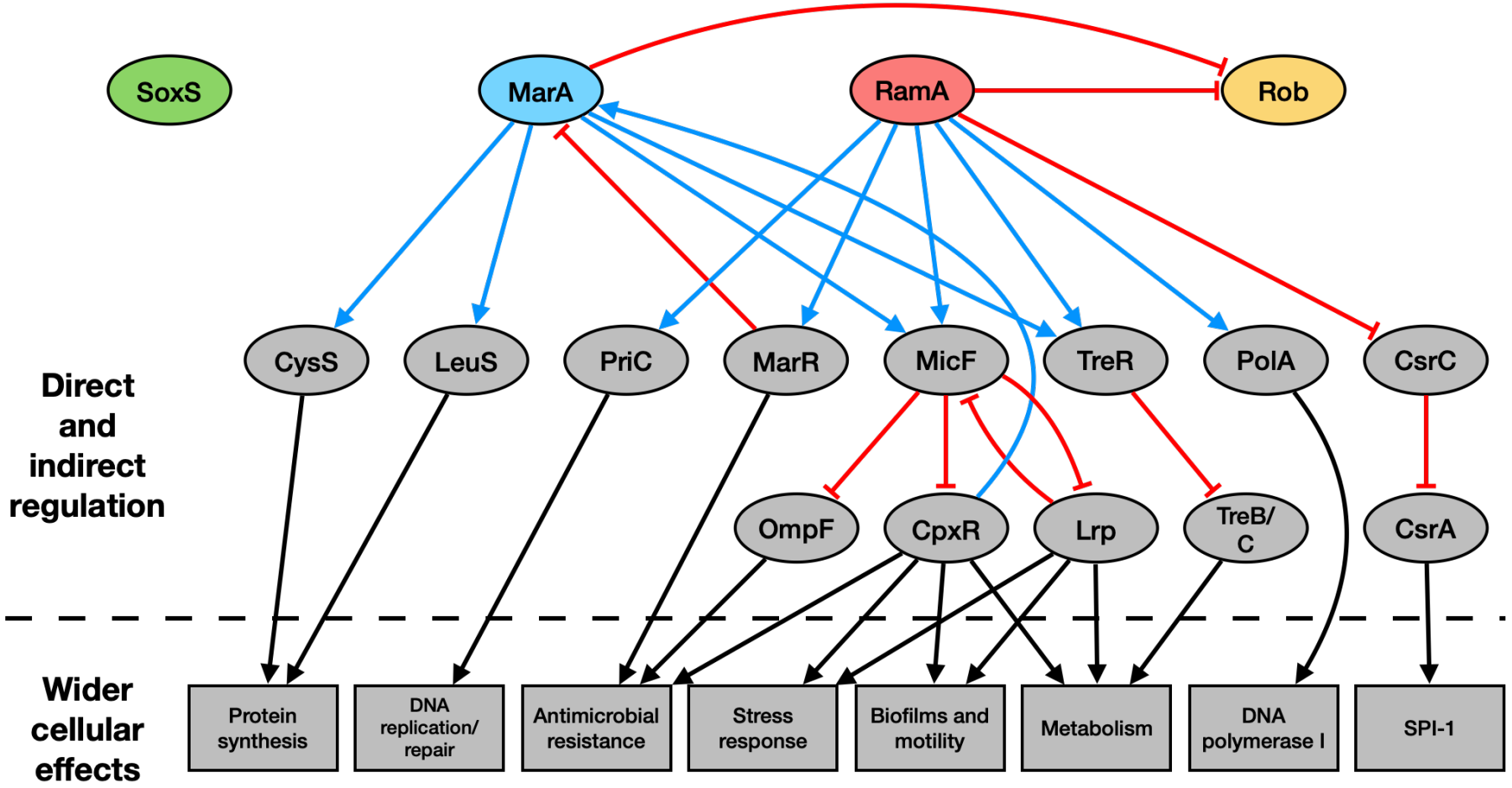


Figure 6.1 Hypothetical regulatory network in response to ectopic production of SoxS, MarA, and RamA in *Salmonella* SL1344. The hypothetical regulatory network controlled by SoxS, MarA, and RamA in SL1344 is determined from the data presented in Tables 3.1 and

4.1-6. Blue arrows represent activation and red arrows represent repression. For clarity, self-regulation has been omitted and black arrows show general effects (both activation and repression) on the cellular processes shown in grey boxes. Ovals represent proteins or small RNAs, the arrows linking them represent effects on their expression mediated by transcriptional or translational regulation of associated genes. Cappable-seq was not done in a strain of SL1344 ectopically producing Rob so no network could be built.

Further to this, the hierarchical nature of these transcription factors also cannot be determined from these results and should be studied in future work. This could be done using both ChIP-seq and Cappable-seq in various knockout combinations of these transcription factors to observe how this regulatory network shifts and responds to the absence of one or more of these transcription factors.

Returning to the ChIP-seq data, of the TFs studied, SoxS bound the most targets. One binding peak unique to SoxS was in the intergenic region between *csgBAC* and *csgDEFG*. This binding peak was interesting as MarA and SoxS are known to indirectly repress *csgD* expression in *E. coli*, through RprA activity (Kettles *et al.*, 2019, Tschowri *et al.*, 2012). Two clear binding sites for SoxS were identified in this complex intergenic region. SoxS binding site 1, was observed to be close to a so-called hotspot of TF binding (Ogasawara *et al.*, 2010a, Ogasawara *et al.*, 2010b) and also sits within the binding site of the activator MlrA; leading to the hypothesis that repression of *csgDEFG* expression by SoxS would occur by steric hindrance of MlrA. Upon further investigation, it was determined that SoxS directly represses the expression of *csgDEFG* through SoxS binding site 2. This observation was seen in both *in vitro* and *in vivo* experiments. As SoxS binding site 2 is -287 bases upstream of the TSS and not close to any major features of note, the mechanism of repression is assumed to be due to RNAP locking. These results agree with previous work by Baugh (2014). They also provide a direct link between observations that the inactivation of efflux induces the overexpression of *marA*, *soxS*, and *ramA*, and the repression of biofilm formation in *Salmonella* (Webber *et al.*, 2009, Holden and Webber, 2020). However, further research is required to confirm this conclusion

or establish the specific mechanism of repression by SoxS, and how widespread this phenomenon is. Regulation of *csgD* expression by SoxS has been observed directly in *Salmonella* (here) and indirectly in *E. coli* (Kettles *et al.*, 2019), so it is likely that these observations will also be seen in other Enterobacteriaceae family members. If this mechanism of SoxS-mediated biofilm regulation is widespread amongst the Enterobacteriaceae or wider community, these results would provide novel avenues of research into biofilm regulation and antimicrobial resistance.

These results are potentially of clinical relevance. With the identification of SoxS as a repressor of biofilm formation, future development of novel antimicrobials targeting SoxS-mediated resistance will need to address the effect of such treatment on biofilm formation. If a therapy reduces the effect of SoxS then biofilm production could increase as a consequence, posing an additional clinical challenge. In such a therapy, biofilm dispersing agents could be given as adjuvant therapies. This could be of benefit in multiple ways, by both dispersing biofilm formation, providing the therapeutic easier access to the bacteria, as well as reducing a potential physiological response to the therapeutic, which could cause an increase in biofilm production.

In addition, the interplay between SoxS and biofilm regulation could be exploited in industrial settings. *Salmonella* biofilms are known to form on a variety of abiotic substances commonly found in environments such as meat and food processing plants, farms, and kitchens, as well

as on plant matter (see Steenackers *et al.* (2012) for a full review). As such, application of these results to biofilm dispersal in these environments would reduce environmental carriage of *Salmonella* and, therefore, prevent infections and the need for treatments with antimicrobials—reducing the development and spread of AMR. In response to nitric oxide stress, *Salmonella* respond with the dispersal of biofilms (Barraud *et al.*, 2015); this could be exploited by exposing environmental biofilms in slaughter houses or industrial meat processing plants and kitchens to nitric oxide. In its gaseous form, nitric oxide, when compared to the convection of solutes used in cleaning products, might also be better able to penetrate deep into the biofilm and facilitate its dissolution. This treatment could be combined with other cleaning or biofilm removal methods to enhance their efficacy and reduce the chance of persistent or recurrent biofilm formation.

In summary, the results obtained here also contribute to the global picture of biofilm regulation in *Salmonella*. The link between MarA and SoxS and biofilm formation in *E. coli* has been previously studied; and the ability of MarA to induce a substantial lifestyle switch in response to unfavourable conditions was shown (Kettles *et al.*, 2019). Counterintuitively, MarA, the activator for the MDR phenotype represses biofilm formation, a condition known to be protective and withstand antimicrobial stresses (Kettles *et al.*, 2019, Holden and Webber, 2020). As biofilm expression was reduced by MarA, it was suggested that *E. coli* use this to favour a short-term survival strategy and provide the option to migrate away from the environment if conditions deteriorate further. The results presented in this study bridge the gap between how *E. coli* and *Salmonella* utilise the global regulators of antimicrobial stress to

repress biofilm formation. Furthering the notion that the repression of biofilms by these global stress response regulators provides a survival mechanism against antimicrobials. It is hypothesised that if planktonic bacteria are subjected to antimicrobial stress, biofilm formation is repressed by direct binding of SoxS to the *csgDEFG* intergenic region. This counterintuitive repression of an antimicrobial resistance mechanism by a regulator of antimicrobial resistance may be of benefit to the bacteria, as formation of a biofilm at this time would be energetically costly and insufficient to aid survival. Therefore, repression of biofilm formation by SoxS could allow the bacteria escape the harmful environment.

References

ALDRED, K. J., KERNS, R. J. & OSHEROFF, N. 2014. Mechanism of quinolone action and resistance. *Biochemistry,* 53**,** 1565-1574.

ALEKSHUN, M. N., LEVY, S. B., MEALY, T. R., SEATON, B. A. & HEAD, J. F. 2001. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. *Nat Struct Biol,* 8**,** 710-4.

ALTHOUSE, C., PATTERSON, S., FEDORKA-CRAY, P. & ISAACSON, R. E. 2003. Type 1 fimbriae of Salmonella enterica serovar Typhimurium bind to enterocytes and contribute to colonization of swine in vivo. *Infection and immunity,* 71**,** 6446-6452.

ALTIER, C. 2005. Genetic and environmental control of Salmonella invasion. *Journal of microbiology,* 43**,** 85-92.

ÁLVAREZ-ORDÓÑEZ, A., PRIETO, M., BERNARDO, A., HILL, C. & LÓPEZ, M. 2012. The Acid Tolerance Response of Salmonella spp.: An adaptive strategy to survive in stressful environments prevailing in foods and the host. *Food Research International,* 45**,** 482-492.

ARTSIMOVITCH, I. & LANDICK, R. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 7090-7095.

AUSTIN, J. W., SANDERS, G., KAY, W. W. & COLLINSON, S. K. 1998. Thin aggregative fimbriae enhance Salmonella enteritidis biofilm formation. *FEMS microbiology letters,* 162**,** 295-301.

AVITAL, G., AVRAHAM, R., FAN, A., HASHIMSHONY, T., HUNG, D. T. & YANAI, I. 2017. scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing. *Genome Biology,* 18**,** 200.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res,* 37**,** W202-8.

BAJAJ, V., LUCAS, R. L., HWANG, C. & LEE, C. A. 1996. Co-ordinate regulation of Salmonella typhimurium invasion genes by environmental and regulatory factors is mediated by control of hilA expression. *Molecular microbiology,* 22**,** 703-714.

BAKOWSKI, M. A., CIRULIS, J. T., BROWN, N. F., FINLAY, B. B. & BRUMELL, J. H. 2007. SopD acts cooperatively with SopB during Salmonella enterica serovar Typhimurium invasion. *Cellular microbiology,* 9**,** 2839-2855.

BALASUBRAMANIAN, R., IM, J., LEE, J.-S., JEON, H. J., MOGENI, O. D., KIM, J. H., RAKOTOZANDRINDRAINY, R., BAKER, S. & MARKS, F. 2019. The global burden and epidemiology of invasive non-typhoidal Salmonella infections. *Human Vaccines & Immunotherapeutics,* 15**,** 1421-1426.

BALKIN, A. S., PLOTNIKOV, A. O., GOGOLEVA, N. E., GOGOLEV, Y. V., DEMCHENKO, K. N. & CHERKASOV, S. V. 2021. Cappable-Seq Reveals Specific Patterns of Metabolism and Virulence for Salmonella Typhimurium Intracellular Survival within Acanthamoeba castellanii. *International journal of molecular sciences,* 22**,** 9077.

BAQUERO, F., COQUE, T. M., GALÁN, J. C. & MARTINEZ, J. L. 2021. The Origin of Niches and Species in the Bacterial World. *Frontiers in Microbiology,* 12.

BAR-NAHUM, G., EPSHTEIN, V., RUCKENSTEIN, A. E., RAFIKOV, R., MUSTAEV, A. & NUDLER, E. 2005. A ratchet mechanism of transcription elongation and its control. *Cell,* 120**,** 183-93.

BARNHART, M. M. & CHAPMAN, M. R. 2006. Curli biogenesis and function. *Annu. Rev. Microbiol.,* 60**,** 131-147.

BARRAUD, N., J KELSO, M., A RICE, S. & KJELLEBERG, S. 2015. Nitric oxide: a key mediator of biofilm dispersal with applications in infectious diseases. *Current pharmaceutical design,* 21**,** 31-42.

BAUGH, S. 2014. *The role of multidrug efflux pumps in biofilm formation of Salmonella enterica serovar Typhimurium.* University of Birmingham.

BAUGH, S., EKANAYAKA, A. S., PIDDOCK, L. J. & WEBBER, M. A. 2012. Loss of or inhibition of all multidrug resistance efflux pumps of Salmonella enterica serovar Typhimurium results in impaired ability to form a biofilm. *Journal of Antimicrobial Chemotherapy,* 67**,** 2409-2417.

BERGER, C. N., SODHA, S. V., SHAW, R. K., GRIFFIN, P. M., PINK, D., HAND, P. & FRANKEL, G. 2010. Fresh fruit and vegetables as vehicles for the transmission of human pathogens. *Environmental microbiology,* 12**,** 2385-2397.

BERLANGA, M. & GUERRERO, R. 2016. Living together in biofilms: the microbial cell factory and its biotechnological implications. *Microbial Cell Factories,* 15**,** 165.

BERTELSEN, L. S., PAESOLD, G. N., MARCUS, S. L., FINLAY, B. B., ECKMANN, L. & BARRETT, K. E. 2004. Modulation of chloride secretory responses and barrier function of intestinal epithelial cells by the Salmonella effector protein SigD. *American Journal of Physiology-Cell Physiology,* 287**,** C939-C948.

BJARNSHOLT, T. 2013. The role of bacterial biofilms in chronic infections. *APMIS,* 121**,** 1-58.

BJARNSHOLT, T., BUHLIN, K., DUFRÊNE, Y., GOMELSKY, M., MORONI, A., RAMSTEDT, M., RUMBAUGH, K., SCHULTE, T., SUN, L. & ÅKERLUND, B. 2018. Biofilm formation–what we can learn from recent developments. Wiley Online Library.

BLATTER, E. E., ROSS, W., TANG, H., GOURSE, R. L. & EBRIGHT, R. H. 1994. Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell,* 78**,** 889-96.

BLISKA, J. B. & VAN DER VELDEN, A. W. 2012. Salmonella "sops" up a preferred electron receptor in the inflamed intestine. *MBio,* 3**,** e00226-12.

BOONYOM, R., KARAVOLOS, M., BULMER, D. & KHAN, C. 2010. Salmonella pathogenicity island 1 (SPI-1) type III secretion of SopD involves N-and C-terminal signals and direct binding to the InvC ATPase. *Microbiology,* 156**,** 1805-1814.

BORUKHOV, S. & NUDLER, E. 2008. RNA polymerase: the vehicle of transcription. *Trends Microbiol,* 16**,** 126-34.

BRENNER, F. W., VILLAR, R. G., ANGULO, F. J., TAUXE, R. & SWAMINATHAN, B. 2000. Salmonella nomenclature. *Journal of clinical microbiology,* 38**,** 2465-2467.

BROWN, P. K., DOZOIS, C. M., NICKERSON, C. A., ZUPPARDO, A., TERLONGE, J. & CURTISS, R., 3RD 2001. MlrA, a novel regulator of curli (AgF) and extracellular matrix synthesis by Escherichia coli and Salmonella enterica serovar Typhimurium. *Mol Microbiol,* 41**,** 349-63.

BROWNING, D. F. & BUSBY, S. J. 2004. The regulation of bacterial transcription initiation. *Nat Rev Microbiol,* 2**,** 57-65.

BROWNING, D. F. & BUSBY, S. J. W. 2016. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology,* 14**,** 638.

BROWNING, D. F., BUTALA, M. & BUSBY, S. J. W. 2019. Bacterial Transcription Factors: Regulation by Pick "N" Mix. *Journal of Molecular Biology,* 431**,** 4067-4077.

BRZÓSTKOWSKA, M., RACZKOWSKA, A. & BRZOSTEK, K. 2012. OmpR, a response regulator of the two-component signal transduction pathway, influences inv gene expression in Yersinia enterocolitica O9. *Frontiers in cellular and infection microbiology,* 2**,** 153-153.

BURGESS, R. R. & JENDRISAK, J. J. 1975. A procedure for the rapid, large-scall purification of Escherichia coli DNA-dependent RNA polymerase involving Polymin P precipitation and DNA-cellulose chromatography. *Biochemistry,* 14**,** 4634-8.

BUSBY, S. & EBRIGHT, R. H. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell,* 79**,** 743-6.

BUSHNELL, D. A., CRAMER, P. & KORNBERG, R. D. 2002. Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. *Proc Natl Acad Sci U S A,* 99**,** 1218-22.

BUTLER, M. S. & BUSS, A. D. 2006. Natural products--the future scaffolds for novel antibiotics? *Biochem Pharmacol,* 71**,** 919-29.

BUTTON, J. E. & GALÁN, J. E. 2011. Regulation of chaperone/effector complex synthesis in a bacterial type III secretion system. *Molecular microbiology,* 81**,** 1474-1483.

CADENA, M., KELMAN, T., MARCO, M. L. & PITESKY, M. 2019. Understanding Antimicrobial Resistance (AMR) Profiles of Salmonella Biofilm and Planktonic Bacteria Challenged with Disinfectants Commonly Used During Poultry Processing. *Foods,* 8**,** 275.

CAI, S. J. & INOUYE, M. 2002. EnvZ-OmpR Interaction and Osmoregulation in Escherichia coli *. *Journal of Biological Chemistry,* 277**,** 24155-24161.

CARVER, T., HARRIS, S. R., BERRIMAN, M., PARKHILL, J. & MCQUILLAN, J. A. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics,* 28**,** 464-9.

CARVER, T., THOMSON, N., BLEASBY, A., BERRIMAN, M. & PARKHILL, J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics,* 25**,** 119-20.

CDC, C. F. D. C. A. P. 2019. *Antibiotic Resistance Threats Report 2019* [Online]. Available: https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf [Accessed 24.11.2021].

CDC, T. C. F. D. C. A. P. 2017. *About antimicrobial resistance* [Online]. Available: https://www.cdc.gov/drugresistance/about.html [Accessed 29 Jul 2019].

CHMIELEWSKI, R. & FRANK, J. 2003. Biofilm formation and control in food processing facilities. *Comprehensive reviews in food science and food safety,* 2**,** 22-32.

CHOONG, F. X., BÄCK, M., FAHLÉN, S., JOHANSSON, L. B. G., MELICAN, K., RHEN, M., NILSSON, K. P. R. & RICHTER-DAHLFORS, A. 2016. Real-time optotracing of curli and cellulose in live Salmonella biofilms using luminescent oligothiophenes. *npj Biofilms and Microbiomes,* 2**,** 16024.

CLARK, L., PERRETT, C. A., MALT, L., HARWARD, C., HUMPHREY, S., JEPSON, K. A., MARTINEZ-ARGUDO, I., CARNEY, L. J., LA RAGIONE, R. M., HUMPHREY, T. J. & JEPSON, M. A. 2011. Differences in Salmonella enterica serovar Typhimurium strain invasiveness are associated with heterogeneity in SPI-1 gene expression. *Microbiology (Reading, England),* 157**,** 2072-2083.

COHEN, S. P., YAN, W. & LEVY, S. B. 1993. A multidrug resistance regulatory chromosomal locus is widespread among enteric bacteria. *J Infect Dis,* 168**,** 484-8.

CORREIA, S., POETA, P., HÉBRAUD, M., CAPELO, J. L. & IGREJAS, G. 2017. Mechanisms of quinolone action and resistance: where do we stand? *Journal of Medical Microbiology,* 66**,** 551-559.

CROOKS, G. E., HON, G., CHANDONIA, J. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator. *Genome Res,* 14**,** 1188-90.

CRUMP, J. A., LUBY, S. P. & MINTZ, E. D. 2004. The global burden of typhoid fever. *Bulletin of the World Health Organization,* 82**,** 346-353.

CRUMP, J. A. & MINTZ, E. D. 2010. Global trends in typhoid and paratyphoid Fever. *Clin Infect Dis,* 50**,** 241-6.

CRUMP, J. A. & OO, W. T. 2021. Salmonella Typhi Vi polysaccharide conjugate vaccine protects infants and children against typhoid fever. *The Lancet,* 398**,** 643-644.

CRUMP, J. A. & WAIN, J. 2017. Salmonella. *In:* QUAH, S. R. (ed.) *International Encyclopedia of Public Health (Second Edition).* Oxford: Academic Press.

CUYPERS, W. L., JACOBS, J., WONG, V., KLEMM, E. J., DEBORGGRAEVE, S. & VAN PUYVELDE, S. 2018. Fluoroquinolone resistance in Salmonella: insights by whole-genome sequencing. *Microbial genomics,* 4**,** e000195.

DARWIN, K. H. & MILLER, V. L. 2001. Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in Salmonella typhimurium. *The EMBO journal,* 20**,** 1850-1862.

DAVEY, M. E. & O'TOOLE, G. A. 2000. Microbial biofilms: from ecology to molecular genetics. *Microbiology and molecular biology reviews,* 64**,** 847-867.

DAVIES, D. 2003. Understanding biofilm resistance to antibacterial agents. *Nature reviews Drug discovery,* 2**,** 114-122.

DEACON, R., LUMB, M., PERRY, J., CHANARIN, I., MINTY, B., HALSEY, M. & NUNN, J. 1980. Inactivation of methionine synthase by nitrous oxide. *Eur J Biochem,* 104**,** 419-23.

DELIHAS, N. & FORST, S. 2001. MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors11Edited by D. Draper. *Journal of Molecular Biology,* 313**,** 1-12.

DEMPLE, B. 1996. Redox signaling and gene control in the Escherichia coli soxRS oxidative stress regulon — a review. *Gene,* 179**,** 53-57.

DEY, S., CHAKRAVARTY, A., GUHA BISWAS, P. & DE GUZMAN, R. N. 2019. The type III secretion system needle, tip, and translocon. *Protein science : a publication of the Protein Society,* 28**,** 1582-1593.

DÉZIEL, E., GOPALAN, S., TAMPAKAKI, A. P., LÉPINE, F., PADFIELD, K. E., SAUCIER, M., XIAO, G. & RAHME, L. G. 2005. The contribution of MvfR to Pseudomonas aeruginosa pathogenesis and quorum sensing circuitry regulation: multiple quorum sensing-regulated genes are modulated without affecting lasRI, rhlRI or the production of N-acyl-L-homoserine lactones. *Molecular microbiology,* 55**,** 998-1014.

DRECKTRAH, D., KNODLER, L. A., GALBRAITH, K. & STEELE-MORTIMER, O. 2005. The Salmonella SPI1 effector SopB stimulates nitric oxide production long after invasion. *Cellular microbiology,* 7**,** 105-113.

DUVAL, V. & LISTER, I. M. 2013. MarA, SoxS and Rob of Escherichia coli - Global regulators of multidrug resistance, virulence and stress response. *International journal of biotechnology for wellness industries,* 2**,** 101-124.

EBRIGHT, R. H. 2000. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol,* 304**,** 687-98.

EL MOUALI, Y., GEROVAC, M., MINEIKAITĖ, R. & VOGEL, J. 2021. In vivo targets of Salmonella FinO include a FinP-like small RNA controlling copy number of a cohabitating plasmid. *Nucleic Acids Research,* 49**,** 5319-5335.

ELLERMEIER, J. R. & SLAUCH, J. M. 2007. Adaptation to the host environment: regulation of the SPI1 type III secretion system in Salmonella enterica serovar Typhimurium. *Current opinion in microbiology,* 10**,** 24-29.

ETTWILLER, L., BUSWELL, J., YIGIT, E. & SCHILDKRAUT, I. 2016. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics,* 17**,** 199.

FÀBREGA, A. & VILA, J. 2013. Salmonella enterica Serovar Typhimurium Skills To Succeed in the Host: Virulence and Regulation. *Clinical Microbiology Reviews,* 26**,** 308-341.

FEKLISTOV, A. & DARST, S. A. 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell,* 147**,** 1257-69.

FEKLÍSTOV, A., SHARON, B. D., DARST, S. A. & GROSS, C. A. 2014. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annual Review of Microbiology,* 68**,** 357-376.

FERREIRA, C., PEREIRA, A., MELO, L. & SIMÕES, M. 2010. Advances in industrial biofilm control with micro-nanotechnology. *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology,* 2**,** 845-854.

FIERER, J. & GUINEY, D. G. 2001. Diverse virulence traits underlying different clinical outcomes of Salmonella infection. *J Clin Invest,* 107**,** 775-80.

FOOKES, M., SCHROEDER, G. N., LANGRIDGE, G. C., BLONDEL, C. J., MAMMINA, C., CONNOR, T. R., SETH-SMITH, H., VERNIKOS, G. S., ROBINSON, K. S., SANDERS, M., PETTY, N. K., KINGSLEY, R. A., BÄUMLER, A. J., NUCCIO, S.-P., CONTRERAS, I., SANTIVIAGO, C. A., MASKELL, D., BARROW, P., HUMPHREY, T., NASTASI, A., ROBERTS, M., FRANKEL, G., PARKHILL, J., DOUGAN, G. & THOMSON, N. R. 2011. Salmonella bongori Provides Insights into the Evolution of the Salmonellae. *PLOS Pathogens,* 7**,** e1002191.

FORTUNE, D. R., SUYEMOTO, M. & ALTIER, C. 2006. Identification of CsrC and characterization of its role in epithelial cell invasion in Salmonella enterica serovar Typhimurium. *Infect Immun,* 74**,** 331-9.

FRYE, J. & JACKSON, C. 2013. Genetic mechanisms of antimicrobial resistance identified in Salmonella enterica, Escherichia coli, and Enteroccocus spp. isolated from U.S. food animals. *Frontiers in Microbiology,* 4.

GAL-MOR, O., BOYLE, E. C. & GRASSL, G. A. 2014. Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. *Frontiers in Microbiology,* 5.

GALLEGOS, M. T., SCHLEIF, R., BAIROCH, A., HOFMANN, K. & RAMOS, J. L. 1997. Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev,* 61**,** 393-410.

GARCÍA-GIL, A., GALÁN-ENRÍQUEZ, C. S., PÉREZ-LÓPEZ, A., NAVA, P., ALPUCHE-ARANDA, C. & ORTIZ-NAVARRETE, V. 2018. SopB activates the Akt-YAP pathway to promote Salmonella survival within B cells. *Virulence,* 9**,** 1390-1402.

GEERTZ, M., TRAVERS, A., MEHANDZISKA, S., SOBETZKO, P., CHANDRA-JANGA, S., SHIMAMOTO, N. & MUSKHELISHVILI, G. 2011. Structural coupling between RNA polymerase composition and DNA supercoiling in coordinating transcription: a global role for the omega subunit? *mBio,* 2**,** e00034-11.

GENTRY, D. R. & BURGESS, R. R. 1989. rpoZ, encoding the omega subunit of Escherichia coli RNA polymerase, is in the same operon as spoT. *J Bacteriol,* 171**,** 1271-7.

GERSTEL, U. & RÖMLING, U. 2003. The csgD promoter, a control unit for biofilm formation in Salmonella typhimurium. *Research in Microbiology,* 154**,** 659-667.

GOURSE, R. L., ROSS, W. & GAAL, T. 2000. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol,* 37**,** 687-95.

GRAINGER, D. C., AIBA, H., HURD, D., BROWNING, D. F. & BUSBY, S. J. 2007. Transcription factor distribution in Escherichia coli: studies with FNR protein. *Nucleic Acids Res,* 35**,** 269-78.

GRANTCHAROVA, N., PETERS, V., MONTEIRO, C., ZAKIKHANY, K. & RÖMLING, U. 2010. Bistable expression of CsgD in biofilm development of Salmonella enterica serovar typhimurium. *Journal of bacteriology,* 192**,** 456-466.

GRIFFITH, K. L., FITZPATRICK, M. M., KEEN, E. F., 3RD & WOLF, R. E., JR. 2009. Two functions of the C-terminal domain of Escherichia coli Rob: mediating "sequestration-dispersal" as a novel off-on switch for regulating Rob's activity as a transcription activator and preventing degradation of Rob by Lon protease. *Journal of molecular biology,* 388**,** 415-430.

GRIFFITH, K. L., SHAH, I. M., MYERS, T. E., O'NEILL, M. C. & WOLF, R. E., JR. 2002. Evidence for "pre-recruitment" as a new mechanism of transcription activation in Escherichia coli: the large excess of SoxS binding sites per cell relative to the number of SoxS molecules per cell. *Biochem Biophys Res Commun,* 291**,** 979-86.

GROVE, A. 2013. MarR family transcription factors. *Current biology,* 23**,** R142-R143.

GRUBER, T. M. & GROSS, C. A. 2003. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology,* 57**,** 441-466.

GU, M. & IMLAY, J. A. 2011. The SoxRS response of Escherichia coli is directly activated by redox-cycling drugs rather than by superoxide. *Molecular microbiology,* 79**,** 1136-1150.

GUERON, M. & LEROY, J. L. 1995. Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol,* 261**,** 383-413.

GUNN, J. S., MARSHALL, J. M., BAKER, S., DONGOL, S., CHARLES, R. C. & RYAN, E. T. 2014. Salmonella chronic carriage: epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol,* 22**,** 648-55.

GUSAROV, I. & NUDLER, E. 1999. The mechanism of intrinsic transcription termination. *Mol Cell,* 3**,** 495-504.

HAMMAR, M., ARNQVIST, A., BIAN, Z., OLSÉN, A. & NORMARK, S. 1995. Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12. *Mol Microbiol,* 18**,** 661-70.

HARAGA, A., OHLSON, M. B. & MILLER, S. I. 2008. Salmonellae interplay with host cells. *Nature Reviews Microbiology,* 6**,** 53-66.

HARRISON, E. & BROCKHURST, M. A. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology,* 20**,** 262-267.

HAUGEN, S. P., ROSS, W. & GOURSE, R. L. 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol,* 6**,** 507-19.

HAWLEY, D. K. & MCCLURE, W. R. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic acids research,* 11**,** 2237-2255.

HAYCOCKS, J. R. J., WARREN, G. Z. L., WALKER, L. M., CHLEBEK, J. L., DALIA, T. N., DALIA, A. B. & GRAINGER, D. C. 2019. The quorum sensing transcription factor AphA directly regulates natural competence in Vibrio cholerae. *PLOS Genetics,* 15**,** e1008362.

HENSEL, M., HINSLEY, A. P., NIKOLAUS, T., SAWERS, G. & BERKS, B. C. 1999. The genetic basis of tetrathionate respiration in Salmonella typhimurium. *Molecular microbiology,* 32**,** 275-287.

HIDALGO, E., LEAUTAUD, V. & DEMPLE, B. 1998. The redox-regulated SoxR protein acts from a single DNA site as a repressor and an allosteric activator. *EMBO J,* 17**,** 2629-36.

HOISETH, S. K. & STOCKER, B. A. D. 1981. Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature,* 291**,** 238-239.

HOLDEN, E. R. & WEBBER, M. A. 2020. MarA, RamA, and SoxS as Mediators of the Stress Response: Survival at a Cost. *Frontiers in Microbiology,* 11.

HOLMQVIST, E., UNOSON, C., REIMEGÅRD, J. & WAGNER, E. G. 2012. A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Mol Microbiol,* 84**,** 414-27.

HU, S., YU, Y., ZHOU, D., LI, R., XIAO, X. & WU, H. 2018. Global transcriptomic Acid Tolerance Response in Salmonella Enteritidis. *LWT,* 92**,** 330-338.

IGARASHI, K., FUJITA, N. & ISHIHAMA, A. 1991. Identification of a subunit assembly domain in the alpha subunit of Escherichia coli RNA polymerase. *J Mol Biol,* 218**,** 1-6.

IRVING, S. E., CHOUDHURY, N. R. & CORRIGAN, R. M. 2021. The stringent response and physiological roles of (pp)pGpp in bacteria. *Nature Reviews Microbiology,* 19**,** 256-271.

JAIR, K. W., MARTIN, R. G., ROSNER, J. L., FUJITA, N., ISHIHAMA, A. & WOLF, R. E., JR. 1995. Purification and regulatory properties of MarA protein, a transcriptional activator of Escherichia coli multiple antibiotic and superoxide resistance promoters. *Journal of bacteriology,* 177**,** 7100-7104.

JAIR, K. W., YU, X., SKARSTAD, K., THONY, B., FUJITA, N., ISHIHAMA, A. & WOLF, R. E., JR. 1996. Transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the Escherichia coli origin of chromosomal replication. *J Bacteriol,* 178**,** 2507-13.

JEON, Y. H., YAMAZAKI, T., OTOMO, T., ISHIHAMA, A. & KYOGOKU, Y. 1997. Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. *J Mol Biol,* 267**,** 953-62.

JOHNSON, R., BYRNE, A., BERGER, C. N., KLEMM, E., CREPIN, V. F., DOUGAN, G. & FRANKEL, G. 2017. The type III secretion system effector SptP of Salmonella enterica serovar Typhi. *Journal of bacteriology,* 199**,** e00647-16.

JOHNSON, R., MYLONA, E. & FRANKEL, G. 2018. Typhoidal Salmonella: Distinctive virulence factors and pathogenesis. *Cellular Microbiology,* 20**,** e12939.

JOHNSTON, I. 2017. *The Contribution of Nitric Oxide Detoxification and Nitrous Oxide Production to Salmonella Pathogenesis.* PhD thesis, University of East Anglia.

KANIGA, K., URALIL, J., BLISKA, J. B. & GALÁN, J. E. 1996. A secreted protein tyrosine phosphatase with modular effector domains in the bacterial pathogen Salmonella typhimurlum. *Molecular microbiology,* 21**,** 633-641.

KAPLAN, J. B., RAGUNATH, C., VELLIYAGOUNDER, K., FINE, D. H. & RAMASUBBU, N. 2004. Enzymatic detachment of Staphylococcus epidermidis biofilms. *Antimicrobial agents and chemotherapy,* 48**,** 2633-2636.

KENNEY, L. J. 2002. Structure/function relationships in OmpR and other winged-helix transcription factors. *Current Opinion in Microbiology,* 5**,** 135-141.

KETTLES, R. A. 2019. *Regulatory mechanisms of MarA, the activator of multiple antibiotic resistance.* PhD Thesis, The University of Birmingham.

KETTLES, R. A., TSCHOWRI, N., LYONS, K. J., SHARMA, P., HENGGE, R., WEBBER, M. A. & GRAINGER, D. C. 2019. The Escherichia coli MarA protein regulates the ycgZ-ymgABC operon to inhibit biofilm formation. *Molecular microbiology,* 112**,** 1609-1625.

KHAJANCHI, B. K., XU, J., GRIM, C. J., OTTESEN, A. R., RAMACHANDRAN, P. & FOLEY, S. L. 2019. Global transcriptomic analyses of Salmonella enterica in Iron-depleted and Iron-rich growth conditions. *BMC Genomics,* 20**,** 490.

KOLB, A., KOTLARZ, D., KUSANO, S. & ISHIHAMA, A. 1995. Selectivity of the Escherichia coli RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic acids research,* 23**,** 819-826.

KOMBADE, S. & KAUR, N. 2021. Pathogenicity Island in Salmonella.

KOO, B.-M., RHODIUS, V. A., CAMPBELL, E. A. & GROSS, C. A. 2009. Mutational analysis of Escherichia coli σ28 and its target promoters reveals recognition of a composite −10 region, comprised of an 'extended −10' motif and a core −10 element. *Molecular Microbiology,* 72**,** 830-843.

KORNER, H., SOFIA, H. J. & ZUMFT, W. G. 2003. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev,* 27**,** 559-92.

KORZHEVA, N., MUSTAEV, A., KOZLOV, M., MALHOTRA, A., NIKIFOROV, V., GOLDFARB, A. & DARST, S. A. 2000. A structural model of transcription elongation. *Science,* 289**,** 619-25.

KRÖGER, C., COLGAN, A., SRIKUMAR, S., HÄNDLER, K., SIVASANKARAN, SATHESH K., HAMMARLÖF, DISA L., CANALS, R., GRISSOM, JOE E., CONWAY, T., HOKAMP, K. & HINTON, JAY C. D. 2013. An Infection-Relevant Transcriptomic Compendium for Salmonella enterica Serovar Typhimurium. *Cell Host & Microbe,* 14**,** 683-695.

KROGER, C., DILLON, S. C., CAMERON, A. D., PAPENFORT, K., SIVASANKARAN, S. K., HOKAMP, K., CHAO, Y., SITTKA, A., HEBRARD, M., HANDLER, K., COLGAN, A., LEEKITCHAROENPHON, P., LANGRIDGE, G. C., LOHAN, A. J., LOFTUS, B., LUCCHINI, S., USSERY, D. W., DORMAN, C. J., THOMSON, N. R., VOGEL, J. & HINTON, J. C. 2012. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc Natl Acad Sci U S A,* 109**,** E1277-86.

KURTZ, J. R., GOGGINS, J. A. & MCLACHLAN, J. B. 2017. Salmonella infection: Interplay between the bacteria and host immune system. *Immunology letters,* 190**,** 42-50.

KUZNEDELOV, K., KORZHEVA, N., MUSTAEV, A. & SEVERINOV, K. 2002a. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J,* 21**,** 1369-78.

KUZNEDELOV, K., MINAKHIN, L., NIEDZIELA-MAJKA, A., DOVE, S. L., ROGULJA, D., NICKELS, B. E., HOCHSCHILD, A., HEYDUK, T. & SEVERINOV, K. 2002b. A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. *Science,* 295**,** 855-7.

KWON, H. J., BENNIK, M. H. J., DEMPLE, B. & ELLENBERGER, T. 2000. Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. *Nature Structural Biology,* 7**,** 424-430.

LAMBERTE, L. E., BANIULYTE, G., SINGH, S. S., STRINGER, A. M., BONOCORA, R. P., STRACY, M., KAPANIDIS, A. N., WADE, J. T. & GRAINGER, D. C. 2017. Horizontally acquired AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. *Nature Microbiology,* 2**,** 16249.

LANDICK, R. 2001. RNA polymerase clamps down. *Cell,* 105**,** 567-70.

LAPTENKO, O., LEE, J., LOMAKIN, I. & BORUKHOV, S. 2003. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J,* 22**,** 6322-34.

LARSON, M. H., GREENLEAF, W. J., LANDICK, R. & BLOCK, S. M. 2008. Applied force reveals mechanistic and energetic details of transcription termination. *Cell,* 132**,** 971-82.

LATASA, C., ROUX, A., TOLEDO-ARANA, A., GHIGO, J. M., GAMAZO, C., PENADÉS, J. R. & LASA, I. 2005. BapA, a large secreted protein required for biofilm formation and host colonization of Salmonella enterica serovar Enteritidis. *Molecular microbiology,* 58**,** 1322-1339.

LAWRENCE, M., HUBER, W., PAGÈS, H., ABOYOUN, P., CARLSON, M., GENTLEMAN, R., MORGAN, M. T. & CAREY, V. J. 2013. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology,* 9**,** e1003118.

LEDEBOER, N. A., FRYE, J. G., MCCLELLAND, M. & JONES, B. D. 2006. Salmonella enterica serovar Typhimurium requires the Lpf, Pef, and Tafi fimbriae for biofilm formation on HEp-2 tissue culture cells and chicken intestinal epithelium. *Infection and immunity,* 74**,** 3156-3169.

LI, L., DAI, X., WANG, Y., YANG, Y., ZHAO, X., WANG, L. & ZENG, M. 2017. RNA-seq-based analysis of drug-resistant Salmonella enterica serovar Typhimurium selected in vivo and in vitro. *PloS one,* 12**,** e0175234-e0175234.

LI, Z. & DEMPLE, B. 1994. SoxS, an activator of superoxide stress genes in Escherichia coli. Purification and interaction with DNA. *J Biol Chem,* 269**,** 18371-7.

LIM, J. S., SHIN, M., KIM, H.-J., KIM, K. S., CHOY, H. E. & CHO, K. A. 2014. Caveolin-1 mediates Salmonella invasion via the regulation of SopE-dependent Rac1 activation and actin reorganization. *The Journal of infectious diseases,* 210**,** 793-802.

LIN, Z., ZHANG, Y.-G., XIA, Y., XU, X., JIAO, X. & SUN, J. 2016. Salmonella enteritidis effector AvrA stabilizes intestinal tight junctions via the JNK pathway. *Journal of Biological Chemistry,* 291**,** 26837-26849.

LINDBLAD, W. J. 2008. Considerations for determining if a natural product is an effective wound-healing agent. *Int J Low Extrem Wounds,* 7**,** 75-81.

LIU, M. Y., GUI, G., WEI, B., PRESTON, J. F., 3RD, OAKFORD, L., YÜKSEL, U., GIEDROC, D. P. & ROMEO, T. 1997. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. *J Biol Chem,* 272**,** 17502-10.

LÖBER, S., JÄCKEL, D., KAISER, N. & HENSEL, M. 2006. Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. *International Journal of Medical Microbiology,* 296**,** 435-447.

LODGE, J., FEAR, J., BUSBY, S., GUNASEKARAN, P. & KAMINI, N. R. 1992. Broad host range plasmids carrying the Escherichia coli lactose and galactose operons. *FEMS Microbiol Lett,* 74**,** 271-6.

LONETTO, M., GRIBSKOV, M. & GROSS, C. A. 1992. The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol,* 174**,** 3843-9.

LOU, L., ZHANG, P., PIAO, R. & WANG, Y. 2019. Salmonella Pathogenicity Island 1 (SPI-1) and Its Complex Regulatory Network. *Frontiers in Cellular and Infection Microbiology,* 9**,** 270.

MACLENNAN, C. A., MARTIN, L. B. & MICOLI, F. 2014. Vaccines against invasive Salmonella disease: current status and future directions. *Human vaccines & immunotherapeutics,* 10**,** 1478-1493.

MADAN BABU, M. & TEICHMANN, S. A. 2003. Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Research,* 31**,** 1234-1244.

MADDOCKS, S. E. & OYSTON, P. C. F. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology (Reading),* 154**,** 3609-3623.

MAGNUSSON, L. U., FAREWELL, A. & NYSTROM, T. 2005. ppGpp: a global regulator in Escherichia coli. *Trends Microbiol,* 13**,** 236-42.

MANUELA, R., WILSON, R. P., SEBASTIAN, E. W. & ANDREAS, J. B. 2008. Clinical pathogenesis of typhoid fever. *The Journal of Infection in Developing Countries,* 2.

MARTIN, R. G., GILLETTE, W. K., MARTIN, N. I. & ROSNER, J. L. 2002. Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in Escherichia coli. *Mol Microbiol,* 43**,** 355-70.

MARTIN, R. G., GILLETTE, W. K. & ROSNER, J. L. 2000. Promoter discrimination by the related transcriptional activators MarA and SoxS: differential regulation by differential binding. 35**,** 623-634.

MARTIN, R. G., JAIR, K. W., WOLF, R. E., JR. & ROSNER, J. L. 1996. Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in Escherichia coli. *J Bacteriol,* 178**,** 2216-23.

MARTIN, R. G. & ROSNER, J. L. 1995. Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. *Proc Natl Acad Sci U S A,* 92**,** 5456-60.

MARTIN, R. G. & ROSNER, J. L. 2001. The AraC transcriptional activators. *Curr Opin Microbiol,* 4**,** 132-7.

MARTIN, R. G. & ROSNER, J. L. 2002. Genomics of the marA/soxS/rob regulon of Escherichia coli: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol Microbiol,* 44**,** 1611-24.

MARTINEZ-HACKERT, E. & STOCK, A. M. 1997. Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol,* 269**,** 301-12.

MARTINS, M., MCCUSKER, M., AMARAL, L. & FANNING, S. 2011. Mechanisms of antibiotic resistance in Salmonella: efflux pumps, genetics, quorum sensing and biofilm formation. *Letters in Drug Design & Discovery,* 8**,** 114-123.

MCCARTHY, D. J., CHEN, Y. & SMYTH, G. K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research,* 40**,** 4288-4297.

MCGHIE, E. J., HAYWARD, R. D. & KORONAKIS, V. 2004. Control of actin turnover by a salmonella invasion protein. *Molecular cell,* 13**,** 497-510.

MCINTOSH, A., MEIKLE, L. M., ORMSBY, M. J., MCCORMICK, B. A., CHRISTIE, J. M., BREWER, J. M., ROBERTS, M. & WALL, D. M. 2017. SipA activation of caspase-3 is a decisive mediator of host cell survival at early stages of Salmonella enterica serovar Typhimurium infection. *Infection and immunity,* 85**,** e00393-17.

MCLEAN, R. J., WHITELEY, M., STICKLER, D. J. & FUQUA, W. C. 1997. Evidence of autoinducer activity in naturally occurring biofilms. *FEMS microbiology letters,* 154**,** 259-263.

MCMURRY, L. M. & LEVY, S. B. 2010. Evidence that regulatory protein MarA of Escherichia coli represses rob by steric hindrance. *J Bacteriol,* 192**,** 3977-82.

MEDALLA, F., GU, W., FRIEDMAN, C. R., JUDD, M., FOLSTER, J., GRIFFIN, P. M. & HOEKSTRA, R. M. 2021. Increased Incidence of Antimicrobial-Resistant Nontyphoidal Salmonella Infections, United States, 2004-2016. *Emerg Infect Dis,* 27**,** 1662-1672.

MEKLER, V., KORTKHONJIA, E., MUKHOPADHYAY, J., KNIGHT, J., REVYAKIN, A., KAPANIDIS, A. N., NIU, W., EBRIGHT, Y. W., LEVY, R. & EBRIGHT, R. H. 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell,* 108**,** 599-614.

MINAKHIN, L., BHAGAT, S., BRUNNING, A., CAMPBELL, E. A., DARST, S. A., EBRIGHT, R. H. & SEVERINOV, K. 2001a. Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A,* 98**,** 892-7.

MINAKHIN, L., CAMARERO, J. A., HOLFORD, M., PARKER, C., MUIR, T. W. & SEVERINOV, K. 2001b. Mapping the molecular interface between the sigma(70) subunit of E. coli RNA polymerase and T4 AsiA. *J Mol Biol,* 306**,** 631-42.

MITTELMAN, M. W. 1998. Structure and functional characteristics of bacterial biofilms in fluid processing operations. *Journal of dairy science,* 81**,** 2760-2764.

MOONEY, R. A., DARST, S. A. & LANDICK, R. 2005. Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell,* 20**,** 335-45.

MOONEY, R. A., DAVIS, S. E., PETERS, J. M., ROWLAND, J. L., ANSARI, A. Z. & LANDICK, R. 2009. Regulator trafficking on bacterial transcription units in vivo. *Molecular cell,* 33**,** 97-108.

MÜLLER-HILL, B. 2011. *The lac Operon: A Short History of a Genetic Paradigm*, De Gruyter.

MURAKAMI, K. S. & DARST, S. A. 2003. Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology,* 13**,** 31-39.

MURAKAMI, K. S., MASUDA, S., CAMPBELL, E. A., MUZZIN, O. & DARST, S. A. 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science,* 296**,** 1285-90.

MURRAY, C. J. L., IKUTA, K. S., SHARARA, F., SWETSCHINSKI, L., ROBLES AGUILAR, G., GRAY, A., HAN, C., BISIGNANO, C., RAO, P., WOOL, E., JOHNSON, S. C., BROWNE, A. J., CHIPETA, M. G., FELL, F., HACKETT, S., HAINES-WOODHOUSE, G., KASHEF HAMADANI, B. H.,

KUMARAN, E. A. P., MCMANIGAL, B., AGARWAL, R., AKECH, S., ALBERTSON, S., AMUASI, J., ANDREWS, J., ARAVKIN, A., ASHLEY, E., BAILEY, F., BAKER, S., BASNYAT, B., BEKKER, A., BENDER, R., BETHOU, A., BIELICKI, J., BOONKASIDECHA, S., BUKOSIA, J., CARVALHEIRO, C., CASTAÑEDA-ORJUELA, C., CHANSAMOUTH, V., CHAURASIA, S., CHIURCHIÙ, S., CHOWDHURY, F., COOK, A. J., COOPER, B., CRESSEY, T. R., CRIOLLO-MORA, E., CUNNINGHAM, M., DARBOE, S., DAY, N. P. J., DE LUCA, M., DOKOVA, K., DRAMOWSKI, A., DUNACHIE, S. J., ECKMANNS, T., EIBACH, D., EMAMI, A., FEASEY, N., FISHER-PEARSON, N., FORREST, K., GARRETT, D., GASTMEIER, P., GIREF, A. Z., GREER, R. C., GUPTA, V., HALLER, S., HASELBECK, A., HAY, S. I., HOLM, M., HOPKINS, S., IREGBU, K. C., JACOBS, J., JAROVSKY, D., JAVANMARDI, F., KHORANA, M., KISSOON, N., KOBEISSI, E., KOSTYANEV, T., KRAPP, F., KRUMKAMP, R., KUMAR, A., KYU, H. H., LIM, C., LIMMATHUROTSAKUL, D., LOFTUS, M. J., LUNN, M., MA, J., MTURI, N., MUNERA-HUERTAS, T., MUSICHA, P., MUSSI-PINHATA, M. M., NAKAMURA, T., NANAVATI, R., NANGIA, S., NEWTON, P., NGOUN, C., NOVOTNEY, A., NWAKANMA, D., OBIERO, C. W., OLIVAS-MARTINEZ, A., OLLIARO, P., OOKO, E., et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*.

NIKAIDO, E., YAMAGUCHI, A. & NISHINO, K. 2008. AcrAB multidrug efflux pump regulation in Salmonella enterica serovar Typhimurium by RamA in response to environmental signals. *J Biol Chem,* 283**,** 24245-53.

O'TOOLE, G., KAPLAN, H. B. & KOLTER, R. 2000. Biofilm formation as microbial development. *Annual Reviews in Microbiology,* 54**,** 49-79.

O'TOOLE, G. A. & KOLTER, R. 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Molecular microbiology,* 30**,** 295-304.

O'NEILL, J. 2016. Tackling drug-resistant infections globally: final report and recommendations. *The review on antimicrobial resistance.*

OCHSNER, U. A., VASIL, M. L., ALSABBAGH, E., PARVATIYAR, K. & HASSETT, D. J. 2000. Role of the Pseudomonas aeruginosa oxyR-recG operon in oxidative stress defense and DNA repair: OxyR-dependent regulation of katB-ankB, ahpB, and ahpC-ahpF. *Journal of bacteriology,* 182**,** 4533-4544.

OGASAWARA, H., HASEGAWA, A., KANDA, E., MIKI, T., YAMAMOTO, K. & ISHIHAMA, A. 2007. Genomic SELEX search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *J Bacteriol,* 189**,** 4791-9.

OGASAWARA, H., ISHIZUKA, T., HOTTA, S., AOKI, M., SHIMADA, T. & ISHIHAMA, A. 2020. Novel regulators of the csgD gene encoding the master regulator of biofilm formation in Escherichia coli K-12. *Microbiology,* 166**,** 880-890.

OGASAWARA, H., ISHIZUKA, T., YAMAJI, K., KATO, Y., SHIMADA, T. & ISHIHAMA, A. 2019. Regulatory role of pyruvate-sensing BtsSR in biofilm formation by Escherichia coli K-12. *FEMS Microbiology Letters,* 366.

OGASAWARA, H., SHINOHARA, S., YAMAMOTO, K. & ISHIHAMA, A. 2012. Novel regulation targets of the metal-response BasS–BasR two-component system of Escherichia coli. *Microbiology,* 158**,** 1482-1492.

OGASAWARA, H., YAMADA, K., KORI, A., YAMAMOTO, K. & ISHIHAMA, A. 2010a. Regulation of the Escherichia coli csgD promoter: interplay between five transcription factors. *Microbiology,* 156**,** 2470-2483.

OGASAWARA, H., YAMAMOTO, K. & ISHIHAMA, A. 2010b. Regulatory role of MlrA in transcription activation of csgD, the master regulator of biofilm formation in Escherichia coli. *FEMS Microbiology Letters,* 312**,** 160-168.

OGASAWARA, H., YAMAMOTO, K. & ISHIHAMA, A. 2011. Role of the Biofilm Master Regulator CsgD in Cross-Regulation between Biofilm Formation and Flagellar Synthesis. *Journal of Bacteriology,* 193**,** 2587-2597.

ÖSTERBERG, S., PESO-SANTOS, T. D. & SHINGLER, V. 2011. Regulation of Alternative Sigma Factor Use. *Annual Review of Microbiology,* 65**,** 37-55.

OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics,* 12**,** 87-98.

P. STOODLEY, K. SAUER, D. G. DAVIES & COSTERTON, J. W. 2002. Biofilms as Complex Differentiated Communities. *Annual Review of Microbiology,* 56**,** 187-209.

PAGET, M. S. & HELMANN, J. D. 2003. The σ70family of sigma factors. *Genome Biology,* 4**,** 203.

PARK, D., LARA-TEJERO, M., WAXHAM, M. N., LI, W., HU, B., GALÁN, J. E. & LIU, J. 2018. Visualization of the type III secretion mediated Salmonella–host cell interface using cryo-electron tomography. *eLife,* 7**,** e39514.

PARK, J. S. & ROBERTS, J. W. 2006. Role of DNA bubble rewinding in enzymatic transcription termination. *Proc Natl Acad Sci U S A,* 103**,** 4870-5.

PEDERSEN, A. G., JENSEN, L. J., BRUNAK, S., STÆRFELDT, H.-H. & USSERY, D. W. 2000. A DNA structural atlas for Escherichia coli. *Journal of molecular biology,* 299**,** 907-930.

PÉREZ-RUEDA, E. & COLLADO-VIDES, J. 2000. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic acids research,* 28**,** 1838-1847.

PHE, P. H. E. 2018. *Enteric fever (typhoid and paratyphoid) England, Wales and Northern Ireland: 2017* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/761348/Enteric_fever_annual_report_2017.pdf [Accessed 23.11.2021].

PHILIPS, S. J., CANALIZO-HERNANDEZ, M., YILDIRIM, I., SCHATZ, G. C., MONDRAGÓN, A. & O'HALLORAN, T. V. 2015. TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science (New York, N.Y.),* 349**,** 877-881.

PIDDOCK, L. J. V. 2006. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clinical microbiology reviews,* 19**,** 382-402.

PLUMBRIDGE, J. 2002. Regulation of gene expression in the PTS in Escherichia coli: the role and interactions of Mlc. *Current Opinion in Microbiology,* 5**,** 187-193.

POMPOSIELLO, P. J. & DEMPLE, B. 2000. Identification of SoxS-regulated genes in Salmonella enterica serovar typhimurium. *Journal of bacteriology,* 182**,** 23-29.

PRIBNOW, D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proceedings of the National Academy of Sciences,* 72**,** 784-788.

PRIGENT-COMBARET, C., BROMBACHER, E., VIDAL, O., AMBERT, A., LEJEUNE, P., LANDINI, P. & DOREL, C. 2001. Complex regulatory network controls initial adhesion and biofilm formation in Escherichia coli via regulation of the csgD gene. *Journal of bacteriology,* 183**,** 7213-7223.

PROUTY, A., SCHWESINGER, W. & GUNN, J. 2002. Biofilm formation and interaction with the surfaces of gallstones by Salmonella spp. *Infection and immunity,* 70**,** 2640-2649.

PU, Y., ZHAO, Z., LI, Y., ZOU, J., MA, Q., ZHAO, Y., KE, Y., ZHU, Y., CHEN, H., BAKER, MATTHEW A. B., GE, H., SUN, Y., XIE, XIAOLIANG S. & BAI, F. 2016. Enhanced Efflux Activity Facilitates Drug Tolerance in Dormant Bacterial Cells. *Molecular Cell,* 62**,** 284-294.

QADRI, F., KHANAM, F., LIU, X., THEISS-NYLAND, K., BISWAS, P. K., BHUIYAN, A. I., AHMMED, F., COLIN-JONES, R., SMITH, N., TONKS, S., VOYSEY, M., MUJADIDI, Y. F., MAZUR, O., RAJIB, N. H., HOSSEN, M. I., AHMED, S. U., KHAN, A., RAHMAN, N., BABU, G., GREENLAND, M., KELLY, S., IREEN, M., ISLAM, K., O'REILLY, P., SCHERRER, K. S., PITZER, V. E., NEUZIL, K. M., ZAMAN, K., POLLARD, A. J. & CLEMENS, J. D. 2021. Protection by vaccination of children against typhoid fever with a Vi-tetanus toxoid conjugate vaccine in urban Bangladesh: a cluster-randomised trial. *The Lancet,* 398**,** 675-684.

RAMIREZ, M. S. & TOLMASKY, M. E. 2010. Aminoglycoside modifying enzymes. *Drug Resistance Updates,* 13**,** 151-171.

RAVCHEEV, D. A., KHOROSHKIN, M. S., LAIKOVA, O. N., TSOY, O. V., SERNOVA, N. V., PETROVA, S. A., RAKHMANINOVA, A. B., NOVICHKOV, P. S., GELFAND, M. S. & RODIONOV, D. A. 2014. Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Frontiers in Microbiology,* 5.

RAY-SONI, A., BELLECOURT, M. J. & LANDICK, R. 2016. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual Review of Biochemistry,* 85**,** 319-347.

REEVES, M. W., EVINS, G. M., HEIBA, A. A., PLIKAYTIS, B. D. & FARMER, J. J., 3RD 1989. Clonal nature of Salmonella typhi and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of Salmonella bongori comb. nov. *Journal of clinical microbiology,* 27**,** 313-320.

REVYAKIN, A., LIU, C., EBRIGHT, R. H. & STRICK, T. R. 2006. Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science,* 314**,** 1139-1143.

RICCI, V., TZAKAS, P., BUCKLEY, A. & PIDDOCK, L. J. 2006. Ciprofloxacin-resistant Salmonella enterica serovar Typhimurium strains are difficult to select in the absence of AcrB and TolC. *Antimicrob Agents Chemother,* 50**,** 38-42.

RICHARDSON, J. P. 2002. Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta,* 1577**,** 251-260.

ROBERTS, J. & PARK, J. S. 2004. Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr Opin Microbiol,* 7**,** 120-5.

ROBERTS, J. W. 2019. Mechanisms of Bacterial Transcription Termination. *J Mol Biol*.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-140.

ROBISON, K., MCGUIRE, A. M. & CHURCH, G. M. 1998. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome. *J Mol Biol,* 284**,** 241-54.

RÖMLING, U., BIAN, Z., HAMMAR, M., SIERRALTA, W. D. & NORMARK, S. 1998. Curli fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation. *Journal of bacteriology,* 180**,** 722-731.

RÖMLING, U., PESEN, D. & YARON, S. 2007. Biofilms of Salmonella enterica. *Salmonella: Molecular Biology and Pathogenesis. Horizon Bioscience, UK*.

RÖMLING, U., SIERRALTA, W. D., ERIKSSON, K. & NORMARK, S. 1998. Multicellular and aggregative behaviour of Salmonella typhimurium strains is controlled by mutations in the agfD promoter. *Molecular microbiology,* 28**,** 249-264.

ROSS, W., VRENTAS, C. E., SANCHEZ-VAZQUEZ, P., GAAL, T. & GOURSE, R. L. 2013. The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. *Molecular cell,* 50**,** 420-429.

ROWLEY, G., HENSEN, D., FELGATE, H., ARKENBERG, A., APPIA-AYME, C., PRIOR, K., HARRINGTON, C., FIELD, S. J., BUTT, J. N., BAGGS, E. & RICHARDSON, D. J. 2012. Resolving the contributions of the membrane-bound and periplasmic nitrate reductase systems to nitric oxide and nitrous oxide production in Salmonella enterica serovar Typhimurium. *Biochem J,* 441**,** 755-62.

SAKATA-SOGAWA, K. & SHIMAMOTO, N. 2004. RNA polymerase can track a DNA groove during promoter search. *Proc Natl Acad Sci U S A,* 101**,** 14731-5.

SAUER, K., CAMPER, A. K., EHRLICH, G. D., COSTERTON, J. W. & DAVIES, D. G. 2002. Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. Am Soc Microbiol.

SCALLAN, E., HOEKSTRA, R. M., ANGULO, F. J., TAUXE, R. V., WIDDOWSON, M.-A., ROY, S. L., JONES, J. L. & GRIFFIN, P. M. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerging infectious diseases,* 17**,** 7.

SCHELL, M. A. 1993. Molecular biology of the LysR family of transcriptional regulators. *Annual review of microbiology,* 47**,** 597-627.

SCHIKORA, A., GARCIA, A. V. & HIRT, H. 2012. Plants as alternative hosts for Salmonella. *Trends in Plant Science,* 17**,** 245-249.

SCHMIDT, M. C. & CHAMBERLIN, M. J. 1987. nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *J Mol Biol,* 195**,** 809-18.

SEO, S. W., KIM, D., SZUBIN, R. & PALSSON, B. O. 2015. Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in Escherichia coli K-12 MG1655. *Cell Rep,* 12**,** 1289-99.

SERENO, M., ZIECH, R., DRUZIANI, J., PEREIRA, J. & BERSOT, L. 2017. Antimicrobial susceptibility and biofilm production by Salmonella sp. strains isolated from frozen poultry carcasses. *Brazilian Journal of Poultry Science,* 19**,** 103-108.

SHAH, D. H. 2014. RNA sequencing reveals differences between the global transcriptomes of Salmonella enterica serovar enteritidis strains with high and low pathogenicities. *Applied and environmental microbiology,* 80**,** 896-906.

SHARMA, A. K., DHASMANA, N., DUBEY, N., KUMAR, N., GANGWAL, A., GUPTA, M. & SINGH, Y. 2017a. Bacterial Virulence Factors: Secreted for Survival. *Indian journal of microbiology,* 57**,** 1-10.

SHARMA, P., HAYCOCKS, J. R. J., MIDDLEMISS, A. D., KETTLES, R. A., SELLARS, L. E., RICCI, V., PIDDOCK, L. J. V. & GRAINGER, D. C. 2017b. The multiple antibiotic resistance operon of enteric bacteria controls DNA repair and outer membrane integrity. *Nat Commun,* 8**,** 1444.

SHIMADA, T., KATAYAMA, Y., KAWAKITA, S., OGASAWARA, H., NAKANO, M., YAMAMOTO, K. & ISHIHAMA, A. 2012. A novel regulator RcdA of the csgD gene encoding the master regulator of biofilm formation in Escherichia coli. *MicrobiologyOpen,* 1**,** 381-394.

SILBERGELD, E. K., GRAHAM, J. & PRICE, L. B. 2008. Industrial food animal production, antimicrobial resistance, and human health. *Annu Rev Public Health,* 29**,** 151-69.

SIMON, R., PRIEFER, U. & PÜHLER, A. 1983. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology,* 1**,** 784-791.

SINGH, S. S. & GRAINGER, D. C. 2013. H-NS Can Facilitate Specific DNA-binding by RNA Polymerase in AT-rich Gene Regulatory Regions. *PLOS Genetics,* 9**,** e1003589.

SINGH, S. S., SINGH, N., BONOCORA, R. P., FITZGERALD, D. M., WADE, J. T. & GRAINGER, D. C. 2014. Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev,* 28**,** 214-9.

SKARSTAD, K., THÖNY, B., HWANG, D. S. & KORNBERG, A. 1993. A novel binding protein of the origin of the Escherichia coli chromosome. 268**,** 5365-70.

SMITS, W. K., HOA, T. T., HAMOEN, L. W., KUIPERS, O. P. & DUBNAU, D. 2007. Antirepression as a second mechanism of transcriptional activation by a minor groove binding protein. *Molecular microbiology,* 64**,** 368-381.

SOBEL, M. L., NESHAT, S. & POOLE, K. 2005. Mutations in PA2491 (mexS) promote MexT-dependent mexEF-oprN expression and multidrug resistance in a clinical strain of Pseudomonas aeruginosa. *Journal of bacteriology,* 187**,** 1246-1253.

SOISSON, S. M., MACDOUGALL-SHACKLETON, B., SCHLEIF, R. & WOLBERGER, C. 1997. Structural basis for ligand-regulated oligomerization of AraC. *Science,* 276**,** 421-5.

SOLANO, C., GARCÍA, B., VALLE, J., BERASAIN, C., GHIGO, J. M., GAMAZO, C. & LASA, I. 2002. Genetic analysis of Salmonella enteritidis biofilm formation: critical role of cellulose. *Molecular microbiology,* 43**,** 793-808.

SOO, V. W. C. & WOOD, T. K. 2013. Antitoxin MqsA Represses Curli Formation Through the Master Biofilm Regulator CsgD. *Scientific Reports,* 3**,** 3186.

SREY, S., JAHID, I. K. & HA, S.-D. 2013. Biofilm formation in food industries: A food safety concern. *Food Control,* 31**,** 572-585.

SRIKANTH, C., WALL, D. M., MALDONADO-CONTRERAS, A., SHI, H. N., ZHOU, D., DEMMA, Z., MUMY, K. L. & MCCORMICK, B. A. 2010. Salmonella pathogenesis and processing of secreted effectors by caspase-3. *Science,* 330**,** 390-393.

STANAWAY, J. D., PARISI, A., SARKAR, K., BLACKER, B. F., REINER, R. C., HAY, S. I., NIXON, M. R., DOLECEK, C., JAMES, S. L., MOKDAD, A. H., ABEBE, G., AHMADIAN, E., ALAHDAB, F., ALEMNEW, B. T. T., ALIPOUR, V., ALLAH BAKESHEI, F., ANIMUT, M. D., ANSARI, F., ARABLOO, J., ASFAW, E. T., BAGHERZADEH, M., BASSAT, Q., BELAYNEH, Y. M. M., CARVALHO, F., DARYANI, A., DEMEKE, F. M., DEMIS, A. B. B., DUBEY, M., DUKEN, E. E., DUNACHIE, S. J., EFTEKHARI, A., FERNANDES, E., FOULADI FARD, R., GEDEFAW, G. A., GETA, B., GIBNEY, K. B., HASANZADEH, A., HOANG, C. L., KASAEIAN, A., KHATER, A., KIDANEMARIAM, Z. T., LAKEW, A. M., MALEKZADEH, R., MELESE, A., MENGISTU, D. T., MESTROVIC, T., MIAZGOWSKI, B., MOHAMMAD, K. A., MOHAMMADIAN, M., MOHAMMADIAN-HAFSHEJANI, A., NGUYEN, C. T., NGUYEN, L. H., NGUYEN, S. H., NIRAYO, Y. L., OLAGUNJU, A. T., OLAGUNJU, T. O., POURJAFAR, H., QORBANI, M., RABIEE, M., RABIEE, N., RAFAY, A., REZAPOUR, A., SAMY, A. M., SEPANLOU, S. G., SHAIKH, M. A., SHARIF, M., SHIGEMATSU, M., TESSEMA, B., TRAN, B. X., ULLAH, I., YIMER, E. M., ZAIDI, Z., MURRAY, C. J. L. & CRUMP, J. A. 2019a. The global burden of

non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Infectious Diseases,* 19**,** 1312-1324.

STANAWAY, J. D., REINER, R. C., BLACKER, B. F., GOLDBERG, E. M., KHALIL, I. A., TROEGER, C. E., ANDREWS, J. R., BHUTTA, Z. A., CRUMP, J. A., IM, J., MARKS, F., MINTZ, E., PARK, S. E., ZAIDI, A. K. M., ABEBE, Z., ABEJIE, A. N., ADEDEJI, I. A., ALI, B. A., AMARE, A. T., ATALAY, H. T., AVOKPAHO, E. F. G. A., BACHA, U., BARAC, A., BEDI, N., BERHANE, A., BROWNE, A. J., CHIRINOS, J. L., CHITHEER, A., DOLECEK, C., EL SAYED ZAKI, M., ESHRATI, B., FOREMAN, K. J., GEMECHU, A., GUPTA, R., HAILU, G. B., HENOK, A., HIBSTU, D. T., HOANG, C. L., ILESANMI, O. S., IYER, V. J., KAHSAY, A., KASAEIAN, A., KASSA, T. D., KHAN, E. A., KHANG, Y.-H., MAGDY ABD EL RAZEK, H., MELKU, M., MENGISTU, D. T., MOHAMMAD, K. A., MOHAMMED, S., MOKDAD, A. H., NACHEGA, J. B., NAHEED, A., NGUYEN, C. T., NGUYEN, H. L. T., NGUYEN, L. H., NGUYEN, N. B., NGUYEN, T. H., NIRAYO, Y. L., PANGESTU, T., PATTON, G. C., QORBANI, M., RAI, R. K., RANA, S. M., RANABHAT, C. L., ROBA, K. T., ROBERTS, N. L. S., RUBINO, S., SAFIRI, S., SARTORIUS, B., SAWHNEY, M., SHIFERAW, M. S., SMITH, D. L., SYKES, B. L., TRAN, B. X., TRAN, T. T., UKWAJA, K. N., VU, G. T., VU, L. G., WELDEGEBREAL, F., YENIT, M. K., MURRAY, C. J. L. & HAY, S. I. 2019b. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Infectious Diseases,* 19**,** 369-381.

STEENACKERS, H., HERMANS, K., VANDERLEYDEN, J. & DE KEERSMAECKER, S. C. 2012. Salmonella biofilms: an overview on occurrence, structure, regulation and eradication. *Food Research International,* 45**,** 502-531.

STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000a. Two-component signal transduction. *Annu Rev Biochem,* 69**,** 183-215.

STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000b. Two-component signal transduction. *Annual review of biochemistry,* 69**,** 183-215.

STOODLEY, P., CARGO, R., RUPP, C. J., WILSON, S. & KLAPPER, I. 2002. Biofilm material properties as related to shear-induced deformation and detachment phenomena. *Journal of Industrial Microbiology and Biotechnology,* 29**,** 361-367.

SU, L. H. & CHIU, C. H. 2007. Salmonella: clinical importance and evolution of nomenclature. *Chang Gung Med J,* 30**,** 210-9.

SUTHERLAND, C. & MURAKAMI, K. S. 2018. An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. *EcoSal Plus,* 8**,** 10.1128/ecosalplus.ESP-0004-2018.

TANNER, J. R. & KINGSLEY, R. A. 2018. Evolution of Salmonella within Hosts. *Trends in Microbiology,* 26**,** 986-998.

TAYLOR, S. J. & WINTER, S. E. 2020. Salmonella finds a way: Metabolic versatility of Salmonella enterica serovar Typhimurium in diverse host environments. *PLOS Pathogens,* 16**,** e1008540.

THOTA, S. S. & CHUBIZ, L. M. 2019. Multidrug Resistance Regulators MarA, SoxS, Rob, and RamA Repress Flagellar Gene Expression and Motility in Salmonella enterica Serovar Typhimurium. *J Bacteriol,* 201.

TORRES, M. J., SIMON, J., ROWLEY, G., BEDMAR, E. J., RICHARDSON, D. J., GATES, A. J. & DELGADO, M. J. 2016. Chapter Seven - Nitrous Oxide Metabolism in Nitrate-Reducing

Bacteria: Physiology and Regulatory Mechanisms. *In:* POOLE, R. K. (ed.) *Advances in Microbial Physiology.* Academic Press.

TOULOKHONOV, I. & LANDICK, R. 2003. The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Mol Cell,* 12**,** 1125-36.

TOULOKHONOV, I. & LANDICK, R. 2006. The role of the lid element in transcription by E. coli RNA polymerase. *J Mol Biol,* 361**,** 644-58.

TSCHOWRI, N., LINDENBERG, S. & HENGGE, R. 2012. Molecular function and potential evolution of the biofilm-modulating blue light-signalling pathway of Escherichia coli. 85**,** 893-906.

TURSI, S. A., PULIGEDDA, R. D., SZABO, P., NICASTRO, L. K., MILLER, A. L., QIU, C., GALLUCCI, S., RELKIN, N. R., BUTTARO, B. A., DESSAIN, S. K. & TÜKEL, Ç. 2020. Salmonella Typhimurium biofilm disruption by a human antibody that binds a pan-amyloid epitope on curli. *Nature Communications,* 11**,** 1007.

UKHSA, U. H. S. A. 2021. *Non-typhoidal Salmonella data 2010 to 2019* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1026208/salmonella-annual-report-2019.pdf [Accessed 23.11.21].

VASSYLYEV, D. G., VASSYLYEVA, M. N., ZHANG, J., PALANGAT, M., ARTSIMOVITCH, I. & LANDICK, R. 2007. Structural basis for substrate loading in bacterial RNA polymerase. *Nature,* 448**,** 163-8.

VERDEROSA, A. D., TOTSIKA, M. & FAIRFULL-SMITH, K. E. 2019. Bacterial Biofilm Eradication Agents: A Current Review. *Frontiers in Chemistry,* 7.

VISHWAKARMA, V. 2020. Impact of environmental biofilms: Industrial components and its remediation. *Journal of Basic Microbiology,* 60**,** 198-206.

VONAESCH, P., SELLIN, M. E., CARDINI, S., SINGH, V., BARTHEL, M. & HARDT, W. D. 2014. The S almonella T yphimurium effector protein SopE transiently localizes to the early SCV and contributes to intracellular replication. *Cellular microbiology,* 16**,** 1723-1735.

WANG, D., BUSHNELL, D. A., WESTOVER, K. D., KAPLAN, C. D. & KORNBERG, R. D. 2006. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell,* 127**,** 941-54.

WARMAN, E. A., FORREST, D., GUEST, T., HAYCOCKS, J. J. R. J., WADE, J. T. & GRAINGER, D. C. 2021. Widespread divergent transcription from bacterial and archaeal promoters is a consequence of DNA-sequence symmetry. *Nature Microbiology,* 6**,** 746-756.

WATANABE, S., KITA, A., KOBAYASHI, K. & MIKI, K. 2008. Crystal structure of the [2Fe-2S] oxidative-stress sensor SoxR bound to DNA. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 4121-4126.

WEATHERSPOON-GRIFFIN, N., YANG, D., KONG, W., HUA, Z. & SHI, Y. 2014. The CpxR/CpxA Two-component Regulatory System Up-regulates the Multidrug Resistance Cascade to Facilitate Escherichia coli Resistance to a Model Antimicrobial Peptide *. *Journal of Biological Chemistry,* 289**,** 32571-32582.

WEBBER, M. A., BAILEY, A. M., BLAIR, J. M., MORGAN, E., STEVENS, M. P., HINTON, J. C., IVENS, A., WAIN, J. & PIDDOCK, L. J. 2009. The global consequence of disruption of the AcrAB-TolC efflux pump in Salmonella enterica includes reduced expression of SPI-1 and other attributes required to infect the host. *Journal of bacteriology,* 191**,** 4276-4285.

WEBBER, M. A. & PIDDOCK, L. J. 2001. Absence of mutations in marRAB or soxRS in acrB-overexpressing fluoroquinolone-resistant clinical and veterinary isolates of Escherichia coli. *Antimicrob Agents Chemother,* 45**,** 1550-2.

WESTON, N., SHARMA, P., RICCI, V. & PIDDOCK, L. J. V. 2018. Regulation of the AcrAB-TolC efflux pump in Enterobacteriaceae. *Research in Microbiology,* 169**,** 425-431.

WHITE, A., GIBSON, D., COLLINSON, S., BANSER, P. & KAY, W. 2003. Extracellular polysaccharides associated with thin aggregative fimbriae of Salmonella enterica serovar Enteritidis. *Journal of bacteriology,* 185**,** 5398-5407.

WHITE, A. P., GIBSON, D. L., GRASSL, G. A., KAY, W. W., FINLAY, B. B., VALLANCE, B. A. & SURETTE, M. G. 2008. Aggregation via the red, dry, and rough morphotype is not a virulence adaptation in Salmonella enterica serovar Typhimurium. *Infection and immunity,* 76**,** 1048-1058.

WHITE, A. P., GIBSON, D. L., KIM, W., KAY, W. W. & SURETTE, M. G. 2006. Thin aggregative fimbriae and cellulose enhance long-term survival and persistence of Salmonella. *Journal of bacteriology,* 188**,** 3219-3227.

WHITTLE, E. E., MCNEIL, H. E., TRAMPARI, E., WEBBER, M., OVERTON, T. W. & BLAIR, J. M. A. 2021. Efflux Impacts Intracellular Accumulation Only in Actively Growing Bacterial Cells. *mBio,* 12**,** e0260821-e0260821.

WHO, T. W. H. O. 2020. *Lack of new antibiotics threatens global efforts to contain drug-resistant infections* [Online]. Available: https://www.who.int/news/item/17-01-2020-lack-of-new-antibiotics-threatens-global-efforts-to-contain-drug-resistant-infections [Accessed 07/02/2022].

WHO, W. H. O. 2017. *WHO publishes list of bacteria for which new antibiotics are urgently needed* [Online]. Available: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed [Accessed 24.11.21].

WICKHAM, H. 2016. *ggplot2: Elegant Graphics for Data Anlysis*, Springer-Verlag New York.

WINTER, S. E., THIENNIMITR, P., WINTER, M. G., BUTLER, B. P., HUSEBY, D. L., CRAWFORD, R. W., RUSSELL, J. M., BEVINS, C. L., ADAMS, L. G., TSOLIS, R. M., ROTH, J. R. & BÄUMLER, A. J. 2010. Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature,* 467**,** 426-429.

WOOD, M. W., JONES, M. A., WATSON, P. R., SIBER, A. M., MCCORMICK, B. A., HEDGES, S., ROSQVIST, R., WALLIS, T. S. & GALYOV, E. E. 2000. The secreted effector protein of Salmonella dublin, SopA, is translocated into eukaryotic cells and influences the induction of enteritis. *Cellular microbiology,* 2**,** 293-303.

WOOD, T. K., KNABEL, S. J. & KWAN, B. W. 2013. Bacterial Persister Cell Formation and Dormancy. *Applied and Environmental Microbiology,* 79**,** 7116-7121.

WOSTEN, M. M. 1998. Eubacterial sigma-factors. *FEMS Microbiol Rev,* 22**,** 127-50.

WRAY, C. & SOJKA, W. J. 1978. Experimental Salmonella typhimurium infection in calves. *Res Vet Sci,* 25**,** 139-43.

WU, H., JONES, R. M. & NEISH, A. S. 2012. The Salmonella effector AvrA mediates bacterial intracellular survival during infection in vivo. *Cellular microbiology,* 14**,** 28-39.

WU, J. & WEISS, B. 1992. Two-stage induction of the soxRS (superoxide response) regulon of Escherichia coli. *J Bacteriol,* 174**,** 3915-20.

YAKHNIN, A. V., MURAKAMI, K. S. & BABITZKE, P. 2016. NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble. *J Biol Chem,* 291**,** 5299-308.

YAMASAKI, S., NIKAIDO, E., NAKASHIMA, R., SAKURAI, K., FUJIWARA, D., FUJII, I. & NISHINO, K. 2013. The crystal structure of multidrug-resistance regulator RamR with multiple drugs. *Nature Communications,* 4**,** 2078.

YANG, X., ZHANG, Z., HUANG, Z., ZHANG, X., LI, D., SUN, L., YOU, J., PAN, X. & YANG, H. 2019. A putative LysR-type transcriptional regulator inhibits biofilm synthesis in Pseudomonas aeruginosa. *Biofouling,* 35**,** 541-550.

YANG, Y., DARBARI, V. C., ZHANG, N., LU, D., GLYDE, R., WANG, Y.-P., WINKELMAN, J. T., GOURSE, R. L., MURAKAMI, K. S., BUCK, M. & ZHANG, X. 2015. TRANSCRIPTION. Structures of the RNA polymerase-σ54 reveal new and conserved regulatory strategies. *Science (New York, N.Y.),* 349**,** 882-885.

YAU, S., LIU, X., DJORDJEVIC, S. P. & HALL, R. M. 2010. RSF1010-like plasmids in Australian Salmonella enterica serovar Typhimurium and origin of their sul2-strA-strB antibiotic resistance gene cluster. *Microb Drug Resist,* 16**,** 249-52.

YE, Z., PETROF, E. O., BOONE, D., CLAUD, E. C. & SUN, J. 2007. Salmonella effector AvrA regulation of colonic epithelial cell inflammation by deubiquitination. *The American journal of pathology,* 171**,** 882-892.

ZAKIKHANY, K., HARRINGTON, C. R., NIMTZ, M., HINTON, J. C. & RÖMLING, U. 2010. Unphosphorylated CsgD controls biofilm formation in Salmonella enterica serovar Typhimurium. *Molecular microbiology,* 77**,** 771-786.

ZENG, L. R. & XIE, J. P. 2011. Molecular basis underlying LuxR family transcription factors and function diversity and implications for novel antibiotic drug targets. *J Cell Biochem,* 112**,** 3079-84.

ZHANG, G., CAMPBELL, E. A., MINAKHIN, L., RICHTER, C., SEVERINOV, K. & DARST, S. A. 1999. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell,* 98**,** 811-24.

ZHOU, D., MOOSEKER, M. S. & GALÁN, J. E. 1999. An invasion-associated Salmonella protein modulates the actin-bundling activity of plastin. *Proceedings of the National Academy of Sciences,* 96**,** 10176-10181.

7. Appendix

7.1 Python Code

7.1.1 RNA_seq_analysis_multiple_samples_final.py

```
# Made by Alistair in Python 3.9

# The first thing to check is that the .gtf file has 9 columns, with column 4 being
the start base of TSS and column 7 being the strand info (+ or -) AND that the
first line of the file is NOT data but is some form of words or other.  The code is
programmed to skip the first line of the .gtf file so you may miss your first TSS
if you don't check this.  You can open a .gtf file with a txt editor (you can even
open them in your favourite bioinformatics program Dave, excel!).  If the data is
on the first line, just add a line and write anything there.  In fact, this program
should run on any file that is tab separated (like a csv or something) as long as
the 4th and 7th columns are as described here.

import pandas as pd
import numpy as np
import os
import Bio
from Bio import SeqIO
import time

if __name__ == '__main__':
    def rnaseq_coordinates(file):
    # Import .gtf file and convert to dataframe.
        dataframe = pd.read_csv(file, sep = "\t", index_col=False, engine='python',
header=None, skiprows=1)

        print(".gtf file is:")
        print(dataframe)
        print(".GTF file has been printed\n")

        # Extract only base and strand info and rename columns to 'base' and
'strand'
        workingdf = pd.read_csv(filename, sep = "\t", index_col=False, usecols=
[3,6], engine='python', header=None, skiprows=1, names=["base", "strand"])

        print("Strand info, top is + strand and bottom is - strand")
        print(workingdf, "\n", workingdf.shape)

        # Iterate over dataframe and minus the desired bases upstream from base
value and add the desired bases downstream if strand is "+" and minus the desired
bases downstream and plus the desired bases upstream if strand is "-".  This gives
the ability to alter the amount of bases upstream and downstream of each TSS if the
original amount is not sufficient for subsequent analysis.
```

```
        # Also scans the genome fasta file and extracts the coordinates defined by
the upstream and downstream variables to generate a fasta file using the
coordinates as the name of each entry
        # Also reverse complements the sequences on the - strand

        genome_coordinates = []
        upstream = 1
        downstream = 1
        coordinates = []
        start_coord = 1
        end_coord = 1
        extracted_sequence = []



        for label, row in workingdf.iterrows():
            if row["strand"] == "+":
                print("\nin", filename, "TSS", row["base"],"is on the top strand,
therefore:")
                upstream = row["base"] - int(bases_upstream)
                downstream = row["base"] + int(bases_downstream)
                print("100 bases upstream is:\n", upstream, "\n50 bases downstream
is:\n", downstream, "\n")
                coordinates = str(upstream)+'-'+str(downstream)
                genome_coordinates.append(coordinates)
                print("extracting sequence from positions", upstream, "-",
downstream, "...")
                for seq in SeqIO.parse(reference_genome, "fasta"):
                    start_coord = upstream
                    end_coord = downstream
                    sequence = seq.seq[start_coord - 1:end_coord]
                    title = ">" + str(start_coord) + "_" + str(end_coord)
                    extracted_sequence.append(title)
                    extracted_sequence.append(sequence)
                    print("extracted sequence is: ", str(sequence))
            elif row["strand"] == "-":
                print("\nin", filename, "TSS", row["base"],"is on the bottom
strand, therefore:")
                upstream = row["base"] - int(bases_downstream)
                downstream = row["base"] + int(bases_upstream)
                print("100 bases upstream is:\n", upstream, "\n50 bases downstream
is:\n", downstream, "\n")
                coordinates = str(upstream)+'-'+str(downstream)
                genome_coordinates.append(coordinates)
                print("extracting sequence from positions", upstream, "-",
downstream, "...")
                print("as this TSS is on the bottom strand, reverse complenting the
extracted sequence...")
                for seq in SeqIO.parse(reference_genome, "fasta"):
```

```
                    start_coord = upstream
                    end_coord = downstream
                    rev_sequence = seq.seq[start_coord -
1:end_coord].reverse_complement()
                    title = ">" + str(start_coord) + "_" + str(end_coord)
                    extracted_sequence.append(title)
                    extracted_sequence.append(rev_sequence)
                    print("extracted and reverse complemented sequence is: ",
str(rev_sequence))


        # Create pathname and filename for genome coordinates output file and for
output fasta sequences
        path = directory.decode("utf-8")
        fileout = os.path.splitext(filename)[0] + '_TSS_coordinates.txt'
        sequenceout = os.path.splitext(filename)[0] +
'_TSS_coordinates_sequence.fasta'
        coordinate_output = str(path) + "/" + str(fileout)
        sequence_output = str(path) + "/" + str(sequenceout)

        # Save each list of genome coordinates as a .txt file
        final_extracted_sequence = np.array(extracted_sequence, dtype=object)
        np.savetxt(coordinate_output, genome_coordinates, fmt = '%-1.16s', newline
= ",")

        # Save fasta sequences as fasta file
        np.savetxt(sequence_output, final_extracted_sequence, fmt = '%-1.200s')
        # end of function


    # Ask for user specified number of bases to flank each TSS (upstream and
downstream)
    bases_upstream = input("How many bases upstream of each TSS would you like?
(enter number only): ")
    bases_downstream = input("How many bases downstream of each TSS would you like?
(enter number only): ")

    # Ask for a reference genome (supplied in FASTA format)
    reference_genome = input("Please paste the path to your FASTA reference genome
here: ")
    print("fasta file is: ", reference_genome)

    # Ask for a path to the folder containing the .gtf files to be analysed
    directory = os.fsencode(input("Enter path to folder with all .gtf files here:
"))
    print("path to folder is: ", directory)

    # run code defined above on all files that end with .gtf in the folder
specified
```

```
    os.chdir(directory)
    start_time = time.time()
    for file in os.listdir(directory):
        filename = os.fsdecode(file)
        if filename.endswith(".gtf"):
            rnaseq_coordinates(filename)
            print("\nTSS coordinates and flanking sequences from ", filename, "
have been extracted" )
        else:
            continue



    print("\nAll .gtf files analysed, coordinates found, and", bases_upstream,
"bases upstream and", bases_downstream, "bases downstream have been extracted from
the reference genome and saved in a fasta file\n")

    end_time = time.time()

    print(f"Program ran in {((end_time - start_time)/60)} minutes")
```

Note: this script (and to a certain degree, 7.1.2/3) are not efficient and can take a while to run. For example, I ran the above script on 8 .gtf files and the script took just over an hour to complete. Be aware of this if you run this script. The script can run in the background but may be interrupted if your computer goes to sleep or the screensaver starts.

7.1.2 RNA_seq_combine_replicates.py

```
# Made by Alistair in Python 3.9

# Script to combine individual replicate .gtf files to identify TSSs found in both
replicates
# Written and tested using Python 3.8

import os
import Bio
from Bio import SeqIO
from Bio.Seq import Seq
import pandas as pd
import numpy as np
import csv
from functools import reduce
import time

if __name__ == '__main__':
```

```
    # Input one .gtf file for each replicate and extract only the columns 'base'
and 'strand'; then ask for the name of the samples, the reference genome, and the
number of bases to extract
    df1 = pd.read_csv(input(r"Enter the path to the first replicate here: "), sep =
"\t", index_col=False, engine='python', header=None, skiprows=1, usecols= [3,6],
names= ["base", "strand"])
    df2 = pd.read_csv(input(r"Enter the path to the second replicate here: "), sep
= "\t", index_col=False, engine='python', header=None, skiprows=1, usecols= [3,6],
names= ["base", "strand"])
    name = input("What is the name of these samples? ")
    reference_genome = (input(r"Enter path to reference genome FASTA file here: "))
    bases_upstream = input("How many bases upstream of each TSS would you like?
(enter number only): ")
    bases_downstream = input("How many bases downstream of each TSS would you like?
(enter number only): ")
    print("replicate 1 is :\n", df1, "\nreplicate 2 is:\n", df2)

    os.chdir(input(r"Enter the path to the output folder here: "))

    start_time = time.time()

    # Merge both .gtf files and count the total number of TSSs found in both
replicates.  This script also includes the TSSs that are found at the same base but
on different strands and extracts them to a seperate csv files
    count = 0
    notequal = 0
    not_equal_output = []

    merged_df = pd.merge(df1, df2, on=["base"])
    print(merged_df)
    merged_df.dropna(inplace=True)
    print(merged_df)

    for label, row in merged_df.iterrows():
        if row["strand_x"] == row["strand_y"]:
            count += 1
        elif row["strand_x"] != row["strand_y"]:
            notequal += 1
            not_equal_output.append(row)

    total = count + notequal
    not_equal_df = pd.DataFrame(not_equal_output)
    print("\nnot equal output is: \n", not_equal_df)
    print(not_equal_df.shape)
    print(merged_df)
    print("total number of rows: ", str(total))
    print("number of rows that are equal: ", str(count))
    print("number of rows that are not equal: ", str(notequal))
```

```
    # Iterate over merged replicates and extract the sequence upstream and
downstream of each TSS and extract the -10 element sequence as well.
    genome_coordinates = []
    upstream = 1
    downstream = 1
    coordinates = []
    start_coord = 1
    end_coord = 1
    extracted_sequence = []
    extracted_minus10 = []

    for label, row in merged_df.iterrows():
        if row["strand_x"] and row["strand_y"] == "+":
            upstream = row["base"] - int(bases_upstream)
            downstream = row["base"] + int(bases_downstream)
            coordinates = str(upstream) +  "-" + str(downstream)
            genome_coordinates.append(coordinates)
            minus10_start = row["base"]
            minus10_end = row["base"]
            for seq in SeqIO.parse(reference_genome, "fasta"):
                start_coord = upstream
                end_coord = downstream
                sequence = seq.seq[start_coord -1:end_coord]
                title = ">" + str(start_coord) + "_" + str(end_coord)
                extracted_sequence.append(title)
                extracted_sequence.append(sequence)
                minus10 = seq.seq[minus10_start - 17:minus10_end]
                title = title + "_-10_element"
                extracted_minus10.append(title)
                extracted_minus10.append(minus10)
                print("coordinates and -10 element extracted from top strand")
        elif row["strand_x"] and row["strand_y"] == "-":
            upstream = row["base"] - int(bases_downstream)
            downstream = row["base"] + int(bases_upstream)
            coordinates = str(upstream) + "-" + str(downstream)
            genome_coordinates.append(coordinates)
            minus10_start = row["base"]
            minus10_end = row["base"]
            for seq in SeqIO.parse(reference_genome, "fasta"):
                start_coord = upstream
                end_coord = downstream
                rev_sequence = seq.seq[start_coord -1:
end_coord].reverse_complement()
                title = ">" + str(start_coord) + "_" + str(end_coord)
                extracted_sequence.append(title)
                extracted_sequence.append(rev_sequence)
                minus10 = seq.seq[minus10_start - 1:minus10_end +
16].reverse_complement()
                title = title + "_-10_element"
```

```
                extracted_minus10.append(title)
                extracted_minus10.append(minus10)
                print("coordinates and -10 element extracted from reverse strand")

    # Create pathname and filename for genome coordinates output file and for
output fasta sequences
    fileout = name + '_combined_replicates'
    sequenceout = fileout + '_coordinates_sequence.fasta'
    minus10out = fileout + "_-10_elements.fasta"
    TSS_output = fileout + "_coordinates.txt"
    sequence_output = fileout + sequenceout
    minus10_output = fileout + minus10out

    # Save each list of genome coordinates as a .txt file and save fasta sequences
as fasta file
    final_extracted_sequence = np.array(extracted_sequence, dtype=object)
    final_minus10 = np.array(extracted_minus10, dtype = object)
    np.savetxt(TSS_output, genome_coordinates, fmt = '%-1.16s', newline = ",")
    np.savetxt(sequenceout, final_extracted_sequence, fmt = '%-1.200s')
    np.savetxt(minus10out, final_minus10, fmt = '%-1.200s')



    # Save the list of TSSs (with strand information for each replicate) to a csv
file
    merged_name = name + "_merged_replicates.txt"
    merged_df.to_csv(merged_name, sep='\t', index=False)
    not_equal_name = name + "_TSSs_opposite.txt"
    not_equal_df.to_csv(not_equal_name, sep='\t', index=False)

    end_time = time.time()
    final_time = (end_time - start_time)/60
    print()
    print(merged_df)
    print(name, "replicates have been merged, sequence and -10 extracted")
    print("total number of rows: ", str(total))
    print("number of rows that are equal: ", str(count))
    print("number of rows that are not equal: ", str(notequal))
    print("program finished with a time of", final_time, "minutes")
```

7.1.3 RNA_seq_total_combined_TSS.py

```
# Made by Alistair in Python 3.9
```

```
# Script to combine all gtf files and then filter any duplicate rows to give single
list of total TSSs within all files/replicates
# Script will also extract the sequence of the start sites and -10 elements and
generate FASTA files
# Written and tested in Python 3.8

import pandas as pd
import numpy as np
import os
import Bio
from Bio import Seq
from Bio import SeqIO
import time
import glob

if __name__ == '__main__':

    # Input path to directory and set as working directory, enter path to reference
genome FASTA file and specify how many bases either side of the TSS to extract
    directory = os.fsencode(input(r"Enter path to folder with all .gtf files here:
"))
    print("path to folder is: ", directory)
    os.chdir(directory)
    reference_genome = (input(r"Enter path to reference genome FASTA file here: "))
    bases_upstream = input("How many bases upstream of each TSS would you like?
(enter number only): ")
    bases_downstream = input("How many bases downstream of each TSS would you like?
(enter number only): ")

    start_time = time.time()

    # Get all files that end with .gtf
    filenames = glob.glob("*.gtf")

    # Generate a list of dataframes from each file in filenames, use only the 4th
and 7th column and call them base and strand
    list_of_dataframes = [pd.read_csv(filename, sep = "\t", index_col=False,
engine='python', header=None, skiprows=1, usecols= [3,6], names= ["base",
"strand"]) for filename in filenames]
    print(list_of_dataframes)

    # Combine all dataframes into one dataframe
    combined_df = pd.concat(list_of_dataframes, ignore_index=True)
    print("\nAll .gtf files have been combined here: ")
    print(combined_df)
    print("\nTotal number of TSSs across all files is: ", len(combined_df), "\n")

    # Strip all duplicate start sites from the combined dataframe
    total_df = combined_df.drop_duplicates().sort_values("base")
```

```
    print(total_df)
    print("\ntotal number of unique TSSs across all files is:", len(total_df),
"\n")


    # Iterate through the dataframe and extract sequence from fasta using defined
upstream and downstream coordinates.  Also extract 16 bases upstream of TSS for -10
analysis
    genome_coordinates = []
    upstream = 1
    downstream = 1
    coordinates = []
    start_coord = 1
    end_coord = 1
    extracted_sequence = []
    extracted_minus10 = []

    for label, row in total_df.iterrows():
        if row["strand"] == "+":
            upstream = row["base"] - int(bases_upstream)
            downstream = row["base"] + int(bases_downstream)
            coordinates = str(upstream) +  "-" + str(downstream)
            genome_coordinates.append(coordinates)
            minus10_start = row["base"]
            minus10_end = row["base"]
            for seq in SeqIO.parse(reference_genome, "fasta"):
                start_coord = upstream
                end_coord = downstream
                sequence = seq.seq[start_coord -1:end_coord]
                title = ">" + str(start_coord) + "_" + str(end_coord)
                extracted_sequence.append(title)
                extracted_sequence.append(sequence)
                minus10 = seq.seq[minus10_start - 17:minus10_end]
                title = title + "_-10_element"
                extracted_minus10.append(title)
                extracted_minus10.append(minus10)
                print("Coordinates and -10 element extracted from TSS on top
strand")
        elif row["strand"] == "-":
            upstream = row["base"] - int(bases_downstream)
            downstream = row["base"] + int(bases_upstream)
            coordinates = str(upstream) + "-" + str(downstream)
            genome_coordinates.append(coordinates)
            minus10_start = row["base"]
            minus10_end = row["base"]
            for seq in SeqIO.parse(reference_genome, "fasta"):
                start_coord = upstream
                end_coord = downstream
```

```
                rev_sequence = seq.seq[start_coord -1:
end_coord].reverse_complement()
                title = ">" + str(start_coord) + "_" + str(end_coord)
                extracted_sequence.append(title)
                extracted_sequence.append(rev_sequence)
                minus10 = seq.seq[minus10_start - 1:minus10_end +
16].reverse_complement()
                title = title + "_-10_element"
                extracted_minus10.append(title)
                extracted_minus10.append(minus10)
                print("Coordinates and -10 element extracted from TSS on bottom
strand")



    # Create pathname and filename for genome coordinates output file and for
output FASTA sequences
    path = directory.decode("utf-8")
    fileout = 'total_combined_TSSs'
    sequenceout = fileout + '_coordinates_sequence.fasta'
    minus10out = fileout + "_-10_elements.fasta"
    out_df = str(path) + "/" + str(fileout) + ".txt"
    TSS_output = str(path) + "/" + str(fileout) + "_coordinates.txt"
    sequence_output = str(path) + "/" + str(sequenceout)
    minus10_output = str(path) + "/" + str(minus10out)

    # Save each list of genome coordinates as a .txt file and save fasta sequences
as FASTA file
    total_df.to_csv(out_df, sep="\t", index=False)

    final_extracted_sequence = np.array(extracted_sequence, dtype=object)
    final_minus10 = np.array(extracted_minus10, dtype = object)
    np.savetxt(TSS_output, genome_coordinates, fmt = '%-1.16s', newline = ",")
    np.savetxt(sequence_output, final_extracted_sequence, fmt = '%-1.200s')
    np.savetxt(minus10out, final_minus10, fmt = '%-1.200s')


    end_time = time.time()
    total_time = (end_time - start_time)/60

    print("\nAll .gtf files from folder:", directory, "have been combined and
duplicates removed. There are", len(total_df), "unique TSSs", "\nAll sequences and
-10 elements have also been extracted. \nTotal time of program was:", total_time,
"minutes")
```

7.1.4 generate_EdgeR_inputs.py

```
# Made by Alistair in Python 3.9

# Script to generate csv files for use with EdgeR in R.  Takes both replicates of a
condition and both replicates of the control and keeps only those TSS found within
all four samples.
# Written and tested in Python 3.8

import os
import pandas as pd
import numpy as np
import csv
from functools import reduce
import time

if __name__ == '__main__':

    # Get all input files and use only the name, base, and strand columns.  Also
ask user for the name of the samples
    df1 = pd.read_csv(input(r"Enter the path to the first sample replicate: "), sep
= "\t", index_col=False, engine='python', header=None, skiprows=1, usecols=
[2,3,6], names= ["name", "base", "strand"])
    df2 = pd.read_csv(input(r"Enter the path to the second sample replicate: "),
sep = "\t", index_col=False, engine='python', header=None, skiprows=1, usecols=
[2,3,6], names= ["name", "base", "strand"])
    df3 = pd.read_csv(input(r"Enter the path to the first control replicate: "),
sep = "\t", index_col=False, engine='python', header=None, skiprows=1, usecols=
[2,3,6], names= ["name", "base", "strand"])
    df4 = pd.read_csv(input(r"Enter the path to the second control replicate: "),
sep = "\t", index_col=False, engine='python', header=None, skiprows=1, usecols=
[2,3,6], names= ["name", "base", "strand"])
    name = input("What is the name of these samples? ")
    os.chdir(input(r"Enter the path to the folder you want your results saved in
here: "))

    start_time = time.time()

    print(df1)
    print(df2)
    print(df3)
    print(df4)

    # Merge both replicates of the condition and merge both replicates of the
control, then merge both together to give one list of all TSS in all files
    merged_1_2_df = df1.merge(df2, on="base").rename(columns={'name_x':
'rep1_name', 'base': 'base', 'strand_x': 'rep1_strand', 'name_y': 'rep2_name',
'strand_y': 'rep2_strand'})
    print(merged_1_2_df)
```

```
    merged_3_4_df = df3.merge(df4, on="base").rename(columns={'name_x':
'rep3_name', 'base': 'base', 'strand_x': 'rep3_strand', 'name_y': 'rep4_name',
'strand_y': 'rep4_strand'})
    print(merged_3_4_df)
    merged_df = merged_1_2_df.merge(merged_3_4_df, on="base")
    print(merged_df)

    # Rename each TSS (removes the accession number) and create a new column for
the coverage at each TSS
    # Note that the coverage must be used for EdgeR and is different to the RRS
generated by the Ettwiler scripts, this section may not work/be relevant if you
havent used the Ettwiler scripts
    merged_df['rep1_name'] = merged_df.rep1_name.str[12::]
    merged_df['rep1_coverage'] = merged_df['rep1_name'].str.split('_').str[1]
    merged_df['rep1_strand'] = merged_df.rep1_name.str[-2:-1:]
    merged_df['rep2_name'] = merged_df.rep2_name.str[12::]
    merged_df['rep2_coverage'] = merged_df['rep2_name'].str.split('_').str[1]
    merged_df['rep2_strand'] = merged_df.rep2_name.str[-2:-1:]
    merged_df['rep3_name'] = merged_df.rep3_name.str[12::]
    merged_df['rep3_coverage'] = merged_df['rep3_name'].str.split('_').str[1]
    merged_df['rep3_strand'] = merged_df.rep3_name.str[-2:-1:]
    merged_df['rep4_name'] = merged_df.rep4_name.str[12::]
    merged_df['rep4_coverage'] = merged_df['rep4_name'].str.split('_').str[1]
    merged_df['rep4_strand'] = merged_df.rep4_name.str[-2:-1:]
    merged_df['ID'] = 'TSS_' + merged_df['base'].astype(str) + '_' +
merged_df['rep1_strand'].astype(str)
    merged_df['strand_all'] = merged_df['rep1_strand'] + merged_df['rep2_strand'] +
merged_df['rep3_strand'] + merged_df['rep4_strand']
    print(merged_df)

    # Count number of TSSs that are found on the same strand in all replicates, and
those whose strands are not conserved
    count = 0
    not_equal_count = 0

    for label, row in merged_df.iterrows():
        if row['strand_all'] == '++++':
            count += 1
        elif row['strand_all'] == '----':
            count +=1
        else:
            not_equal_count +=1

    print("\nNumber of TSSs equal in all replicates is: ", str(count))
    print("\nNumber of TSSs not equal is: ", str(not_equal_count))

    # Save each replicates coverage into a list for that replicate if the TSS is
found on the same strand for all four samples.  Uses the same name (ID) for each
file
```

```
    rep1_both = []
    rep2_both = []
    rep3_both = []
    rep4_both = []

    for label, row in merged_df.iterrows():
        if row['strand_all'] == '++++':
            rep1_both.append(row[['ID', 'rep1_coverage']])
            rep2_both.append(row[['ID', 'rep2_coverage']])
            rep3_both.append(row[['ID', 'rep3_coverage']])
            rep4_both.append(row[['ID', 'rep4_coverage']])
        if row['strand_all'] == '----':
            rep1_both.append(row[['ID', 'rep1_coverage']])
            rep2_both.append(row[['ID', 'rep2_coverage']])
            rep3_both.append(row[['ID', 'rep3_coverage']])
            rep4_both.append(row[['ID', 'rep4_coverage']])

    # Convert lists to a dataframe and rename columns to TSS_ID and counts
    rep1_both_df = pd.DataFrame(rep1_both).rename(columns={'ID': 'TSS_ID',
'rep1_coverage': 'counts'})
    print("Replicate 1 start sites and coverage is: \n", rep1_both_df)
    rep2_both_df = pd.DataFrame(rep2_both).rename(columns={'ID': 'TSS_ID',
'rep2_coverage': 'counts'})
    print("Replicate 2 start sites and coverage is: \n", rep2_both_df)
    rep3_both_df = pd.DataFrame(rep3_both).rename(columns={'ID': 'TSS_ID',
'rep3_coverage': 'counts'})
    print("Replicate 3 start sites and coverage is: \n", rep3_both_df)
    rep4_both_df = pd.DataFrame(rep4_both).rename(columns={'ID': 'TSS_ID',
'rep4_coverage': 'counts'})
    print("Replicate 4 start sites and coverage is: \n", rep4_both_df)


    # Make file names and paths to save files.
    replicate1_name = '_replicate_1'
    replilcate2_name = '_replicate_2'
    control1_name = '_control_1'
    control2_name = '_control_2'
    EdgeR = '_EdgeR_input.csv'

    final_rep1 = name + replicate1_name + EdgeR
    final_rep2 = name + replilcate2_name + EdgeR
    final_rep3 = name + control1_name + EdgeR
    final_rep4 = name + control2_name + EdgeR

    # Save files to current directory
    rep1_both_df.to_csv(final_rep1, index=False)
    rep2_both_df.to_csv(final_rep2, index=False)
    rep3_both_df.to_csv(final_rep3, index=False)
    rep4_both_df.to_csv(final_rep4, index=False)
```

```
    end_time = time.time()
    final_time = end_time - start_time
    print("\nAll .gtf files for", name, "replicates and controls have been
analysed.  \nAll files have been merged and only start sites found within all 4
files and found on the same strand in all replicates have been kept.  \nEach start
site has been named and the coverage for each replicate has been saved to a .csv
file for analysis with EdgeR.\nTime taken was: ", final_time, "seconds")
```

7.1.5 extract_ChIP-seq_coordinates.py

```
# Made by Alistair in Python 3.9

# Script to take ChIP seq peaks centres and extract 100 or 150 bases either side.
Will use this to determine which TSSs from the RNA-seq data lie within the ChIP
targets.
# Written and tested in Python 3.8

import pandas as pd
import numpy as np
import os
from pandas.core.frame import DataFrame

if __name__ == '__main__':

    os.chdir(input(r'Enter the path to the output folder here: '))
    chip_peaks = pd.read_csv(input(r'Enter path to the ChIP results table as .csv
file: '), index_col=False)

    print(chip_peaks)

    # Create new columns based on the Peak Centre column.  Finds the genome
coordinates for either 100 or 150 bases up and downstream
    chip_peaks['100_start'] = chip_peaks.apply(lambda row: row['Peak Centre'] -
100, axis=1)
    chip_peaks['100_end'] = chip_peaks.apply(lambda row: row['Peak Centre'] + 100,
axis=1)
    chip_peaks['150_start'] = chip_peaks.apply(lambda row: row['Peak Centre'] -
150, axis=1)
    chip_peaks['150_end'] = chip_peaks.apply(lambda row: row['Peak Centre'] + 150,
axis=1)

    print(chip_peaks)
```

```
    # Extract coordinates for the targets of each transcripton factor.  Two files
for each TF, one with 100 bases either side and the other with 150 bases
    marcount = 0
    mar100_coords = []
    mar150_coords = []
    soxcount = 0
    sox100_coords = []
    sox150_coords = []
    ramcount = 0
    ram100_coords = []
    ram150_coords = []
    totalcount = 0
    total100 = []
    total150 = []


    for label, row in chip_peaks.iterrows():
        if 'MarA' in row['Binding Protein']:
            marcount += 1
            mar100_coords.append(row[['100_start', '100_end', 'Gene(s)1']])
            mar150_coords.append(row[['150_start', '150_end', 'Gene(s)1']])
        if 'SoxS' in row['Binding Protein']:
            soxcount += 1
            sox100_coords.append(row[['100_start', '100_end', 'Gene(s)1']])
            sox150_coords.append(row[['150_start', '150_end', 'Gene(s)1']])
        if 'RamA' in row['Binding Protein']:
            ramcount += 1
            ram100_coords.append(row[['100_start', '100_end', 'Gene(s)1']])
            ram150_coords.append(row[['150_start', '150_end', 'Gene(s)1']])

    for label, row  in chip_peaks.iterrows():
        totalcount += 1
        total100.append(row[['100_start', '100_end', 'Gene(s)1']])
        total150.append(row[['150_start', '150_end', 'Gene(s)1']])


    # Convert all lists of coordinates to a dataframe and creates a new column to
number each line
    mar100_df = DataFrame(mar100_coords).rename(columns={'100_start': 'start',
'100_end': 'end', 'Gene(s)1': 'genes'})
    mar100_df.insert(0, 'MarA peak', np.arange(len(mar100_df))+1)
    mar150_df = DataFrame(mar150_coords).rename(columns={'150_start': 'start',
'150_end': 'end', 'Gene(s)1': 'genes'})
    mar150_df.insert(0, 'MarA peak', np.arange(len(mar150_df))+1)

    sox100_df = DataFrame(sox100_coords).rename(columns={'100_start': 'start',
'100_end': 'end', 'Gene(s)1': 'genes'})
    sox100_df.insert(0, 'SoxS peak', np.arange(len(sox100_df))+1)
```

```
    sox150_df = DataFrame(sox150_coords).rename(columns={'150_start': 'start',
'150_end': 'end', 'Gene(s)1': 'genes'})
    sox150_df.insert(0, 'SoxS peak', np.arange(len(sox150_df))+1)

    ram100_df = DataFrame(ram100_coords).rename(columns={'100_start': 'start',
'100_end': 'end', 'Gene(s)1': 'genes'})
    ram100_df.insert(0, 'RamA peak', np.arange(len(ram100_df))+1)
    ram150_df = DataFrame(ram150_coords).rename(columns={'150_start': 'start',
'150_end': 'end', 'Gene(s)1': 'genes'})
    ram150_df.insert(0, 'RamA peak', np.arange(len(ram150_df))+1)

    total100_df = DataFrame(total100).rename(columns={'100_start': 'start',
'100_end': 'end', 'Gene(s)1': 'genes'})
    total100_df.insert(0, 'Total peaks', np.arange(len(total100_df))+1)
    total150_df = DataFrame(total150).rename(columns={'150_start': 'start',
'150_end': 'end', 'Gene(s)1': 'genes'})
    total150_df.insert(0, 'Total peaks', np.arange(len(total150_df))+1)

    # Save as a csv file
    mar100_df.to_csv("MarA ChIP peaks 100.csv", index=False)
    mar150_df.to_csv("MarA ChIP peaks 150.csv", index=False)

    sox100_df.to_csv("SoxS ChIP peaks 100.csv", index=False)
    sox150_df.to_csv("SoxS ChIP peaks 150.csv", index=False)

    ram100_df.to_csv("RamA ChIP peaks 100.csv", index=False)
    ram150_df.to_csv("RamA ChIP peaks 150.csv", index=False)

    total100_df.to_csv("Total ChIP peaks 100.csv", index=False)
    total150_df.to_csv("Total ChIP peaks 150.csv", index=False)

    print('\nMarA count is: ', marcount)
    print('\nSoxS count is: ', soxcount)
    print('\nRamA count is: ', ramcount)
    print("\nTotal count is: ", totalcount)
```

7.1.6 merge_controls.py

```
# Made by Alistair in Python 3.9

# Script to merge two .gtf files (can use any tab seperated file really but will
need a few modifications) and keep only start sites found in both replicates on the
same strand
# Written and tested in Python 3.8
```

```
import pandas as pd
import numpy as np
from pandas.core.frame import DataFrame

if __name__ == '__main__':

    # Import both .gtf files
    df1 = pd.read_csv(input(r"Enter path to replicate one .gtf here: "), sep =
"\t", index_col=False, engine='python', header=None, skiprows=1, usecols= [2,3,6],
names= ["name", "base", "strand"])
    df2 = pd.read_csv(input(r"Enter path to replicate two .gtf here: "), sep =
"\t", index_col=False, engine='python', header=None, skiprows=1, usecols= [2,3,6],
names= ["name", "base", "strand"])


    print(df1)
    print(df2)

    # Combine both replicates and keep only TSSs that are found in both
    merged_1_2_df = df1.merge(df2, on="base").rename(columns={'name_x':
'rep1_name', 'base': 'base', 'strand_x': 'rep1_strand', 'name_y': 'rep2_name',
'strand_y': 'rep2_strand'})
    print(merged_1_2_df)

    merged_df = merged_1_2_df

    # Rename samples as TSS_[base]_[strand] and merge the strand information into
one column
    merged_df['rep1_name'] = merged_df.rep1_name.str[12::]
    merged_df['rep2_name'] = merged_df.rep1_name.str[12::]
    merged_df['TSS_ID'] = 'TSS_' + merged_df['base'].astype(str) + '_' +
merged_df['rep1_strand'].astype(str)
    merged_df['strand_all'] = merged_df['rep1_strand'] + merged_df['rep2_strand']

    equalcount = 0
    not_equal = 0

    final_ls = []

    # Iterate over dataframe and extract only TSSs that are found on the same
strand in both replicates
    for label, row in merged_df.iterrows():
        if row['strand_all'] == '++':
            equalcount +=1
            final_ls.append(row)
        elif row['strand_all'] == '--':
            equalcount +=1
            final_ls.append(row)
        else:
```

```
            not_equal += 1

    # Convert list to dataframe for saving
    final_df = DataFrame(final_ls)

    header = ['TSS_ID', 'start', 'strand']

    # Save only TSS_ID, base, and strand columns as .csv file
    final_df.rename(columns={'base':'start', 'rep1_strand': 'strand'},
inplace=True)
    final_df.to_csv(r"/Users/alistair/RNA-seq/bidirectional promoters/transcription
factors/RamA_TSSs.csv", columns=header, index=False)

    print(final_df)
    print("Number of TSS that are found on the same strand in both replicates is:
\n", str(equalcount))
    print("Number of TSS that are found on different strands is:: \n",
str(not_equal))
```

7.1.7 bidirectional_analysis_final.py

```
# Made by Alistair in Python 3.9

# Script to identify bidirectional promoters
# Takes a csv file (the one used to design the script was the output of the
merge_controls.py script) that has 3 columns (name, base, strand)
# Other csv files can be used as long as the TSSs are in a column called 'base' and
the strand is represented as a '+' or '-' and in a column called 'strand'
# Written and tested using Python 3.8

import pandas as pd
import copy
import time
import os

if __name__ == '__main__':

    controls_df = pd.read_csv(input(r"Enter path to list of start sites here (as a
csv file): "))
    sample_name = input("What is the name of the sample? ")
    os.chdir(input(r"Enter path to output folder here: "))

    # Create filenames for outputs
    final_name = sample_name + " all bidirectional TSS.csv"
    half_up_name = sample_name + " half bidirectional upstream TSS.csv"
```

```
    half_down_name = sample_name + " half bidirectional downstream TSS.csv"
    non_bd_name = sample_name + " TSS excluding bidirectional.csv"
    non_bd_plus_name = sample_name + " all plus strand non-bidirectional.csv"
    non_bd_minus_name = sample_name + " all minus strand non-bidirectional.csv"
    summary_name = sample_name + " bidirectional analysis summary.txt"

    start_time = time.time()

    # Convert to a dictionary of dictionaries
    df_dict = controls_df.to_dict('index')

    # Convert to a list of dictionaries (a little easier to work with)
    df_list = []
    for index in df_dict:
          df_list.append(df_dict[index])

    print(df_list[0:1:])

    matches_list = []
    pair_count = 0

    # Iterate through list of dictionaries and compare each 'row' to the 'rows'
below it.  When that difference becomes greater than 25, break and move on to the
next row.
    # Then calculate the difference between the starting row and the query rows.
If the difference is between 7 and 25 (inclusive) and the upstream TSS is on the '-
' strand AND the downstream TSS is on the '+' strand, call a bidirectional pair.
    # For each pair, add the values of pair count (to count the number of
bidirectional pairs) and difference to the keys "bidirectional pair" and
"difference".
    # Then take both these rows and append them to a new list, with the starting
row and then the query (ensures upstream start site is before/above the downstream
start site).
    for start_index, start_row in enumerate(df_list):
        for index, row in enumerate(df_list[start_index:]):
            diff = row['start'] - start_row['start']
            if 7 <= diff <= 25:
                if start_row['strand'] == '-':
                    if row['strand'] == '+':
                        pair_count += 1
                        start_row["bidirectional_pair"] = pair_count
                        row["bidirectional_pair"] = pair_count
                        start_row["difference"] = diff
                        row["difference"] = diff
                        start_row["coordinates"] = str(start_row['start']) + "-" +
str(row['start'])
                        matches_list.append(copy.deepcopy(start_row))
                        matches_list.append(copy.deepcopy(row))
            elif diff > 25:
```

```
                break

    # Save list of bidirectional promoters to csv file
    final_df = pd.DataFrame(matches_list)
    final_df.to_csv(final_name, index=False)
    print(final_df)


    # Split list of bidirectional promoter pairs into the upstream or downstream
promoter
    upstream_df = final_df.iloc[::2]
    print(upstream_df)
    upstream_df.to_csv(half_up_name, index=False)

    downstream_df = final_df.iloc[1::2]
    print(downstream_df)
    downstream_df.to_csv(half_down_name, index=False)


    # Now iterate through original list and extract rows that are NOT part of a
bidirectional pair.
    non_bd_list = []

    for i in df_list:
        if i not in matches_list:
            non_bd_list.append(i)

    all_plus = []
    plus_count = 0
    all_minus = []
    minus_count = 0


    # Now go through the non-bidirectional start sites and seperate into plus or
minus strand also
    for index, row in enumerate(non_bd_list):
        if row['strand'] == "+":
            all_plus.append(row)
            plus_count += 1
        if row['strand'] == "-":
            all_minus.append(row)
            minus_count += 1

    # Save all non-bidirectional results to csv files
    final_non_bd_df = pd.DataFrame(non_bd_list)
    final_non_bd_df.to_csv(non_bd_name, index=False)

    final_plus_df = pd.DataFrame(all_plus)
    final_plus_df.to_csv(non_bd_plus_name, index=False)
```

```
    final_minus_df = pd.DataFrame(all_minus)
    final_minus_df.to_csv(non_bd_minus_name, index=False)

    end_time = time.time()
    total_time = "{:.2f}".format((end_time - start_time))

    output = (
        f"Summary of results for {sample_name}\n\n"
        f"Total number of TSSs is: {len(df_list)}\n"
        f"\nNumber of directional TSSs is: {len(non_bd_list)}\n"
        f"\nNumber of unique start sites on plus strand is: {plus_count}\n"
        f"\nNumber of unique start sites on minus strand is: {minus_count}\n"
        f"\nNumber of unique start sites within a divergent pair: {(len(df_list) -
len(non_bd_list))}\n"
        f"\nNumber of divergent TSS pairs: {pair_count}\n"
        f"Which gives a total number of bidirectional promoters of: {(pair_count *
2)}\n"
        f"Therefore bidirectional promoters make up {round(((((pair_count * 2) /
(len(df_list))) * 100)), 2)}% of TSS in the {sample_name} data set\n"
        f"\nScript run with a time of {total_time} seconds"

    )

    summary = open(summary_name, 'w')
    summary.writelines(output)
    summary.close()


    print(output)
```

7.1.8 extract_bidirectional_promoter_sequences.py

```
# Made by Alistair in Python 3.9

# Take coordinates from bidirectional promoter lists and extracts those coordinates
(with user specified flanking regions) into a fasta file.

import pandas as pd
import numpy as np
import Bio
from Bio import SeqIO
import time
import copy

if __name__ == '__main__':
```

```
    # Ask for input file and number of bases to flank bidirectional pair
    reference_genome = input(r"Enter path to reference genome FASTA file here: ")
    inputfile = input(r"Enter path to bidirectional promoters as csv file here: ")
    flanking = int(input(r"Enter how many bases upstream and downstream of the
bidirectional pair you want included (number only): "))
    start_time = time.time()
    bidirectional_df = pd.read_csv(inputfile)

    # Take input file and convert to dictionary
    bd_dict = bidirectional_df.to_dict('index')

    # Convert dictionary to list of dictionary as this is easier to work with
    bd_list = []
    for index in bd_dict:
        bd_list.append(bd_dict[index])

    # Remove every other line of list as first line of pair contains all relevant
information including the coordinates, but TSSs are now only on the "-" strand
    half_list = copy.deepcopy(bd_list[::2])

    bidirectional_coordinates = []
    seven = []
    ten = []
    twelve = []
    eighteen = []
    twenty_three = []
    start_coord = 1
    end_coord = 1
    extension = 25 + flanking

    # Iterate through the list of bidirectional promoters and extract their
sequence from the reference genome.
    # Takes the upstream start site and extracts a user specified number of bases
upstream as well as downstream from the start site + 25.
    # This is because weblogo requires all sequences to be the same length, so 25
bases is added to the start site value first.
    # Because this is the furthest distance away that a bidirecional promoter is
called. The flanking value is then added to the 25 to give the end coordinate.
    # Note that as all the first TSS in the bidirectional pair is on the "-"
strand, all sequences should be reverse complemented
    for index, row in enumerate(half_list):
        start_coord = row['base'] - flanking
        end_coord = row['base'] + extension
        title = ">SL1344 bidirectional pair " + str(row['bidirectional_pair']) + "
with coordinates " + str(row['coordinates']) + " plus " + (str(extension)) + "
bases up and downstream"
        for seq in SeqIO.parse(reference_genome, "fasta"):
            sequence = seq.seq[start_coord -1:end_coord].reverse_complement()
```

```
            bidirectional_coordinates.append(title)
            bidirectional_coordinates.append(sequence)
        if row['difference'] == 7:
            last_coord = row['coordinates'].split('-')[1]
            end_coord = int(last_coord) + flanking
            start_coord = row['base'] - flanking
            title = ">SL1344_bidirectional_pair_" + str(row['bidirectional_pair'])
+ "_7_bases_spacing with coordinates " + str(row['coordinates']) + " plus " +
str(flanking) + " bases up and downstream"
            for seq in SeqIO.parse(reference_genome, "fasta"):
                sequence = seq.seq[start_coord -1:end_coord].reverse_complement()
                seven.append(title)
                seven.append(sequence)
        if row['difference'] == 10:
            last_coord = row['coordinates'].split('-')[1]
            end_coord = int(last_coord) + flanking
            start_coord = row['base'] - flanking
            title = ">SL1344_bidirectional_pair_" + str(row['bidirectional_pair'])
+ "_10_bases_spacing with coordinates " + str(row['coordinates']) + " plus " +
str(flanking) + " bases up and downstream"
            for seq in SeqIO.parse(reference_genome, "fasta"):
                sequence = seq.seq[start_coord -1:end_coord].reverse_complement()
                ten.append(title)
                ten.append(sequence)
        if row['difference'] == 12:
            last_coord = row['coordinates'].split('-')[1]
            end_coord = int(last_coord) + flanking
            start_coord = row['base'] - flanking
            title = ">SL1344_bidirectional_pair_" + str(row['bidirectional_pair'])
+ "_12_bases_spacing with coordinates " + str(row['coordinates']) + " plus " +
str(flanking) + " bases up and downstream"
            for seq in SeqIO.parse(reference_genome, "fasta"):
                sequence = seq.seq[start_coord -1:end_coord].reverse_complement()
                twelve.append(title)
                twelve.append(sequence)
        if row['difference'] == 18:
            last_coord = row['coordinates'].split('-')[1]
            end_coord = int(last_coord) + flanking
            start_coord = row['base'] - flanking
            title = ">SL1344_bidirectional_pair_" + str(row['bidirectional_pair'])
+ "_18_bases_spacing with coordinates " + str(row['coordinates']) + " plus " +
str(flanking) + " bases up and downstream"
            for seq in SeqIO.parse(reference_genome, "fasta"):
                sequence = seq.seq[start_coord -1:end_coord].reverse_complement()
                eighteen.append(title)
                eighteen.append(sequence)
        if row['difference'] == 23:
            last_coord = row['coordinates'].split('-')[1]
            end_coord = int(last_coord) + flanking
```

```
            start_coord = row['base'] - flanking
            title = ">SL1344_bidirectional_pair_" + str(row['bidirectional_pair'])
+ "_23_bases_spacing with coordinates " + str(row['coordinates']) + " plus " +
str(flanking) + " bases up and downstream"
            for seq in SeqIO.parse(reference_genome, "fasta"):
                sequence = seq.seq[start_coord -1:end_coord].reverse_complement()
                twenty_three.append(title)
                twenty_three.append(sequence)


    final_list = np.array(bidirectional_coordinates, dtype=object)
    np.savetxt(r"/Users/alistair/RNA-
seq/hinton_paper/weblogos/extracted_sequences/Hinton_474_bidirectional_coordinates.
fasta", final_list, fmt = '%-1.200s')

    final_seven = np.array(seven, dtype=object)
    np.savetxt(r"/Users/alistair/RNA-
seq/hinton_paper/weblogos/extracted_sequences/Hinton_474_bidirectional_coordinates_
7_spacing_with_flanking.fasta", final_seven, fmt = '%-1.200s')

    final_ten = np.array(ten, dtype=object)
    np.savetxt(r"/Users/alistair/RNA-
seq/hinton_paper/weblogos/extracted_sequences/Hinton_474_bidirectional_coordinates_
10_spacing_with_flanking.fasta", final_ten, fmt = '%-1.200s')

    final_twelve = np.array(twelve, dtype=object)
    np.savetxt(r"/Users/alistair/RNA-
seq/hinton_paper/weblogos/extracted_sequences/Hinton_474_bidirectional_coordinates_
12_spacing_with_flanking.fasta", final_twelve, fmt = '%-1.200s')

    final_eighteen = np.array(eighteen, dtype=object)
    np.savetxt(r"/Users/alistair/RNA-
seq/hinton_paper/weblogos/extracted_sequences/Hinton_474_bidirectional_coordinates_
18_spacing_with_flanking.fasta", final_eighteen, fmt = '%-1.200s')

    final_twenty_three = np.array(twenty_three, dtype=object)
    np.savetxt(r"/Users/alistair/RNA-
seq/hinton_paper/weblogos/extracted_sequences/Hinton_474_bidirectional_coordinates_
23_spacing_with_flanking.fasta", final_twenty_three, fmt = '%-1.200s')

    end_time = time.time()
    final_time = (end_time - start_time)
    print("\nAll bidirectional promoter sequences extracted with ", str(flanking),
" bases either side.")
    print("\nTime taken is ", final_time, " seconds")
```