Global Gene Expression in ETEC H10407

by

Prateek Sharma



A thesis submitted to the University of Birmingham
for the degree of
Doctor of Philosophy

School of Biosciences

The University of Birmingham
August, 2017

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

ETEC (Enterotoxigenic *Escherichia coli*) is a gram negative bacterium that is responsible for acute watery diarrhoea in humans. ETEC is a significant cause of mortality; over 180,000 annual deaths reported by WHO (Qadri *et al.,* 2005) and high morbidity in infants. Virulence factors of ETEC have been characterized but underlying gene regulatory mechanisms are not yet understood. Moreover, antibiotic resistance in bacteria is a serious problem worldwide. Multiple antibiotic resistance in bacteria can be driven by transcriptional regulators in the AraC/XylS family. The *E. coli mar* regulon is considered a paradigm for such systems. The *mar* locus consists of 3 genes; *marR*, *marA*, and *marB*. A transcriptional activator encoded by *marA* enhances drug resistance by binding to "marbox" sequences at target promoters. The best characterised MarA targets encode the AcrAB-TolC drug efflux pump.

In this work, we have defined a molecular mechanism controlling expression of ETEC heat stable enterotoxin. We show that the cAMP receptor protein (CRP) directly activates expression of heat-stable toxin and Histone-like nucleoid structuring (H-NS) can exclude CRP from the activation-binding site. We also show that heat-stable toxin expression can be controlled by osmo-metabolic flux; CRP and H-NS allow the toxin gene promoter to respond to both glucose and salt. These conditions are encountered on host cell attachment and during oral rehydration therapy. We also identify 33 MarA targets to reveal novel mechanisms of antibiotic resistance.

Acknowledgements

“Knowing is not enough; we must apply. Willing is not enough; we must do.” - Goethe

The journey of PhD studies has been “the most enriching experience of my life”. I always believed on the above quote that kept me motivated. However, my PhD would have been impossible without people who accompanied, supported, motivated and encouraged me all the time during my PhD. I would like to extend my sincere thanks and appreciations to all those lovely people.

Firstly I would like to express my sincere gratitude to my supervisor, Dr. David C Grainger for the continuous support from the start date and until the very end of my PhD. I am very grateful for his trust, support and providing me an opportunity to come to the UK for my PhD. The ideas and suggestions he provided me were always brilliant. He looked after me by providing his suggestions almost every time I got stuck in my research. The invaluable guidance he provided me, have helped me a lot for my research and also in writing of this thesis. I could not have imagined having a better guide and mentor for my PhD. I feel really lucky and happy to have him as my supervisor, who is really talented, enthusiatatic, motivated and always ready for the next step forward. I can proudly say that he is my ideal who keeps me determined and motivated to achieve success in life.

Similar, profound gratitude goes to Dr. David Lee, Dr. Anne-Marie Krachler and Dr. Apoorva Bhatt, not only for their insightful comments and encouragement but also for

the hard questions which incented me to widen my research knowledge in many aspects. Other sincere thanks go to Prof. Laura Piddock, who gave me access to her laboratory and for her kind guidance. I also thank Dr. Vito Ricci from Piddock Lab, Dr. Ed Bevan and Dr. Laura Sellars for their kind help in some of my experiments.

I thank Dr. James Haycocks who have been a great colleague and a friend. He not only provided help by training me when I started my PhD but also is ready to help me with his kind suggestions whenever I am in need. Special mention goes to Dr. Gemma Warren for helping me by providing comments and proofreading my thesis.

I would like to thank Dr. Shivani Singh for always helping me like a sister throughtout my time in the UK. I also thank Mr. Sunindra Nath and Dr. Kiran Chintakayala for being such good friends who have always been supportive. I also thank Dr. Lisa Lamberte, Jainaba Manneh, Rachel Kettles, Emily Warman, Alistair Middlemiss, Tom Guest and all my labmates for keeping me motivated, their support throughout and making the lab friendly, full of fun and a wonderful place. I thank Fayyaz for being a wonderful and a great friend throughout my time in the UK, who is always available to help. I thank Anmol, Yasir, Hrishiraj, Asma, Panchali, Alessandro and Ajit for always being there as my friends. I would like to thank everyone at the School of Biosciences who have ever helped me in some way or the other. I appreciate the help provided by the reseach facilities of School of Biosciences in particular Dr. Norman Day and Lorraine Wallace. I thank the wonderful study and research atmosphere provided by the University of Birmingham.

I am also grateful to the Darwin Trust of Edinburgh for funding me and for providing the financial support throughout my PhD without which, it's almost impossible to complete my PhD.

Finally, I would like to thank the most important people in the world, my family: my Papa Ji and Mummy. I thank them for their unbelieveable support, motivation and for just being there with me. I can't thank my parents enough for whatever they have done for me. I will always be indebted for that. I would like to thank my little sister Miss. Prakriti Sharma for helping me preparing some of my figures, for her kind support all the time and being a great friend in my life. I would like to thank my wife Mrs. Neha Dwivedi for always being there with me as my partner. She helped me in many of my decisions and always been supportive throughout.

I would have not completed my PhD without the support of my family and my supervisor. Hence, I would like to dedicate this thesis to my family and my supervisor. Thank you so much for everything.

- Prateek Sharma

Table of Contents

Index of Figures 1

Index of Tables 6

Chapter 1. Introduction 7

1.1 Gene Regulation in *Escherichia coli* 8

1.2 Transcriptional Regulation 8

1.2.1 RNA polymerase 8

1.2.2 Domain Organisation of Sigma Factors and Alternate Sigma factors 11

1.2.3 Transcription Initiation 15

1.3 Transcription Factors required to activate or repress gene expression 17

1.3.1 Global and Specific Regulators 17

1.3.2 CRP; a conventional global regulator 19

1.3.3 H-NS; an unconventional global regulator 21

1.3.4 AraC; a specific regulator 22

1.4 Mechanisms of Transcription Activation 25

1.4.1 Class I activation 25

1.4.2 Class II activation 25

1.4.3 Class III activation: 26

1.5 Transcription repression 26

1.5.1 Repression by steric hindrance: 26

1.5.2 Repression by DNA looping: 27

1.5.3 Repression by modulation of an activator protein: 27

1.6 Transcription regulation by modulation of RNAP activity 27

1.7 Transcription regulation by altering the preference of RNAP binding 32

1.8 Diarrhoea and Enterotoxigenic *Escherichia coli* (ETEC)..... 35

1.9 Heat labile LT toxins..... 40

1.9.1 Secretion of LT toxin requires a Type 2 secretion system (T2SS)..... 43

1.10 Heat stable ST toxins are small highly structured enterotoxins 45

1.11 Colonizaton factors (CFs)..... 46

1.12 The Multiple Antibiotic Resistance operon of *E. coli*..... 46

1.13 Objectives of this study..... 55

Chapter 2. Materials and Methods.....57

2.1 General antibiotics and growth media used..... 58

2.1.1 Antibiotics (Stock Solutions)..... 58

2.1.2 Bacterial Growth Media..... 58

2.1.2.1 Solid media..... 58

2.1.2.2 Liquid media..... 59

2.2 Bacterial Strains and Plasmids..... 60

2.2.1 Bacterial strains and growth conditions..... 60

2.2.2 Plasmids..... 60

2.3 Plasmid DNA preparation using QIAprep Spin Mini/Maxi prep kit (QIAgen)..... 60

2.4 Phenol/Chloroform extraction of DNA..... 66

2.4.1 Reagents and chemicals used: 66

2.4.2 *Method* 66

2.5 Ethanol precipitation 66

2.5.1 Reagents and chemicals used: 66

2.5.2 *Method*: 66

2.6 Gel electrophoresis 67

2.6.1 Reagents and chemicals used: 67

2.6.2 Agarose gel electrophoresis 67

2.6.3 Polyacrylamide gel electrophoresis 68

2.7 DNA manipulations 68

2.7.1 Polymerase Chain Reaction (PCR) 72

2.7.2 Megaprimer PCR 72

2.7.3 Colony PCR 72

2.8 DNA purification using QIAquick PCR Purification kit (QIAgen) 72

2.9 DNA restriction digest 74

2.10 DNA Ligation 74

2.11 DNA transformation of bacterial cells 75

2.11.1 Preparation of $CaCl_2$ competent cells 75

2.11.2 Transformation of chemically competent cells with plasmid DNA 75

2.12 DNA Sequencing 76

2.13 β-galactosidase assays 76

2.13.1 *Reagents and chemicals used:* 76

2.13.2 *Method:* 76

2.14 Radiolabelling of DNA fragments 78

2.14.1 *Reagents and chemicals used* 78

2.14.2 *Method* 78

2.15 Maxam-Gilbert sequencing reactions 78

2.15.1 *Reagents and chemicals used* 78

2.15.2 *Method* 79

2.16 DNA sequencing gel electrophoresis 79

2.16.1 *Reagents and chemicals used* 79

2.16.2 *Method* 79

2.17 DNAse I footprinting 80

2.17.1 *Reagents and chemicals used* 80

2.17.2 *Method* 80

2.18 *In vitro* transcription assays 81

2.18.1 *Reagents and chemicals used* 81

2.18.2 *Method* 81

2.19 Microscopy 82

2.19.1 Reagents and chemicals used 82

2.19.2 Preparation of microscope slides 82

2.19.3 Imaging and analysis 82

2.20 Growth assays 83

2.21 Minimum inhibitory concentration (MIC) determination 83

2.21.1 *Materials, reagents and chemicals used* 83

2.21.2 Method 84

2.22 Accumulation Assay 84

2.22.1 *Materials, reagents and chemicals used* 84

2.22.2 *Method* 85

2.23 Efflux Assay 86

2.24 Hydrocarbon assay 86

2.24.1 *Materials, reagents and chemicals used* 86
2.24.2 *Method* 86
2.25 Crystal Violet Assay 87
2.26 Chromatin Immuno Precipitation (ChIP) 87
2.26.1 *Materials, reagents and chemicals used* 87
2.26.2 *Method* 89
2.26.3 ChIP analysis by PCR 90
2.26.4 ChIP analysis by Ilumina MiSeq sequencing 90
2.27 Agencourt AMPure XP magnetic bead clean up 92
2.28 Bioinformatic analysis of sequence reads 92
2.29 Bioinformatic identification of MarA binding sites 95
2.30 Phylogenetic analysis of the marbox 95
2.31 Pulsed-Field Gel Electrophoresis 96
2.31.1 Reagents and chemicals used 96
2.31.2 DNA Preparation 97

Chapter 3. Molecular Mechanisms controlling expression of ETEC H10407 enterotoxins. 99

3.1 Introduction 100
3.2 The *estA2* promoter is repressed by H-NS and activated by CRP 100
3.3 The *estA2 and estA1* are similarly regulated by H-NS 103
3.4 The *eltAB* operon is directly repressed by H-NS 104

3.5 Similar H-NS binding on all the enterotoxin promoters of ETEC H10407 104

3.6 The *estA1* and *estA2* promoters of ETEC H10407 respond to Glucose and Salt 108

3.7 The *eltAB* promoter also responds to glucose and salts 111

3.8 Regulation of ETEC enterotoxins by H-NS and CRP is conserved in other strains 111

3.9 Discussion 116

Chapter 4. Identification and verification of MarA binding sites across the ETEC H10407 genome 122

4.1 Introduction 123

4.2 MarA binding sites across the whole genome of ETEC H10407 123

4.3 *In vitro* characterization of MarA bound targets 126

4.4 The *mar* regulon is conserved amongst the family of enteric bacteria 126

4.5 Phenotypic landscape of the mar regulon 132

Chapter 5. MarA-dependent regulation of *xseA* is associated with DNA repair and quinolone resistance 137

5.1 Introduction 138

5.2 Characterisation of P*xseA* 138

5.2.1 Binding of MarA to *xseA* promoter (P*xseA*) 138

5.3 P*xseA* is a MarA activated promoter 141

5.4 Innate resistance to ciprofloxacin requires MarA mediated *xseA* expression 141

5.5 *xseA* determinants for DNA binding, catalysis and *XseB* interactions are required for complementation 145

5.6 Sub-inhibitory concentration of ciprofloxacin introduces double stranded breaks in the cells that lack MarA controlled *xseA* 145

5.7 Discussion 149

Chapter 6. MarA-dependent regulation of *mlaFEDCB* is associated with altered outer membrane barrier function and tetracycline resistance. 154

6.1 Introduction 155

6.2 Binding of MarA to *mlaFEDCB* 155

6.3 MarA activates transcription of *mlaFEDCB in vitro* and *in vivo* 155

6.4 Innate resistance to doxycycline requires MarA mediated *mlaFEDCB* expression 159

6.5 Defective barrier functions but normal efflux was observed in the cells lacking *mlaE* 162

6.6 Surface hydrophobicity increased in the cells lacking *mlaFEDCB* ... 167

6.7 Discussion 167

Chapter 7. Final conclusions and general discussions. 172

7.1 Final conclusions and General discussions 173

7.1.1 Overview 173

7.1.2 Understanding why chromatin immunoprecipitation does not identify all binding sites for CRP and MarA 174

7.1.3 Potential for overlapping regulation of *mlaFEDCB* and *xseA* by SoxS and Rob 176

7.1.4 The Rob paradox and its implications for this work 178

7.1.5 Why were targets such as *mlaFEDCB* and *xseA* not found in previous attempts to map the MarA regulon? 180

7.1.6 Comparison of MarA binding sites in ETEC and *Escherichia coli* K-12 181

7.1.7 Wider roles for DNA repair systems in modulating the effects of quinolones on gram negative bacteria 182

7.1.8 Future work 184

References 186

Appendix 221

Abbreviations 222

Index of Figures

Chapter 1:

Figure 1.1: Central dogma of molecular biology 9

Figure 1.2: Cartoon representation of RNA polymerase holoenzyme. 12

Figure 1.3: Domain organisation of σ factors. 14

Figure 1.4: The transcription cycle. 18

Figure 1.5: Ribbon structure of CRP dimer. 20

Figure 1.6: H-NS structural properties. 23

Figure 1.7: Domain organisation of single subunit of AraC. 24

Figure 1.8: Cartoon representation of transcription activation mechanisms 29

Figure 1.9: Cartoon representation of the mechanism of transcription repression. 31

Figure 1.10: Cartoon representation to show the regulation by promoter mimic. 33

Figure 1.11: Cartoon representation showing that some factors can remodel parts of RNAP 34

Figure 1.12: Representation of ETEC showing the genetic and virulent components. . 37

Figure 1.13: Synthesis and secretion of Heat Labile Toxin (LT) and Heat Stable Toxin (ST). 38

Figure 1.14: Representation of the model of pathogenesis in ETEC. 39

Figure 1.15: The structures of LT toxin. 41

Figure 1.16: Regulation of *eltAB* by Glucose 42

Figure 1.17: Type 2 secretion system that helps in secreting LT 44

Figure 1.18: The structure of the toxic domain of ST toxin. 47

Figure 1.19: Sequence alignments showing the homology between the endogenous peptides and the bacterial toxins. 48

Figure 1.20: Cartoon structure to illustrate the activation of *marRAB*. 51

Figure 1.21: Cartoon representation of several instrinsic resistance mechanisms. 52

Figure 1.22: Structure and binding site of MarA 53

Figure 1.23: Cartoon representation of the interaction of MarA with αCTD of RNAP. .. 54

Chapter 2:

Figure 2.1: pRW50 plasmid map. 63

Figure 2.2: pSR plasmid map. 64

Figure 2.3: pBR322 plasmid map. 65

Figure 2.4: Cartoon representation of ChIP-seq. 93

Chapter 3:

Figure 3.1: Defined hypothesis about *in vivo* regulation of enterotoxins by H-NS. 101

Figure 3.2: Investigation of LacZ activities (miller units) from different promoter derivatives of *estA2*. 102

Figure 3.3: Investigation of the LacZ activity from promoter derivative of *estA1*. 105

Figure 3.4: Investigation of the LacZ activities from different promoter derivatives of *eltAB* 106

Figure. 3.5: The various enterotoxin promoters of ETEC H10407 are repressed by H-NS. 107

Figure 3.6: Effect of titrating glucose and salts individually over the *estA1* and *estA2* promoter. 109
Figure 3.7: Control of production of ST toxin by glucose. 110
Figure 3.8: Production of ST toxin of ETEC H10407 is controlled by an osmo-metabolic flux. 112
Figure 3.9: The response from glucose needs an intact CRP site over the P*estA2* promoter. 113
Figure 3.10: Effect of titrating glucose and salts individually over the *eltAB* promoter. 114
Figure 3.11: Production of LT toxin of ETEC H10407 is controlled by an osmo-metabolic flux. 115
Figure 3.12: Transcription levels of enterotoxins of ETEC E24377A 117
Figure 3.13: Transcription level changes in the transcriptome of ETEC E24377A post intestinal cell attachment. 118
Figure 3.14: A model proposed for the control of enterotoxins expression of ETEC. . 119

Chapter 4:

Figure 4.1: MarA and RNAP σ70 binding 124
Figure 4.2: DNA sequence motifs derived from MarA binding sites. 125
Figure 4.3: Location of MarA and RNA polymerase binding peaks with respect to genes. 128
Figure 4.4: In vitro verification of MarA binding to the identified DNA targets. 130

Figure 4.5: Conservation of MarA regulon amongst the enterobacteriacea. 131

Figure 4.6: Phenotypic landscape of the MarA regulon. 133

Chapter 5:

Figure 5.1: MarA and σ70 binding across the xseA and regulatory sequence of *xseA*. 139

Figure 5.2: Binding of MarA to the *xseA* gene regulatory region. 140

Figure 5.3: DNAse I footprint of MarA binding to the xseA gene regulatory region. ... 142

Figure 5.4: The *xseA* promoter activity requires an intact marbox *in vivo*. 143

Figure 5.5: Genetic complementation of *xseA* phenotype. 144

Figure 5.6: An intact marbox is required at the upstream of the promoter for genetic complementation of *xseA* phenotype. 146

Figure 5.7: Domain organisation of XseA. 147

Figure 5.8: The key determinants of *xseA* are required for its proper functioning. 148

Figure 5.9: Ciprofloxacin induced damage of *E. coli* cells in the absence of MarA mediated *xseA* expression. 151

Figure 5.10: Ciprofloxacin induced DNA damage in the absence of *xseA* expression. 152

Chapter 6:

Figure 6.1: MarA and RNA polymerase bind together across the *mla* regulatory region. 156

Figure 6.2: MarA binds fragments of *mlaFEDCB* in the presence of the marbox. 157

Figure 6.3: DNAse I footprint of MarA binding to the regulatory region of *mlaFEDCB.* 158

Figure 6.4: *In vitro* transcription assay revealed the presence of three promoters in the regulatory region of *mlaFEDCB*. 160

Figure 6.5: The *mlaFEDCB* activity requires an intact marbox *in vivo*..... 161

Figure 6.6: Genetic complementation of *mlaF*::kan phenotype. 163

Figure 6.7: An intact marbox is required for genetic complementation of *mlaF*::kan phenotype. 164

Figure 6.8: Cells lacking *mlaE* are defective for barrier function. 165

Figure 6.9: Cells lacking *mlaE* are not defective for drug efflux..... 166

Figure 6.10: Cells that lack *mlaE* have increased surface hydrophobicity. 168

Figure 6.11: Cells that lack *mlaE* have increased surface hydrophobicity. 169

Chapter 7:

Figure 7.1: Overlap in the genome-wide DNA binding profiles of MarA in ETEC H10407 and SoxS in *E. coli* K-12..... 179

Figure 7.2: Comparison of the *xseA* and *recBCD* gene regulatory systems..... 183

Index of Tables

Table 1.1: Different classes of CFs:..... 49

Table 2.1: Strains used in the study..... 61

Table 2.2: Plasmids used in the study. 62

Table 2.3: Oligonucleotides used in this study..... 69

Table 2.4: Typical thermocycler settings used in the study..... 73

Table 4.1: MarA binding sites identified by ChIPseq……127

Table 7.1: Interactions of MarA and SoxS with previously proposed targets……177

Chapter 1. Introduction

1.1 Gene Regulation in *Escherichia coli*

Bacteria respond to external stimuli using a variety of molecular pathways. A common response is to modify patterns of gene expression. This can be done by mediating a series of catalytic reactions. Hence, information encoded by DNA is transcribed to form a RNA message that is translated to form a protein. This is known as the central dogma of molecular biology (Figure 1.1). Transcription, the first step in gene expression, involves copying DNA to make mRNA. This requires an enzyme called RNA polymerase. Translation requires ribosomes that assemble proteins using mRNA as a template. (Ishihama, 2012, Shimizu, 2013). Gene regulation is the process of increasing or decreasing the efficiency of gene expression in response to a stimulus. Thus, the earliest gene expression control step is transcription initiation (Sharma and Chatterji, 2010).

1.2 Transcriptional Regulation

1.2.1 RNA polymerase

In most organisms, RNA polymerase is a multi subunit enzyme that can be imagined as a crab claw. A channel runs from the centre of the claws to the active centre cleft (Chakraborty *et al.,* 2012). The active site of the enzyme is found at the base of the cleft created by the claws. The active site contains 2 magnesium ions that are responsible for addition of nucleotides to the 3' end of the polynucleotide RNA chain (Cramer *et al.,* 2001). In Gram negative bacteria, RNAPs are approximately 400 kDa in size. The core enzyme is composed of the $\alpha_2\beta\beta'\omega$ subunits (Murakami, 2015).

The identical α subunits each have a larger amino-terminal domain (αNTD) and a smaller carboxy-terminal domain (αCTD). The α subunit domains are connected by a

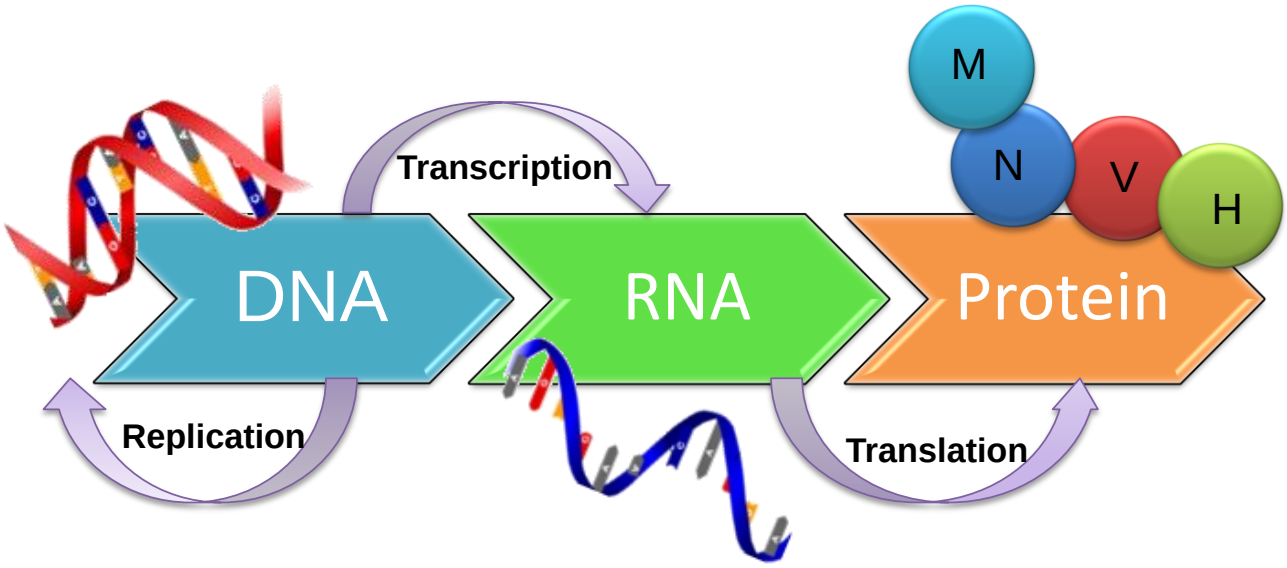


Figure 1.1: Central dogma of molecular biology: Genes are expressed based on the central dogma, where DNA (blue) is copied to make RNA (mRNA) (green) by the process of transcription and RNA then gets translated to form proteins (orange) by the process of translation.

13 amino acid flexible linker (Estrem *et al.,* 1999). The αCTD assists in DNA binding at promoters and also interacts with transcription factors (Ishihama, 1992). A group of amino acid residues, known as the '265 determinant', allow αCTD to contact the DNA at UP elements. The key residue, Arg-265, is a part of helix-hairpin-helix motif. This makes contact with the minor groove of DNA (Ross and Gourse, 2005). The β and β' subunits form the catalytic centre for RNA synthesis and are held together by αNTD. Moveable structures close to the active centre play a role in translocation of RNAP and in selecting nucleoside triphosphates. These structures are referred to as the bridge helix and trigger looping (Darst, 2001, Vassylyev *et al.,* 2002). The ω subunit is a 91 amino acid protein that binds the β' subunits and helps in its maturation (Zhang *et al.,* 1999, Mathew and Chatterji, 2006). The ω subunit also plays an important role in regulating RNA polymerase in response to guanosine tetraphosphate (ppGpp); a small molecule that plays a role in intracellular signalling, and changes in gene expression based on environmental signals (Chatterji *et al.,* 2007, Zuo *et al.,* 2013, Ross *et al.,* 2013, Vrentas *et al.,* 2005). Although the RNA polymerase core enzyme is catalytically active it cannot bind DNA specifically (Vassylyev *et al.,* 2002). This property is provided by the binding of the dissociable σ subunit (Vassylyev *et al.,* 2002). The σ subunit has 2 main functions: recognition of promoter sequences and facilitation of DNA melting at the transcription start sites (Finn *et al.,* 2000, Browning and Busby, 2004, Saecker *et al.,* 2011). In all bacteria one sigma factor recognises most promoters under normal growth conditions and is known as the housekeeping sigma factor (σ^{70} in *Escherichia coli*). In cases where bacteria respond to other signals, the house keeping sigma factor is

supplemented with alternate sigma factors. A cartoon representation of the RNA polymerase holoenzyme is shown in Figure 1.2.

1.2.2 Domain Organisation of Sigma Factors and Alternate Sigma factors

The domain organisation of sigma factors was first determined by proteolysis (Murakami, 2015). Thus, after proteolysis, individual σ domains could be isolated. Structures were determined for stable domains using X-Ray Crystallography and NMR. The first domain structure was obtained from *E. coli* group 1 σ^{70} in 1996 (Malhotra *et al.,* 1996). The 1.2 to 2.4 regions in the N-terminal domain of the sigma factor provided details of -10 recognition by σ region 2.4 and melting of promoter DNA by region 2.3 (Malhotra *et al.,* 1996). A more complete view of sigma factors was provided in the case of *Thermus aquaticus*, where sigma was divided into 2 proteolytic fragments. One of the fragments has domain 2 (region 1.2 - 2.4) and domain 3 (regions 3.0 - 3.1) and the other fragment has domain 4 (regions 4.1 - 4.2) (Campbell *et al.,* 2002). The modular structure of sigma factors is shown in Figure 1.3.

The genome of *E. coli* encodes 6 alternative σ factors that can compete with σ^{70}, in stress conditions, for binding core RNAP (Gruber and Gross, 2003). The highest number of genomic binding sites are for σ^{70} (1643 binding sites) in *E. coli* (Cho *et al.,* 2014). Different stresses can induce different sigma factors. These stresses include changes in temperature (σ^{32}), nutrient balance (σ^{38}), and protein folding (σ^{24}). The alternate sigma factors have several roles, for example:

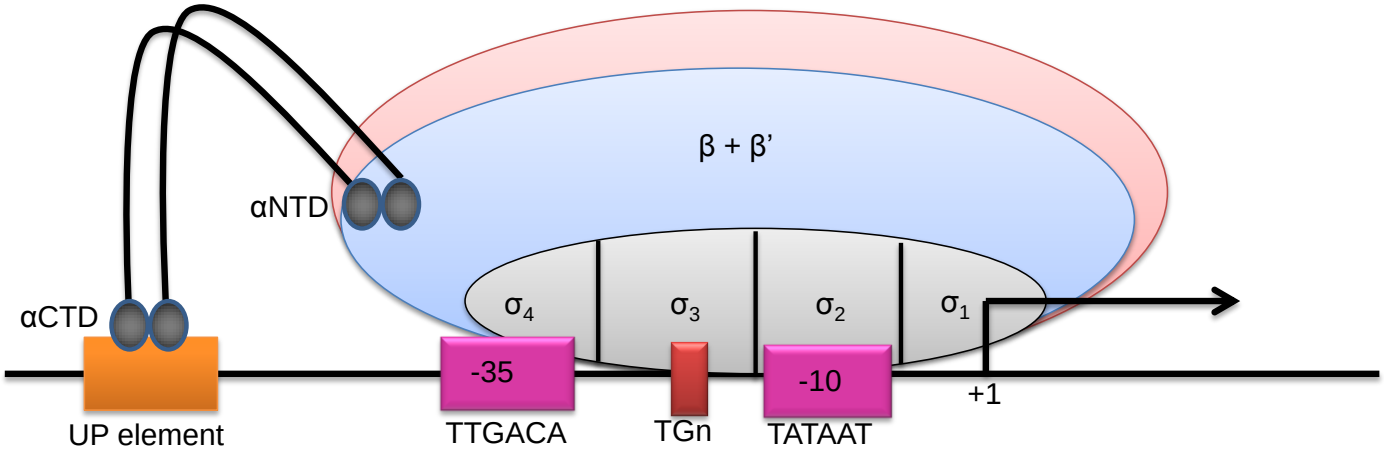


Figure 1.2: Cartoon representation of RNA polymerase holoenzyme. RNA polymerase associated with the upstream sequences. αNTD and αCTD are the Amino terminal and Carboxy terminal domains. -10 and -35 are promoter recognition sequences. Adapted from (Browning and Busby, 2004).

- σ^{38} controls the gene expression in starved cells and is abundant during the stationary phase of the bacterial growth cycle (Weber *et al.,* 2005).
- σ^{32} is the heat shock factor that is active when there is a change in temperature, pH, osmolarity or nutrients (Jenkins *et al.,* 1991).
- σ^{28} controls the expression of flagella related genes (Liu and Matsumura, 1996).
- σ^{24} responds to many stresses e.g., periplasmic protein stress, heat shock, oxidative stress etc. (Raina *et al.,* 1995, Erickson and Gross, 1989, De Las Penas *et al.,* 1997).
- σ^{19} FecI is an Iron starvation sigma factor that controls the expression of genes involved in ferric citrate transportation (Pressler *et al.,* 1988).
- σ^{54} is structurally unrelated to the other sigma factors. It is involved in the response to nitrogen limitation. The activity of σ^{54} is always regulator dependent (Carmona and Magasanik, 1996). This involves an ATP hydrolysis event mediated by AAA+ ATPase transcriptional activators known as enhancer binding proteins. This allows DNA melting and transcription initiation (Wedel *et al.,* 1990, Lee *et al.,* 2003). Consequently, σ^{54} pursuits tight control of gene expression.

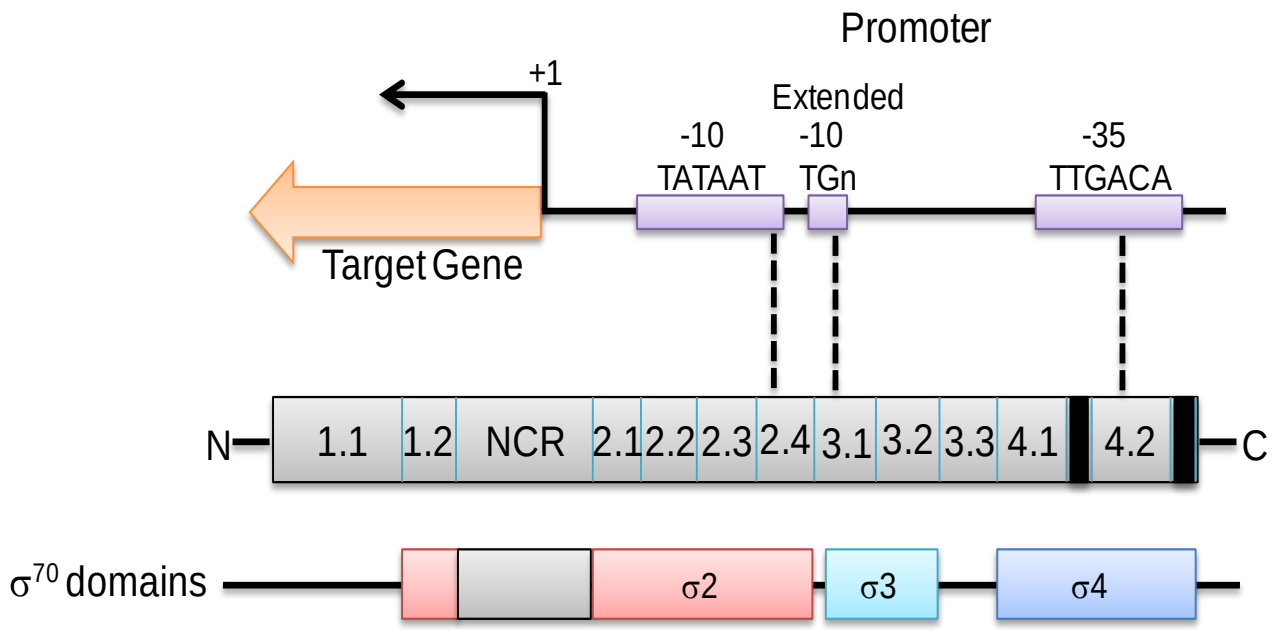


Figure 1.3: Domain organisation of σ factors. Structural domains of σ^{70} are represented in this figure. All four conserved regions, their subregions and the non conserved region (NCR (grey)) are shown. The promoter -10, -35 and extended -10 elements are purple rectangles. Interaction between σ subunits and the DNA elements are shown by black broken lines. Adapted from (Llamas *et al.,* 2014).

1.2.3 Transcription Initiation

1.2.3.1 Promoter Recognition and Binding

The RNAP σ subunit allows specific binding to DNA sequences called promoters (Burgess *et al.,* 1969). Hence the efficiency of transcription is related to promoter sequence. The -10 (TATAAT) and -35 (TTGACA) hexamers are located at -10 and -35 base pair upstream of the transcription start site respectively (Hawley and McClure, 1983). These elements are recognised by the σ_2 and σ_4 domains of the σ^{70} (Campbell *et al.,* 2002, Murakami *et al.,* 2002). Some promoters have an extended -10 element; the sequence TGn is found immediately upstream of the -10 hexamer. This extended TGn is recognized by σ3 domain of σ^{70}. UP elements recognised by the αCTD are 20 bp in length, AT rich and located at different distances upstream of -35 hexamers (Browning and Busby, 2004).

1.2.3.2 Open Complex formation (Isomerisation)

The interaction of RNAP with its promoter -10 element results in melting of the DNA upstream of the transcription start site (the open complex) (Browning and Busby, 2004, Browning and Busby, 2016). This process is also known as isomerisation (Borukhov *et al.,* 2005). The A11 and the T7 bases of the -10 elements ($T^{-12}A^{-11}T^{-10}A^{-9}A^{-8}T^{-7}$) are flipped out from the base stack and buried in a region of σ_2 (Zhang *et al.,* 2012). Consequently the A11 and T7 bases are highly conserved in -10 hexamers (Shultzaberger *et al.,* 2007). An initiation complex can now form with the addition of nucleoside triphosphates (NTPs). The σ_1 domain of the σ subunit contains negatively

charged amino acids that prevent spontaneous melting of DNA unless σ^{70} is properly associated with a promoter (Miropolskaya *et al.,* 2012).

1.2.3.3 Abortive Initiation (DNA scrunching)

Once the initiating nucleotide is positioned in the active site, a phosphodiester bond can form between this initial and an adjacent nucleoside. This leads to the synthesis of template dependent RNA transcripts. During initiation the template strand is pulled into RNAP in a process known as DNA scrunching. Hence, rather than RNAP moving, the RNA chain begins to grow, the complex is held at the promoter and the RNA:DNA duplex is pulled into the enzyme. This can result in abortive initiation producing small RNAs of 8-9 bp in length (Murakami *et al.,* 2002, Revyakin *et al.,* 2006). Recently, Zuo and Steitz solved the structure of a σ^{70} transcription initiation complex. The crystal structure provides details of promoter binding and transcript initiation. It was also shown that there is a bridging interaction between the 5' triphosphate of RNA and σ3.2 loop. This interaction suggested a function for σ^{70} in stabilizing short RNA:DNA hybrids during the early steps of transcription (Zuo *et al.,* 2013).

1.2.3.4 Promoter Escape

When RNA production extends beyond 9 nucleotides RNAP detaches from the DNA and escapes the promoter. The new RNA chain leaves the exit channel that accommodates 5 nucleotides of RNA. After promoter escape conformational changes occur in the holoenzyme resulting in dissociation of σ factor from the core RNA polymerase (Hernandez *et al.,* 1996, Roberts *et al.,* 1998). After escaping the promoter, RNAP is said to form an elongation complex (Browning and Busby, 2016).

The transcription elongation complex is very stable (Krummel and Chamberlin, 1989) and translocates at a speed of 30-100 nucleotides per second (Levin *et al.,* 1987). Transcription continues until a termination signal is encountered, RNA polymerase is freed from the DNA, and can rebind sigma (Vogel and Jensen, 1994, Murakami and Darst, 2003). The transcription cycle is illustrated in Figure 1.4.

1.3 Transcription Factors required to activate or repress gene expression

Many promoters have sequences that are suboptimal. At these promoters, regulatory proteins, referred to as transcription factors, bind specific DNA sequences and help RNAP to initiate transcription. Hence, transcription activators can activate gene expression. Conversely, a repressor is a protein that binds DNA or RNA to inhibit gene expression. Promoters have specific sequences of DNA (referred to as operators) that bind transcription factors. Binding of transcription factors is permitted by a structural motif in the transcription factor. The specificity for DNA targeting by transcription factors is further enhanced by dimerisation or multimerisation (Browning and Busby, 2016).

1.3.1 Global and Specific Regulators

Transcription factors can be characterised as global or local regulators (Perrenoud and Sauer, 2005). Half of the genes of *E. coli* are regulated by at least one of the 7 global transcription factors, which include CRP (cyclic AMP receptor protein), FNR (fumarate and nitrate reductase regulatory protein), H-NS (Histone like nucleoid structural protein), FIS (Factor for inversion and stimulation), IHF (Integration host factor), ArcA (Aerobic respiration regulatory protein) and Lrp (Leucine-responsive regulatory protein). There are about 300 transcription factors in the genome of *E. coli,* equating to 8% of the genes (Browning and Busby, 2004, Perez-Rueda and Collado-Vides, 2000).

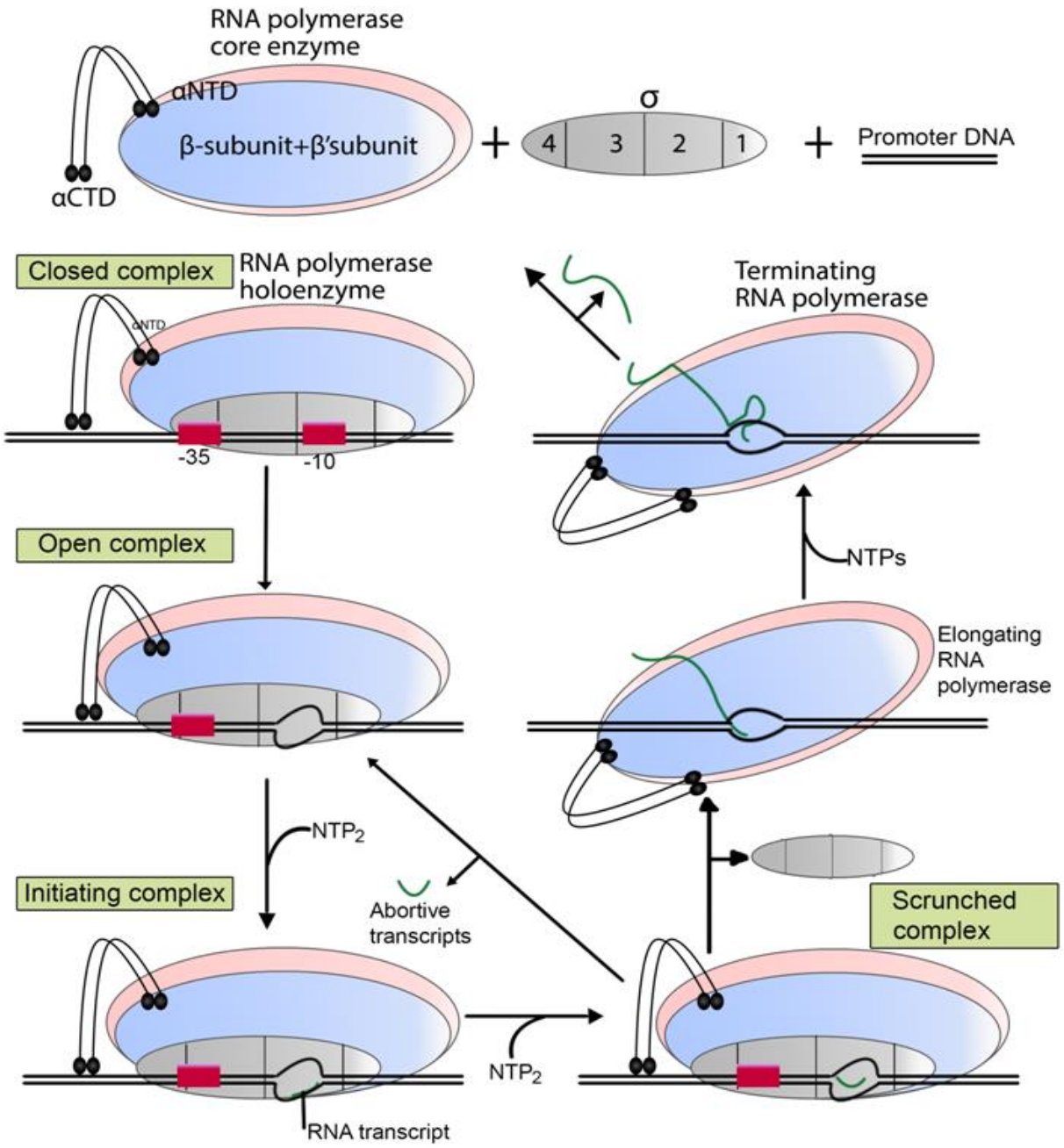


Figure 1.4: The transcription cycle. The process involves 4 steps: 1. Conversion of core RNAP into holoenzyme by binding a sigma factor. 2. Binding the promoter region forming the initiation complex. 3. Initial polymerisation by the formation of phosphodiester bonds. 4. Elongation. The nascent RNA is shown by a solid green line. Adapted from (Borukhov *et al.,* 2005, Browning and Busby, 2016).

1.3.2 CRP; a conventional global regulator

CRP has a regulon of around 200 genes consisting of many catabolic enzymes. In some pathogens, CRP also regulates various virulence factors. CRP is a 47 kDa molecule that binds DNA as a dimer (Figure 1.5) (Lawson *et al.,* 2004). The dimer of CRP contains two identical subunits (A and B) and each of the subunits is comprised of two domains: the N- terminal domain that is required for cAMP binding and dimerisation, and the C-terminal domain required for DNA binding (Busby and Ebright, 1999). Dimerisation requires a hydrophobic interaction between amino acids of an alpha helix in each N-terminal domain, known as the C helix. Precisely, the C, E and F helix residues (112-133, 168-176 and 180-191 respectively) form a helix-turn-helix motif that binds the DNA. The recognition helix from this motif is inserted into the major groove of the DNA where it contacts specific bases (Lawson *et al.,* 2004). 3'-5'- cyclic adenosine monophosphate (cAMP), binds CRP at a site in each monomer to control CRP activity. The cAMP binding site is formed by the C helix and β roll motifs consisting of β strands. Specific residues needed for cAMP binding are Arg-185, Arg-180, and Glu-181. cAMP binding brings the F helix into optimal position to interact with the DNA major groove. CRP-cAMP binds a consensus sequence TGTGAN_6TCACA which has two-fold symmetry (Lawson *et al.,* 2004). CRP binding to the DNA induces a 90° bend (Schultz *et al.,* 1991). CRP contacts RNAP via series of 3 activating regions (AR1-3) on the surface of the protein (Benoff *et al.,* 2002). AR1 is present in the CRP C-terminal domain and contacts αCTD of RNAP. AR2 is in the CRP N terminal domain contacts αNTD. AR3 also in the CRP N-terminal domain can contact σ^{70} (Busby and Ebright, 1999).

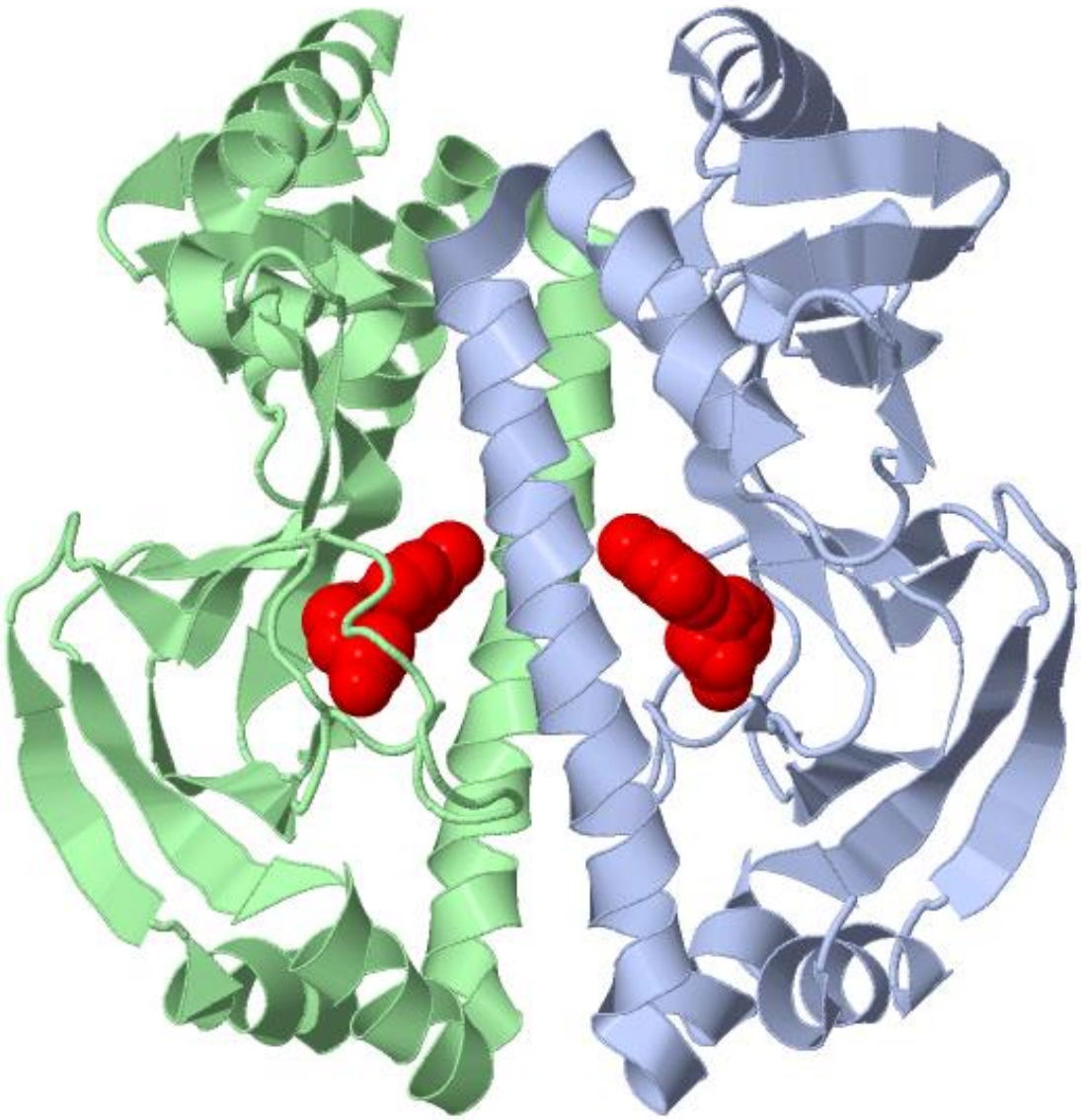


Figure 1.5: Ribbon structure of CRP dimer. The A subunit is in green and B subunit is in grey. The cAMP in red is bound to the 2 CRP monomers. (PDB 1G6N).

1.3.3 H-NS; an unconventional global regulator

Histone-like nucleoid structuring (H-NS) protein is an abundant (>20,000 copies a cell), extensively studied DNA binding protein found in many Gram negative bacteria. H-NS binds to around 5% of the genome in *E. coli* and these loci are often horizontally transferred (Ueda *et al.,* 2013). These genes can have an adverse effect on the host unless bound by H-NS. Hence, H-NS influences genome evolution and chromosome organisation as well as transcription (Rimsky, 2004, Winardhi *et al.,* 2014, Dorman, 2007, Dorman, 2014, Singh *et al.,* 2014, Lamberte *et al.,* 2017). H-NS binds AT-rich regions of DNA by recognising a nucleation site that is distinguished by the presence of a central T-A base step (Bhat *et al.,* 2014). Subsequently, H-NS oligomerises across and bends the double helix (Pon *et al.,* 2005). If H-NS forms distal clusters on the DNA, these can interact resulting in DNA bridging and genome condensation (Wang *et al.,* 2011). H-NS is a universal repressor of gene expression and is expressed in all growth phases of *E. coli* (Gallant *et al.,* 2004). Repression can involve trapping of RNAP in a DNA loop resulting from DNA bridging or complete exclusion of RNAP from the DNA (Dame *et al.,* 2002). In *E. coli,* the *proU* promoter is the best studied example for repression by H-NS (Nagarajavel *et al.,* 2007). In many enterobacteria, the function of H-NS is dependent on temperature and osmolarity (Yang *et al.,* 2005).

E. coli H-NS is a 137 amino acid protein comprised of 2 domains joined by a flexible linker (Rimsky, 2004). The N-terminal domain is 83 amino acids long and contains four α helices (α1, α2, α3 and α4). These permit self association by “head to head” or “tail to tail” contacts (Rimsky, 2004, Arold *et al.,* 2010). The helices α1- α3 form the head to head contact surface that contains a coiled coil made up of α1 and α2 (Esposito *et al.,*

2002). The C-terminal part of α3, and the α4 helix, is responsible for the tail to tail contact and forms a helix-turn-helix motif (Rimsky, 2004). The C-terminal domain of H-NS contains 2 β sheets (β1and β2), an α5 helix and a 3_{10} helix (Gordon *et al.,* 2010). A small loop between the β2 and α5 helix is the key determinant that binds DNA (Figure 1.6). The helix contains a side chain, R114, within a sequence of amino acids (TWTXGRXP) that forms an AT hook. The R114 side chain contacts the minor grove of the DNA (Gordon *et al.,* 2010). Head to head and tail to tail interaction results in filamentation. DNA bridging occurs if the AT hook motif in alternate promoters contact separate sections of DNA (Arold *et al.,* 2010).

1.3.4 AraC; a specific regulator

AraC is a regulator of L-arabinose metabolism. AraC has a C-terminal domain and an N-terminal domain. These domains are connected by a flexible linker (Mahon *et al.,* 2010, Seedorff and Schleif, 2011). The C-terminal domain is responsible for DNA binding and the N-terminal domain recognises L-arabinose.

Figure 1.7 shows the domain organisation of AraC. The broader AraC family is a large group of transcription factors; 830 family members have already been identified amongst diverse Gram negative bacteria (Egan, 2002, Tramonti *et al.,* 2002, Alekshun and Levy, 1999). Many of the family members are also specific regulators but some regulate larger sets of genes. AraC family members are identified by a 100 amino acid stretch that constitutes the DNA binding domain. The residues form two HTH motifs that bind the promoter DNA (Martin and Rosner, 2001, Grainger *et al.,* 2003, Rodgers and Schleif, 2009, Munson *et al.,* 2001, Morin *et al.,* 2010). If present, the N-terminal domain of most AraC family members is responsible for signal recognition and dimerization

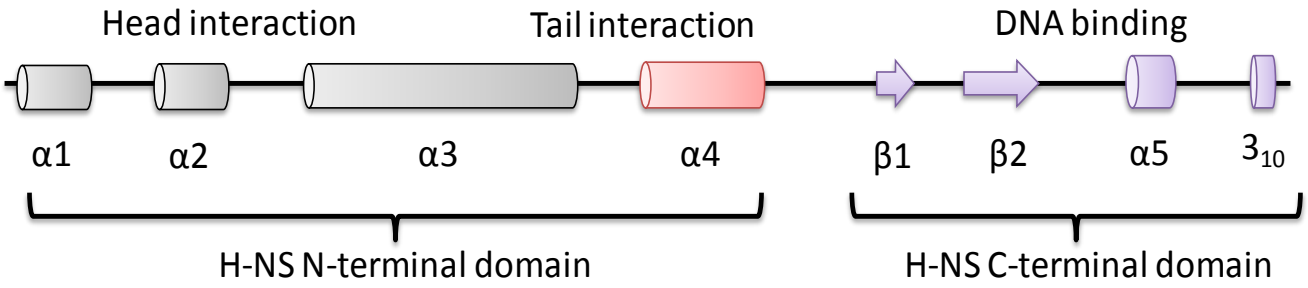


Figure 1.6: H-NS structural properties. The figure shows the 2 domains of H-NS. The NTD is composed of four α-helices (cylinders) and has 2 determinants for self association of H-NS. The determinants responsible for head-to-head interaction are shown in grey, and the one for tail-to-tail interaction is shown in red. The CTD of H-NS is the DNA binding domain composed of 2 β sheets, purple arrows, an α helix and a 3_{10} helix, shown as purple cylinders. Figure adapted from (Grainger, 2016).

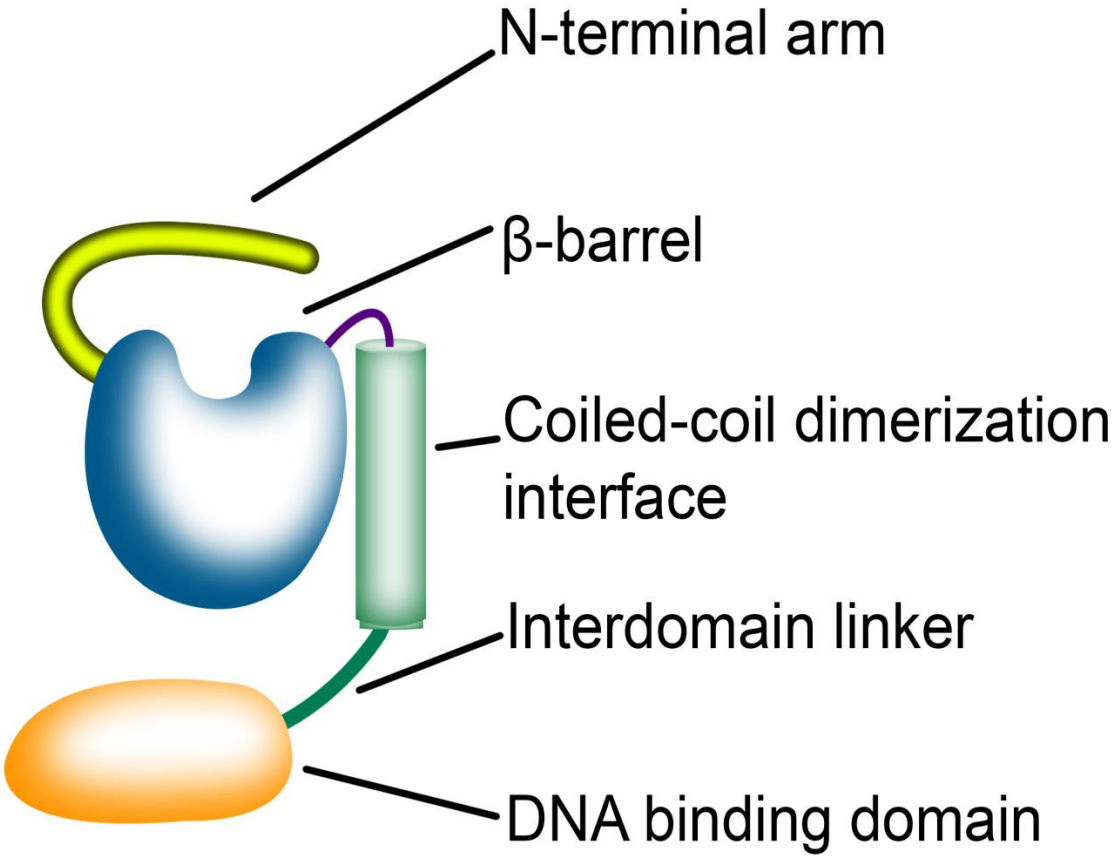


Figure 1.7: Domain organisation of single subunit of AraC. Strucural characteristics of a AraC homodimer are shown. Orange denotes the DNA binding domain, the N-terminal domain is in fluorescent green, the β - barrel is in blue, the coiled coil in light green, and the interdomain linker in dark green.

(Ruiz *et al.,* 2003, Parra and Collins, 2012).

1.4 Mechanisms of Transcription Activation

The basic rules for transcription activation have been defined at simple promoters regulated by a single transcription factor. There are 3 major classes of simple transcription activation defined on the basis of transcription factor binding location and contacts with RNAP (Busby and Ebright, 1999, Browning and Busby, 2004).

1.4.1 Class I activation: The transcription activator protein is bound on the DNA sequence upstream, and separated from the promoter -35 element. By the direct interaction between the activator and αCTD, the RNAP is recruited to the promoter (Chen *et al.,* 2003). One example is the *lac* promoter activated by CRP (Landis *et al.,* 1999). Note that, because the linker joining the αCTD and αNTD is flexible, αCTD can bind at different locations upstream of class I promoters. The residues in αCTD required at class I promoters are grouped and termed the 261, 265 and 268 determinants. The αCTD 265 determinant aids binding of αCTD to the minor groove of the DNA as described previously. Class I activation is more efficient if a promoter UP element is present (Busby and Ebright, 1999, Tang *et al.,* 1994, Gaal *et al.,* 1996). Figure 1.8A. illustrates class I activation.

1.4.2 Class II activation: Class II transcription activation, involves activator binding to a target that overlaps the promoter -35 element. The activator can contact both the sigma and alpha subunits of RNAP (Browning and Busby, 2004). An example of class II activation is the *E. coli*, *narG* promoter. In this example, αCTD, and domain 4 of σ^{70}, are contacted by FNR and this facilitates recruitment of RNAP (Blake *et al.,* 2002). Similarly, at the *galP1* promoter, RNAP is contacted by AR1 and AR2 of CRP. AR1

interacts with 287 determinant of αCTD and so stabilises UP element binding by the 265 deteminant of αCTD. AR2 binds the αNTD (residues 162 to 165) to further enhance promoter binding (Busby and Ebright, 1999). The Figure 1.8B illustrates the class II activation.

1.4.3 Class III activation: In this class of activation, the transcription activator is required to bind at multiple sites. Hence, class III promoters are activated by a combination of class I and class II mechanisms (Busby and Ebright, 1999, Browning and Busby, 2004). An illustration of activation by re-orientation of binding sites is shown in Figure 1.8C. An example for class III activation, provided by a mixture of mutiple factors, occurs at the *E. coli melAB* promoter (Grainger *et al.,* 2003). MelR binds two pairs of sites (1,1' and 2,2') as a dimer. These sites are separated by a CRP site (Grainger *et al.,* 2004b). When CRP is present in the cell, MelR binds to the 2' site more tightly and activates transcription by contacting σ^{70} (Grainger *et al.,* 2004a). CRP binding depends on sites 1 and 2 both being prebound by MelR. Hence, two signals (melibiose and cAMP) are integrated to regulate transcription (Wade *et al.,* 2001). Figure 1.8D. shows MelR type regulation. Complex activation can also involve modulation of promoter DNA conformation by an activator.

1.5 Transcription repression

Transcriptional repressors function via three major mechanisms (Browning and Busby, 2004).

1.5.1 Repression by steric hindrance: This mechanism prevents RNAP binding to the promoter. Hence, the promoter contains an operator binding site recognised by the repressor near the core promoter sequence. An example is the LacI repressor bound at

the *lac* promoter. LacI binds to additional operators to completely repress transcription of *lac ZYA* genes (Bell and Lewis, 2001, Wilson *et al.,* 2007). The Figure 1.9A. shows the repression by steric hindrance.

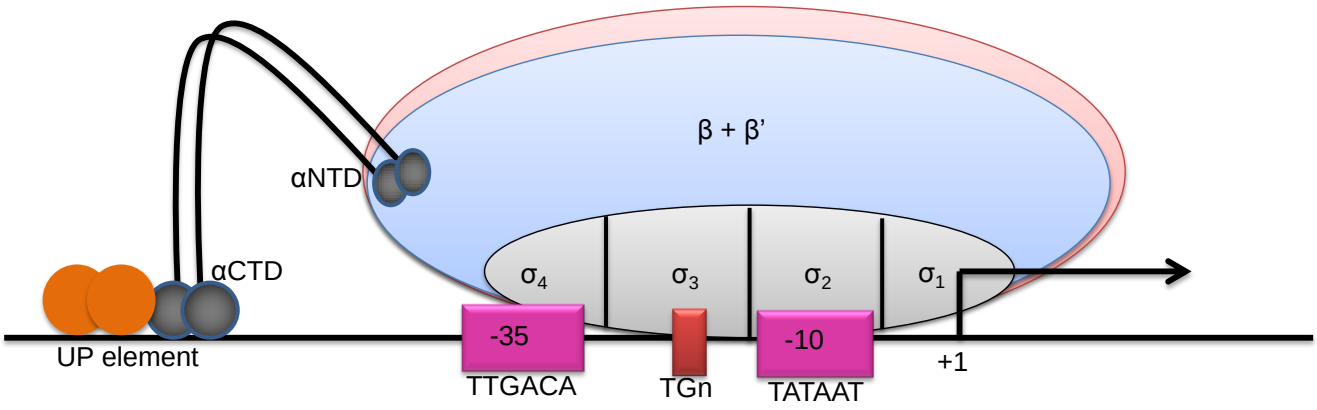
1.5.2 Repression by DNA looping: Transcription is blocked by looping of DNA so that RNAP cannot bind or escape the promoter. An example is the *gal* promoter, bound by GalR at 2 operators (O_E and O_I). Loops are formed by interactions between GalR molecules bound at distant sites (Lewis and Adhya, 2002). This interaction is weak and an additional factor, HU, stabilises the loop (Semsey *et al.,* 2002, Swint-Kruse and Matthews, 2009). Figure 1.9B. shows repression by looping.

1.5.3 Repression by modulation of an activator protein: Sometimes repressors act as anti-activators. This is evident for CytR (repressor) and CRP (activator) that interact to regulate genes involved in pyrimidine metabolism. CytR binds between 2 CRP dimers and blocks the RNAP αCTD access to the AR1 (Valentin-Hansen *et al.,* 1996, Shin *et al.,* 2001). Figure 1.9C. shows the repression by modulation of an activator protein.

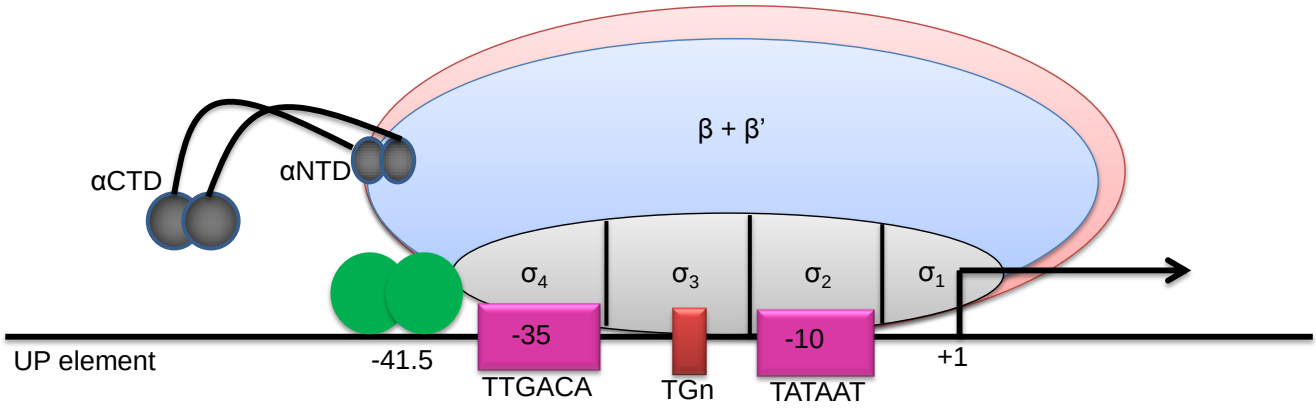
1.6 Transcription regulation by modulation of RNAP activity

Some regulators control transcription by binding to RNAP but not promoter DNA. Hence, these regulators can target both transcription initiation and elongation. For example, the *E. coli* transcription factors Nus (N-Utilization substance factor) A and G can control the speed of RNA chain elongation (Burns *et al.,* 1998) (Artsimovitch and Landick, 2000). NusG is able to couple transcription and translation by acting as a linker between the ribosomes and RNAP (Burmann *et al.,* 2010, Washburn and Gottesman, 2015).

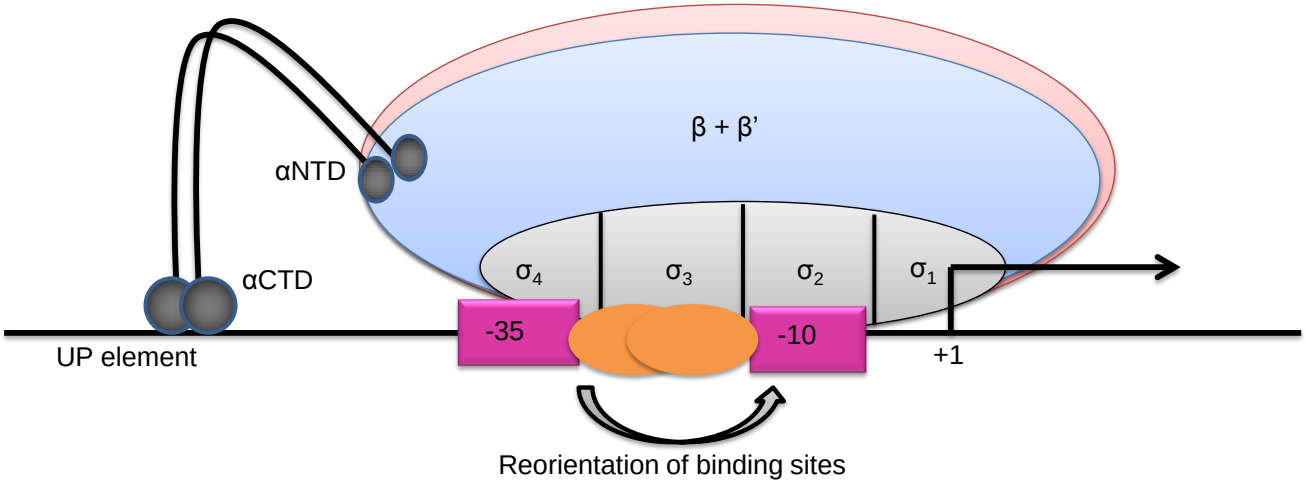
A Class I Activation



B Class II Activation



C Class III Activation



D MelR type regulation

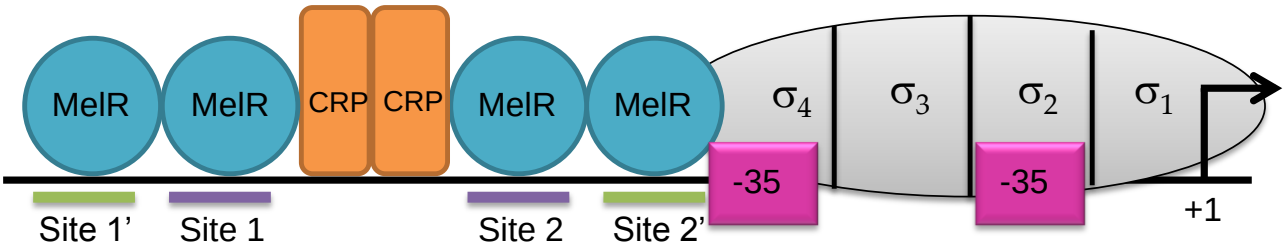
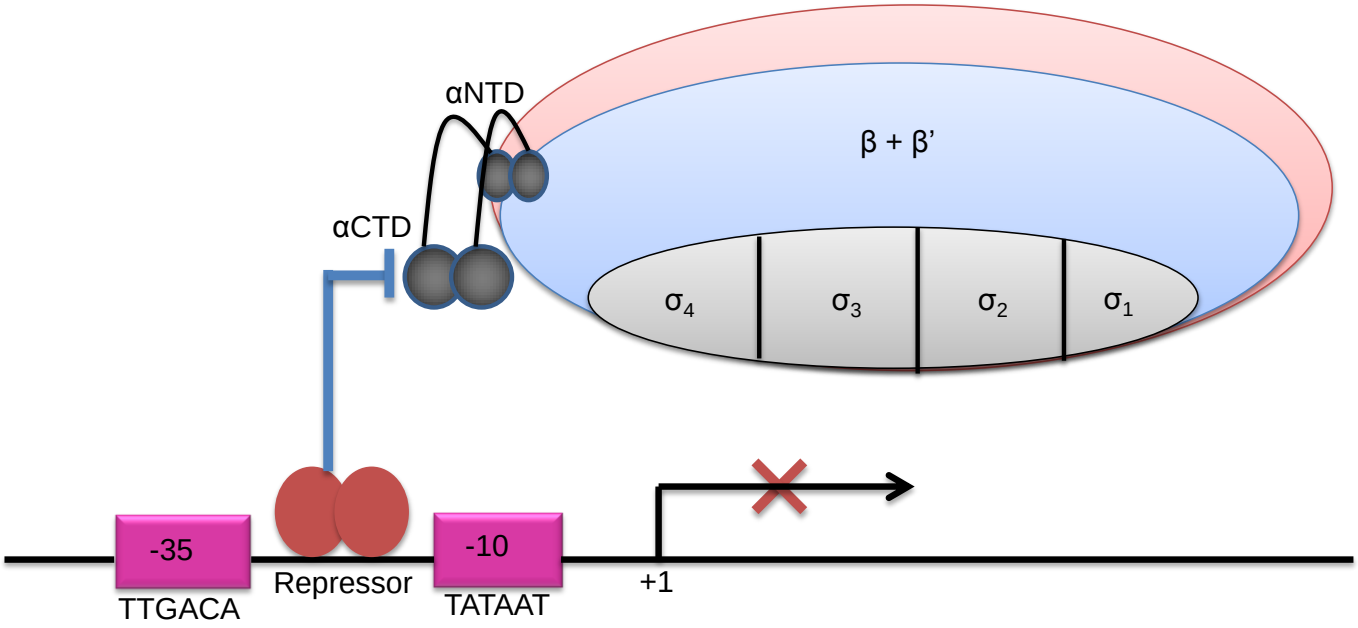
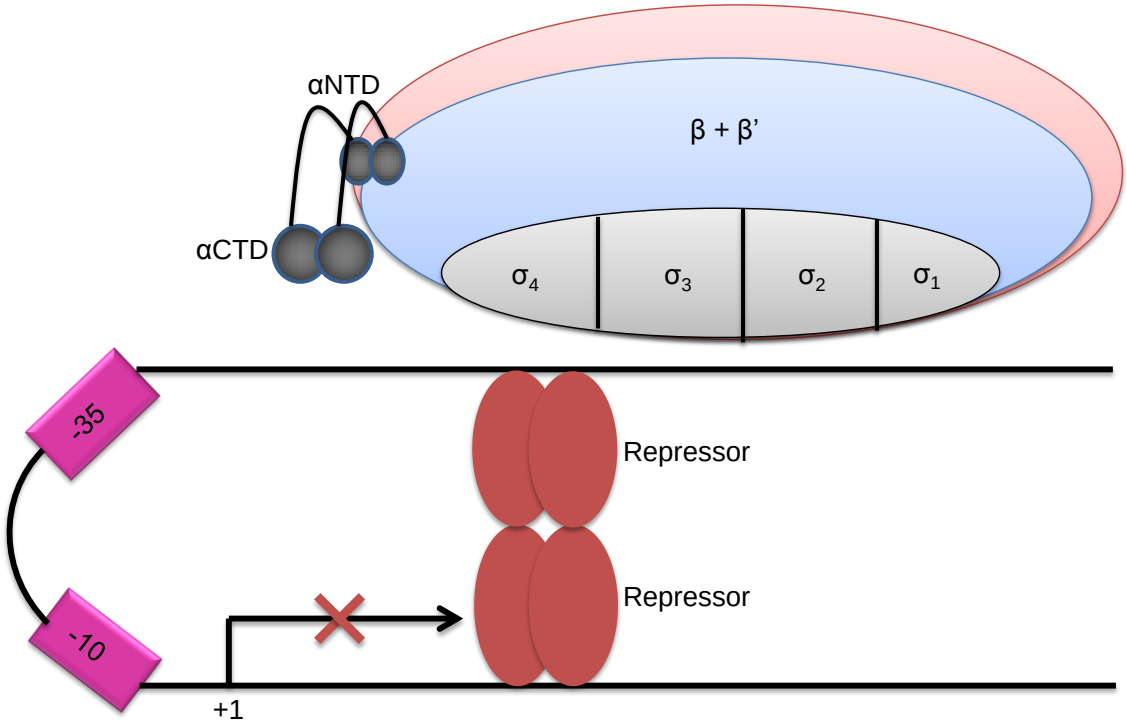


Figure 1.8: Cartoon representation of transcription activation mechanisms A. Class I activation: transcription factor binds upstream and recruits RNA polymerase by contacting αCTD. B. Class II activation: an activator is bound adjacent to the promoter and contacts domain 4 of sigma. C. Class III activation:the activator changes the conformation of the promoter. D. Co-operative binding of two activators; MelR (blue circles) and CRP (orange rectangles) drive transcription of *melAB* operon. MelR contacts domain four of σ^{70}. Blue circle represents MelR, CRP dimer is represented by orange rectangles grey oval represents the sigma. A, B and C were adapted from Browning & Busby, 2004. D, adapted from (Barnard *et al.,* 2004).

A Repression by steric hindrance



B Repression by looping



C Repression by modulation of an activator

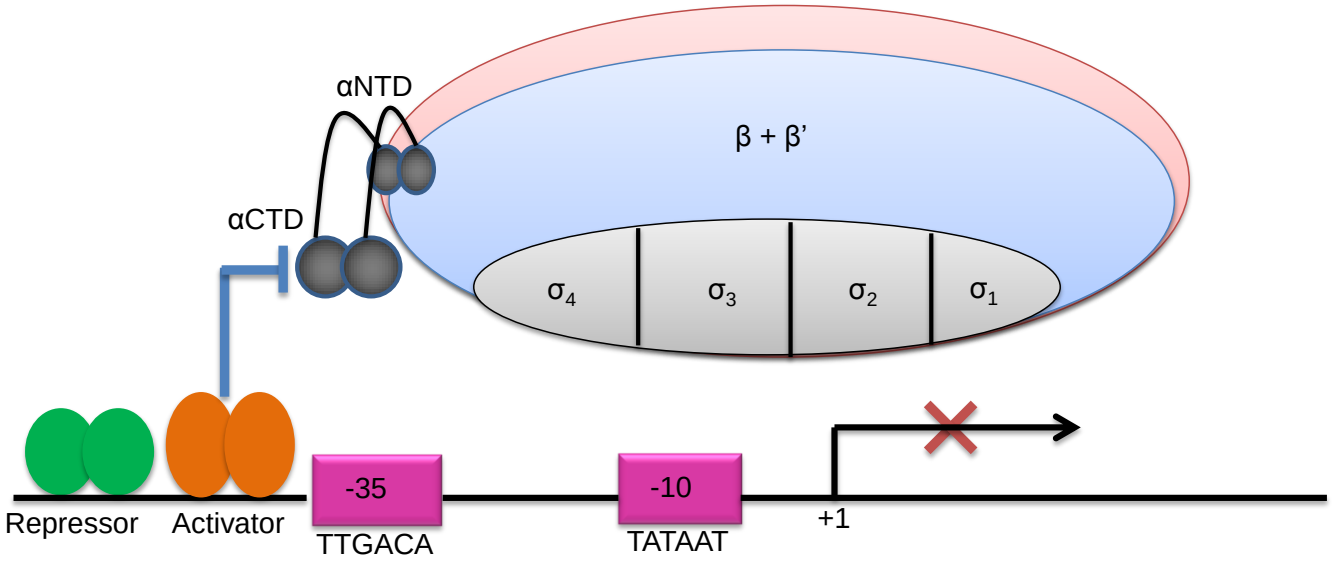


Figure 1.9: Cartoon representation of the mechanism of transcription repression. A. Repression by steric hindrance: the promoter site is blocked by a repressor and hence the recognition of the promoter by RNAPs doesn't occur. B. Repression by looping: the intervening promoters are repressed due to looping. C. Transcription repression by modulation of an activator: the repressor protein is bound to the activator, preventing it from functioning properly by cutting off the promoter recognition by the RNAPs holoenzyme (Adapted from Browning and Busby, 2004)

Not all transcription regulators are proteins. For example, in response to slow growth synthesis of 6S RNA, a 180 nt non-coding transcript, occurs. The 6S RNA forms a complex with σ^{70} and hence hinders binding to DNA (Figure 1.10) (Cavanagh and Wassarman, 2014). The 6S RNA achieves this by mimicking promoter DNA targets for housekeeping RNAPs. Hence, the amount of free σ^{38} increases relative to σ^{70} bound by 6s RNA in the cell. A change in overall transcription results (Cavanagh and Wassarman, 2014).

In *Actinomycetes,* RbpA and CarD stabilise the transcription initiation complex (Flentie *et al.,* 2016). Hence, domain 2 of σ^{70} is directly bound by RbpA in such a way that RbpA is positioned to make stable contacts with the upstream end of unwound DNA in the open complex (Hubin *et al.,* 2015). CarD and RbpA function equitably, whereby the C-terminal domain of CarD makes stabilising contacts that are complementary to RbpA, and the N-terminal domain of CarD makes contact with the RNAP β subunit (Bae *et al.,* 2015).

1.7 Transcription regulation by altering the preference of RNAP binding

Some regulatory factors remodel parts of RNAP to alter DNA binding preference at the promoter. A well documented example is T4 phage AsiA protein; a factor that appropriates *E. coli* RNAP to transcribe T4 phage genes (T4 doesn't code for its own RNAP). AsiA is expressed during the early part of the T4 life cycle. Once expressed, AsiA, binds and positions domain 4 of σ^{70} in such a way that it cannot bind housekeeping promoters (Lambert *et al.,* 2004, Gregory *et al.,* 2004). Instead, domain 4 of σ^{70} targets the T4 phage encoded protein known as MotA (Figure 1.11A). MotA is an activator expressed in the mid T4 life cycle that interacts with the C-terminal domain of

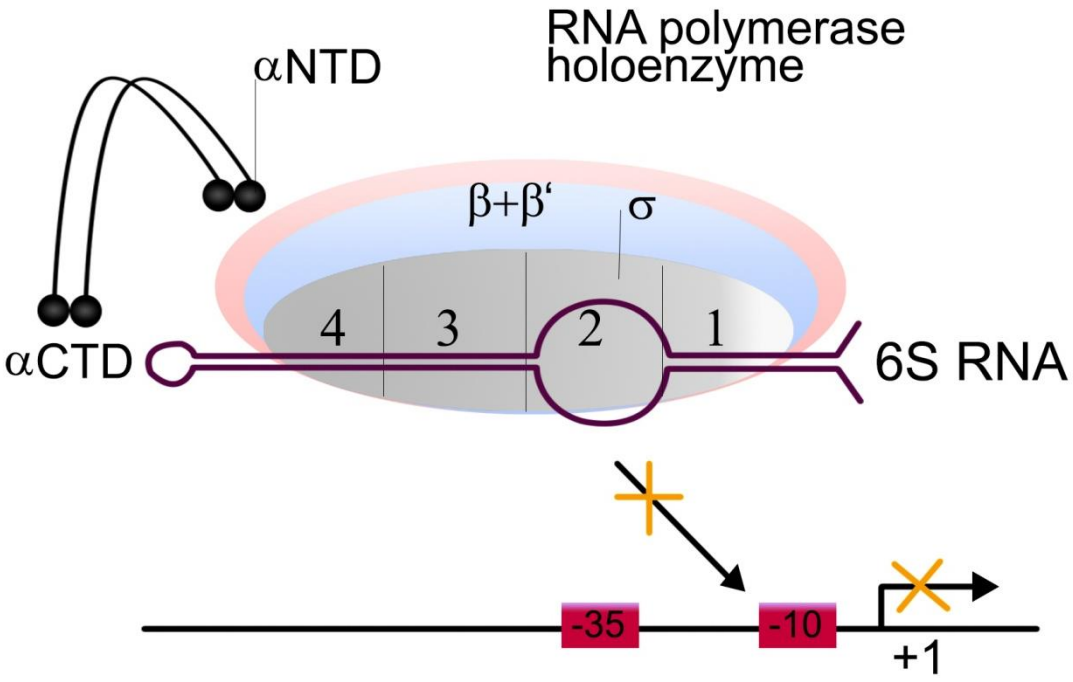


Figure 1.10: Cartoon representation to show the regulation by promoter mimic. 6S RNA (violet hair pin loop structure) mimics the promoter and hence can alter the binding of RNAP and of σ^{70} associated RNAP.

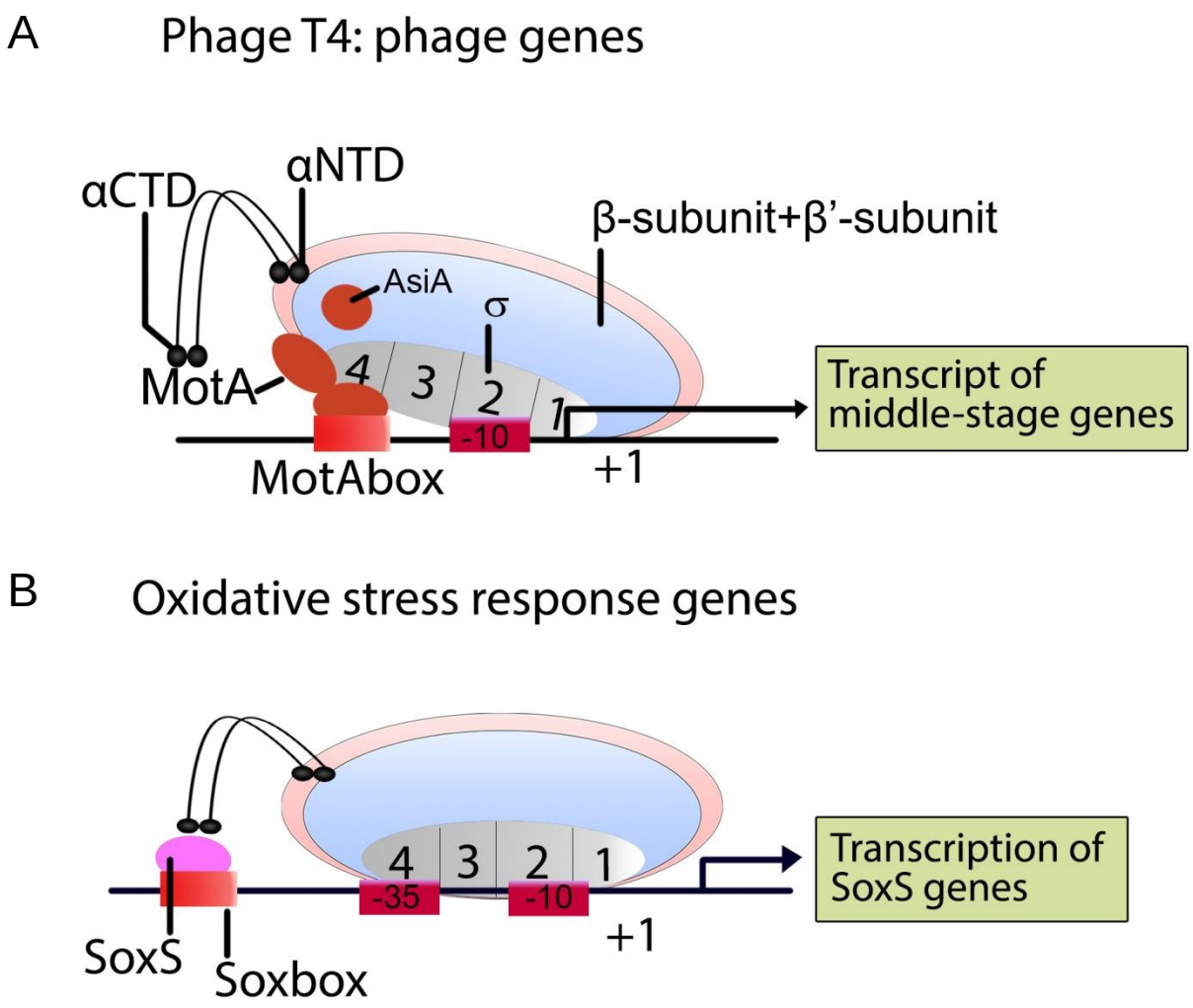


Figure 1.11: Cartoon representation showing that some factors can remodel parts of RNAP to alter A. DNA binding preference (AsiA / MotA) and others by B. RNAP appropriation (SoxS).

σ^{70}, which is normally bound by the β flap and the -35 element. In this way, σ^{70} is hindered from binding the host promoters after T4 invasion (Baxter *et al.,* 2006, Hinton, 2010). Some structural and biochemical studies shows that AsiA has a novel fold that helps in binding tightly with σ^{70} and remaining bound to the holoenzyme during the expression of phage genes (Urbauer *et al.,* 2002, Minakhin *et al.,* 2003).

There is evidence that some bacterial transcription factors work by RNAP appropriation. This is best documented for the *mar/sox/rob* group of regulators (Figure 1.11B). For instance, in response to the oxidative stress, the amount of a regulator called SoxS increases (Demple, 1996, Pomposiello and Demple, 2001). Although SoxS can bind a so called sox-box, which is present at Sox*S* regulated promoters. SoxS first makes a contact with αCTD of RNAP (Griffith *et al.,* 2002). After making this contact SoxS-RNAP can bind to SoxS controlled promoters with increased specificity (Shah and Wolf, 2004).

Similarly, the closely related protein MarA binds αCTD via the 265 determinant normally involved in αCTD:DNA interaction (Shah and Wolf, 2004). Thus, this complex has a preference for promoters with a MarA binding site rather than an UP element.

1.8 Diarrhoea and Enterotoxigenic *Escherichia coli* (ETEC)

The human gut is home to a vast diversity of bacterial species (Savage *et al*, 2016). These species include *E. coli* and, for a long time, this was thought to be the most prevalent bacterium in the intestine (Tenaillon *et al.,* 2010). However, *E. coli* make up only 0.1% of the total bacterial population in the intestine (Tenaillon *et al.,* 2010). The most dominant bacteria in the intestines are obligate anaerobes like *Bacteroides fragilis* (Berg, 1996). Whilst most *E. coli* are harmless, some *E. coli* acquire transposons, pathogenicity islands or other genetic elements that encode factors that can cause

disease (Croxen *et al.,* 2013). Hence, all *E. coli* share homology in the core genome but diverse virulence factors are encoded by pathogenic strains (Croxen *et al.,* 2013). Diarrhoeagenic *E. coli* (DEC) falls into 6 pathotypes; enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), enterohemmorhagic *E. coli* (EHEC), diffusely-adhering *E. coli* (DAEC) and enetroaggregative (EAEC) (Kaper *et al.,* 2004). Of these various pathogens, ETEC causes diseases that are more prevalent; 600 million episodes are reported every year resulting in 800,000 deaths (Qadri *et al.,* 2005). Children are most likely to die as a result of ETEC infection but visitors to areas where ETEC is endemic (e.g., Asia, Africa and Latin America) are also affected (Gaastra and Svennerholm, 1996, Shah *et al.,* 2009).

The prototypical ETEC strain, H10407, has a genome comprising a 5.1 Mbp chromosome and 4 plasmids (Figure 1.12). Two large plasmids (p948 and p666) encode a copy each of heat stable toxin (ST) and p666 encodes heat labile toxin (LT). The two small plasmids (p58 and p52) encode factors that might be responsible for maintaining stability of the larger plasmids (Crossman *et al.,* 2010, Liu *et al.,* 2015). Disease caused by ETEC and *Vibrio cholera* are often indistinguishable. The similarities between ETEC and *Vibrio cholera* mediated disease are due to the similar toxins produced during infection. For example, LT of ETEC and cholera toxin (CTX) of *Vibrio cholera* are 80% identical (Sanchez and Holmgren, 2005). Models for the synthesis and secretion of toxins by ETEC are shown in Figure 1.13. An overview of pathogenesis is shown in Figure 1.14. The role of major toxins is described in more detail below.

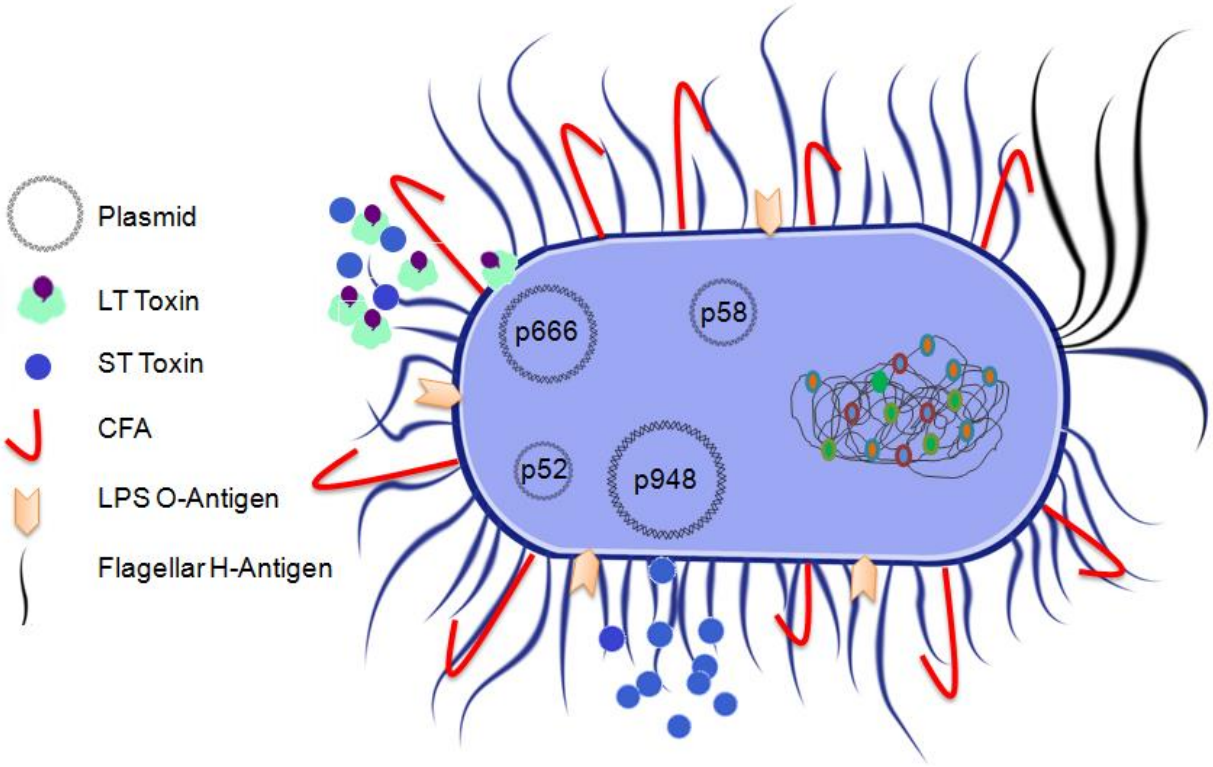


Figure 1.12: Representation of ETEC showing the genetic and virulent components. Cartoon of *E. coli* H10407 showing the chromosome bound by transcription factors. This figure also shows larger plasmids (p666 and p948) which carry the genes that encode the virulence factors and the smaller plasmids (p52and p58). Based on Crossman *et al.,* 2010.

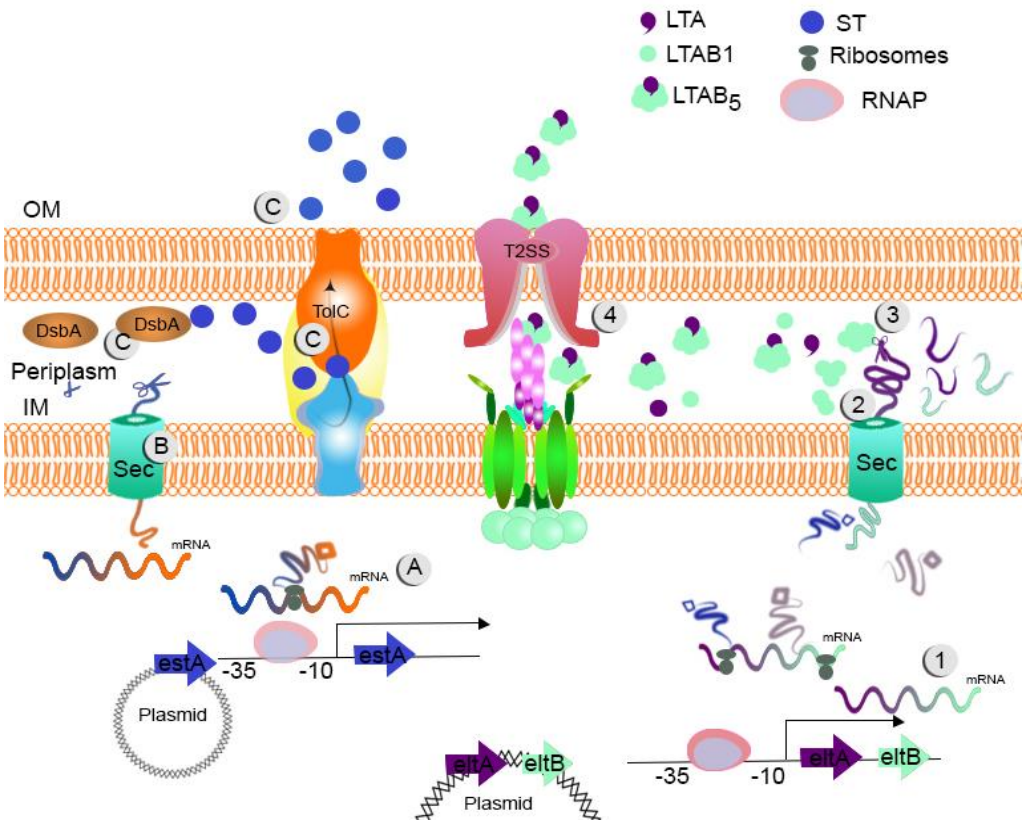


Figure 1.13: Synthesis and secretion of Heat Labile Toxin (LT) and Heat Stable Toxin (ST). For LT (steps 1-4): 1. LTA and LTB are synthesized (by *eltAB*) and transported from the cytoplasm to the periplasm with signal peptides. 2. Mature subunits undergo proteolysis of signal peptides. 3. The mature subunits are released in the periplasm and holotoxin starts to assemble. This assembly is mediated by LTB pentamerisation and it can even occur without LTA. However, LTA speeds up the process 3-fold. 4. The type 2 secretion system secretes the LT toxin. For ST (steps A-C): A. STa is synthesized (by *estA*) as a pre-pro toxin and the signal sequence is cleaved off by signal peptidases. B. STa is translocated into the periplasm by a SecA-dependent pathway and disulphide bonds start to form. C. DsbA cleaves 53 amino acid pro-STa; mature toxin is then secreted by the TolC channel. Figure based on Sears and Kaper, 1996

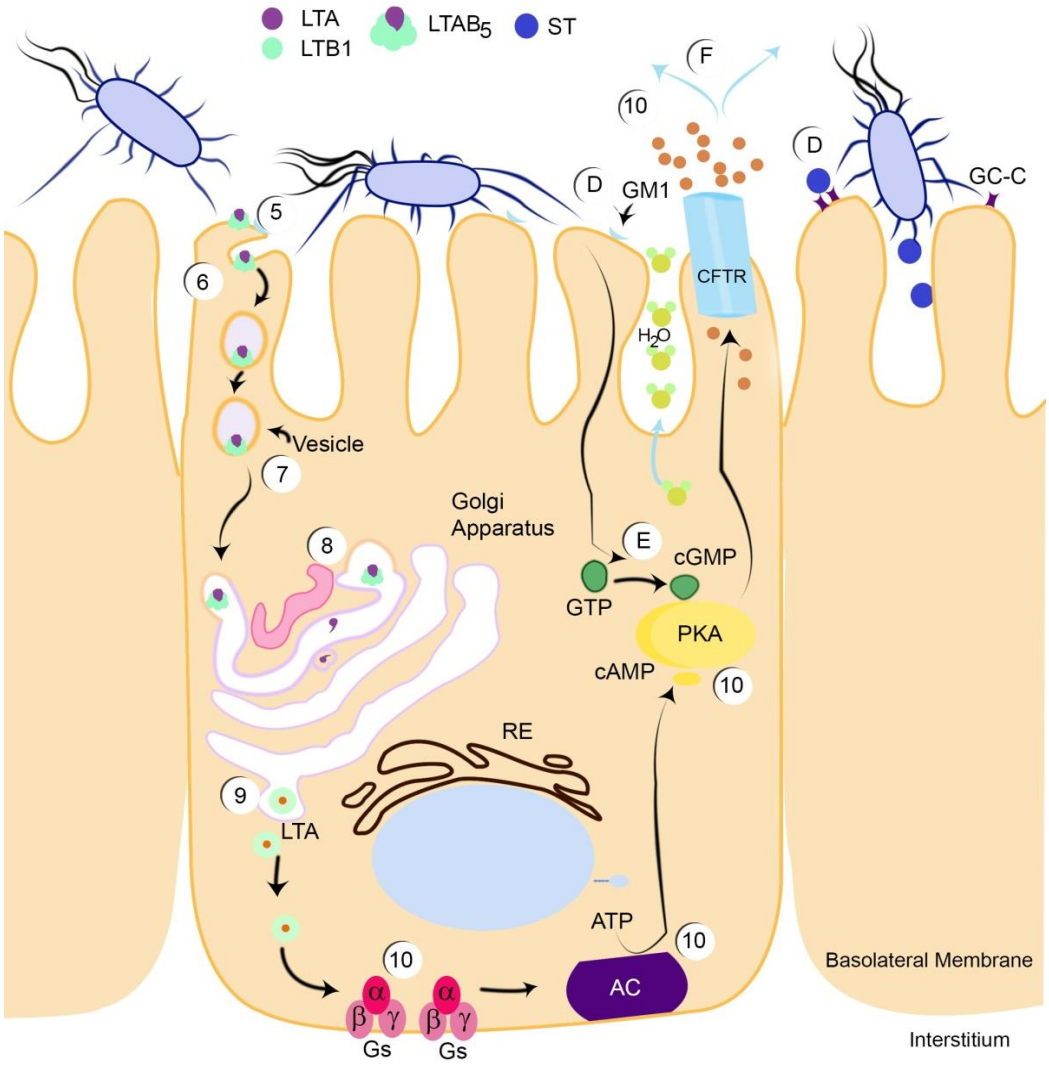


Figure 1.14: Representation of the model of pathogenesis in ETEC. For LT (steps 5 to 10): 5. Recognition and binding of LTB monomer to GM1 receptor. 6. Prior to activation the toxin gets internalized. 7. GM1 receptor associates with cell membrane lipids and mediates toxin endocytosis. 8. Transportation of LTA1 subunit to cytoplasm. 9. Toxin activation by nicking and reduction. 10. LTA1 activates G protein in Gsα that in turn activates the adenylate cyclase (AC). This results in production of cAMP leading to watery diarrhea. The ETEC binds to the brush border epithelium of the intestine by using the pathogenesis factors and various genes. For ST (Steps D to F): D. Recognition and binding of STa on epithelium and activation of GC-C receptor. E. cGMP starts to form and induces the CFTR membrane. F. Accumulation of water and ions in intestine leading to secretory diarrhea.

1.9 Heat labile LT toxins

LT is a hexameric toxin whose crystal structure was solved in 1991 (Sixma *et al.,* 1991). LT is an AB5 toxin consisting of 5 B subunits (11.8 kDa) and 1 A subunit (27.2 kDa). The AB5 toxins include pertussis, anthrax, and shiga toxin (Odumosu *et al.,* 2010). The two domains of the A subunit contain cysteine residues that can form disulphide bonds. The B subunit can bind to human GM-1 ganglioside, AB blood antigens and lipopolysaccharide (Sears and Kaper, 1996, Mudrak and Kuehn, 2010). The bacterial LeoA protein can also help to secrete LT by interacting with human Gs cells (Brown and Hardwidge, 2007). Thus, the B subunit and LeoA allow the A subunit (which has catalytic activity) to enter the cytoplasm of human epithelium with the help of the endoplasmic reticulum (ER). The structure of LT is shown in Figure 1.15. This toxin activates human adenylate cyclase and overproduction of cAMP results. Accumulation of cAMP activates protein kinase A, phosphorylating the cystic fibrosis transmembrane conductance regulator (CFTR) on the mucosal membrane. This results in loss of salt ions and water into the gut lumen (Sears and Kaper, 1996, Fleckenstein *et al.,* 2010). Previous study measured osmolality of less than equal to 100 mosmol/kg hypertonic and average glucose concentration of 0.4 to 24 mM in animals (Ferraris *et al.,* 1990). The *eltAB* operon encodes the LT toxin and expression requires a promoter upstream of the operon called P*eltAB*. H-NS represses P*eltAB* (Yang *et al.,* 2005) and, when ETEC cells enter the host, environmental factors trigger *eltAB* transcription (Sahl and Rasko, 2012). One such trigger may be the availability of glucose. Previously it was suggested that CRP binds directly to P*eltAB* to influence transcription but subsequent work suggested the effect was indirect (Haycocks *et al.,* 2015). Figure 1.16 shows the mechanism of *eltAB* operon regulation by glucose.

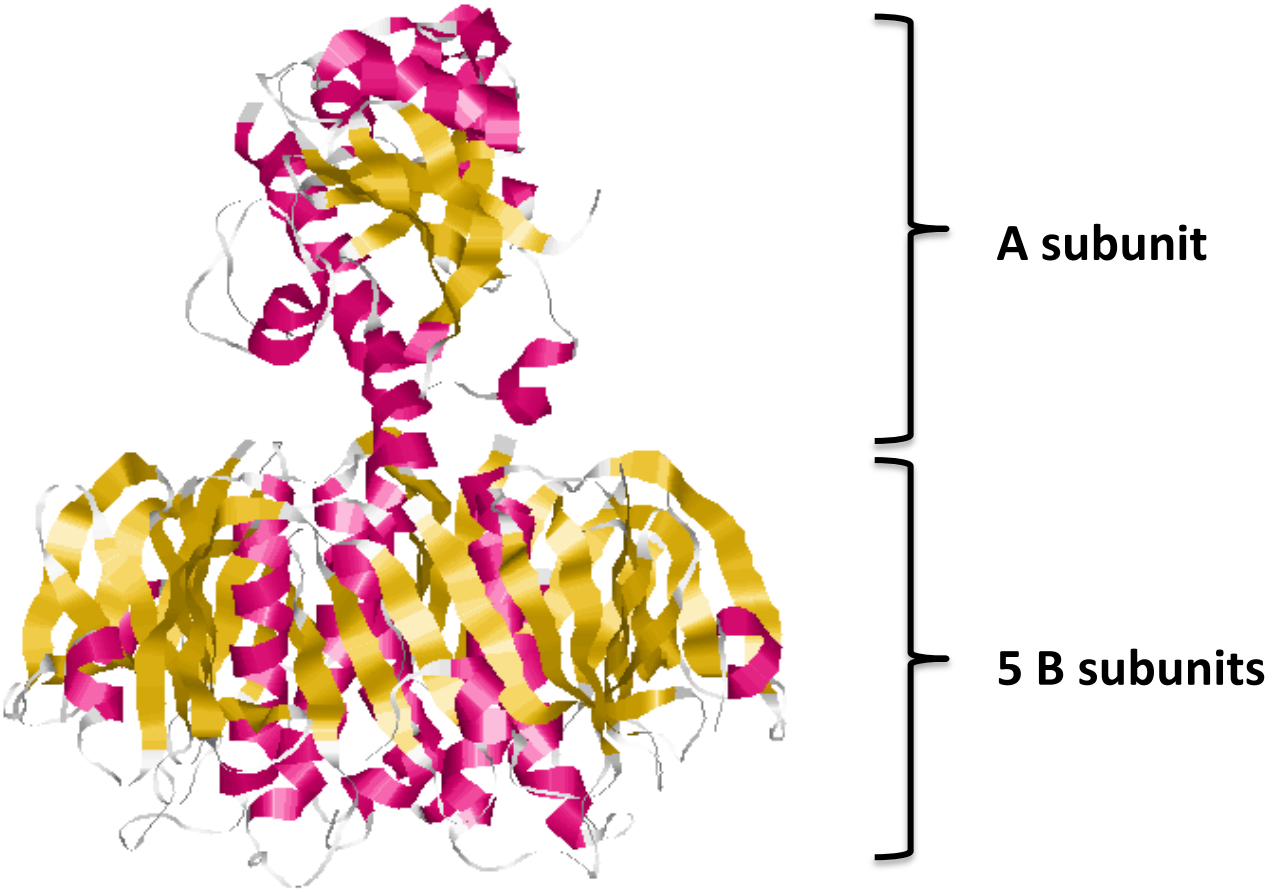


Figure 1.15: The structures of LT toxin. The ribbon structure of LT toxin consisting of 1 A subunit and 5 B subunits. The structure is obtained from Rasmol PDB viewer using the PDB file 1LTA.

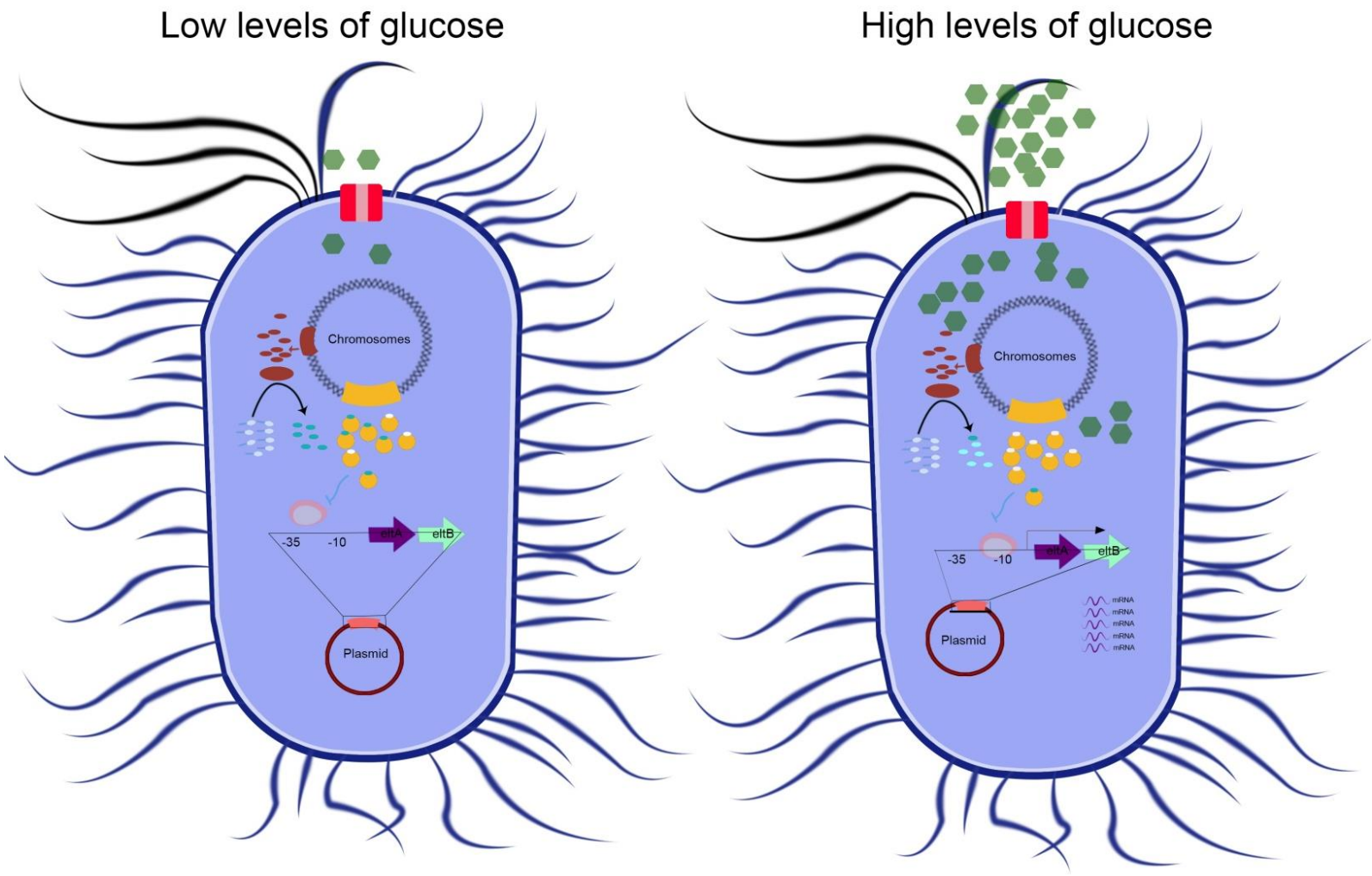


Figure 1.16: Regulation of *eltAB* by Glucose: bacterial cells shown in blue are grown in the presence of limited and adequate amount of glucose, the expression of *eltAB* is driven in the presence of adequate amount of glucose, seen in the top right figure. This is because the Phosphotransferase system (PTS) pathway helps in the activation of CRP which in turns activates the *eltAB.*

1.9.1 Secretion of LT toxin requires a Type 2 secretion system (T2SS)

The T2SS machinery comprises 13 proteins encoded by *gspC-O* operon (sometimes involving *gspA, gspB, gspS* and also the Sec and TAT components). The genome sequencing of ETEC H10407 revealed two operons, one having alpha (α) designated as T2SSα and the other having beta (β) designated as T2SSβ. T2SSα have roles in *in vivo* expression & activation while T2SSβ in LT secretion (Yang *et al.,* 2007). The cartoon structure of the proteins in the T2SS is shown in Figure 1.17.

The type 2 secretion system contains four important parts: 1) a periplasmic pseudopilus, 2) an inner membrane platform, 3) an outer membrane complex and 4) an ATPase present in the cytoplasm. The inner platform is formed by GspC, L, F and M. These proteins interact with the pseudopilus (GspG and other mini pseudopilins GspH, I, J and K), the GspD (secretin) and cytoplasmic ATPase GspE. ATP induced conformational changes cause extension of the pseudopilus. This change in the pseudopilus pushes proteins from the outer membrane channel (Korotkov *et al.,* 2012, Douzi *et al.,* 2012). T2SS substrate is regulated by GspC and is specified through the post synaptic drosophila/ zonula occludens-I protein (PDZ) and the homology domain.

GspC also act as a mediator, which helps to connect the outer membrane complex (consisting of GspD and forms a D complex) and the inner membrane complex (Gerard-Vincent *et al.,* 2002). Zinc containing secretion ATPase GspE, forms a hexamer on the inner membrane platform. The mechanism by which, GspE supplies power to T2SS is unknown. However, the energy required for the assembly or disassembly of the pseudopilus is thought to be derived by the hydrolysis of ATP. GspE interacts with GspL, which in turn interacts with GspG that forms the major pseudopilin

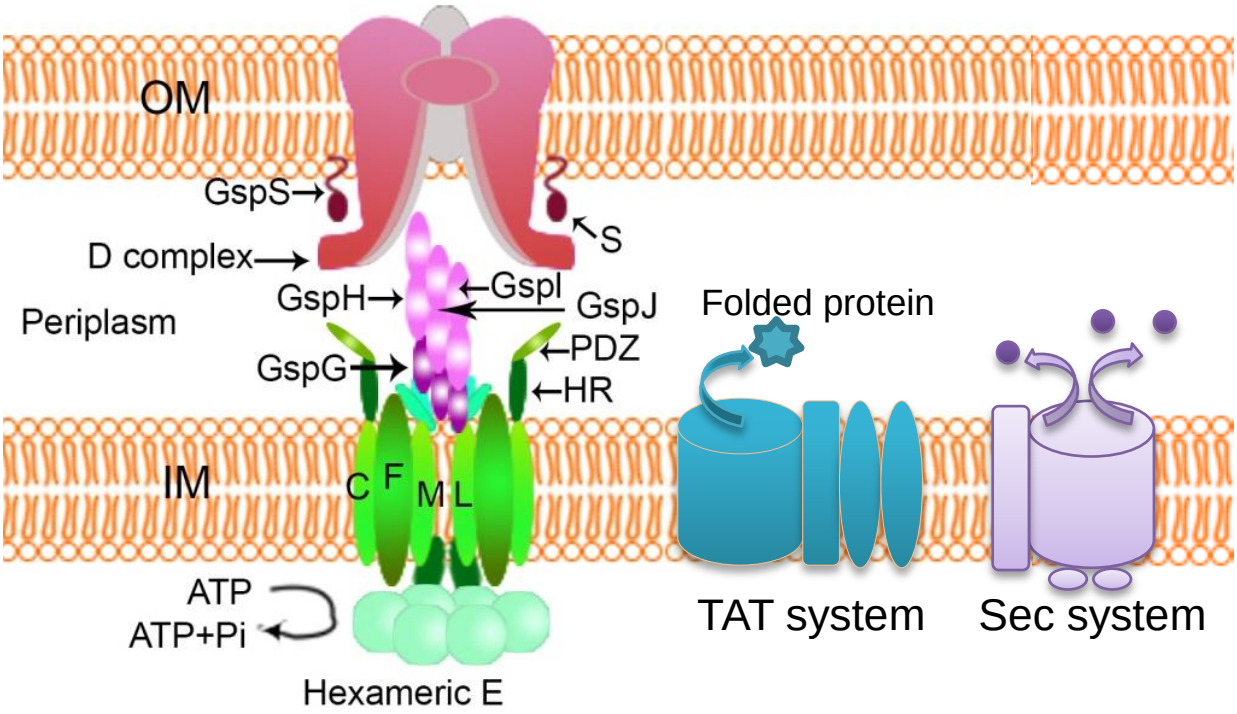


Figure 1.17: Type 2 secretion system that helps in secreting LT: T2SS is formed by the 3 major defined structures; the ATPase, light green; the periplasmic helical proteins, magenta; and the outer membrane complex embedded within the outer membrane, pink. This figure also shows the TAT and Sec components in blue and purple respectively.

(Gray *et al.,* 2011). T2SS cannot function in the absence of ATPase (Possot *et al.,* 2000).

The pseudopilus (inside the periplasm) is composed of 5 proteins: GspG, H, I, J and K. Since they share sequence identity to the type I pilin, it is termed a pseudopilus. An arrowhead is formed by 3 proteins; GspK, I and J. GspK interacts directly with the secretin (Forest, 2008). The outer membrane part of T2SS is composed of 12-14 subunits of GspD which is referred as secretin. The secretin functions as a pore (on the outer membrane) to secrete the LT toxin (Strozen *et al.,* 2012).

1.10 Heat stable ST toxins are small highly structured enterotoxins

ETEC H10407 encodes two variants of the heat stable toxin; STa/ST1 and STb/ST2. These toxins are cysteine rich small peptides of around 18-19 amino acids. The proteins are spiral in shape and held in this conformation by 3 disulphide bonds (Tauschek *et al.,* 2002, Myers *et al.,* 2013). ST toxins are non immunogenic (Fleckenstein *et al.,* 2010) and remain active at 95-105^{o}C even after one hour of boiling (Nair and Takeda, 1998, Weiglmeier *et al.,* 2010).

Initially expressed as a pre-pro peptide of 72 amino acids, the unprocessed ST toxin contains a signal sequence for sec-dependent transport. The signal sequence is removed in the periplasmic space to generate a pro-peptide. In the periplasm, DsbA, a disulphide bond isomerase, catalyses Cys-Cys bond formation in the pro-peptide. DsbA also cleaves the 53 amino acid pro-peptide and the mature toxin is then exported via TolC. Figures 1.13 and Figure 1.14 show the formation and translocation of ST toxin.

Inside the host ST increases the levels of cGMP by binding to the guanylate cyclase C receptor (Rao, 1985). This leads to phosphorylation of CFTR protein by CAMP dependent kinase (Chao *et al.,* 1994). Hence, increased sodium and chloride ions are expelled into the gut lumen. Figure 1.18 shows the structure of the toxic domain of heat stable toxins. The toxin resembles human hormones including guanylin, uroguanylin and lymphoguanylin (Forte *et al.,* 1999). Figure 1.19 shows an alignment of these hormones with bacterial ST toxin derivatives.

The two variants of ST in ETEC H10407 are encoded by different versions of the *estA* gene. However, nothing is known about regulation of heat stable toxin expression.

1.11 Colonizaton factors (CFs)

Gut colonization by ETEC requires colonization factors. These are plasmid encoded surface antigens (Fleckenstein *et al.,* 2010, Madhavan and Sakellaris, 2015). Based on morphology, the four main types of CF are designated as fimbrial, helical, afimbrial and fibrillar. The most commonly found CFs in ETEC are CFA/I, CS1-7, CS14, CS17 and CS21, however around 50% of the ETEC strains lack any of these CFs (Isidean *et al.,* 2011). Table 1.1 lists different ETEC colonization factors andtheir morphology. The genes encoding each CF are usually organised as an operon acquired by horizontal gene transfer. The operon typically encodes 4 genes; a periplasmic chaperon, a major fimbrial subunit, a minor subunit which is located at the fimbrial apex and outer membrane usher protein.

1.12 The Multiple Antibiotic Resistance operon of *E. coli*

E. coli strains, including ETEC H10407, encode the multiple antibiotic resistance operon, *marRAB* (Cohen *et al.,* 1993a). The first gene in the operon, *marR,* encodes an

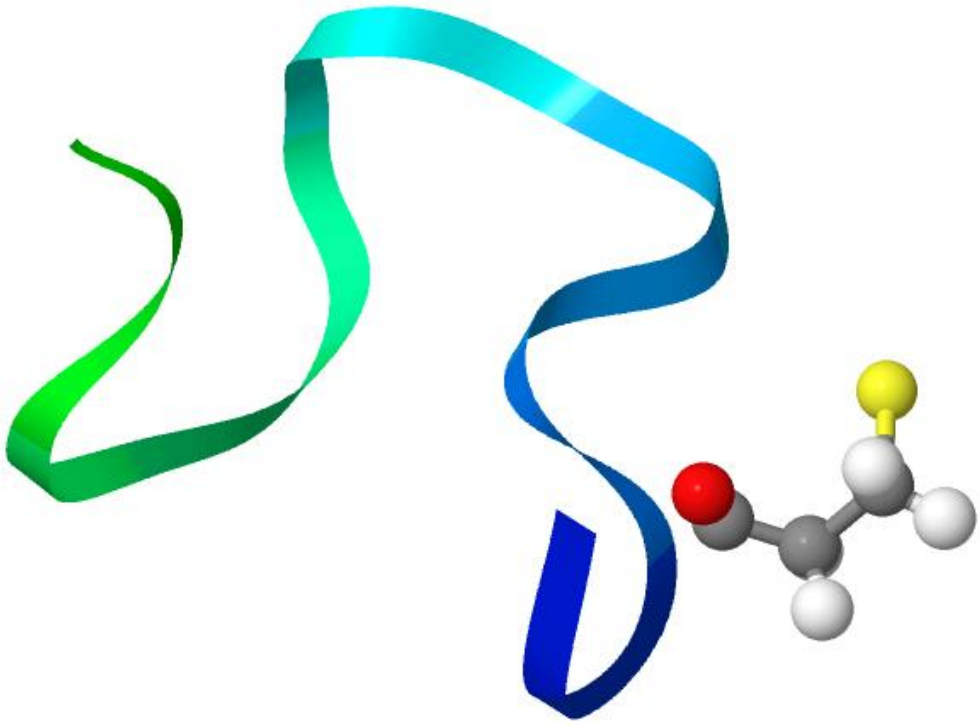


Figure 1.18: The structure of the toxic domain of ST toxin. Shown in ribbon is the biological assembly structure of the toxic domain of heat stable enterotoxin. The rainbow colour of the ribbon structure shows the different amino acids in the structure (PDB: 1ETN).

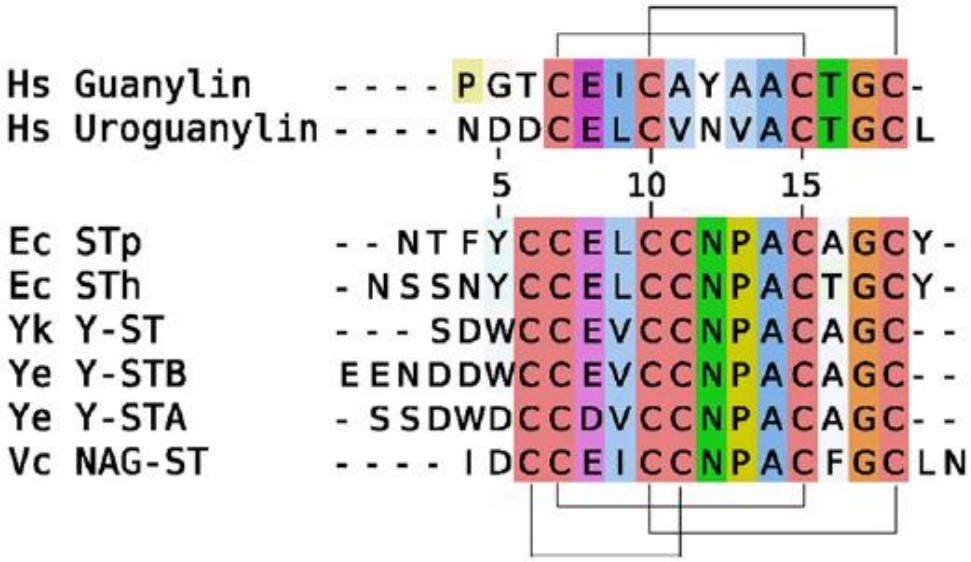


Figure 1.19: Sequence alignments showing the homology between the endogenous peptides and the bacterial toxins. Sequence alignment of human GC-C receptor ligands (first two sequences) and putative bacterial GC-C receptor binding ligands (last four sequences) using webPRANK. The brackets formed by black solid lines show the formation of the disulphide bond between the residues. The abbreviation labelled in this figure are as Hs, *Homo sapiens*; Ec, *Escherichia coli*; Yk, *Yersinia kristensenii*; Ye, *Yersinia enterocolitica*; Vc, *Vibrio cholerae*.

Table 1.1: Different classes of CFs: CFs are classified on the basis of morphology, where pink denotes fimbrial, yellow denotes fibrillar, blue denotes helical and green denotes afimbrial CFs. This table has been generated based on the study by (Anantha *et al.,* 2004).

Nature based on Morphology	Colonization Factors in ETEC
1) Fimbrial (Pilus) type	CFA/I
	CS1
	CS2
	CS4
	CS8
	CS12
	CS14
	CS17
	CS18
	CS19
	CS20
	CS21
	CS26
2) Fibrillar	CS3
	CS11
	CS13
	CS22
3) Helical	CS5
	CS7
4) Afimbrial	CS6
	CS10
	CS15
	CS23

autorepressor (Alekshun and Levy, 1999). Hence, MarR represses the *marRAB* operon by binding to the operator DNA sequence *marO* (Martin *et al.,* 1996). Transcription of *marRAB* can occur only when repression by MarR is disrupted. This can result from phenolic compounds, antibiotics, oxidative stress or mutation of *marR* or *marO* (Cohen *et al.,* 1993a). Derepression leads to expression of *marA* that encodes a global transcriptional activator. This leads to activation of genes coding for the AcrAB-TolC efflux pumps (Figure 1.20) (Okusu *et al.,* 1996). Thus, increased drug efflux leads to better antibiotic tolerance (Figure 1.21). The complete *mar* regulon (also known as *mar/sox/rob* regulon) is poorly defined. This is partly because the MarA homologs, SoxS and Rob, recognise the same DNA sequence as MarA and regulate transcription similarly (Martin and Rosner, 2001).

MarA, SoxS and Rob all belong to the AraC family (Egan, 2002). MarA and SoxS are small proteins, of 129 and 107 amino acids respectively, formed of a single domain containing 2 HTH motifs connected by a 27 Å rigid helix (Martin *et al.,* 1999, Martin *et al.,* 2002). Unlike most AraC family proteins, MarA and SoxS have no ligand sensing domain. Rob is a bigger 289 amino acid protein (Martin and Rosner, 2001) and differs from MarA and SoxS by virtue of a C-terminal domain that may be involved in ligand binding (Kwon *et al.,* 2000). Structural studies show that MarA binds adjacent sections of the DNA major groove as a monomer at a 20 bp consensus sequence referred to as the marbox (Figure 1.22A and Figure 1.22B) (Martin *et al.,* 2002, Rhee *et al.,* 1998). The marbox is asymmetric and highly degenerate. There are around 10,000 copies of the marbox in the *E. coli* chromosome but most are inactive (Martin *et al.,* 1999).

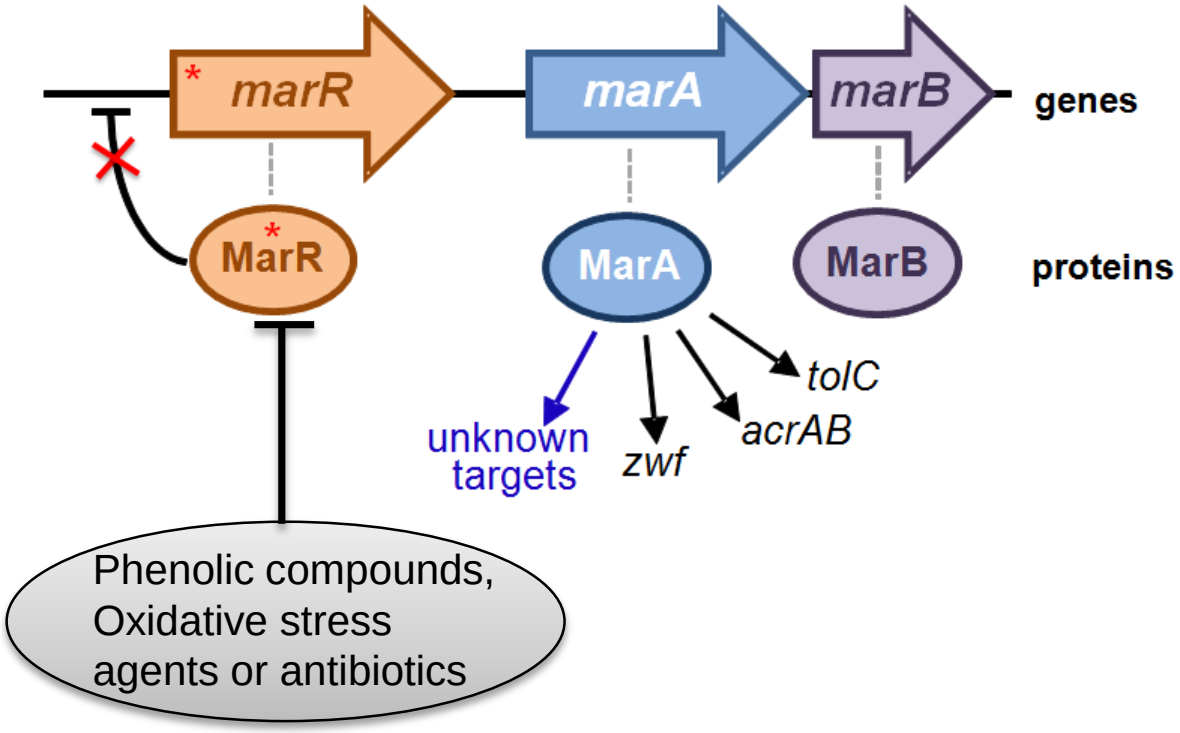


Figure 1.20: Cartoon structure to illustrate the activation of *marRAB*. Cells are stressed by the presence of phenolic compounds or antibiotics (grey oval) etc. This stress causes the repression activity of MarR (Orange) to be blocked and MarA (Blue) is produced, which in turn activates the downstream genes such as *zwf*, *acrAB*, *tolC* etc. However, there are some more unknown targets that MarA can activate.

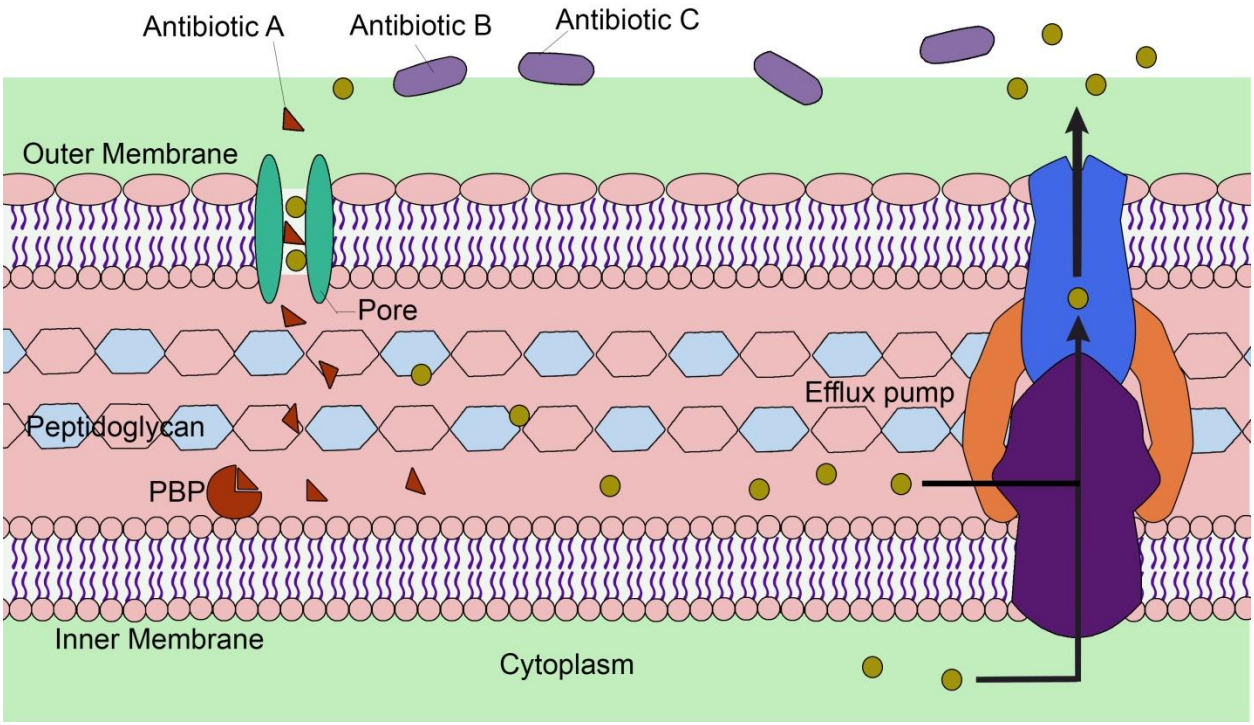
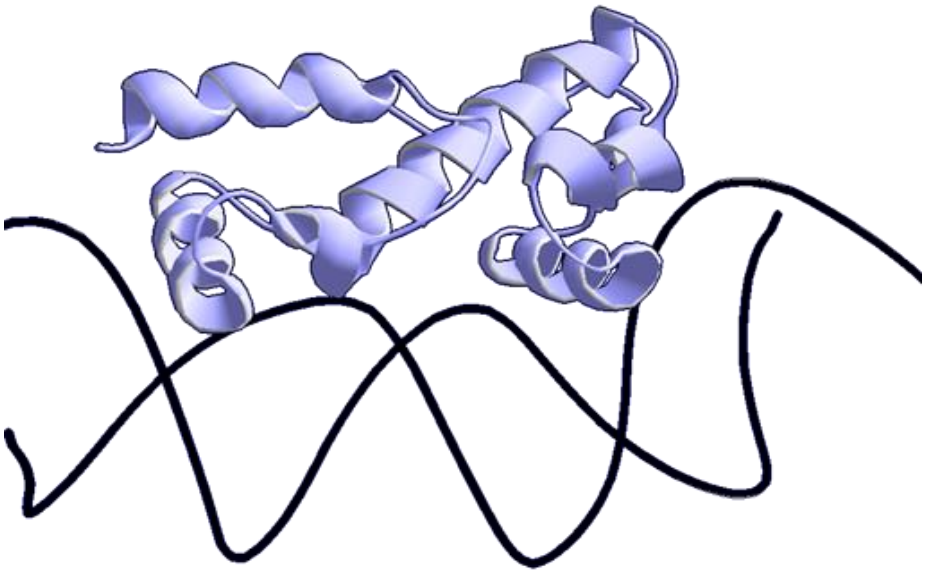


Figure 1.21: Cartoon representation of several instrinsic resistance mechanisms. β-lactam antibiotics targeting a penicillin-binding protein (PBP). Antibiotic A (red triangle) enters the cell and, after reaching its target, inhibits the synthesis of peptidoglycan. Some antibiotics (green circle) can also enter the cytoplasm but is removed by efflux. Antibiotic B and C (purple oval) is not able to reach the target since it cannot cross the outer membrane.

A



B

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

AHRGCACRWWNNRYYAAAHN

H= A,T or C
R= A or G
W= A or T
N= any nucleotide
Y= T or C

Figure 1.22: Structure and binding site of MarA: A) MarA (purple) binds the major grooves of DNA (black) by its DNA binding domain (generated via Rasmol by using PDB 1B10). B) The 20 bp consensus sequence bound by MarA, which is highly degenerate and asymmetric. Adapted from Rhee *et al.,*1998.

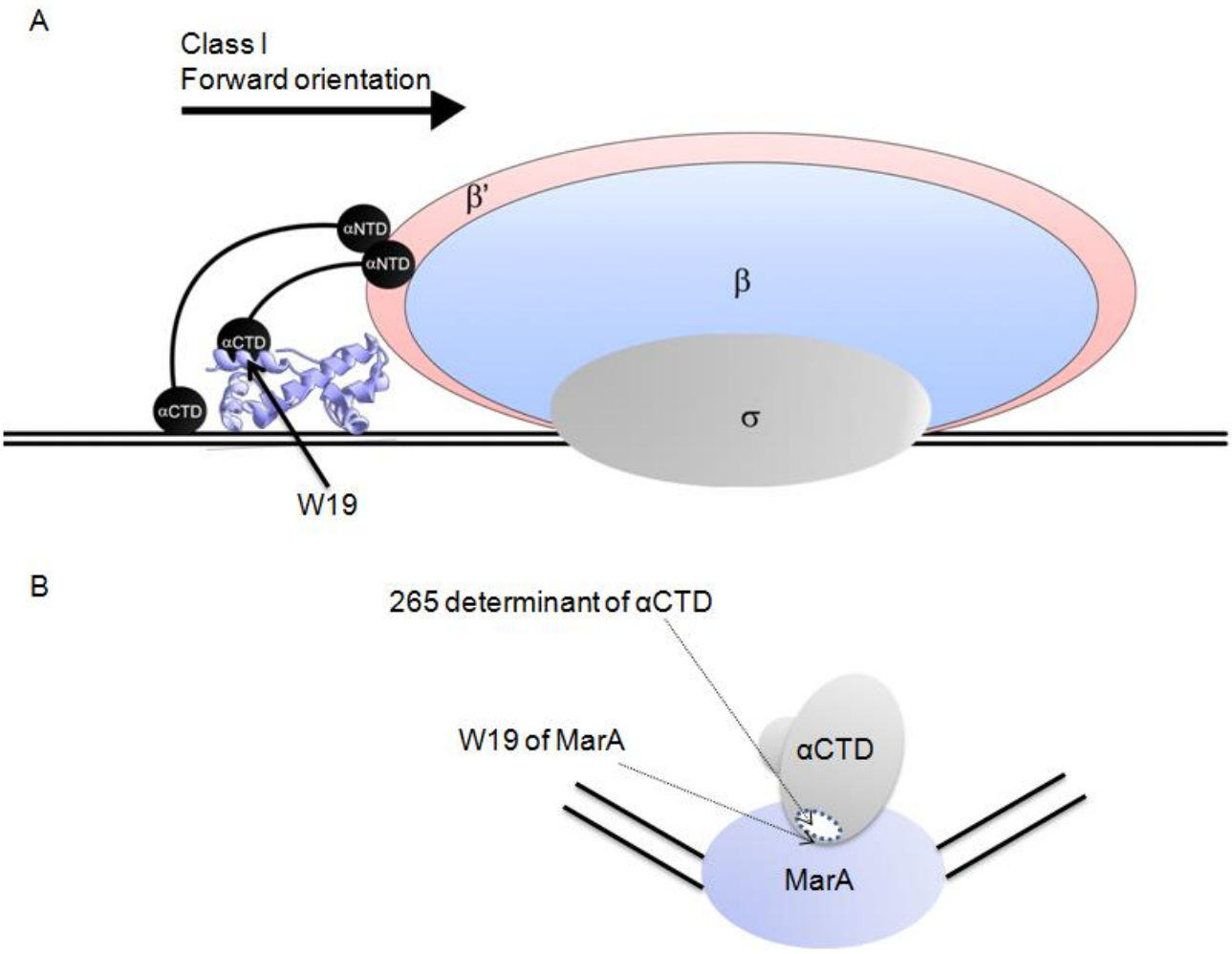


Figure 1.23: Cartoon representation of the interaction of MarA with αCTD of RNAP. A) The Class I *zwf* promoter is activated when MarA interacts with αCTD and recruits the RNAP to the promoter. On binding, the DNA bends by 35°. B) Zoomed in view of the interaction, where the 265 determinant of αCTD is bound to the W19 residue of MarA.

1.13 Objectives of this study

This work involves determining global gene expressions in ETEC H10407. Haycocks *et al.,* 2015 defined the global regulons of CRP and H-NS. They found that genes of major toxins in ETEC H10407 were not bound by CRP *in vivo*. However, previous literature showed that the promoter region of toxin genes was bound by CRP *in vitro.* This was also confirmed by Haycocks *et al.,* 2015 *in vitro*. Interestingly, they found that the toxin genes were bound by H-NS both *in vivo* and *in vitro.* Additionally, Myers *et al.,* 2013 showed an increase of FNR binding sites in Δ*hns* cells. Hence, there is a knowledge gap that whether or not H-NS binds CRP site *in vivo* and play a role in regulating the toxin expression.

Thus, the objectives for Chapter 3 are as follows:

1) To test that the CRP binding on major toxins *estA1, estA2 and eltAB* promoters is occluded by H-NS;
2) To test whether *estA1, estA2 and eltAB* promoters are repressed or activated in the presence of H-NS;
3) To test whether *estA1, estA2 and eltAB* promoters responds to glucose and salts and measure the effect; and
4) To test whether the regulation of major toxin genes (*estA2 and eltAB*) is conserved in other ETEC isolates.

Moreover, *E. coli* strains encode MarRAB that is well implicated in antibiotic resistance. MarA, which is the activator of the major efflux pump AcrAB-TolC, maintains multiple

antibiotic resistance. There is very little known about the regulon and the genes that are regulated by MarA. Hence, we made the following objectives for Chapter 4, 5 and 6:

1) To determine the binding of MarA across whole genome of ETEC H10407 *in vivo* by ChIP-Seq and confirm the binding sites *in vitro* by gel shift assays; and
2) To determine the regulation of some identified targets of MarA; and
3) To determine the mechanisms additional to AcrAB-TolC efflux pump that produces a multiple antibiotic resistance phenotype.

Chapter 2. Materials and Methods

2.1 General antibiotics and growth media used

Sigma-Aldrich Co. Ltd., Fisher Scientific, VWR and Bioline supplied all chemicals. Radioactive nucleotides and Phosphor imager screens were supplied by MP biochemical, Perkin Elmer and Biorad. To prevent decontamination, solutions were sterilised by autoclaving for 15-20 minutes at 121°C as necessary.

2.1.1 Antibiotics (Stock Solutions)

- **Ampicillin**: 100 mg/ml in distilled H_2O (dH_2O) stored at -20°C
- **Tetracycline**: 35 mg/ml in Methanol stored at -20°C
- **Kanamycin**: 50 mg/ml in dH_2O stored at -20°C
- **Doxycycline:** 50 mg/ml in dH_2O stored at -20°C
- **Ciprofloxacin:** 50 mg/ml in dH_2O (acetic acid was added until dissolved) stored at -20°C

Liquid and solid growth media were autoclaved without the antibiotics. The antibiotics were added to media 1:1000.

2.1.2 Bacterial Growth Media

2.1.2.1 Solid media

Solid media was used for plating of bacterial cultures and for short-term storage of strains.

2.1.2.1.1 Reagents and chemicals used:

- **MacConkey-lactose agar**: 52 g/l from Oxoid
- **Nutrient agar** : 20 g/l (Sigma)

2.1.2.1.2 Method:

Solid media was prepared by dissolving the powder in water according to the manufacturer's instructions, followed by autoclaving. After cooling, molten agar was mixed with appropriate concentration of antibiotics and the mixture was poured into a Petri dish next to a lit Bunsen burner. After pouring, plates were dried at room temperature overnight.

MacConkey-lactose agar was used to monitor cell phenotypes (i.e. Lac+ or Lac-) and nutrient agar was used for all other purposes.

2.1.2.2 Liquid media

2.1.2.2.1 Reagents and chemicals used:

- **Lennox-Broth media (LB)**: 20 g/l tryptone, 10 g/l yeast, 10 g/l NaCl
- **M9 minimal media 10 X salts:** 60 g Na_2HPO_4, 30 g KH_2PO_4,5 g NaCl,10 g NH_4Cl in 1 L ddH_2O. pH maintained to 7.2 using NaOH.
- **M9 minimal media:** 10 ml 10 X M9 Salts, 90 ml Sterile ddH_2O, 200 µl 1 M $MgSO_4$,100 µl 0.1 M $CaCl_2$, 500 µl 20% (w/v) Casamino acids, 1.5 ml 20% (w/v) Carbon source.

LB broth and M9 minimal medium was used for liquid bacterial cultures. For selection appropriate antibiotics were used to supplement the growth media.

2.2 Bacterial Strains and Plasmids

2.2.1 Bacterial strains and growth conditions

Strains used in this study are listed in Table 2.1. All the strains were stored at -80°C. On removal, a small amount of frozen culture was used to inoculate 5-10 ml of liquid medium and incubated at 37°C in a shaking incubator overnight. To monitor growth of liquid cultures, optical density (OD) at 650 nm was measured in a Jenway 6300 Spectrophotometer (Thermo Fisher Scientific Inc.).

2.2.2 Plasmids

Plasmids used in this study are listed in Table 2.2. Storage of these plasmids was at -20°C in dH_2O. Plasmid maps are shown in Figures 2.1 to 2.3.

2.3 Plasmid DNA preparation using QIAprep Spin Mini/Maxi prep kit (QIAgen)

LB (5 ml) supplemented with appropriate antibiotics was used to culture strains carrying the desired plasmid DNA. DNA was extracted according to the manufacturer's instructions. Plasmid DNA was eluted in 30 µl sterile dH_2O.

Table 2.1: Strains used in the study. The genotype and the source for each strain is depicted in the table. All strains are K-12 derivatives (until stated).

Strain	Genotype	Source
***E. coli* JCB387**	Δ*nir Δlac*	(Typas and Hengge, 2006)
***E. coli* M182**	*Δlac galK galU strA*	(Busby *et al.,* 1983)
M182Δ*hns*	*Δhns* Kan^R	This study
M182Δ*crp*	MI182Δ*crp39*	(Busby *et al.,* 1983)
***E. coli* MG1655**	$F^-λ^-$rph-1	(Blattner *et al.,* 1997)
***E. coli* BW25113**	rrnB3 Δ*lacZ*4787 hsdR514 Δ*(araBAD)567* Δ*(rhaBAD)568 rph-1*	(Baba *et al.,* 2006)
E. coli* BW25113 *xseA::kan	Derivative of BW25113	(Baba *et al.,* 2006)
E. coli* BW25113 *yrbE::kan	Derivative of BW25113	(Baba *et al.,* 2006)
***E. coli* NCTC 10418**	Clinical isolate	National Collection of Type Cultures
***E. coli* ATCC 25922**	Clinical isolate	American Type Culture Collection
ETEC H10407	Isolated in Dhaka in 1971, Bangladesh, from an adult with cholera-like symptoms	(Sack *et al.,* 1971, Crossman *et al.,* 2010)

Table 2.2: Plasmids used in the study. Plasmids with features and source are depicted in the table.

Plasmid	Feature	Source
pRW50	A low copy number, 17 kb plasmid with *Eco*RI and *Hin*dIII sites used for cloning, promoters upstream of the *lacZ* gene. Encodes Tc^R and has the RK2 origin of replication.	(Lodge *et al.,* 1992)
pSR	A high copy number, 2.6 kb plasmid derived from pBR322. Has *an Eco*RI *and Hin*dIII *site* just upstream of the *λoop* transcription terminator. Encodes for Amp^R and has CoIE1 origin for replication.	(Kolb *et al.,* 1995)
pBR322	A medium copy number, 4.4 kb plasmid. It has a *bla* gene that encodes for AmpR and has pMB1origin of replication.	(Balbas *et al.,* 1986)
pRGM9818	*marA* over expression plasmid, pUC19 derivative encoding *marA* under the control of the *tac* promoter.	(Martin *et al.,* 2000)

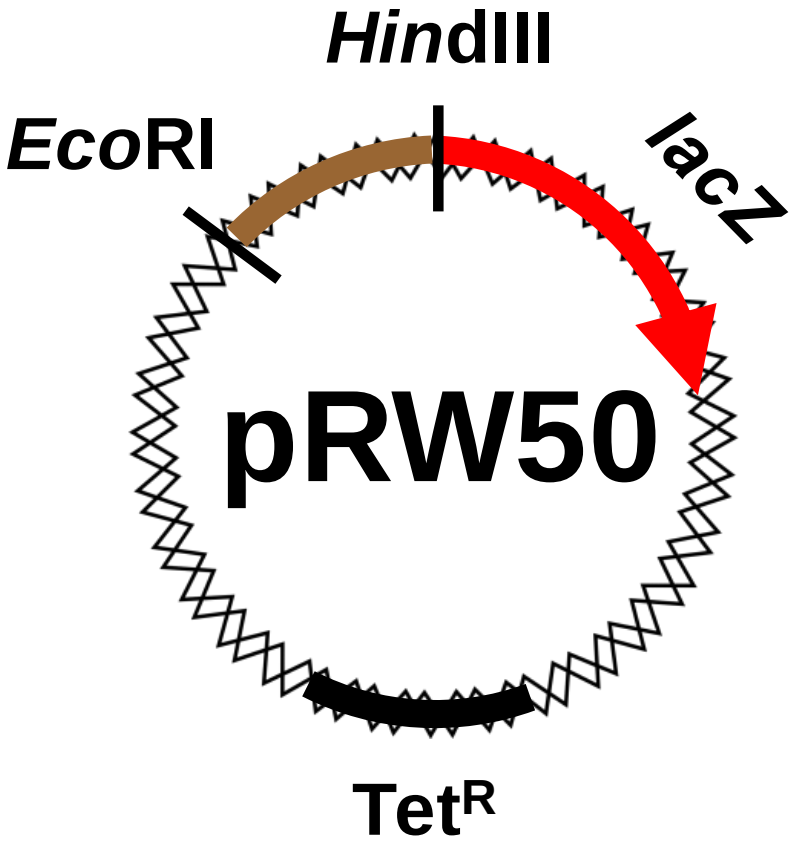


Figure 2.1: pRW50 plasmid map. pRW50 was used as a reporter plasmid in LacZ assays. The brown bar represents the site where promoter DNA was cloned using the *EcoR*1 and *Hind*III sites. The *lacZ* gene is shown in red, and the black bar denotes the tetracycline resistance.

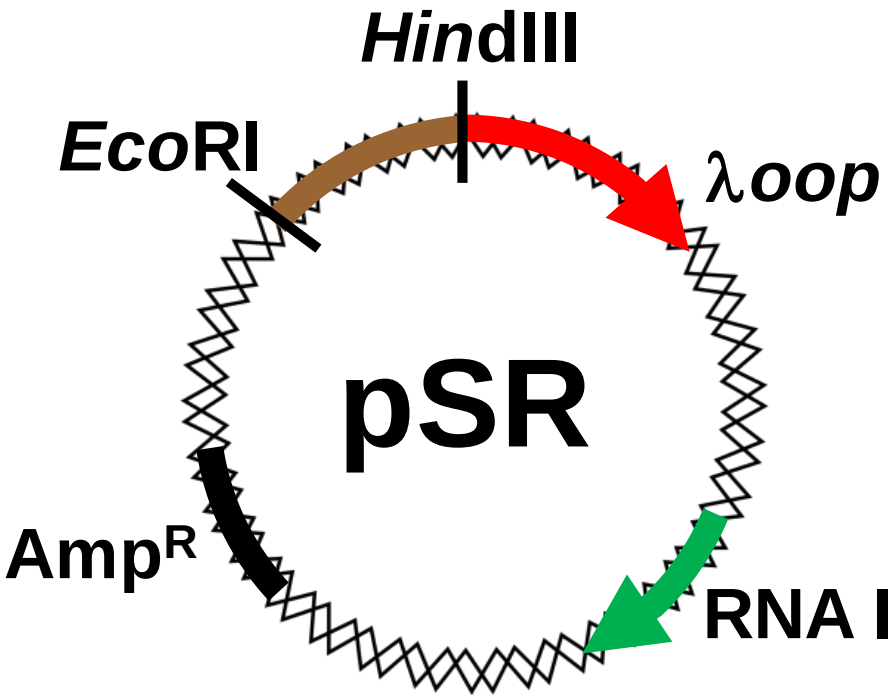


Figure 2.2: pSR plasmid map. pSR was used as a reporter plasmid in *in vitro* transcription assays. The brown bar represents the site where promoter DNA was cloned using the *EcoR*1 and *Hind*III sites upstream of the *λoop* transcription terminator. Green denotes the RNAI that serves as an internal control and the black bar denotes the ampicillin resistance.

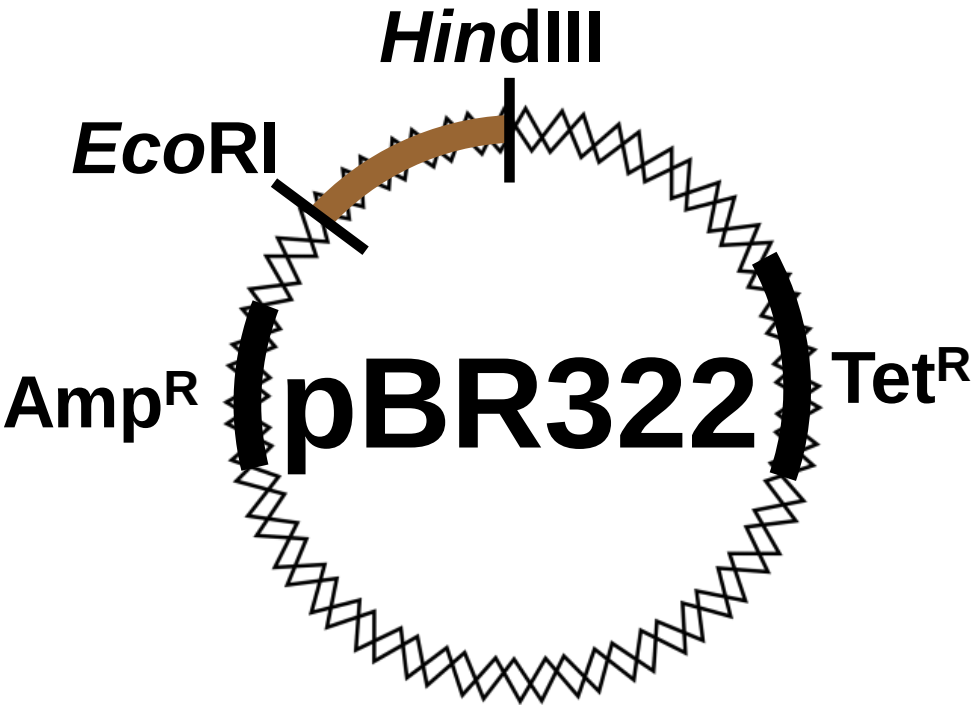


Figure 2.3: pBR322 plasmid map. pBR322 was used to express the target gene in genetic complementation assay. The brown bar represents the site where the promoter or full gene was cloned, and the black bar denotes the ampicillin and tetracycline resistance.

2.4 Phenol/Chloroform extraction of DNA

Phenol/Chloroform extraction was used to purify plasmids and DNA fragments following enzymatic modification.

2.4.1 Reagents and chemicals used:

- **Phenol/Chloroform/Isoamyl alcohol**: pH 6.7/8.0
- **TE buffer**: 10 mM Tris (pH 8.0), 0.5 mM EDTA

2.4.2 *Method*

An equal amount of phenol/chloroform was added to aqueous DNA solutions and mixed for 10 seconds with a vortex mixer. The samples were then centrifuged for 2 minutes at maximum speed in a bench top microfuge. The upper aqueous layer containing DNA was carefully transferred to a sterile tube ensuring no proteins, at the interface between the aqueous and organic layer, were disrupted. DNA in the aqueous layer was further concentrated by precipitation with ethanol before resuspension in a smaller volume.

2.5 Ethanol precipitation

2.5.1 Reagents and chemicals used:

- **Ethanol**: 100% (v/v) stored at -20°C
- **Ethanol**: 70% (v/v) diluted in dH_2O stored at -20°C
- **Sodium acetate**: 3 M pH 5.2

2.5.2 *Method*:

Two volumes of ice cold ethanol, and 0.1 volumes of 3 M sodium acetate (pH 5.2) were added to the DNA solution and incubated at -20°C overnight or at -80°C (20 minutes). After precipitation, samples were centrifuged at maximum speed in a refrigerated

benchtop microfuge for 15 minutes to collect the precipitated DNA. The pellet was washed using 1 ml of ice cold 70% ethanol. The pellet was then recovered by centrifugation at 4°C, as described above, and the supernatant was again discarded. The DNA pellet was dried, at low or medium heat in vacuum manifold and dissolved in the required volume of dH_2O.

2.6 Gel electrophoresis

2.6.1 Reagents and chemicals used:

- **Gel loading dye**: 0.025% (v/v) bromophenol blue; 20% (v/v) glycerol; 0.025% (v/v) xylene cyanol; 1 mM EDTA; 10 mM Tris, pH 7.5
- **Ethidium bromide**: 10 mg/ml
- **DNA ladder**: 100 bp DNA ladder purchased from NEB and diluted 6-fold in DNA loading dye.

2.6.2 Agarose gel electrophoresis

2.6.2.1 *Reagents and chemicals used:*

- **5 X TBE Buffer**: 0.445 M Tris-borate pH 8.3, 10 mM Na_2EDTA (Fisher Scientific)
- **1 X TBE Buffer**: Diluted from 5 X TBE Buffer in dH_2O
- **Agarose solutions**: 1% diluted in 1 X TBE Buffer. This was then heated for a further two minutes on a medium power in a microwave. Furthermore, this was allowed to cool before addition of ethidium bromide (1% (v/v)) and pouring into a gel cast.

2.6.2.2 *Method:*

Samples were mixed in a 1:1 ratio with gel loading dye and loaded onto 1% (w/v) agarose gel containing ethidium bromide. Electrophoresis was done at 120 V for 30 minutes in 1 X TBE buffer. Once the gel running was completed, DNA was visualised with a UV-transilluminator.

2.6.3 Polyacrylamide gel electrophoresis

2.6.3.1 *Reagents and chemicals used:*

- **Acrylamide**: 30% ((w/v)) (Geneflow)
- **N,N,N′,N′-Tetramethylethylenediamine (TEMED) :** Sigma Aldrich
- **Ammonium persulphate (APS) :** Sigma Aldrich

Briefly, 10 ml of 30% acrylamide protogel was mixed with 6 ml of 5X TBE and 44 ml of ddH_2O. 0.05 g of APS and 50 µl of TEMED was then added to the above mixture. The Polyacrylamide gel was then run at 150 V with 0.5 X TBE.

2.6.3.2 *Method:*

DNA-samples were mixed in a 1:1 ratio with gel loading dye. Gels were run at 150 V in 0.5 X TBE. After running the gel was dried and exposed on to the phosphor screen until the next day when it was scanned.

2.7 DNA manipulations

Oligonucleotides used in this study are listed in Table 2.3.

Table 2.3: Oligonucleotides used in this study. The name of the gene or DNA fragment to be amplified is shown on the left and the oligonucleotide sequence (with restriction sites highlighted in bold) in the middle panel of the table.

Oligonucleotides (5' to 3') for amplification* of *estA1 promoter fragments		
*estA*1 up	GGCTGC**GAATTC**ATGAAAATAATATATAAAAAGCGAGTG	This work
*estA1*down	GCCCG**AAGCTT**TTAATAACATCCAGCACAGGCAGGATTAC	This work
P*estA1*up	GGCTGC**GAATTC**TAACATGATGCAACTCACAAAAAAAATA AAAAAATTGCAAAATCCGTTTAACTAATCT	This work
P*estA1*down	GCCCG**AAGCTT**CATGTTACCTCCCGTCATGTTGTTTCACG GATATTTGAGATTAGTTAAACGGATTTTG	This work

Oligonucleotides (5' to 3') for amplification* of *estA2 promoter fragments		
estA2 up	GGCTGC**GAATTC**CCGGGGCGGTTCATTGTTATTTTTTTTGTG	This work
estA2 down	GCCCG**AAGCTT**TTAATAGCACCCGGTACAAGCAGGATTACAAC	This work
P*estA2* up	GGCTGC**GAATTC**TAGTATGATACACATCACAAAAAAATAAAAA AGTTTGCGCAATCGTTCTGATTTTGAT	This work
P*estA2* down	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAAT ATTTAAATCAAAATCAGAACGATTGC	This work

Oligonucleotides (5' to 3') for amplification* of *eltAB promoter fragments		
P*eltAB* up	GGCTGC**GAATTC**TTCTGGTGTGGACTTTCTGGTGCTCCAGGTTGTG	This work
P*eltAB* down	GCCCG**AAGCTT**TTCATTCCGAATCCTGTTATATATGTCAAC	This work
P*eltAB* down	GCCCG**AAGCTT**ACAATCCGGAAAAAGATAACGCCAC	This work

***Oligonucleotides (5' to 3') for amplification* of CRP site mutator fragment of *estA2* 460bp up**		
P*estA2* CRP mutator up	GCAAACTTTTTTATTTTTTAGAGTTGTGTAACTTTCTATAC	This work

Oligonucleotides (5' to 3') for amplification* of *different DNA targets in ChIP-PCR experiment		
P*estA2* up	GTTCTCGCAAGGACGAGAATTTC	This work
P*estA2* down	GCCCGAAGCTTCATATTACCTCCGAAACACGTCGTCCACGAA TATTTAAATCAAAATCAGAACGATTGC	This work
P*eltAB* up	GTTCTCGCAAGGACGAGAATTTC	This work
P*eltAB* down	GCCCGAAGCTTACAATCCGGAAAAAGATAACGCCAC	This work
yabN up	GGTGCGGCTGTCGAACAGTAAATAG	This work
yabN down	GCCGCGCGAATGGGAAACCCTCAG	This work

Oligonucleotides (5' to 3') for amplification of MarA targets for EMSA analysis

thrL up	GGCTGC**GAATTC**GCTTTTCATTCTGACTGCAATG	This work
thrL down	CGCCCG**AAGCTT**CATGGATGTTGTGTACTCTG	This work
leuO/*L* up	GGCTGC**GAATTC**GAAAAGCGTCGGTAGTTAAGCAG	This work
leuO/*L* down	CGCCCG**AAGCTT**CATTAAATCAGCTCCAGATG	This work
degP up	GGCTGC**GAATTC**CTATAAAACGAATCTGAAGAACAC	This work
degP down	GCCCG**AAGCTT**CAGAGAGCGGAGATAACGCCAAAC	This work
lacZ up	GGCTGC**GAATTC**GGAAAGCGGGCAGTGAGCG	This work
lacZ down	CGCCCG**AAGCTT**CATAGCTGTTTCCTGTGTGAAATTG	This work
ybaO up	GGCTGC**GAATTC**TAACGAGATCCCTTCCAGCACCG	This work
ybaO down	CGCCCG**AAGCTT**CATAGCCCTTCCACAGAGAATTTTTTTCTC	This work
pheP up	GGCTGC**GAATTC**GGCGTATAAGCTGATGTGGCTG	This work
pheP down	CGCCCG**AAGCTT**CACGCCTTTCCCCTGTGTGTCTTTTTTGTTGAG	This work
modE<>*acrZ* up	GGCTGC**GAATTC**GTCTTATTGTGACGGAAAACGAACG	This work
modE<>*acrZ* down	GCCCG**AAGCTT**AATAACTCTAACATGGTCAACTCC	This work
ybiV up	GGCTGCGAATTCCGTCGTTAAGAAAAGTACCGTCCAT	This work
ybiV down	GCCCGAAGCTTAATAACTCTAACATGGTCAACTCC	This work
ybjC/*grxA* up	GGCTGC**GAATTC**GCGCGCATACGCTTCCCTCTG	This work
ybjC/*grxA* down	CGCCCG**AAGCTT**CATTATTTCTCTCCTCATAG	This work
ycgZ/*bluF* up	GGCTGC**GAATTC**TTGAACACTAGTTGGCGAAAAATCTTG	This work
ycgZ/*bluF* down	CGCCCG**AAGCTT**CATGCTACGCCTCTGTTAAAAATG	This work
fnr up	GGCTGC**GAATTC**AGGTTATCTTTTGCTGTAAACATTAAACAATTTGTG	This work
fnr down	CGCCCG**AAGCTT**CATAGGTCTGCTCAAGCCGTAATTG	This work
yneO up	GGCTGC**GAATTC**GAAAACTGTTTCTTTCAATAGGA	This work
yneO down	GCCCG**AAGCTT**GCATTAAGCACAACCCTTATTTTATA	This work
marR up	GGCTGC**GAATTC**TTCCGCTTCGGGGTGAAATAGTAG	This work
marR down	CGCCCG**AAGCTT**CATTAGTTGCCCTGGCAAGTAATTAG	This work
yeeF up	GGCTGC**GAATTC**CATAATAATTTTTCTTTAAATGGC	This work
yeeF down	GCCCG**AAGCTT**GGACTTGCCAGTGGCTGGTGGCGG	This work
ompC<>*micF* up	GGCTGC**GAATTC**GGTTAAAATCAATAACTTATTCT	This work
ompC<>*micF* down	GCCCG**AAGCTT**ATTCAGAAATGAATGACGGTAATA	This work
ypeC up	GGCTGC**GAATTC**ATCGTCCGAAGCAACAGCCCC	This work
ypeC down	CGCCCG**AAGCTT**CATTGTTATTCTCCTTCACGATCG	This work
yfeS><*cysM* up	GGCTGC**GAATTC**CGGCGTTTCCTCACGAAAATTAA	This work
yfeS><*cysM* down	GCCCG**AAGCTT**TGCAATATATTGAATTAGCACGAT	This work
nudF<>*tolC* up	GGCTGC**GAATTC**GACTGCCGTTTGAGCAGTCATGTG	This work
nudF<>*tolC* down	GCCCG**AAGCTT**TTCTAGCAGAAGCCGCTACCGCAA	This work
yhbV up	GGCTGC**GAATTC**CTGGCTGGAGATGGCAAAATCGCT	This work
yhbV down	GCCCG**AAGCTT**TCAATCAGAAACTCACCGTTCTC	This work
yidQ/*ibpA* up	GGCTGC**GAATTC**CCCTCAGTCTATGCAATAGACCATAAACTG	This work
yidQ/*ibpA* down	CGCCCG**AAGCTT**CATAATCAATAGCTCCTGAAATCAG	This work
mnmG up	GGCTGC**GAATTC**GTGGATTAATTTACTCAAATAAG	This work
mnmG down	CGCCCG**AAGCTT**CATGGTGATTGCCTCGCATAACG	This work
yihT up	GGCTGC**GAATTC**GATATGGTCATTCAGACGTTGTGA	This work
yihT down	GCCCG**AAGCTT**AAATTCGATCGCGAACAAGCGATC	This work
yiiG up	GGCTGC**GAATTC**GGATGTCTATATCAAGTGCTACAA	This work
yiiG down	GCCCG**AAGCTT**CACTCAGCGAGGTTGGATTCACT	This work
ETEC_4304 up	GGCTGC**GAATTC**GTAAGCCTATAGACCTGAAAGAA	This work
ETEC_4304 down	GCCCG**AAGCTT**GTTCCTTTCCTGTATGGCTGATA	This work
ETEC_4307 up	GGCTGC**GAATTC**ATGGTGGGAATATATACCATAGC	This work
ETEC_4307 down	GCCCG**AAGCTT**AATAGCTTTCTGGCAAAAACACC	This work

yjcB<>yjcC up	GGCTGC**GAATTC**GGCTTGCAGAACACAAAAATGAA	This work
yjcB<>yjcC down	GCCCG**AAGCTT**TTAATGCCACTGAAAAACCACAT	This work
yjjQ/P up	GGCTGC**GAATTC**GTAGATATTTTAAATGTCTCCAGG	This work
yjjQ/P down	CGCCCG**AAGCTT**CATATTTCCTCCACATCCACTG	This work
ETEC_4702 up	GGCTGC**GAATTC**GACAGATGATAATTATTTCATGA	This work
ETEC_4702 down	GCCCG**AAGCTT**GACCGCTAATGCTGTTGTCAGCC	This work
deoB up	GGCTGC**GAATTC**CGAAAGACGAAAACAGCTGGCA	This work
deoB down	GCCCG**AAGCTT**CAGTATACCGTTATTCACTGATA	This work
ETEC_P948_0770up	GGCTGC**GAATTC**GCAGATTCGAGATTAATTTTGGGTC	This work
ETEC_P948_0770down	GCCCG**AAGCTT**TAAAGGATTAGGCAAAAATAGCG	This work
ETEC_3200 up	GGCTGCGAATTCGCCGCCAGACAACAACCCATACTTT	This work
ETEC_3200 down	CGCCCGAAGCTTTGACTATCTCGCGAAAGAGTACACCA	This work

Oligonucleotides (5' to 3') for amplification of *xseA* and regulatory region derivatives

P*xseA*1 up	GGCTGC**GAATTC**GTAACCGATTGCATCTACCCCTTTTTGC	This work
P*xseA*2 -36C up	GGCTGC**GAATTC**AAATCCTTGCTATCCCCGAAGGGCGG	This work
P*xseA* down	GCCCG**AAGCTT**CATGTGAGCGAGATCAAATTCTAAATCAG	This work
P*xseA* & *xseA* –marbox	GGCTGC**GAATTC**CTTGCTATCCCCGAAGGGCGGGTTACTATCG	This work
xseA down	GCCCG**AAGCTT**TTAATGCACCTTTTTACGCGATTTTTTTACTG	This work
F63A up	GTACGCTGCGCGATGGCCCGCAACAGCAACCG	This work
F63A down	CGGTTGCTGTTGCGGGCCATCGCGCAGCGTAC	This work
R205A down	CGAACCGCCGCCGGCCCCGACGATCAAT	This work
Q267STOP up	CCGCTGAAGTAGTGAGCCGTAATTAGCAAGAGTTAC	This work
Q267STOP down	GTAACTCTTGCTAATTACGGCTCACTACTTCAGCGG	This work
P353STOP up	CGGTTAACACAGCGGCTGAATCAGCAAAATTAGCAGCCGAAGATTCATC	This work
P353STOP down	GATGAATCTTCGGCTGCTAATTTTGCTGATTCAGCCGCTGTGTTAACCG	This work

Oligonucleotides (5' to 3') for amplification of *mlaFEDCB* and regulatory region derivatives

P*mlaF*1 up	GGCTGC**GAATTC**TTTATGCGGCTAAAAAGTAAAAC	This work
P*mlaF*2 up	GGCTGC**GAATTC**CACGGCGGGTAATATTCTG	This work
P*mlaE* down	GCCCG**AAGCTT**CATAATTCACCCTTCGTCTTGCG	This work
mlaFEDCB down	GCCCG**AAGCTT**TTAACGAGGCAGAACATCAGCAGGC	This work

Oligonucleotides (5' to 3') for sequencing of plasmid constructs

pRW50 up	GTTCTCGCAAGGACGAGAATTTC	This work
pRW50 down	AATCTTCACGCTTGAGATAC	This work
pBR322 up	GGCTGCGTGCCACCTGACGTCTAAGAAACC	This work
pBR322 down	GCCCGCAGCATCCAGGGTGACGGTGCCGAG	This work
pSR up	CCATATATCAGGGTTATTGTCTC	This work
pSR down	CATCACCGAAACGCGCGAGG	This work

2.7.1 Polymerase Chain Reaction (PCR)

Velocity DNA polymerase and buffer, provided by Bioline, was used with 50-200 ng of template. All oligonucleotides were generated by Thermo Fisher Scientific. In a reaction 100 pM of oligonucleotide, was used from the stock of 100 µM making overall concentration of 2 µM. Deoxynucleotide triphosphates (dNTPs) were from Bioline; where 2 mM (0.5 mM each) were used in a PCR mixture volume of 50 µl. The typical PCR cycle used in the study is shown in Table 2.4.

2.7.2 Megaprimer PCR

Megaprimer PCR is a technique that is used to introduce point mutations in large DNA fragments without useful restriction endonuclease sites near to the base pair being mutated. This PCR involves two reactions, wherein the first reaction generates a long ‘megaprimer’ with a point mutation. In the second reaction this ‘megaprimer’ is used with a second flanking primer to generate a larger DNA fragment, with the point mutation inserted, suitable for cloning.

2.7.3 Colony PCR

To amplify chromosomal DNA, colony PCR was used. A colony was smeared on the inside of the PCR tube before addition of Phusion DNA polymerase, buffer (NEB), oligonucleotides and dNTPs. To lyse the cells, the starting denaturation time was extended to 10 minutes.

2.8 DNA purification using QIAquick PCR Purification kit (QIAgen)

After PCR or restriction digestion the DNA was purified using a QIAquick PCR purification kit (QIAgen) following the manufacturer’s instructions.

Table 2.4: Typical thermocycler settings used in the study. PCR cycle used for amplifying different fragments for the purpose of ligation, inserting mutations using megaprimer PCR and also for colony PCR

Step	Temperature	Duration
Initialization	95°C	1 min
Denaturation	95°C	30 s
Annealing (T_a)	X	30 s
Extension	72°C	30 s
Final elongation	72°C	5 min
Final hold	4°C	Infinite

Note - The annealing temperature (T_a) varied with the different melting temperature (T_m) of different fragments. For this the temperature used was $T_a = T_m - 5°C$ and is represented as X.

This kit can extract up to 1 µg DNA. The kit purifies by the principle of adsorption of the product onto a silica membrane column, washing by high salt and purified DNA was eluted in 30 µl DNase/RNase free water. The purification is performed to remove the unused primers, enzymes and the buffer from the PCR product.

2.9 DNA restriction digest

Restriction enzymes and buffers were used according to the manufacturer's instructions. 1 µg DNA was used with the enzymes and buffers in a final volume of 50 µl with dH_2O. This was followed by incubation for 3 hours at 37°C. During the preparation of vectors for cloning purposes, the digested plasmid DNA was treated with calf alkaline phosphatase (NEB) to remove terminal 5' phosphate groups. The reaction was then followed by incubation for a further 30-60 minutes at 37°C. Subsequent DNA-purification was done by phenol-chloroform extraction (for pRW50), and QIAgen gel extraction or PCR purification kits were used for smaller vectors and insert DNA fragments.

2.10 DNA Ligation

For ligation, 1 µl T4 DNA ligase (New England Biolabs) was used with 100 ng of restriction digested insert DNA and 100 ng of restriction digested, and phosphatase treated, vector DNA. The final volume was 20 µl with dH_2O and T4 DNA ligase buffer added as required (New England Biolabs). Ligation mixes were incubated for 2 hours at room temperature and used to transform chemically competent cells as outlined in the next section. Nutrient agar or MacConkey-lactose plates, supplemented with suitable antibiotics, were used to select transformants. Candidate transformants were cultured overnight. Plasmid was then extracted, and digestion with restriction enzymes was used to check for presence of cloned DNA fragment.

2.11 DNA transformation of bacterial cells

2.11.1 Preparation of $CaCl_2$ competent cells

2.11.1.1 *Reagents and chemicals used:*

- **Calcium Chloride ($CaCl_2$)**: 0.1M
- **Glycerol**: 50% (v/v)

2.11.1.2 *Method:*

5 ml of LB was inoculated with the desired strain, incubated at 37°C overnight, and 1 ml used to inoculate 50 ml of LB broth the next day. This sub-culture was incubated at 37°C with shaking until the culture reached to an OD_{650} of 0.3-0.5. The culture was then poured into a sterile centrifuge tube and left on ice for 10 minutes. Harvesting of cells was done at 4°C for 20 minutes at 2400 x *g*. The cells were re-suspended in 25 ml ice-cold 0.1 M $CaCl_2$ and left on ice for 30 minutes. Cells were harvested again by centrifugation at 2400 x *g* for 10 minutes at 4°C and resuspended in 3.3 ml ice-cold 0.1 M $CaCl_2$ before being stored overnight at 4°C. Glycerol of 15% (v/v) was added to allow storage of the cells at -70°C as aliquots.

2.11.2 Transformation of chemically competent cells with plasmid DNA

50 ng of plasmid DNA was added to 100 µl of competent cells followed by incubation on ice for 1 hour. The cells were heat shocked for 2 minutes at 42°C followed by 3 minutes on ice. 1 ml LB was then added to the cells and they were incubated at 37°C in a shaking incubator for 60 minutes. Harvesting cells was done by centrifugation for 3 minute at 2800 x *g.* The pellet was re-suspended in 100 µl LB before plating on solid media. Agar plates were incubated at 37°C overnight.

2.12 DNA Sequencing

The Functional Genomics and Proteomics Laboratory (University of Birmingham) facility was used to perform “plasmid-to-profile” sequencing whereby 300ng plasmid miniprep was mixed with a 10 µM sequencing primer. The Table 2.3 outlines primers used in this study for the purpose of sequencing inserts in plasmids.

2.13 β-galactosidase assays

2.13.1 *Reagents and chemicals used:*

- **Z-buffer**: 0.75 g KCl, 0.25 g $MgSO_4.7H_2O$, 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_20$, 2.70 ml β-mercaptoethanol
- **Toluene** 100% (v/v)
- **Sodium deoxycholate**: 1% (w/v)
- **ONPG in Z-buffer**: 13 mM (made on day of use).
- **Sodium carbonate**: 1 M
- **Z-buffer**: 0.75 g KCl, 0.25 g $MgSO_4.7H_2O$, 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_20$, 2.7 ml β-mercaptoethanol

2.13.2 *Method:*

β-galactosidase assay is used to monitor *lacZ* expression from the reporter plasmid pRW50 for observing promoter activity. The unit of the activity is the Miller Unit (Miller, 1972).

A fresh colony of each strain was used to inoculate 5 ml culture with the suitable antibiotic. Cultures were incubated at 37°C overnight in a shaking incubator. The next day, 50 µl was sub-cultured in 5 ml of the same medium for each strain/condition. These cultures were placed in a shaking incubator, at 37°C, until an OD_{650} of 0.3-0.5 was

reached. The OD_{650} was recorded and two drops each of toluene and sodium deoxycholate were added to each culture, vortexing for 15 seconds, to lyse cells. To ensure the toluene evaporated efficiently, cultures were put back at 37°C with lids open for 20 to 30 minutes.

To assay for β-galactosidase activity, 2.5 ml of 13 mM 2-Nitrophenyl β-D-galactopyranoside (ONPG) solution was added to 100 µl of each lysate. The resulting solution was incubated until it turned yellow. To halt the experiment 1 ml of 1 M sodium carbonate was added and the OD_{420} of the solution was measured.

The following formula was used to calculate β-galactosidase activity:

$$\text{β-galactosidase activity} = \frac{1000 \text{ x } 2.5 \text{ x } 3.5 \text{ x } OD420}{OD650 \text{ x } 4.5 \text{ x t x v}} \text{ nmol / min / mg bacterial mass}$$

Wherein:

2.5 = factor for conversion of OD_{650} into bacterial mass, based on OD_{650} of 1 being equivalent to 0.4 mg/ml bacteria (dry weight).

3.5 = final assay volume (ml)

1000/4.5 = factor for conversion of OD_{420} into nmol o-nitrophenyl (ONPG), based on 1 nmol/ml ONP having an OD_{420} of 0.0045

t = incubation time (min)

v = volume of lysate added (in ml)

In each experiment, each strain was assayed in triplicate and, once the mean β-galactosidase activity and standard deviation were calculated, data were depicted as bar charts. At least two separate experiments for each strain and set of conditions were

performed. The background beta-galactosidase activity from empty pRW50, was subtracted from the values shown. The standard deviation was always <10% of the mean.

2.14 Radiolabelling of DNA fragments

Radio-labelling of DNA fragments allowed DNA detection in electrophoretic mobility shift assays (EMSA) and DNAseI footprint assays.

2.14.1 *Reagents and chemicals used:*

- **T4 polynucleotide kinase**
- **T4 polynucleotide kinase buffer**: 10 x
- **Tris-EDTA (TE) buffer**: 10 mM Tris-HCl, 1 mM EDTA (pH 8.0)
- **Sephadex G-50**: 12% (v/v) Sephadex G-50 in TE buffer.

2.14.2 *Method:*

The reaction mixture contained 16 µl of DNA fragment (100-400 nM), 1 µl T4 polynucleotide kinase, 2 µl 10 X kinase buffer and 1 µl [γ-^{32}P]-ATP (10 µCi/µl). These were incubated at 37°C for 30 minutes. Unincorporated [γ -^{32}P]-ATP was removed with a 200 µl volume of Sephadex G-50 spin column. The flow through contained labelled DNA fragments and was collected, followed by storage at -20°C.

2.15 Maxam-Gilbert sequencing reactions

2.15.1 *Reagents and chemicals used:*

- **Formic acid**: 100% (v/v)

- **DNAse I blue:** 20 mM urea, 1 mM EDTA, 20 mM NaOH, 0.025% (v/v) of each bromophenol blue and xylene cyanol.
- **Sodium acetate**: 3 M, pH 7.0
- **Piperidine**: 10 M diluted in water before use
- **Ethanol**: 100% (v/v)

2.15.2 *Method:*

Precipitation of DNA followed by using 20 µl sodium acetate and 700 µl 100% ethanol. The DNA pellet was dried and resuspended in 10 M piperidine with incubation for 30 minutes at 90°C. Precipitation of DNA was done again followed by washing, twice with 70% ethanol. The dried pellet was resuspended in 20 µl of loading buffer.

2.16 DNA sequencing gel electrophoresis

2.16.1 *Reagents and chemicals used:*

- 6% Denaturing gel was made using UreaGel (Geneflow) concentrate, diluents and buffer
- N,N,N′,N′-Tetramethylethylenediamine (TEMED)
- 10% ((w/v)) ammonium persulphate (APS) solution (diluted in dH_2O)

2.16.2 *Method:*

The gel mixture was poured with the aim of obtaining a 0.4 mm thick and 40 cm x 30 cm gel. Prior to sample loading, the gels were warmed by electrophoresis at 60 W for 30 minutes. Once run, the gel was soaked in fixing solution for 15 minutes and transferred to a 3 mm whatman filter paper. Gels were dried under vacuum at 80°C° for 45 minutes

and then exposed to a Biorad screen in a light-proof autoradiography cassette overnight.

2.17 DNAse I footprinting

2.17.1 *Reagents and chemicals used:*

- **Binding buffer**: 20 mM Tris pH 7, 10 mM $MgCl_2$, 100 mM EDTA and 120 mM KCl. 12.5 mg/ml Herring sperm DNA
- **DNAse I stop solution**: 1 ml 3 M NaAc pH 7.0 + 200 µl 500 mM EDTA, made up to 10 ml with dH_2O
- **Phenol/Chloroform/Isoamyl alcohol**: pH 6.7/8.0
- **Ethanol**: 100% (v/v) and 70% (v/v)
- **Gel loading buffer**: 40% (v/v) deionised formamide, 5 M urea, 5 mM NaOH, 1 mM EDTA, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF

2.17.2 *Method:*

DNA fragments were prepared from maxi preparations of plasmid pSR, carrying the fragment to be examined. After the maxiprep the plasmid was subjected to restriction digestion; first digested with *Hin*dIII and then treated with alkaline phosphatase. The product was purified using the phenol/chloroform method and a second digestion was done using the *Aat*II enzyme. The desired DNA fragments were purified by gel extraction. The DNA were labelled at the *Hin*dIII end using [γ-^{32}P]-ATP and T4 polynucleotide kinase as described above (10-40 nM). Footprinting reactions contained labelled DNA in binding buffer, 12.5 µg/ml Herring sperm DNA, and purified MarA. After incubation at 37°C° for 20 minutes 2 µl of DNAse I was added (at a concentration

previously determined by calibration) and left for 40 seconds. The reaction was stopped with 200 µl of DNase I stop solution. Next, 200 µl of phenol/chloroform was added, and the sample was vortexed and centrifuged. The aqueous layer was then transferred to a fresh tube and 1 µl of 20 mg/ml glycogen, along with 400 µl ice-cold ethanol, was added. This mixture was left at -80°C for 20 minutes. The samples were then centrifuged at 10,000 x *g* for 15 minutes at 4°C. The DNA pellet was washed with 600 µl ice-cold 70% ethanol and re-centrifuged at 10,000 x *g* for 10 minutes at 4°C. The pellet was dried and re-suspended in 8 µl DNAse loading buffer. Before loading, all the samples were heated at 90°C for 2 minutes.

2.18 *In vitro* transcription assays

2.18.1 *Reagents and chemicals used:*

- ***In vitro* transcription buffer**: 20 mM Tris pH 7.9, 5 mM $MgCl_2$, 500 µM DTT, 50 mM KCl, 100 µg/ml BSA, 200 µM ATP, 200 µM GTP, 200 µM CTP, 10 µM UTP with 5 µCi [α-^{32}P]-UTP.
- **Stop solution**: 40% (v/v) deionised formamide, 1 mM EDTA, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol.

2.18.2 *Method:*

The *in vitro* transcription experiments were done as described previously by Kolb et al. (1995). A Qiagen maxiprep kit was used to purify supercoiled pSR plasmid carrying the regulatory region to be analysed. This template (16 µg/ml) was incubated with *in vitro* transcription buffer. The reaction was started by adding purified *E. coli* RNA polymerase

holoenzyme. Labelled RNA products were analysed on a denaturing polyacrylamide gel.

2.19 Microscopy

2.19.1 Reagents and chemicals used*:*

- **PBS Buffer:** 140 mM NaCl, 2 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, to 1 L in sterile ddH_2O. All components were autoclaved before use.
- **Poly-L-lysine:** 10 mg/ml (Sigma)
- **Hoechst 33258:** 5 µg/ml Hoechst 33258 in PBS containing 40% (v/v) glycerol.

2.19.2 Preparation of microscope slides

Once the desired OD_{650} (0.6 - 0.7) was reached, cells were harvested by centrifugation at 2400 x *g* for 10 minutes from 1.5 ml of culture. The cells were then washed at least 3 times with PBS. Microscope slides were prepared by drying 5 µl poly-L-lysine onto the centre of a slide. The cells were again centrifuged at 2800 x *g* whereby all supernatant was removed and the pellet was resuspended in 10 µl Hoechst 33258. Lastly, cells were incubated at room temperature for 5 minutes before 5 µl was pipetted onto the centre of the slide and a cover slip was then applied.

2.19.3 Imaging and analysis

Slides were imaged using a Nikon Eclipse 90i microscope, Nikon Intensilight C-HGFI lamp, Hamamatsu ORCA ER camera (1344x1024 pixels, pixel size 6.45 µm) and Nikon Plan Apo VC 100X Oil immersion lens (Numerical Aperture 1.4), with a final optical magnification of 100X. A DAPI filter set with excitation filter of 340-380 nm, exposure time of 90 ms, dichroic mirror of 400 and barrier filter of 435-485 nm was used in

microscopy for visualising the Hoechst 33258 stained nucleoid. Microscopy was carried out at room temperature, within 30 minutes of slides being prepared.

2.20 Growth assays

An independent colony of each bacterial strain was used to inoculate 5 ml of LB broth and incubated overnight at 37°C. The next day OD_{650} was recorded. The inoculum OD was then adjusted to 0.05 in LB broth by using the formula OD1 X V1 = OD2 X V2 (where OD1 is the OD_{650} value measured for the overnight culture, V1 is the volume of this culture to be used for adding it to the inoculum, OD2 was constant as 0.05 and V2 is the volume of LB broth in which we are inoculating (5 ml)). V1 was calculated and the required volume was added to 5ml of LB broth (with or without antibiotics) in triplicate. Hence, the OD_{650} of 0.05 and the net volume remains constant. The sub-cultures were then incubated in a shaking incubator at 37°C and, at 40 minute intervals, 200 µl was transferred to a 96 well plate to measure OD_{650}.

2.21 Minimum inhibitory concentration (MIC) determination

2.21.1 *Materials, reagents and chemicals used:*

- Sterile micro-titre tray and lid.
- Multi-channel pipette.
- Multi-stepper pipette.
- Gilson pipettes
- Sterile distilled water (8X 5 ml)
- Sterile broth (2X 10ml, 8X 5 ml).
- Sterile 100 µl and multi-stepper tips.

- 5 ml overnight broth of strains.
- 2X Round sterile Petri-dish.
- Antibiotics solution of appropriate concentration.

2.21.2 Method

MIC broth microtitre double dilution method (Johnson *et al.,* 1999) was used to assess antibiotic sensitivity. These assays were done in a final volume of 100 µl of LB medium in a 96 well round bottomed microtitre plate. Each well contained 10 colony forming units of *E. coli*, 100 µl of LB medium and antibiotics as required. The microtitre plate was covered with a sterile lid and kept in a gently shaking incubator for 24 hours at 37°C. The minimum inhibitory concentration was the lowest antibiotic concentration that prevented bacterial growth. Results were only accepted if the observed MIC for the control NCTC *E. coli* 10418 and ATCC *E. coli* 25922 strains was within 1 doubling dilution of the expected result.

2.22 Accumulation Assay:

2.22.1 *Materials, reagents and chemicals used:*

- Isosensitest broth 10 ml and 250 ml volumes Sterile
- 50 mM Potassium phosphate buffer pH 7.0 (filter sterile)
- 0.1M glycine HCl pH 3 (autoclave)
- Sterile Universals
- 1.5 ml Eppendorfs - labelled with organism, antibiotic and time of sample - in triplicate
- Heated magnetic stirrer

- Sterile universal containing magnetic flea
- Ice refrigerated microcentrifuge
- 500 ml centrifuge buckets
- Doxycycline

2.22.2 *Method*

From an overnight culture, a 4% (v/v) inoculum was prepared in 250 ml LB. This was incubated at 37°C with shaking until OD_{650} value reached 0.7-0.8. Cells were harvested by centrifugation at 3000 x *g* for 20 min at 4°C and resuspended in 10 ml ice cold 50 mM phosphate buffer. This was followed by another centrifugation step at 3000 x *g* for 20 min at 4°C. The cell pellet was resuspended in 10 ml of phosphate buffer. The suspension was transferred to a sterile universal tube containing a magnetic flea and placed in a water-bath at 37°C (unless otherwise stated) on a magnetic stirrer to equilibrate for about 10 minutes. 1 ml of sterile phosphate buffer (ice-cold) was transferred into 1 set of Eppendorf tubes embedded in ice. Next, at time = 0 sec, a 500 µl sample was transferred to a new Eppendorf tube. 35 µg/ml doxycycline was mixed with the sample still in the water bath. In addition, at timed intervals, duplicate 500µl samples were transferred to Eppendorf tubes. Timed intervals were normally 10, 20, 30, 40, 50, 60, 120, 180, 240, 300 seconds. Once the samples were placed on ice, the Eppendorf tubes were centrifuged at maximum speed at 4 °C for 5 minutes after which they were put back on ice. The supernatant was discarded and cell pellet was resuspended in 1 ml of sterile phosphate buffer. This ensures the cells are washed and the surface bound doxycycline was removed. The samples were again centrifuged and the pellet was resuspended in 1 ml of 0.1 M glycine pH 3.0 and left for 2 hours at room

temperature. The centrifugation was again done for 5 minutes at maximum speed, after which the supernatant was transferred to freshly labelled Eppendorf tubes. These tubes were centrifuged for a further 5 minutes at maximum speed at 4 °C. The samples were diluted 1:10 before measuring fluorescence. A fluorescence spectrophotometer was used to measure the fluorescence from the samples at the correct excitation and emission wavelength for the doxycycline which were 425 nm and 520 nm respectively.

2.23 Efflux Assay

For the efflux assay, cells were grown as for accumulation assays. After washing the cells in PBS, the cell suspension had 0.3 OD_{650} units. 5 µg/ml of ethidium bromide and 50 µg/ml of chlorpromazine (CPZ) were added and the suspension was incubated at 25°C for 1 hour. When the cells were loaded with ethidium bromide, they were washed with PBS to remove the excess ethidium bromide and finally suspended in PBS that contains 0.4% (v/v) glucose.

2.24 Hydrocarbon assay

2.24.1 *Materials, reagents and chemicals used:*

- **p-xylene**
- **PUM buffer :** pH 7.1 : 22.2 g K_2HPO_4 • $3H_20$, 7.26 g KH_2PO_4, 1.8 g urea, 0.2 g $MgSO_4$ • $7H_20$ and distilled water to 1000 ml
- **Round bottomed test tubes**

2.24.2 *Method*

We used the method proposed by Rosenberg (Rosenberg, 1984). In this method, the bacterium to be tested is grown at 30°C in LB broth on a shaking incubator. After

reaching the early logarithmic phase, cells were harvested at 2400 x *g* for 15 min and then washed twice with PUM buffer before being resuspended in PUM buffer. 1.2 ml of the suspension was then placed in round bottomed test tubes and different volumes of p-xylene were added (0, 0.1, 0.2 ml). The suspension was incubated at 30°C for 15 min. The tubes were then agitated uniformly for 2 min and left it to rest for about 15 min. This allows the hydrocarbon layer to rise and the aqueous phase was then carefully taken for OD measurement in a cuvette. The absorption was measured at 400 nm.

2.25 Crystal Violet Assay

The crystal violet method to assay surface hydrophobicity was described in (Halder *et al.,* 2015). The bacterial cells were grown in LB broth and were harvested at 2400 x *g* for 5 min at 4°C. The cells were then washed twice and finally resuspended in 0.5 mM potassium phosphate buffer (pH 7.4). The cells were again harvested at 9300 x *g* for 5 min and then incubated in 0.5 mM potassium phosphate buffer with 10 g/ml crystal violet at 37°C for 10 min. The cells were again centrifuged at 9300 x *g* for 15 min. The supernatant was then transferred to a cuvette and absorbance was measured at 590 nm. The crystal violet solution was considered to have an OD_{590} of 100% and the percentage uptake of the crystal violet was calculated by using the formula:

(OD value of sample we are testing / OD value of CV solution) × 100

2.26 Chromatin Immuno Precipitation (ChIP)

2.26.1 *Materials, reagents and chemicals used:*

- **1 X TBS** (20 mM Tris-HCl, pH 7.4, 0.9% (w/v) NaCl)

- **FA lysis buffer (150 mM NaCl)** (50 mM HEPES-KOH, pH 7.0, 150 mM NaCl, 1 mM EDTA, 1% (w/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS).
- **FA lysis buffer (500 mM NaCl)** (50 mM Hepes-KOH, pH 7.0, 500 mM NaCl, 1 mM EDTA, 1% (w/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS).
- **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 EDTA, 0.1% (w/v) SDS).
- **1 X TE buffer** (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).
- **NEXTflex ChIP-seq barcodes** – 24 (supplied by BioOscientific).
- **Agencourt AMpure Magnetic Beads** (supplied by Beckman Coulter) XP DNA clean up kit.
- **Quick Blunting and Quick Ligation kit** (supplied by NEB).
- **Klenow fragment** (3'-5' exo-) (supplied by NEB)
- **Monoclonal anti-H-NS antibody**
- **Monoclonal anti-σ^{70} antibody** (supplied by Neoclone).
- **Monoclonal anti-MarA antibody**
- **Qubit fluorometer / Assay kit** (Life technologies)
- **10 mM Tris-HCl**, pH 7.5 / 8.0
- **Protein A sepharose beads** (GE healthcare), washed with dH_2O and store in a 50% (v/v) slurry with 1 X TBS.
- **Spin-X columns**

2.26.2 *Method*

An overnight culture of the desired bacterial strain was used to inoculate a subculture in 10 ml of growth media. The cultures were incubated until the OD_{650} reached 0.5. Formaldehyde (v/v) was added to cultures and incubated for 20 minutes. 3 M glycine was added to a final concentration of 0.5 M to quench crosslinking for 5 minutes. The cultures were then centrifuged at 2400 x *g* for 5 minutes. Supernatant was discarded and cells were suspended in 10 ml 1X TBS. The cells were again centrifuged at 2400 x *g* for 5 minutes and again resuspended in 1X TBS. Cells were collected by centrifugation and resuspended in 1 ml FA lysis buffer containing 4 mg/ml lysozyme and incubated for at 37°C for 30 minutes. Cells were sonicated to fragment the DNA using a bioruptor. Three 10 minute cycles each with 30 sec on and 30 sec off were performed at 4°C. The lysates were then centrifuged at 1500 x *g* for 5 minutes and the supernatant was transferred to a tube. From the tube 400 µl of the supernatant was mixed with 400 µl of FA lysis buffer. The 800 µl mix was then taken and mixed with 25 µl of the protein A beads (in 50% TBS slurry). 1 µl antibody was then added to the tube that was covered with parafilm and rotated for 90 minutes. Beads were collected by centrifugation at 1500 x *g* for 1 min. Supernatant was removed and 700 µl of FA lysis buffer (150 mM NaCl) was added. The suspension was then transferred to spin-x-columns, rotated for 3 minutes and centrifuged at 1500 x *g* for 1 min. Washing steps were then started and each step was repeated with rotation and centrifugation. The washing was done with 750µl each of the following buffers: FA lysis (150 mM NaCl), FA lysis (500 mM NaCl), ChIP wash buffer, TE (300 µl TE). The spin-X column was then transferred to a new dolphin-nosed tube and 100 µl of ChIP elution buffer was added before incubation at 65°C for 10-20 minutes. After incubation, the tube was immediately

centrifuged at 1500 x *g* for 1 min. The eluted samples were de-crosslinked by incubating the tube at 100°C for 10 minutes. PCR clean up was then done and DNA was then eluted in 30 µl of water. Precipitated DNA was analysed by PCR or Ilumina sequencing.

2.26.3 ChIP analysis by PCR

PCR reactions were prepared using 5 µl of the ChIP isolated DNA and desired oligonucleotides. PCR was set up according to the procedure described earlier in this chapter.

2.26.4 ChIP analysis by Ilumina MiSeq sequencing

Immunoprecipitations were done as described in 2.26.2 with changes at the washing steps. In all wash steps 750 µl solutions were used as stated above. This was followed by rotation for 3 min and centrifugation at 1500 x *g* for 1 min. 150 mM NaCl FA lysis buffer was used unless stated otherwise. The spin-x column containing the beads was washed by twice each by FA lysis buffer and then with 10 mM Tris-HCl, pH 7.5. A blunting kit was used to make the blunt DNA fragments; where 10 µl of the quick blunting buffer, 10 µl of dNTP mix, 80 µl of water and 2 µl of the blunt enzyme was added. After sealing the the lids with parafilm, the columns were rotated for half an hour at room temperature in such a way that the tubes weren't completely inverted. The beads were washed twice each with FA lysis buffer and Tris-HCl pH 8.0. Further, addition of A tail was achieved by resuspending the beads in 10 µl NEB buffer 2, 2 µl 100 mM dATP, 88 µl of water and 1 µl of klenow fragment. After sealing the lids, columns were rotated for 30 minutes at 37°C in such a way that the tubes weren't completely inverted. The beads were again washed twice each with FA lysis buffer and then by Tris-HCl pH 7.5. The fragments were then ligated with the barcodes (NEXTflex

BioOscientific). This was achieved by suspending the beads in 100 µl of 1X ligase buffer, 1 µl of the NEXTflex adapter and 4 µl of the quick ligase supplied in the kit. After sealing the lids, columns were rotated for 15 min over the rotating holder at room temperature in such a way that the tubes weren't inverted. The beads were washed with FA lysis buffer, then with FA lysis buffer with 500mM NaCl, ChIP wash buffer and TE buffer. Finally, a wash with 300 µl of TE buffer was done with rotation. The spin-x column was then moved to dolphin nosed tubes and 100 µl of elution buffer was added before incubation for 10 min at 65°C. Before centrifugation, the column was transferred to a new Eppendorf tube. Centrifugation was done at 1500 x *g* for 1 minute. The eluted sample was incubated for 10 min at 100°C to decrosslink proteins. AMPure magnetic bead clean up was used to clean the eluted and decrosslinked sample that was eluted in 12 µl of water. A test PCR was setup with the NEXTflex supplied primers to check the adapter ligation. The libraries were amplified using PCR and samples were removed at different cycle numbers to analyse the formation of adapter dimers. The PCR was then setup using the number of cycles which gave no adapter dimers. The libraries were purified with AMPure magnetic and loaded on a 1% agarose gel. After running the gel, the DNA was stained with ethidium bromide and visualised under a UV transilluminator. The sample ideally appears as a 200-600bp smear with no adapter dimer. Finally, the sample was cleaned and concentrated using the AMPure magnetic beads and was then quantified using a Qubit 2.0 fluorometer. An equimolar mix of libraries was prepared for multiplexed sequencing using an Ilumina MiSeq 2500. The complete ChIP-seq protocol is illustrated in the Figure 2.4.

2.27 Agencourt AMPure XP magnetic bead clean up

For ChIP-seq experiments, Agencourt AMPure XP magnetic beads, supplied by Beckman Coulter, were used both for DNA purification and for size selection of DNA libraries containing DNA of variable sizes.

2.28 Bioinformatic analysis of sequence reads

The Fastq files obtained from the Ilumina sequencer were converted into Fastq Sanger format using FastqGroomer. The files were then aligned to the Genbank reference sequences (FN649414.1, FN649418, FN649417, FN649416 or FN649415) using BWA for Illumina. The reference sequences correspond to the ETEC H10407 chromosome and plasmids p948, p666, p58 and p52 respectively. In addition, the SAM files were converted to BAM format using SAM-to-BAM. For each experiment, multiBamSummary was used to get coverage per base. Subsequent processing was done using R. Data were normalised to the same average read depth and the mean coverage per base was determined for each pair of replicates. The anti-MarA and anti-MarR immunoprecipitations shared numerous enriched DNA loci likely to be artefacts. DNA sequences of the reads have been submitted in ArrayExpress under accession number E-MTAB-5521.

These spuriously isolated DNA sequences were removed by subtracting MarR binding signal from the equivalent MarA binding value. Negative values, were replaced with zero. The MarR dataset was not examined further. Artemis was used to select the binding peaks for MarA or σ^{70}. Peaks scoring >2.7-fold (for MarA) or >3-fold (for σ^{70}) over background were retained in the “create features from graph” function. Some peaks were called twice because the signal oscillated up and down from the set

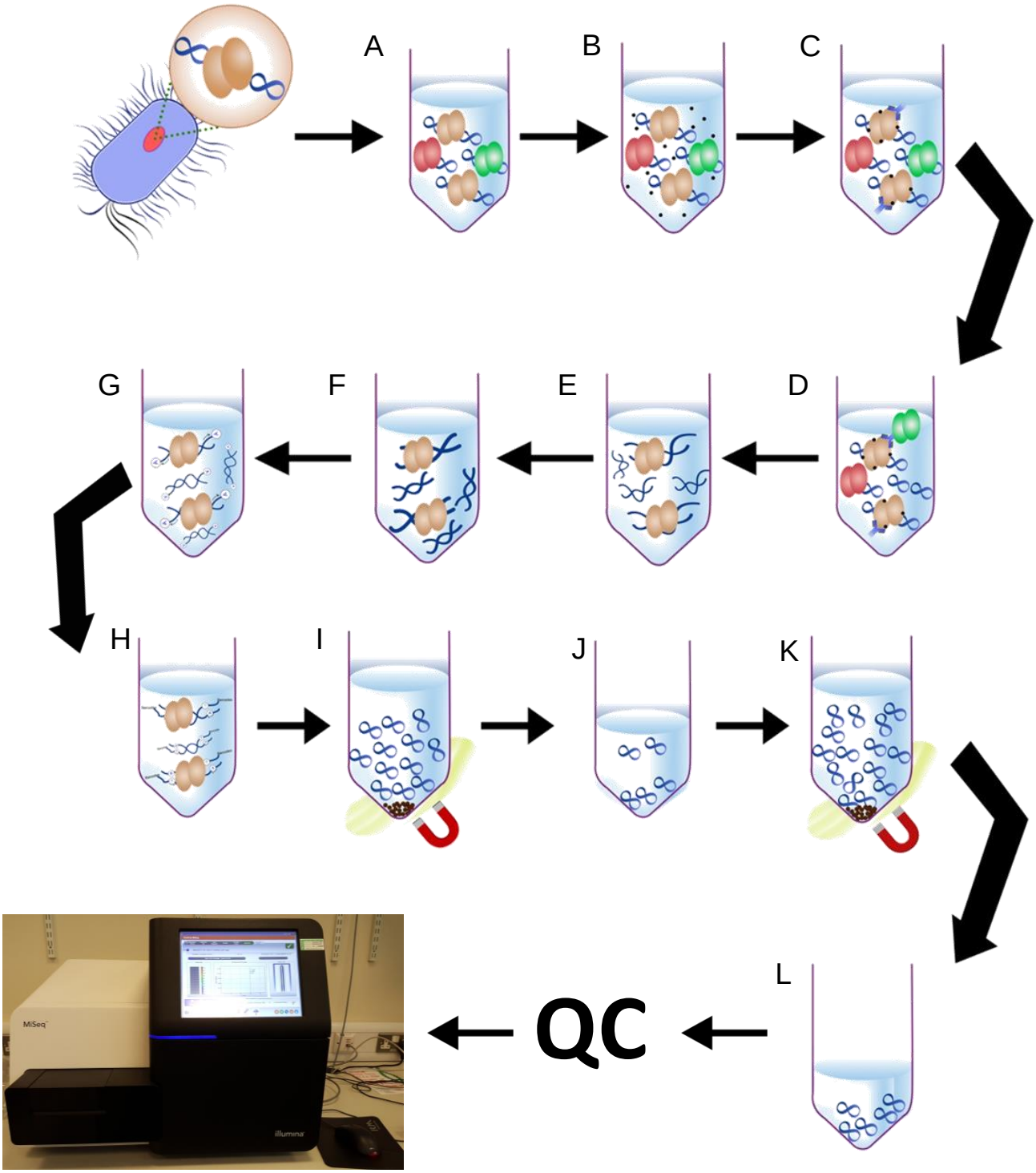


Figure 2.4: Cartoon representation of ChIP-seq. The bacterial cells were fixed by formaldehyde and treated by lysozyme. A. Cells were sonicated. B. Protein A beads

(small black dots) were added. C. Immunoprecipitation was carried out with antibodies (blue Y shaped). D. Washing was performed with different buffers. E. Non specific binding was removed by washing. F. Blunting enzyme was added so the DNA fragments all had blunt ends. G. ‘A’ tail addition was performed using a Klenow fragment, so all the DNA fragments have an A tail at one end. H. The barcodes for sequencing were then ligated onto the A tails and the samples were eluted, decrossslinked by boiling. I. Before the PCR AMPure magnetic beads were used to clean up the samples. J. The samples were then eluted in 12 µl of water. PCR was then set up to amplify the libraries. K. The amplified libraries were again cleaned up and concentrated by AMPure magnetic beads. L. Finally the DNA was eluted in 10 µl of water. The samples were then sent for quality control followed by sequencing on MiSeq 2500.

threshold. Those duplicates were removed. After visual inspection, 4 peaks for MarA binding were added manually. The peak centres were set to be located on the centre of the region passing the cut-off rounded to the nearest integer.

2.29 Bioinformatic identification of MarA binding sites

After defining the MarA peak centres we created a set of genome features in gff file format. Feature boundaries were 100 bp either side of each peak centre. The 201 bp DNA sequence corresponding to each feature was extracted using Artemis and submitted to MEME to search for motifs. Standard settings were used except that the expected number of sites was set to one per sequence and the minimum motif width was set to 15 bp. A single statistically significant motif was recovered and this matched the known MarA binding consensus. To determine the distance between peak centres (for σ^{70}) and binding sites (for MarA) we used the fetch closest non-overlapping feature tool in Galaxy (Afgan *et al.,* 2016). Peaks for σ^{70} and MarA described as overlapping were those centred were within 150 bp of each other.

2.30 Phylogenetic analysis of the marbox

Excluding *E. coli* species, we used BLASTp to search for genomes encoding MarA. We manually removed genomes encoding MarA with changes in the amino acid sequence of either DNA recognition helix; this is indicative of altered DNA binding specificity. Many of the resulting 161 genomes were derived from closely related strains of the same species. These were omitted so that 29 representative genomes remained. Phylogeny was determined using the sequence of the *polA* gene from each organism and BLASTn pairwise alignments. We used BLASTn to search the 29 genome sequences for a match to the 201 bp DNA sequence derived from MarA ChIP-seq

peaks in *E. coli*. If multiple hits were obtained from a single genome only the best match was used. If required, pairwise alignments were optimised manually to remove alignment gaps within marbox sequences. If no sequence match was identified this was scored as “region absent”. An identified marbox was scored as “conserved” if it matched either the equivalent *E. coli* sequence, or the consensus marbox (5'-gcactaattgctaaa-3') in at least 14 of the 15 possible positions. Sites were scored as “conserved with mismatches” if the above criteria were satisfied at 13 of the 15 marbox positions. Sequences falling below this threshold we scored as “not conserved”.

2.31 Pulsed-Field Gel Electrophoresis

2.31.1 Reagents and chemicals used

PIV buffer:

- 1 M NaCl
- 10 mM Tris-HCl (pH 7.6)

EC lysis solution:

- 6 mM Tris-HCl (pH 7.6)
- 1 M NaCl
- 100 mM EDTA
- 0.2% (w/v) deoxycholate
- 1% (v/v) Sarkosyl (N-lauroylsarcosine (Na salt))
- 1 mg/ml lysozyme

- 20 µg/ml RNase

ESP buffer:

- 0.5 M EDTA (pH 9.0 to 9.5) with NaOH adjusted
- 1% (v/v) Sarkosyl
- 50 µg/ml proteinase K

TE (pH 7.5):

- 10 mM Tris-HCl (pH 7.5)
- 1 mM EDTA

2.31.2 DNA Preparation

2.31.2.1 Method

The chromosomes of *E. coli* were prepared using the method proposed by (Heath *et al.,* 1992). Briefly the cells are immobilised in agarose and then incubated and washed with different enzymes and detergents which removes unnecessary materials and the purified DNA remains immobilised in the plugs of agarose.

In this method an overnight culture was set up for the required strain. The OD_{650} was measured the next day and it was adjusted to be 1.7 for all the strains. The culture was then centrifuged at 2400 x *g* for 15 min. The cells were suspended in equal volume of PIV buffer. The cells were again spun as before and suspended in half volume of PIV buffer. The cells were then incubated at 37°C to warm them. In the meantime 1% (w/v) of PFGE grade agarose was prepared in TE buffer and stored at 55°C. Equal volume of

the cells and the agarose were mixed gently. This suspension was dispensed in the agarose plug moulds and left to solidify. Two plugs of each sample can be made from these amounts of cell suspension. After the solidification the agarose plugs were stored in the refrigerator (4°C) for at least 5 minutes before use. Two plugs per sample were transferred into each universal and 2.5 ml of EC lysis solution without any enzyme was added and the universal was gently shaken at room temperature for 15 min. EC lysis solution was then aspirated off the universal. Another 2.5 ml of the EC lysis solution was added, but this time with lysozyme and RNase and incubated at 37°C with gentle shaking overnight. The next day the solution was removed by aspiration. ESP buffer was then added without any enzyme and shaken gently at room temperature for 15 min. The solution was then replaced with ESP buffer containing proteinase K and incubated with shaking gently at 37°C for 48 hours. After 2 days the ESP buffer with proteinase K was removed and the plugs were washed 6 times by 2.5 ml of TE buffer each time incubating it for 15 min, gently shaking at room temperature. After the final wash 2.5 ml of sterile TE was added to each universal and the plugs were stored at 4°C until used. The agarose gel was then prepared for PFGE and the agarose plugs were cut to an appropriate size and loaded by hand (using the support of a scalpel) into the well of the gel. The final lane of the gel was loaded with a slice of *Saccharomyces cervisiae* PFGE size marker. The conditions used for the PFGE run were: switch time of 60 and 120 sec (initial and final respectively), 24 hours of gel run with 6V/cm and 120° angle. The gel was finally stained in ethidium bromide and visualised under a UV transilluminator.

Chapter 3. Molecular Mechanisms controlling expression of ETEC H10407 enterotoxins.

3.1 Introduction

In this chapter I have investigated the control of ETEC enterotoxin expression. I have shown that regulation of ST and LT enterotoxin expression is controlled by both CRP and H-NS. Consistent with this, expression of the toxins is controlled by the extracellular signals of osmolarity and glucose availability; signals sensed by H-NS and CRP respectively.

Previously, Haycocks *et al.* (2015) identified a potential binding site for CRP, in a class I position, upstream of the *estA1* and *estA2* genes (both encoding derivatives of the ST toxin). The authors show that CRP was able to bind *in vitro* using gel shift assays. Whilst CRP bound these sites *in vitro,* binding was not detected globally *in vivo* using ChIP-seq assays. Myers *et al.* (2013) observed a similar phenomenon for FNR, a regulator related to CRP. They show that FNR binding sites are revealed in Δ*hns* cells. Hence we did ChIP-seq assays, set up to investigate the global binding of H-NS, revealed H-NS bound across the *estA1, estA2* genes and promoter regions. Thus, it was suggested that these targets, bound by CRP *in vitro,* were occluded by H-NS *in vivo.* In this chapter I have tested the H-NS occlusion model outlined in Figure 3.1., explained the regulation of the toxins by glucose and salts and also explained the relevance of Oral Rehydation Therapy (ORT) in ETEC mediated diseases.

3.2 The *estA2* promoter is repressed by H-NS and activated by CRP

To examine the effect of H-NS on CRP binding we selected a 460 bp DNA fragment containing the *estA2* gene and regulatory DNA. We also made a derivative of the 460 bp fragment that lacked the CRP site and a shorter 93 bp fragment that contains the *estA2* promoter but not the adjacent *estA2* gene. The various fragments are shown in

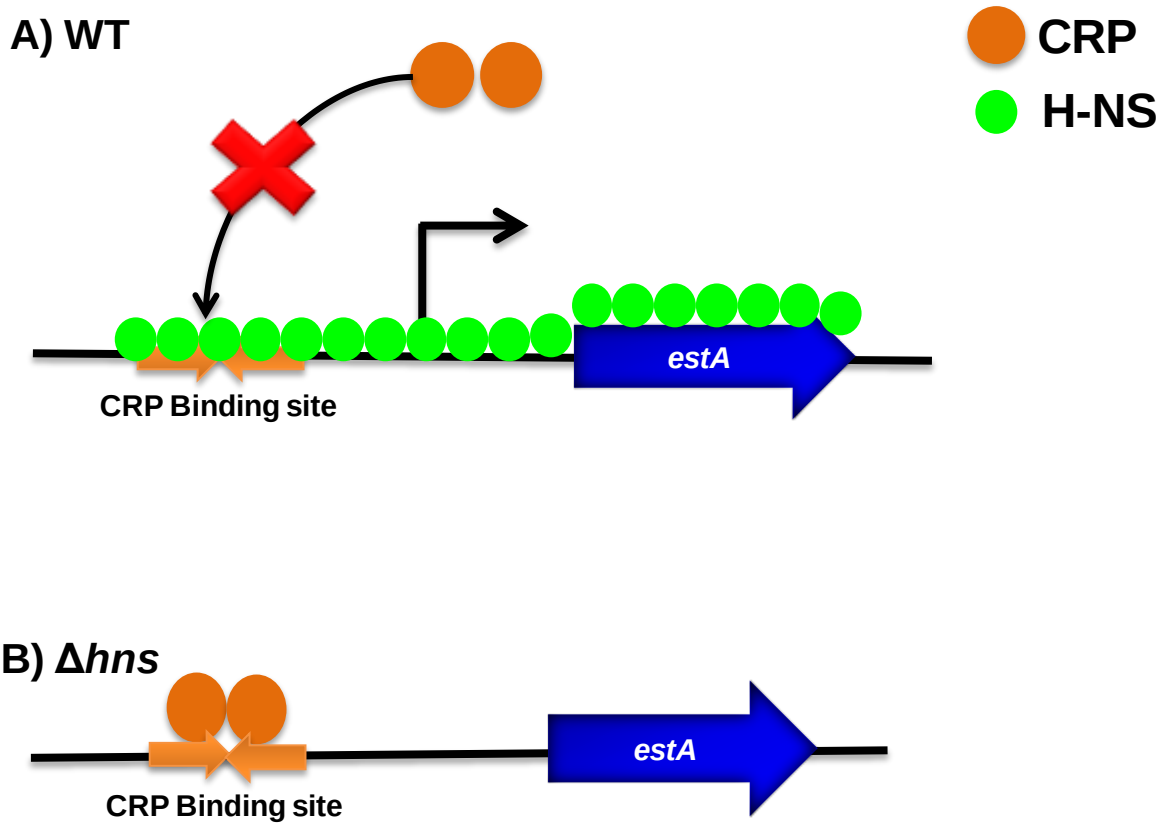
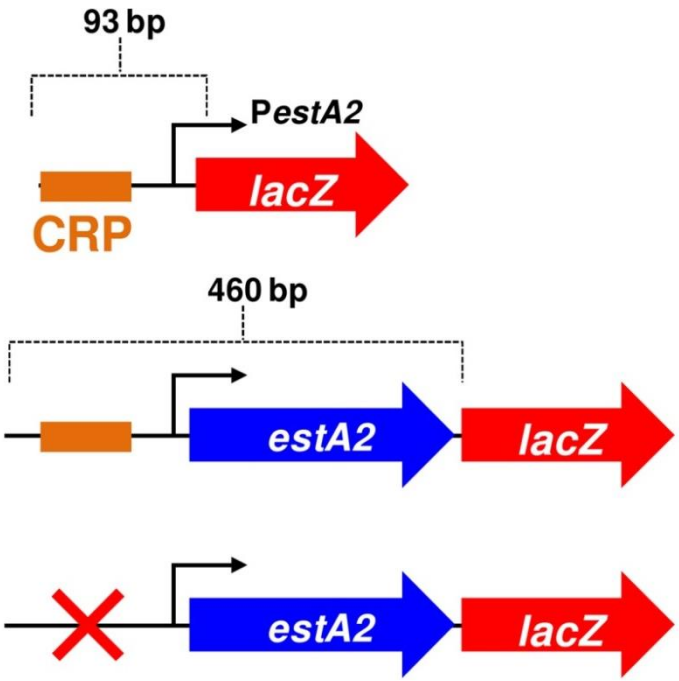


Figure 3.1: Defined hypothesis about *in vivo* regulation of enterotoxins by H-NS. In the figure we show the hypothesis derived for the region of *estA2* gene on p948 plasmid. A. In WT, when H-NS (Fluorescent Green) and CRP (Orange) both are present, H-NS nucleates the region across *estA2* and occludes CRP to regulate the expression of *estA2.* B. In Δ*hns* cells, CRP is the only regulator and is able to bind over the CRP site (Orange arrows) and regulate the gene expression.

A.



B.

LacZ activity (Miller Units)	
WT	**Δ*hns***
607 ± 5	596 ± 20
92 ± 3	465 ± 12
16 ± 4	89 ± 11

Figure 3.2: Investigation of LacZ activities (Miller units) from different promoter derivatives of *estA2*. A. The enterotoxin genes are shown in blue, the promoters are represented by black bent arrows, the CRP sites are shown in orange and the *lacZ* gene is in red. i. *estA2* promoter with 460 bp downstream flanking region containing *estA2*, ii. the 460 bp derivative of *estA2* lacking CRP site and iii *estA2* promoter alone **B.** The LacZ activity for each of the fragments cloned in front of *lacZ* gene over the pRW50 plasmid in M182 (Wild type (WT)) and M182 (H-NS knock out (*hns::kan)*). n=3.

(Figure 3.2A). It has been shown previously that promoters present in short DNA fragments are liberated from repression by H-NS (Singh *et al.,* 2014). Hence, we expected that the promoter with flanking coding DNA would be repressed by H-NS. To test this we cloned the various DNA fragments upstream of *lacZ* in a reporter plasmid pRW50. The LacZ activities obtained using the various plasmids in strain M182 or M182 *hns::kan* are shown in Figure 3.2B. In cells with H-NS, the β-galactosidase activity of the 93 bp fragment is 6-fold higher than the 460 bp fragments. This was expected; the 93 bp fragment contains the *estA2* promoter but no 3' flanking DNA where H-NS can bind to repress P*estA2.* In the absence of H-NS the 93 bp fragment and the wild type 460 bp fragment have similar activity but mutation of the CRP binding site completely eliminates the activity of *PestA2.*

3.3 The *estA2 and estA1* are similarly regulated by H-NS

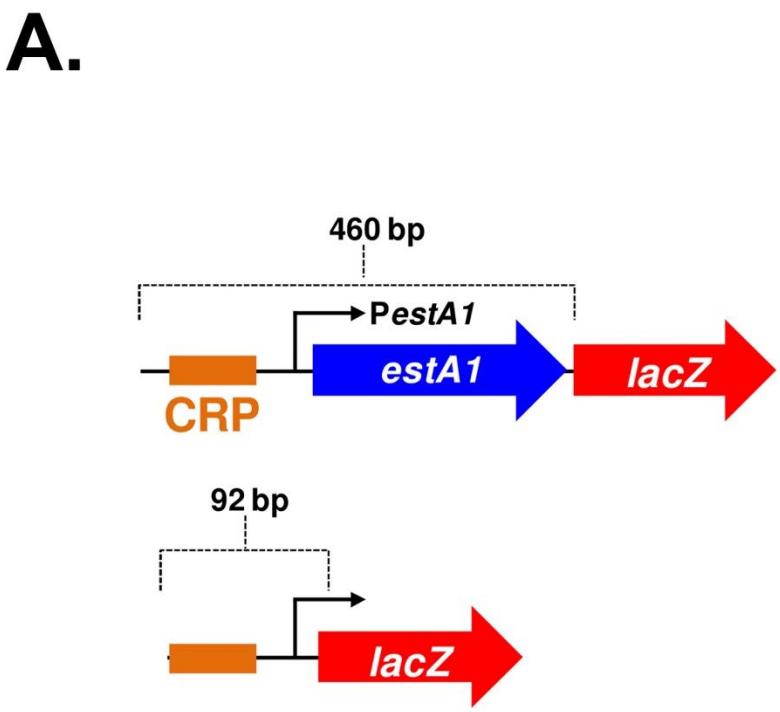
The regulatory region of *estA1* carries a sequence very similar to *PestA2.* We predicted that this would be *PestA1* and reasoned that 3' flanking DNA might again bind H-NS *in vivo.* Hence, we cloned P*estA1* with or without the adjacent *estA1* gene fused to *lacZ* in plasmid pRW50 (Figure 3.3A). M182 and M182 *hns::kan* were transformed with the resulting plasmids and LacZ activities were measured. As predicted, in the presence of flanking DNA, transcription from *PestA1* was repressed by H-NS but not in the absence of flanking DNA (Figure 3.3B). Note that in comparison to P*estA2, PestA1* is poorly active and not dependent on CRP. This is because the CRP site at *PestA1* is misplaced (Haycocks *et al.,* 2015).

3.4 The *eltAB* operon is directly repressed by H-NS

The A and B subunits of the LT toxin are encoded by the *eltAB* operon. The ChIP-seq data of Haycocks *et al* (2015) also suggest binding of H-NS over the *eltAB* operon. Hence, we reasoned that the *eltAB* (*PeltAB*) promoter may also be repressed by H-NS. To test this prediction, we prepared two DNA fragments; a 359 bp fragment containing only *PeltAB,* and a 1127 bp fragment also containing the *eltA* and part of *eltB* (Figure 3.4A). Note that P*eltAB* contains sequences that bind CRP (open orange boxes in Figure 3.4A) but binding is poorly specific (Haycocks *et al.,* 2015). The DNA fragments were cloned in plasmid pRW50 and used to transform M182 and the *hns::kan* derivative. The LacZ activity was determined and the result is illustrated in Figure 3.4B. Consistent with our prediction, LacZ activity dropped 15 fold when the flanking DNA downstream of *PeltAB* was present. There was a relief of this repression in the absence of H-NS (Figure 3.4B).

3.5 Similar H-NS binding on all the enterotoxin promoters of ETEC H10407

The LacZ activity data suggests that H-NS does not bind the short *estA2* and *eltAB* fragments. To test this, H-NS binding to the various fragments was compared using ChIP and PCR. The result of the PCR, is shown in Figure 3.5. As predicted, the longer *estA2* and *eltAB* fragments were enriched in the immunoprecipitations with anti-H-NS (bottom panel). However, no enrichment was seen for the shorter *estA2* and *eltAB* fragments (top panel). The *yabN* locus, that does not bind H-NS, served as a control in the PCR reactions and immunoprecipitations.



B.

LacZ activity (Miller Units)	WT	Δ*hns*
	96 ± 9	321 ± 32
	92 ± 4	81 ± 3

Figure 3.3: Investigation of the LacZ activities (Miller units) from promoter derivative of *estA1*. A. *estA1* gene shown in blue, promoter shown by a black bent arrow, and *lacZ* gene shown in red. i. *estA1* promoter with 460 bp downstream flanking region containing *estA1* and ii P*estA1* on its own. **B.** LacZ activity for *estA2* fragment cloned in front of *lacZ* over pRW50 plasmid in M182 (WT) and M182 *hns::kan.* Representative of n=3.

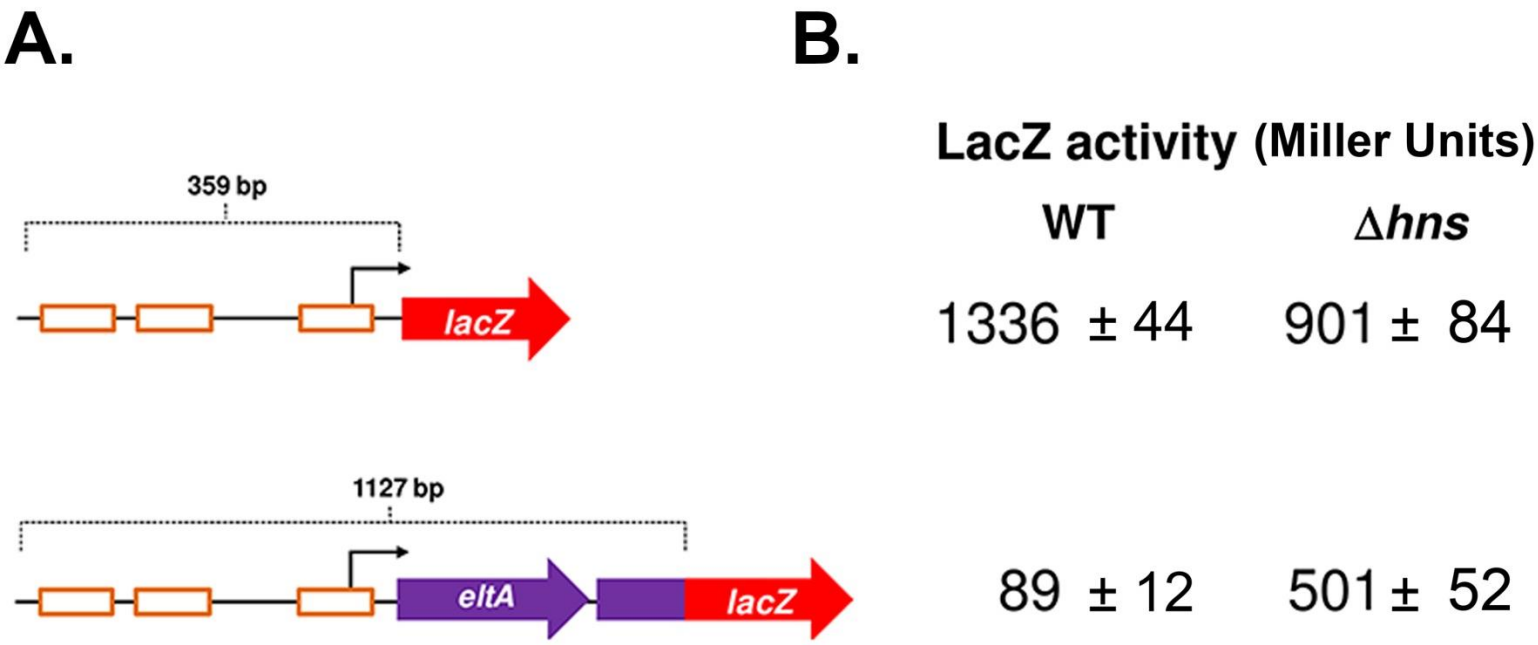


Figure 3.4: Investigation of the LacZ activities (Miller units) from different promoter derivatives of *eltAB*.

A. The enterotoxin genes are shown in purple, their promoters are shown by a black bent arrow, the CRP sites are small orange boxes and the *lacZ* gene is red. i. *eltAB* promoter alone with its CRP sites, ii. *eltAB* promoter with 460 bp downstream flanking region containing *eltAB*. **B.** LacZ activity for each derivative of pRW50 cloned in front of the *lacZ* gene and transformed in M182 (Wild Type (WT)) and M182 (H-NS knockout strain (*hns::kan)*). Representative of n=3.

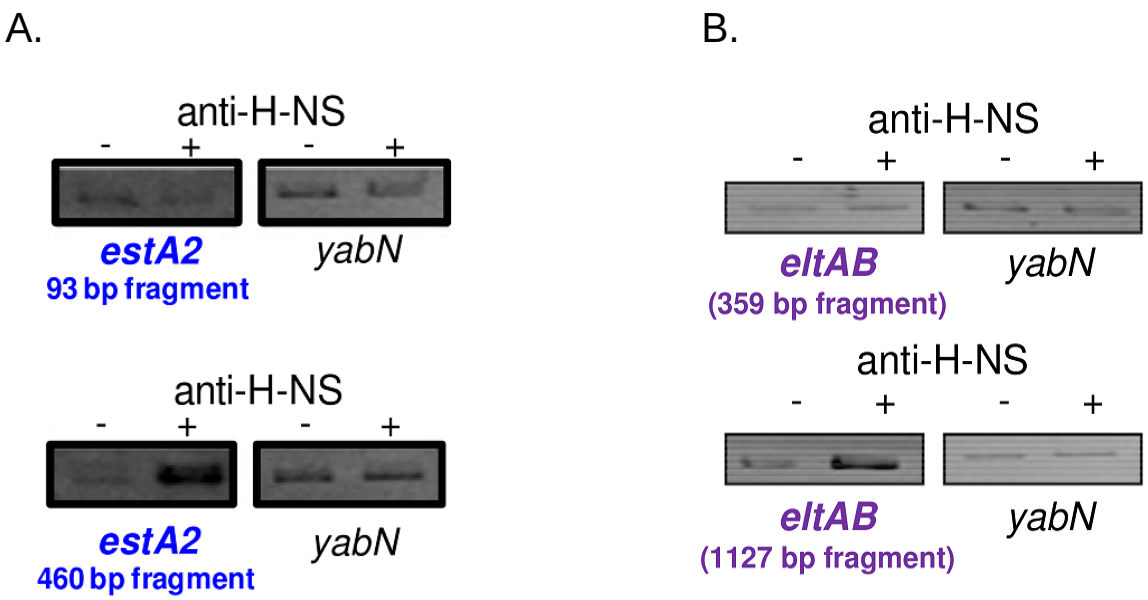


Figure. 3.5: The various enterotoxin promoters of ETEC H10407 are repressed by H-NS. A. H-NS binding to the P*estA2* which includes the gene downstream. B. H-NS binding to the P*eltAB* in the presence of flanking DNA, which includes the gene downstream. ChIP-PCR was used to detect the binding of H-NS to the different promoter fragments. Primers used in ChIP-PCR were able to detect P*estA2* in both derivatives, 93 bp and 460 bp respectively, and were also able to detect P*eltAB* in both derivatives of 359 bp and 1127 bp respectively. *yabN* was used as a control locus and no enrichment observed. Representative of n=3.

3.6 The *estA1* and *estA2* promoters of ETEC H10407 respond to Glucose and Salt

According to our model, CRP is vital for transcription of *estA2* and H-NS represses this transcription. This suggests that *estA2* is regulated by both glucose and osmolarity. The transcriptional response to glucose, controlled via cAMP, is mediated by CRP (Busby & Ebright, 1999). Hence, when glucose is absent, the cAMP level increases and CRP binds DNA (Parkinson *et al.,* 1996). Moreover, the way in which H-NS modulates DNA topology depends on the abundance of divalent cations (i.e. presence of salts) (Atlung and Ingmer, 1997). Hence, H-NS mediated repression should be relieved by increasing osmolarity.

To check the response of P*estA1* and P*estA2* to glucose and salt, we established a range of conditions across which the *estA1* and *estA2* promoters responded to these signals. We used DNA fragments that included the downstream flanking region to ensure that signals sensed by both CRP and H-NS could be integrated. Note that we expected little response from *PestA1* because of the misplaced CRP site (Haycocks *et al.,* 2015). Cells were grown in M9 minimal medium with varying concentrations of glucose or salt (a combination of NaCl and KCl). The LacZ activity data is shown in Figure 3.6A and B. As expected, the activity of P*estA1* was low and the effects of glucose and salt were negligible (Figure 3.6 A). Conversely, a response of *PestA2* to both glucose and salt was detected. Hence, *PestA2* was repressed by increasing concentrations of glucose (orange line) and activated by increasing concentrations of salt (green line) in Figure 3.6B. A cartoon is also shown in Figure 3.7 that depicts the control of *estA2* expression by glucose. Thus, H-NS is able to repress the P*estA2* promoter but activity is always dependent on CRP. We therefore predicted that the

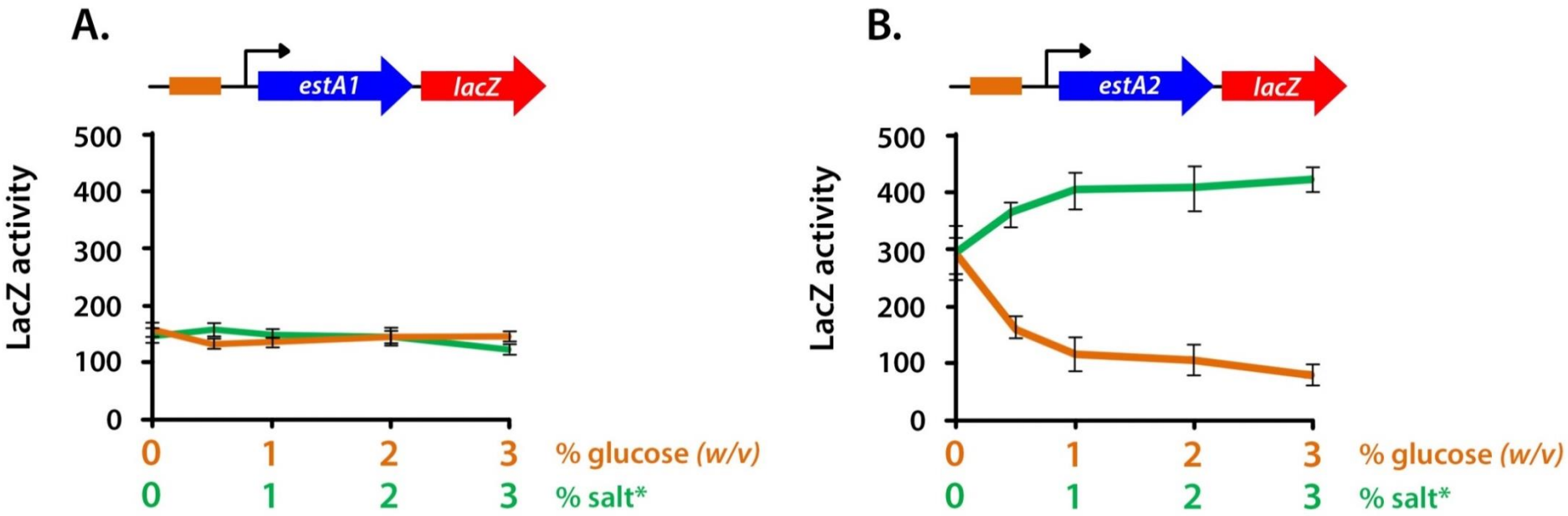


Figure 3.6: Effect of titrating glucose and salts individually over the *estA1* and *estA2* promoter. This figure shows the activity observed as a response of increasing glucose and salts concentration over the different promoter::*lacZ* fusions of A) P*estA1,* B) P*estA2*. These promoters include downstream flanking regions containing enterotoxin genes. Representative of n=3.

*** NaCl and KCl was mixed in a ratio of 3:1 together. The % *w/v* refers to the NaCl concentration, rather than the total concentration of the solution.**

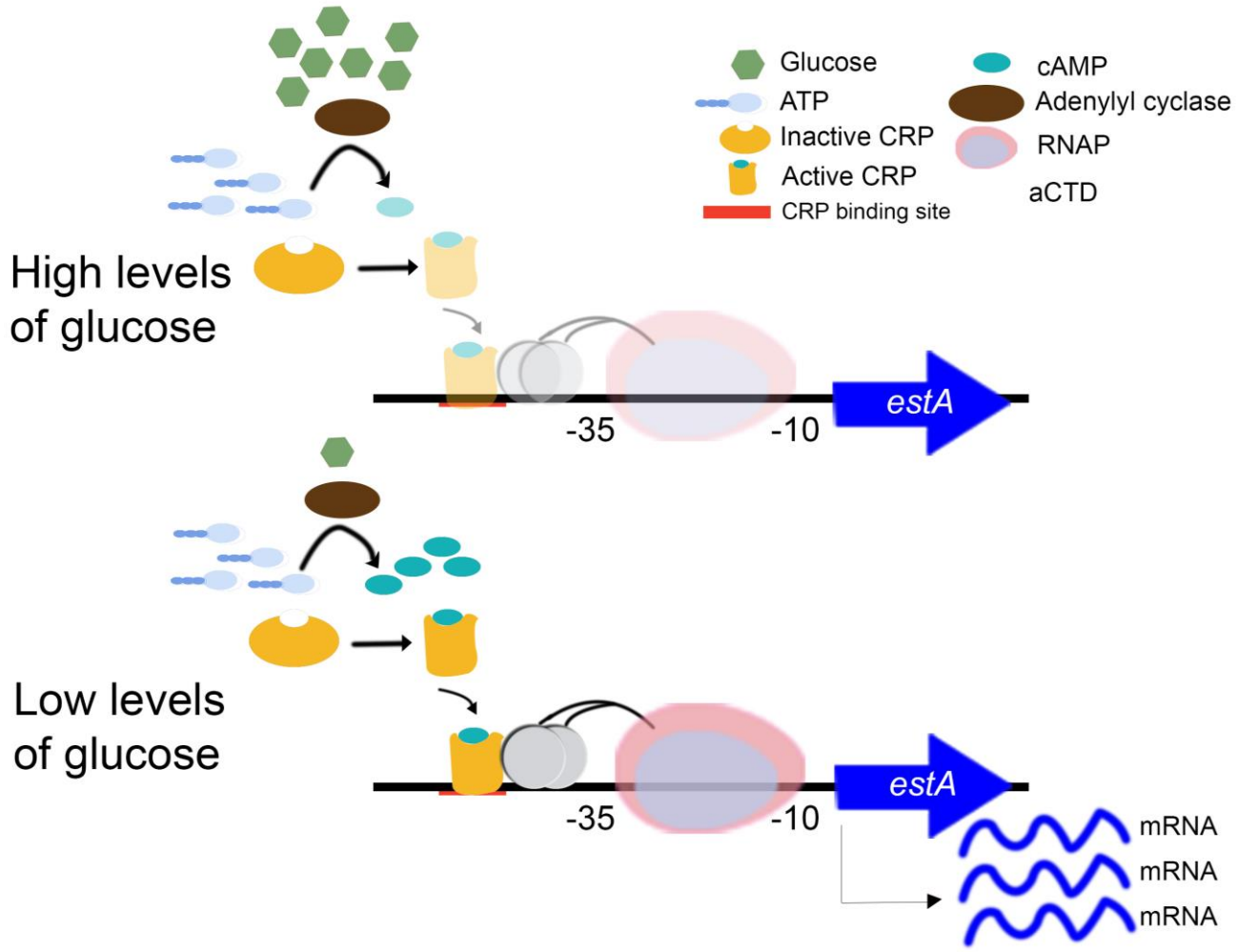


Figure 3.7: Control of production of ST toxin by glucose. A. High glucose levels, no expression of toxin is seen as the CRP is absent and cannot bind the promoter region. B. Low levels of glucose, CRP activates *estA* promoter and toxin is produced shown as blue curve represented as mRNA.

inhibitory effect of glucose must override the stimulatory effect of salt. To investigate this we added the glucose and salt individually or mixed together. Consistent with our hypothesis, the inhibitory effect of glucose was dominant (Figure 3.8A) and occurred even in the absence of H-NS (Figure 3.8B). Conversely, the stimulatory effect of salt required H-NS (compare green bars in Figure 3.8A). To determine if the *PestA2* glucose response required CRP, we compared activity of the 460bp *estA2* fragments with or without the CRPsite. As expected, the glucose effect is present only for the fragment where the CRP site is intact (compare Figure 3.9A and Figure 3.9B).

3.7 The *eltAB* promoter also responds to glucose and salts

The activity of P*eltAB* was found to increase in the presence of increasing concentrations of both the glucose and salt. The enhancement in activity by salt was more prominent compared to the effect of glucose (Figure 3.10). We also tested combinations of glucose and salt in the presence and absence of H-NS. The results suggest that effects of salt and glucose both require H-NS and were not additive (Compare Figure 3.11A and B). Since CRP is unlikely to regulate *PeltAB* directly, the effect of glucose is difficult to interpret. Conversely, the effect of salt is likely mediated by H-NS.

3.8 Regulation of ETEC enterotoxins by H-NS and CRP is conserved in other strains

On examining all available ETEC genomes we found that the promoter sequences of *estA2* and *eltAB* varied. Hence, we set out to determine if the response of *estA2* and *eltAB* to CRP and H-NS was conserved. We selected promoters from ETEC E24377A

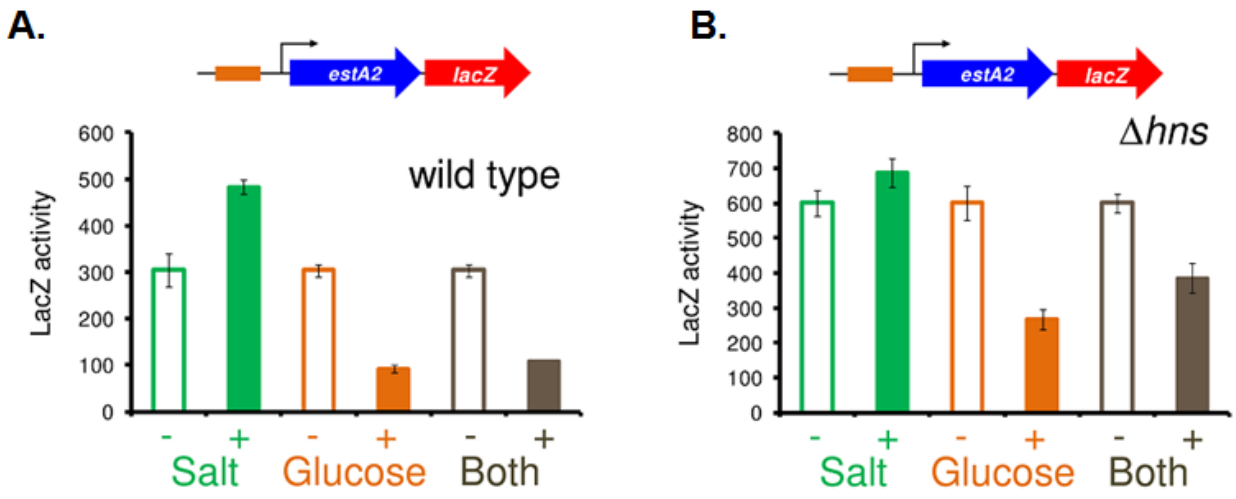


Figure 3.8: Production of ST toxin of ETEC H10407 is controlled by an osmo-metabolic flux. This figure shows the effect of 2% glucose, 60 mM NaCl and 20 mM KCl over the ST production. The effect was measured by β-galactosidase assay. Briefly cells were first grown in M9 minimal media so that glucose and salt concentrations could be more accurately controlled. The measurement was done for lysates of A. M182 (Wild Type) or B. M182 *hns::kan* (Δ*hns)* containing P*estA2*::*lacZ* fusion construct in reporter plasmid pRW50. Representative of n=3.

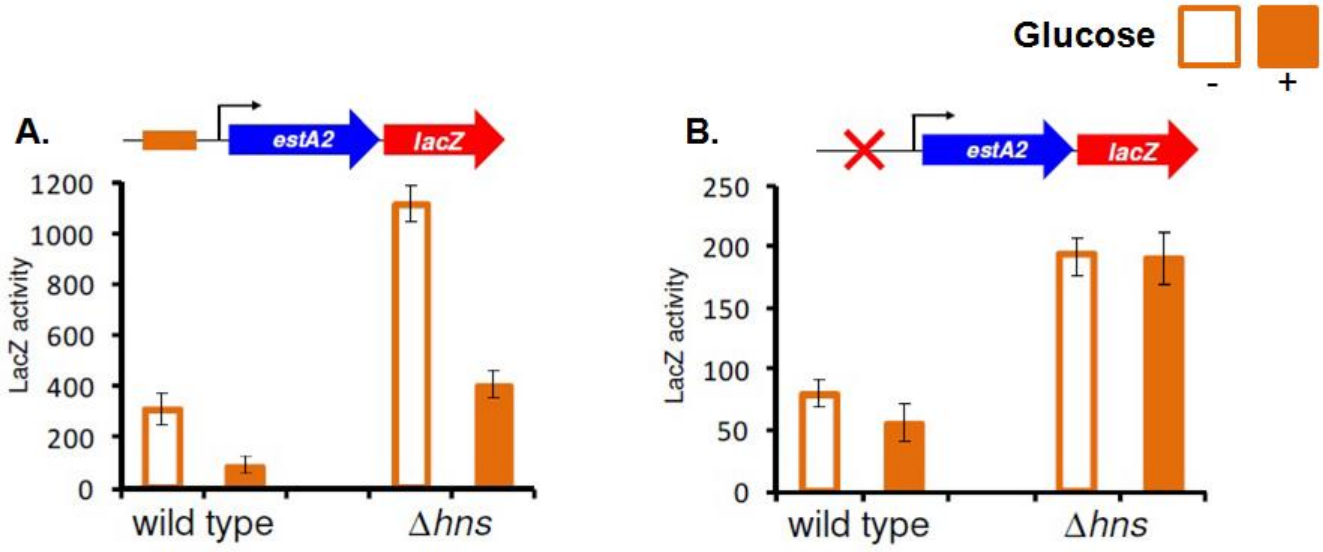


Figure 3.9: The response from glucose needs an intact CRP site over the P*estA2* promoter. A.This figure shows activity of the P*estA2* fragment with the downstream flanking DNA containing *estA2* toxin gene. This activity assay was carried out in M9 minimal media with and without 2% glucose in both WT and Δ*hns* cells. B. This figure shows activity of the P*estA2* fragment including the downstream flanking DNA containing ST toxin gene but lacks the CRP site. Representative of n=3.

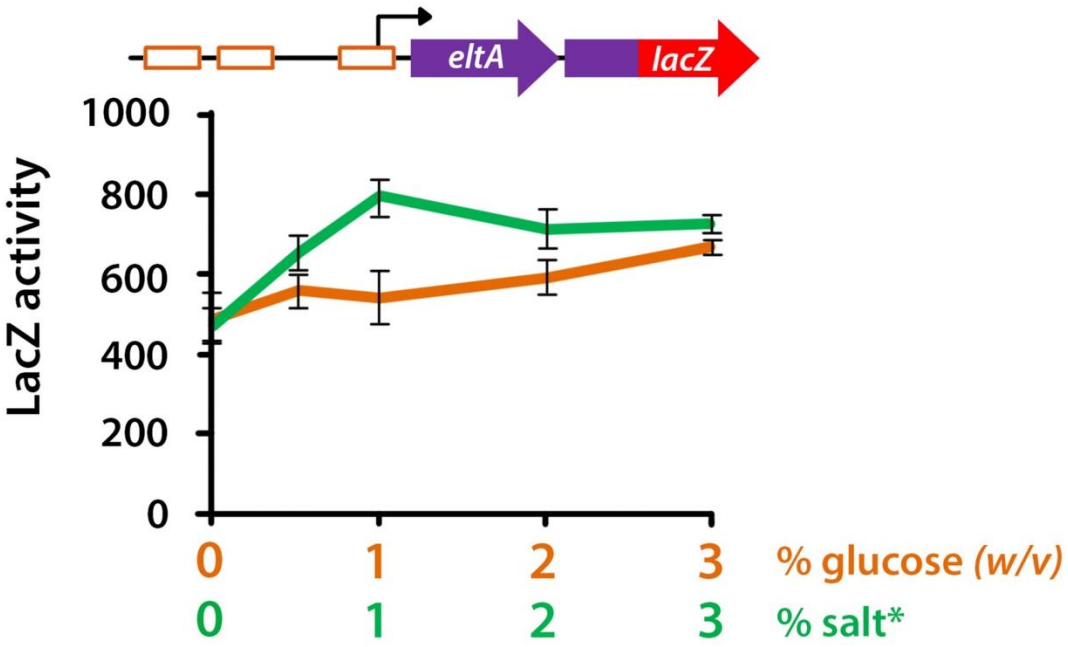


Figure 3.10: Effect of titrating glucose and salts individually over the *eltAB* promoter. This figure shows the activity observed as a response of increasing glucose and salts concentration over the promoter::*lacZ* fusions of P*eltAB* in the reporter plasmid pRW50. The promoters include the downstream flanking regions containing enterotoxin gene. Representative of n=3.

*** NaCl and KCl was mixed in a ratio of 3:1 together. The % *w/v* refers to the NaCl concentration, rather than the total concentration of the solution.**

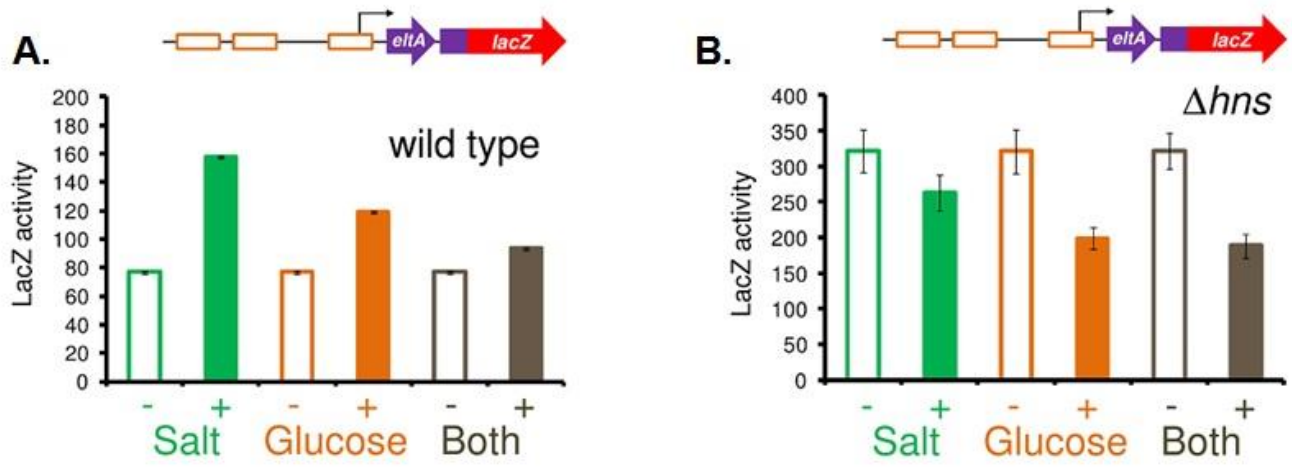


Figure 3.11: Production of LT toxin of ETEC H10407 is controlled by an osmo-metabolic flux. This figure shows effect of 2% glucose, 60 mM NaCl and 20 mM KCl over the LT production. The effect was measured by β-galactosidase assay. Briefly, cells were first grown in M9 minimal media so that the glucose and salt concentrations could be more accurately controlled. The measurement was done for lysates of A. M182 (Wild Type) or B. M182 *hns::kan* (Δ*hns)* containing A) P*estA2*::*lacZ* fusion construct in reporter plasmid pRW50. Representative of n=3.

that has a sequenced genome and transcriptomics data (Sahl and Rasko, 2012, Kansal *et al.,* 2013). I generated ETEC E24377A *estA2* and *eltAB* fragments equivalent to those from H10407 and cloned these in front of *lacZ* gene in pRW50. Promoter activity was then measured in response to CRP and H-NS. The regulation was found to be similar to that of H10407; *estA2* was repressed by H-NS and activated by CRP whilst *eltAB* was repressed by H-NS (Figure 3.12A and B). The global transcriptome response of E24377A to glucose and salts has already been examined upon attachment to epithelial cells (Sahl and Rasko, 2012). I made use of their transcriptomics data and re-analysed it relative to my data. The data in figure 3.13A show changes in transcription of *crp*, *hns*, *eltAB* and *estA2* during attachment. Levels of *hns* transcription changed dramatically after 1 hour post intestinal cell attachment and both *eltA* and *estA2* transcription correspondingly increased. Levels of *crp* did not change during host cell attachment but we did note a variation in *crp* transcript levels between experimental variants. Hence, when these absolute levels of *crp* transcripts were plotted against *estA2* transcript levels, a linear relationship was found. For each data point in Figure 3.13B the absolute level of *hns* mRNA is added in parenthesis. The plot show 2 outliers in the data points, they are from those samples which had increased *hns* expression.

3.9 Discussion

Enterotoxin production can be controlled by CRP and H-NS in ETEC H10407 and our model of regulation is applicable in other ETEC strains. Our model suggests activation of heat-stable toxin expression by CRP and repression of both heat-stable and heat-labile toxin expression by H-NS (Figure 3.14). Thus, glucose and salt affect enterotoxin expression.

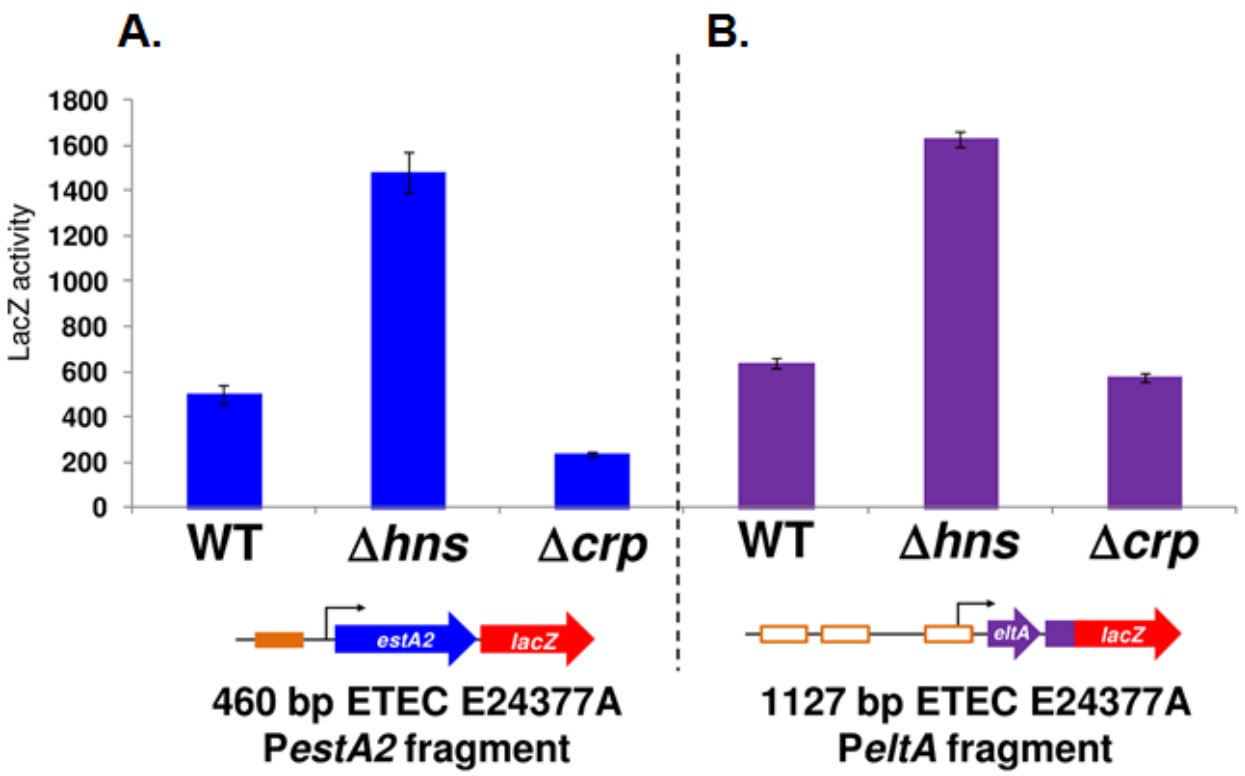


Figure 3.12: Transcription levels of enterotoxins of ETEC E24377A A. Activity measurements by β-galactosidase assay in WT (M182), Δ *hns* and Δ *crp* cells containing the *estA2* fragment cloned in front of lac*Z* in pRW50 plasmid. B) Activity measurements by β-galactosidase assay in WT (M182), Δ *hns* and Δ *crp* cells containing the *eltAB* fragment cloned in front of lac*Z* in pRW50 plasmid. Representative of n=3.

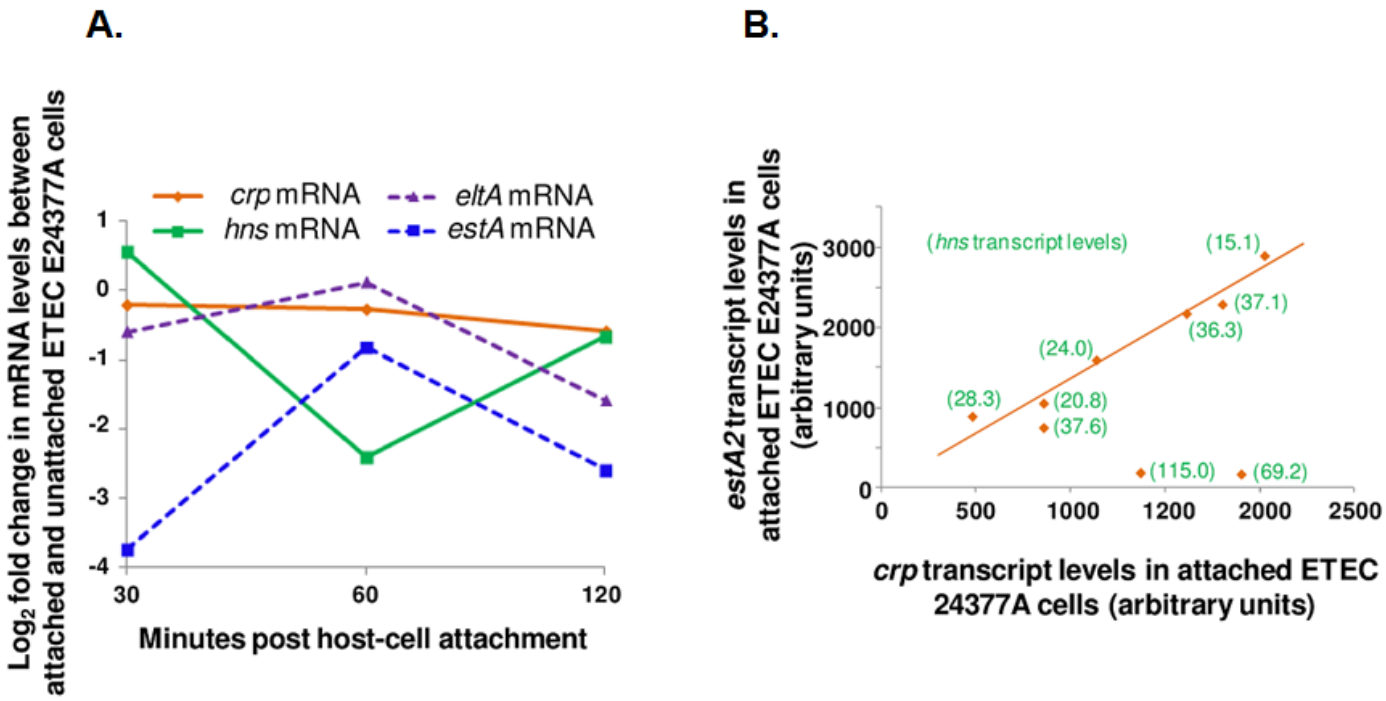
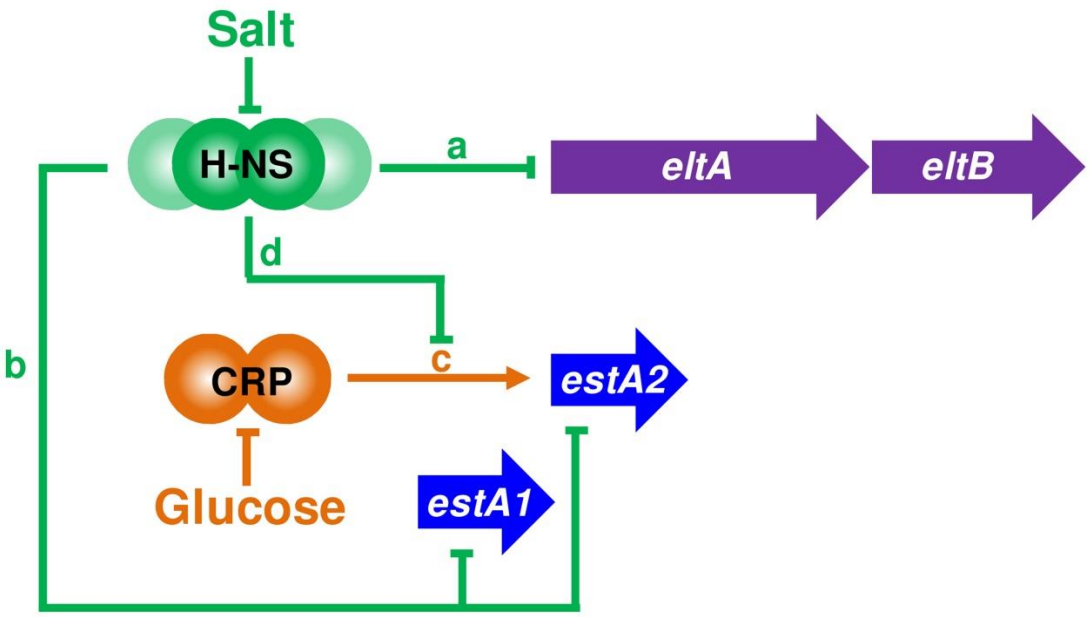


Figure 3.13: Transcription level changes in the transcriptome of ETEC E24377A post intestinal cell attachment. A. Log_2 fold change in transcript levels of enterotoxins (Dashed purple and blue lines for *estA2* and *eltAB* respectively), *crp* (orange) and *hns* (green) during non-attachment and post attachment to the intestinal cells. B) The absolute transcripts were plotted for *crp* and *estA2*, which shows a clear linear relation between them (*hns* transcripts are shown in brackets).



a. *eltAB* is directly repressed by H-NS

b. *estA1* and *estA2* are directly repressed by H-NS.

c. *estA2* is directly activated by CRP.

d. CRP site occluded by H-NS (a case of indirect repression by H-NS).

Figure 3.14: A model proposed for the control of enterotoxins expression of ETEC. The LT and ST can be controlled by the two gene regulators H-NS and CRP. The glucose and salt can be sensed by them and can help in control of ST and LT production.

Gene expression can be modulated by glucose and salts when regulators like CRP and H-NS are involved in regulation. In case of CRP, activation only occurs when there are enough cAMP molecules that are bound to CRP. Catabolite repression is the major inhibitory effect of glucose present in *E. coli.* Glucose present in high concentration prevents the inducer uptake inside the cells resulting in negatively regulating gene expression (Kolb *et al.,* 1993). Tagami *et al* (1995) have shown that glucose reduces the CRP expression from a multicopy plasmid (Tagami *et al.,* 1995). In case of H-NS, using single-molecule studies Amit and coworkers (2003) have shown that independent of H-NS concentration and with increasing concentration of salts, H-NS loses its effect (Amit *et al.,* 2003). Salts disrupt the binding between H-NS and DNA and similar case was also observed in its paralogue StpA (Lim *et al.,* 2012).

The observed regulation pattern in this chapter is pertinent to the disease and its treatment. When ST and LT bind to epithelial cells in small intestine, there is an efflux of water, charged ions and cAMP. This efflux could maintain a positive feedback loop for toxin production. Interestingly, Oral Rehydration Therapy (ORT), which contains necessary electrolytes, is given with water to the patients colonised by ETEC (Nalin *et al.,* 1968). The ORT solutions contain glucose and salt. Glucose is present in ORT at 10 fold higher concentrations than those necessary to repress *estA2* via CRP. Similarly, salt concentrations in ORT are sufficient to induce expression of *estA2* and *eltAB*, though the high glucose concentration present for *estA2* sites in ETEC H10407 appear occluded by H-NS (Haycocks et al, 2015) and similar repression by H-NS was proposed for a CRP homologue FNR (Myers *et al.,* 2013). The authors have shown that in the Δ*hns* background there is a sudden increase of FNR binding sites. Therefore it is

believed that ORT has relevance in diarrhoeal diseases at least that are caused by ETEC strains.

Chapter 4. Identification and verification of MarA binding sites across the ETEC H10407 genome

4.1 Introduction

MarA is an AraC/XylS family transcription factor implicated in multiple-antibiotic resistance (mar). Thus, MarA upregulates genes involved in drug efflux to protect cells from antibiotics. MarA binds degenerate target site and deleting *marA* has pleiotropic effects on gene expression (Veleba *et al.,* 2012). Also the binding site of MarA; the marbox is orientation specific and Jair *et al.* (1995) have shown that inverting the marbox results in complete loss of activity. Due to these reasons, it has been difficult to find additional targets for MarA. Hence, in this chapter, chromatin immunoprecipitation and next generation Ilumina sequencing (ChIP-seq) was used to map MarA and RNAP σ^{70} subunit binding across the ETEC H10407 genome. The ChIP-seq was done in presence of a MarA overexpressing plasmid known as pRGM9818.

4.2 MarA binding sites across the whole genome of ETEC H10407

ChIP-seq was done using ETEC H10407 that shares 3,766 genes with *E. coli* K-12. Hence, ETEC H10407 specific targets of MarA, and sites shared with *E. coli* K-12, should be identified by our analysis. After immunoprecipitation, sequencing and bioinformatic analysis we identified 33 and 839 binding peaks for MarA and σ^{70} respectively. The binding data of MarA (green) and σ^{70} (orange) are plotted against the ETEC chromosome and plasmids in Figure 4.1. In each plot, tracks 1 and 2 are the genes, track 3 is the MarA binding data and track 4 is the σ^{70} binding data. Comparing tracks 3 and 4, shows that MarA bound to very few loci compared to σ^{70}. A DNA sequence motif associated with MarA binding peaks was generated using Multiple Em for Motif Elicitation (MEME) (Figure 4.2A). The ChIP-seq derived motif and known MarA binding sites closely resemble each other (Compare Figure 4.2A and B). The 33 MarA

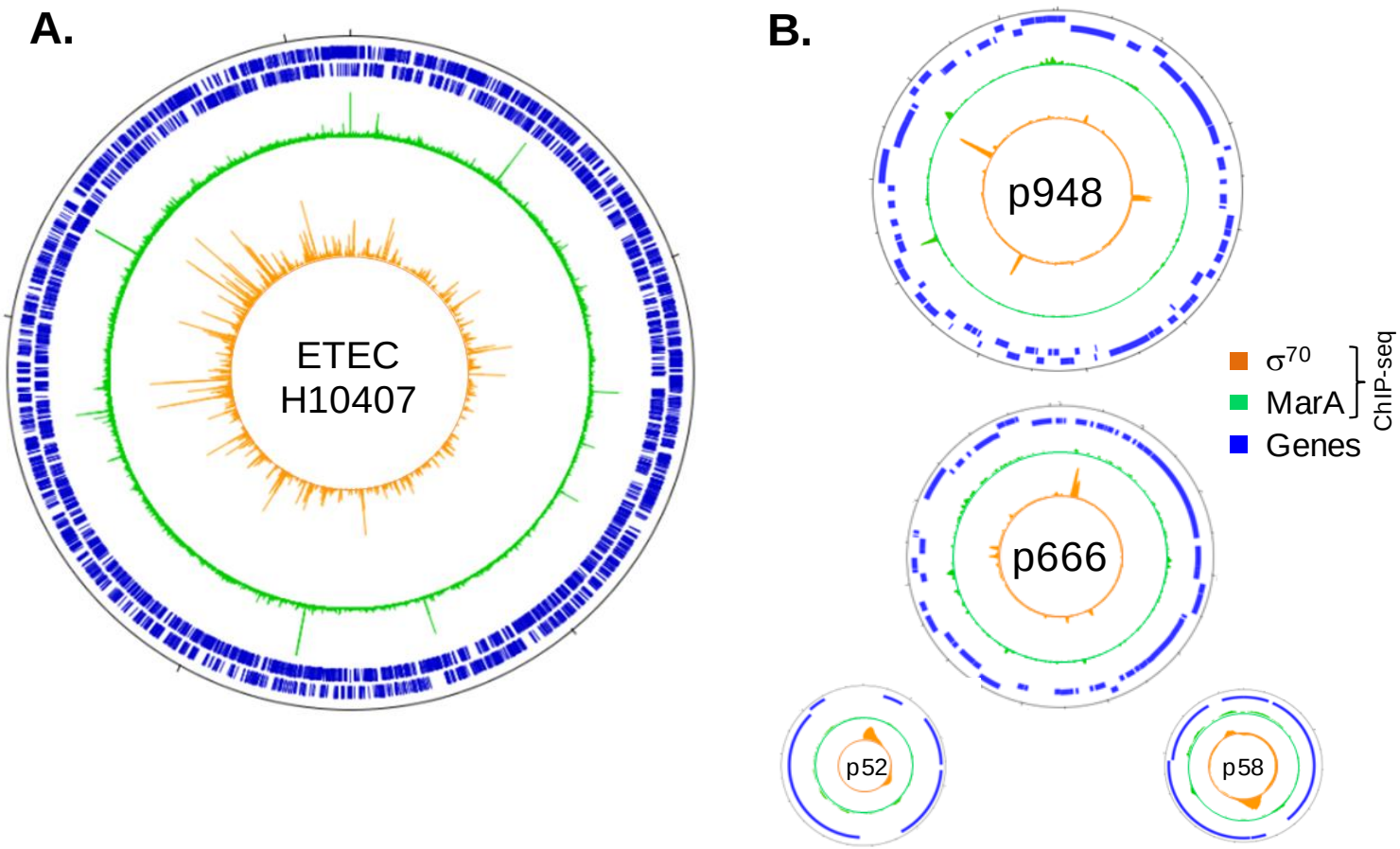


Figure 4.1: MarA and RNAP σ70 binding A. MarA and σ^{70} binding across the chromosome and B. the binding on all the plasmids of ETEC H10407. In each plot, genes are represented by track 1 and 2 in blue, MarA binding data is in green (track 3), and RNAP σ^{70} in orange (track 4).

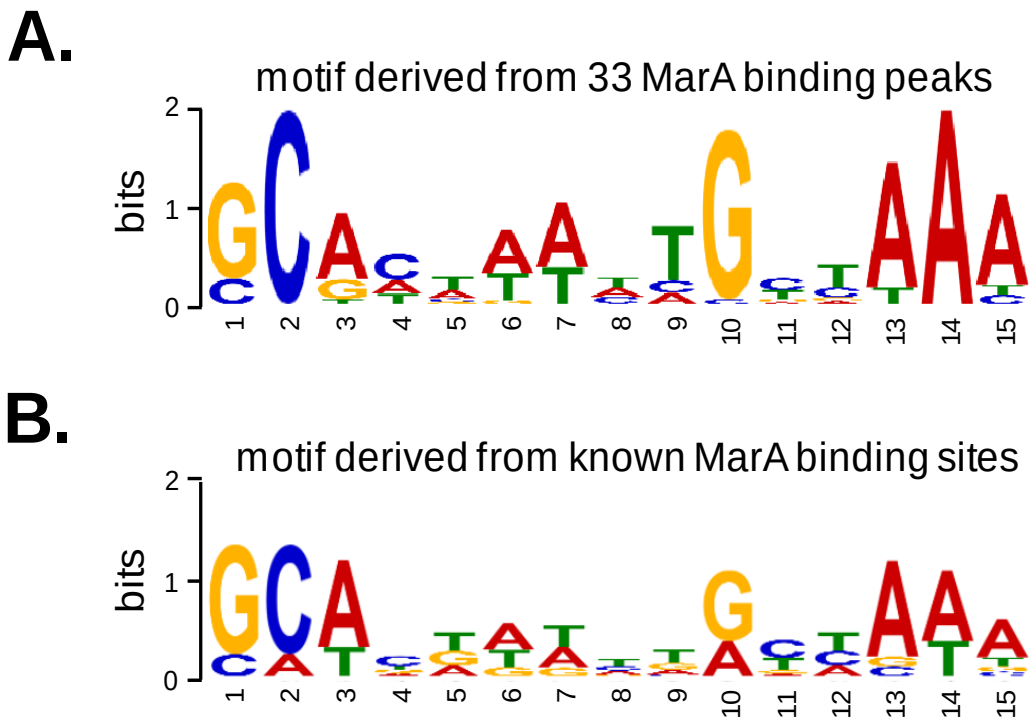


Figure 4.2: DNA sequence motifs derived from MarA binding sites. A. DNA sequence motif generated by the MarA binding data from the ChIP-seq experiment using MEME. B. DNA sequence motif generated by aligning the MarA binding sites listed in the Ecocyc database.

binding targets and associated motifs are shown in Table 4.1. Each binding peak for MarA and σ^{70} was allocated a 100 bp bin. The 100 bp bin is determined by the distance between the peak and the nearest start codon. The result is shown in the form of a histogram in Figure 4.3A. It was found that both MarA and σ^{70} bind the 100 bp upstream of the 5' end of a gene most often. Of the 33 MarA sites, 15 were within 150 bp of a peak for σ^{70} binding. A Venn diagram that illustrates the overlap in MarA and σ^{70} binding is shown in Figure 4.3B.

4.3 *In vitro* characterization of MarA bound targets

To validate the MarA binding data from our ChIP-seq analysis, we tested the binding of purified MarA to DNA fragments derived from each peak. As a control we also tested 5 DNA fragments from elsewhere in the genome including the 92 bp *estA1* fragment. The DNA binding was measured using Electrophoretic Mobility Shift Assays (EMSA). Thus, 130 bp DNA fragments, corresponding to each MarA target, were generated using PCR and then radio labelled. The results are shown in Figure 4.4. Only 1 of the MarA targets (ETEC3200) did not bind the MarA protein *in vitro.* As expected, the control fragments did not bind MarA (bottom panel in Figure 4.4).

4.4 The *mar* regulon is conserved amongst the family of enteric bacteria

In many enteric bacteria overexpression of MarA-like regulators is associated with multidrug resistance. Hence, we next investigated the conservation of MarA binding targets amongst enteric bacteria containing an unambiguous MarA homologue. The results are represented in the Figure 4.5. The MarA regulon of *E. coli* was found to be best conserved amongst other *Escherichia spp*. and *Shigella spp*. In more divergent enterobacteriaceae, a conserved 'core' regulon is apparent. For example, *Klebsiella*

Table 4.1: MarA binding sites identified by ChIP-seq

ChIP peak[a]	MEME site centre[b]	Site Sequence (5'-3')[c]	H10407 genes[d]	MG1655 genes[e]	MarA binding *in vitro*[f]
Chromosomal Targets					
206	161	gcacagacagataaa	ETEC0001	*thrL*	+++
87300	87344	gcacaattagctaat	ETEC0074<>ETEC0075	*leuL*<>*leuO*	+++
184212	184180	gcgttatctgttaat	ETEC0157	*degP*	++
428898	428876	gcataaagtgtaaag	ETEC0400	*lacZ*	++
529310	529338	gcacaaaatgacaaa	ETEC0500	*ybaO*	+++
655542	655557	gcactaaatgttaaa	ETEC0604	*pheP*	+++
846408	846418	ccacgcaaagctgac	ETEC0765<>*acrZ*	*modE*<>*acrZ*	++
963300	963348	cctatgagcgtaaaa	ETEC0889	*ybiV*	+
994036	994060	gcattaattgctaaa	ETEC0916<>ETEC0917	*grxA*<>*ybjC*	+++
1354006	1354020	gcactaattgcaaaa	ETEC1264<>ETEC1265	*ycgF*<>*ycgZ*	+++
1536992	1537038	gcacaaattgtttaa	ETEC1438	*fnr*	++
1717096	1717112	gcactaattgctaaa	ETEC1580	*yneO*	+++
1739002	1739016	ccacgttttgctaaa	ETEC1599<>ETEC1600	*marC*<>*marR*	+++
2321000	2321062	gcactatttgctaaa	ETEC2157	*yeeF*	+++
2538709	2538641	gcactgaatgtcaaa	ETEC2344<>*micF*	*ompC*<>*micF*	++
2727890	2727941	gcatttttgctaaa	ETEC2509	*ypeC*	+++
2755444	2755459	gcaacaactgttaaa	ETEC2533><ETEC2534	*yfeS*><*cysM*	+++
2887268	2887307	gcatttttgcaaaa	ETEC2665<>ETEC2666	*guaB*<>*xseA*	+++
3455714	3455708	ccaatatccggcaaa	ETEC3200	ETEC specific	-
3569696	3569763	gcacgtaacgccaac	ETEC3306<>ETEC3307	*nudF*<>*tolC*	++
3695690	3695710	gcacaatctgcttac	(ETEC3426)	(*yhbV*)	+++
3733124	3733195	ccagctttcgctaac	ETEC3460<>ETEC3461	*mlaF*<>*yrbG*	+++
4289772	4289787	gcacgaaacgttaaa	ETEC3977<>ETEC3978	*ibpA*<>*yidQ*	++
4348148	4348176	gcacgatctgtatac	ETEC4032	*mnmG*	++
4494984	4495025	ccgctttacggtaaa	(ETEC4151)	(*yihT*)	++
4510208	4510184	gcgcgttatgctgac	(ETEC4166)	(*yiiG*)	++
4685066	4685041	aggctaatcgtataa	(ETEC4304)	ETEC specific	+++
4686378	4686377	ccaaaaacaggtaaa	(ETEC4307)	ETEC specific	++
4737304	4737238	gcaataaaagtcacg	ETEC4370<>ETEC4371	*yjcB*<>*yjcC*	++
5066076	5066105	gcatcaaatgataac	ETEC4666<>ETEC4667	*yjjP*<>*yjjQ*	+++
5093964	5093988	ccgataaatgcgaaa	ETEC4702	ETEC specific	++
5132420	5132347	gcaggaagcggcgaa	ETEC4739	*deoB*	++
Plasmid p948 Targets					
65178	65159	gcattttctgtcaaa	ETECp9480770	ETEC specific	+++

[a]Genome coordinate of MarA ChIP-seq peak centre in H10407. Bold type indicates peaks within 150 bp of a σ^{70} binding peak.
[b]Genome coordinate of MarA binding site predicted by MEME.
[c]Sequence of MarA binding site predicted by MEME
[d]Nearest gene to MarA binding site. Some MarA targets were between divergent (<>) and convergent (><) genes. Genes in parentheses indicate that the ChIP-seq peak is located within that gene.
[e]*E. coli* K-12 homologues of ETEC genes in the previous column. *E. coli* K-12 MarA binding sites, listed in the Ecocyc database, are highlighted according to experimental confirmation (solid line) or prediction (dashed line).
[f]*in vitro* binding of purified MarA observed at concentration of 0.3 µM (+++), 1.0 µM (++) or 1.7 µM (+). Note that five control DNA fragments (*cydD*, *ybiS*<>*ybiT*, ETEC2117, *cyoA* and P*estA* did not bind MarA at any of these concentrations (Figure 4.4).

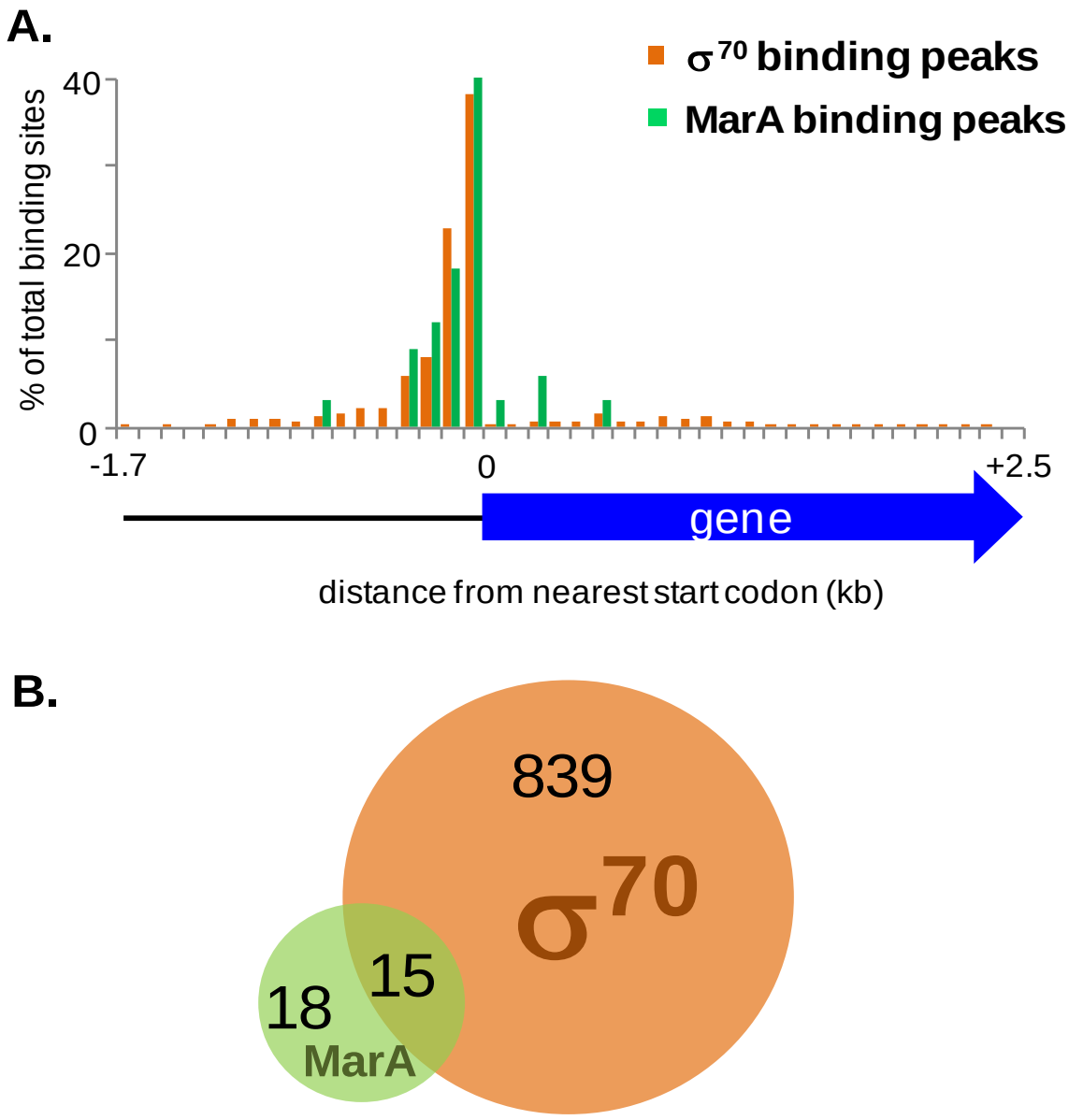
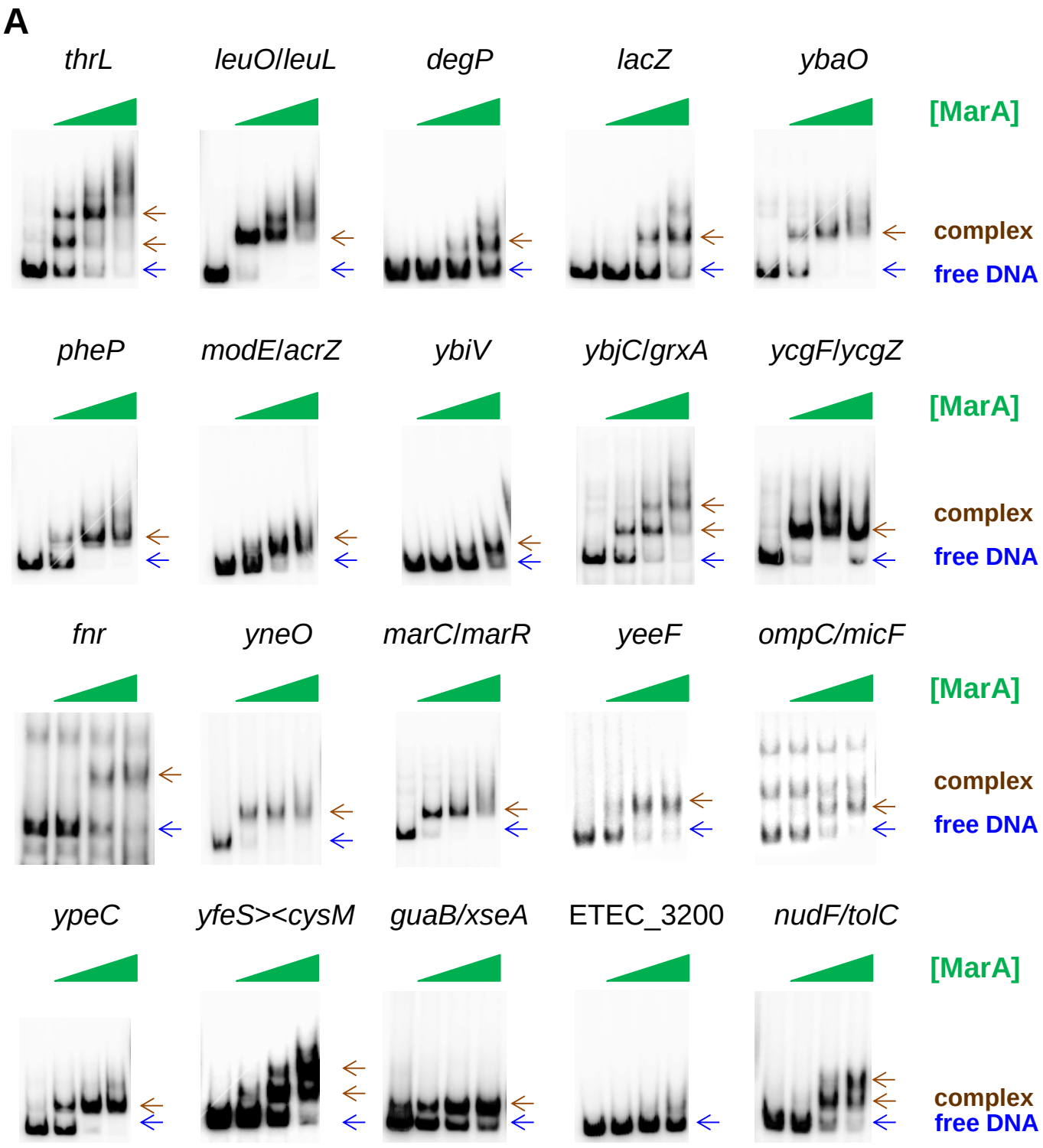


Figure 4.3: Location of MarA and RNA polymerase binding peaks with respect to genes. A. Histogram that depicting the distance between ChIP-seq binding peaks and the nearest 5' end of a gene; MarA binding data is represented in green and RNAP σ^{70} binding data in orange. Each binding peak was allocated to a series of bins occurring at 100 bp intervals. B. Venn diagram depicting the overlap of MarA and σ^{70} binding peaks.


A
thrL
leuO/leuL
degP
lacZ
ybaO
[MarA]
complex
free DNA
pheP
modE/acrZ
ybiV
ybjC/grxA
ycgF/ycgZ
[MarA]
complex
free DNA
fnr
yneO
marC/marR
yeeF
ompC/micF
[MarA]
complex
free DNA
ypeC
yfeS><cysM
guaB/xseA
ETEC_3200
nudF/tolC
[MarA]
complex
free DNA

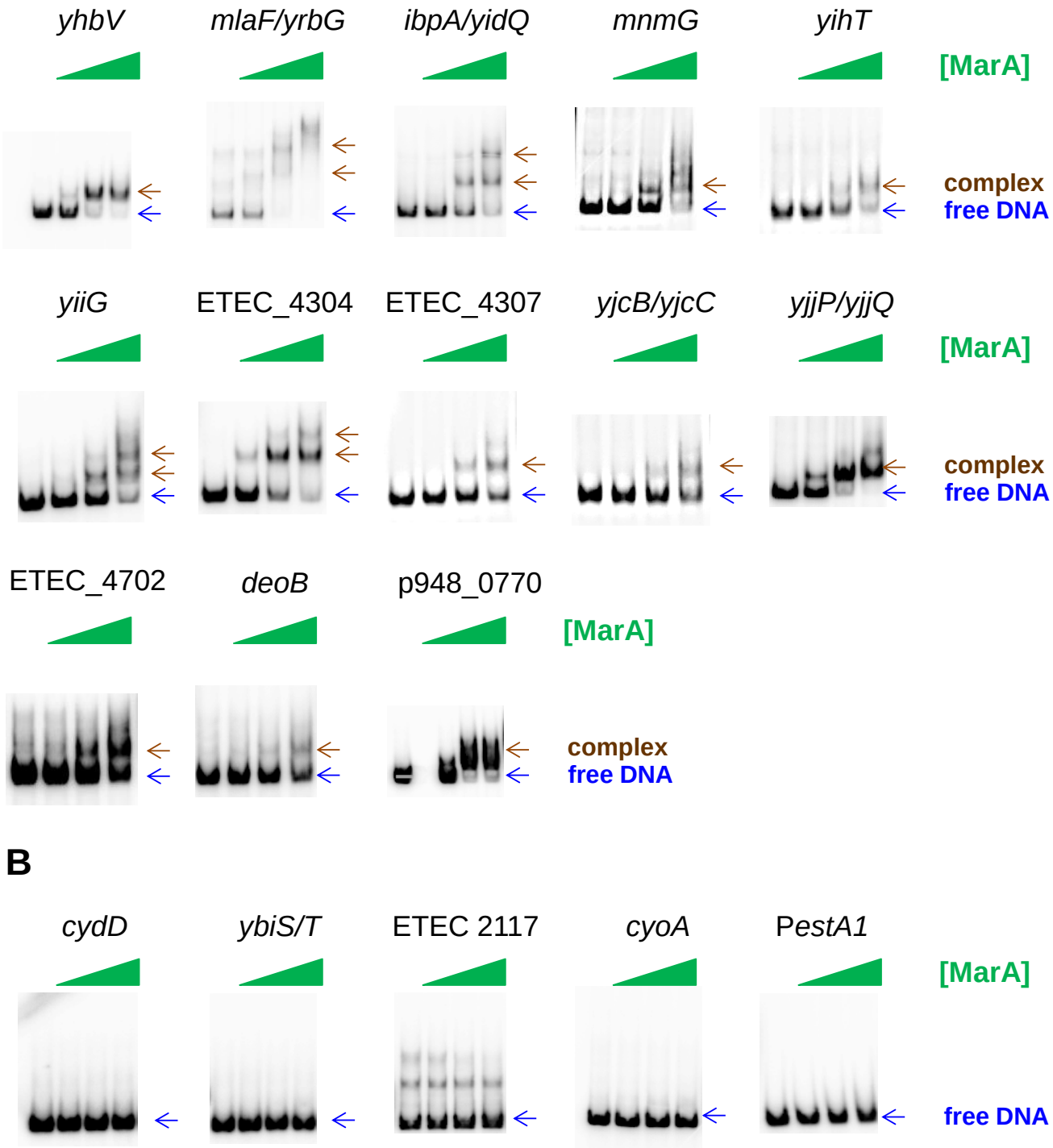


Figure 4.4: In vitro verification of MarA binding to the identified DNA targets.

A. EMSA gels, where MarA bound to various fragments. MarA-DNA complex is shown as a brown arrow and the free DNA as a blue arrow. B. EMSA gels where no MarA binding was observed.

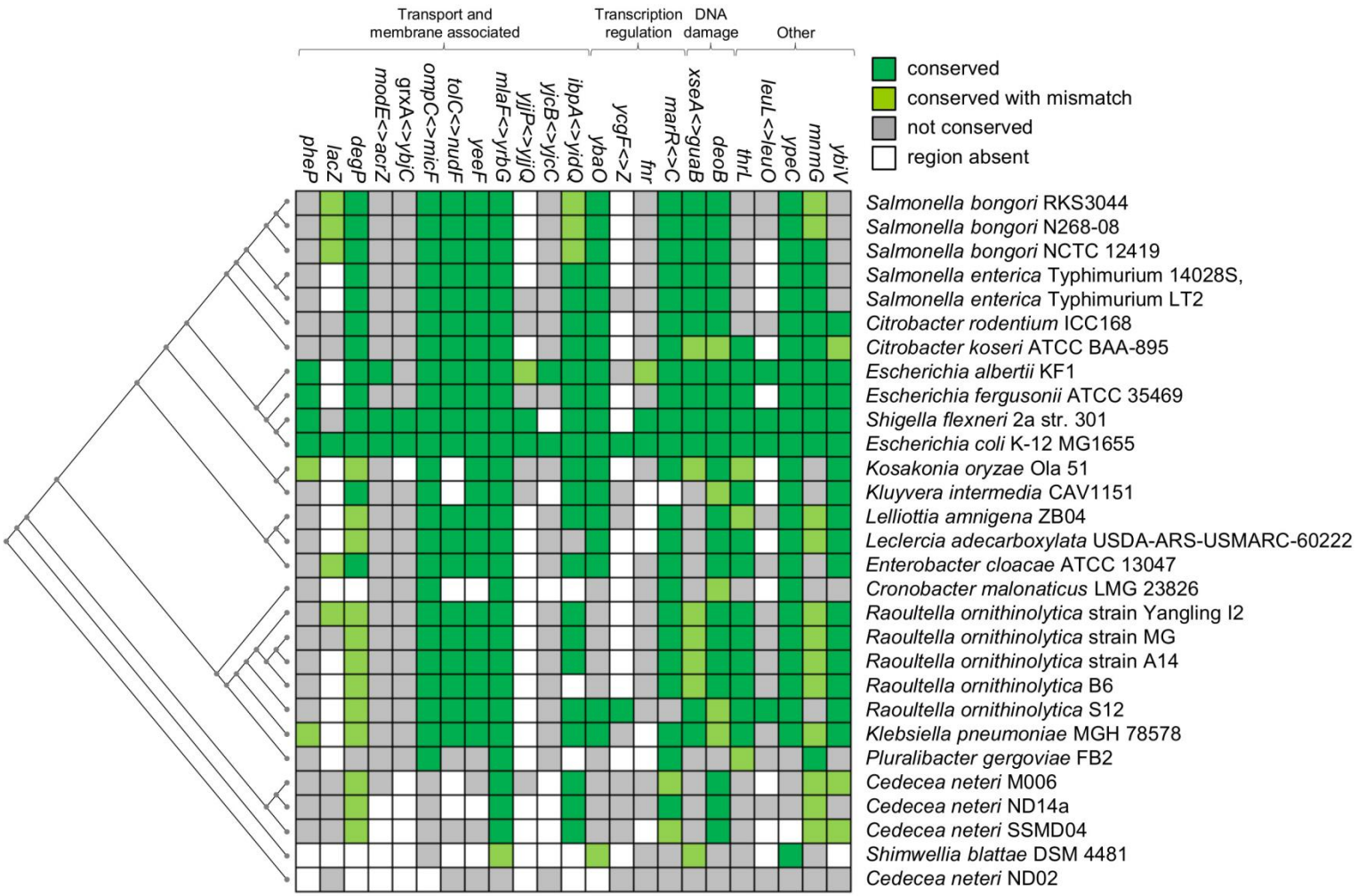


Figure 4.5: Conservation of MarA regulon amongst the enterobacteriacea. Heatmap explaining the conservation of the marboxes identified in ChIP-seq (x-axis) in the genomes of different enteric bacteria (y-axis). The cladogram shows the relationship between different organisms based on *polA* gene sequence.

pneumoniae, *Enterobacter cloacae* and *Raoultella ornithinolytica*, share a similar subset of marbox sequences. These sites include the marbox sequences adjacent to *micF* and *tolC* and the auto-regulatory site present on the upstream of the *marRAB* operon. A site present upstream of the *mlaFEDCB* operon is the best conserved MarA site identified in our data; the sequence is present in all but the most divergent *Cedecea neteri* ND14b genome.

4.5 Phenotypic landscape of the mar regulon

The binding peaks of MarA that are specific to ETEC H10407 reside mostly inside prophage genes and are unlikely to be functional. Hence, we concentrated on the 28 targets that are identical in *E. coli* K-12. This allowed us to make use of the significant resources available for the K-12 strain. To better understand the downstream effects of MarA, we specifically made use of phenotypic analysis data for the *E. coli* K-12 Keio collection. Data are shown as a heatmap in Figure 4.6 (Nichols *et al.,* 2011). In the heatmap, the rows represent the strain of *E. coli* lacking the MarA target gene and columns represent various antibiotics bundled according to the cellular processes inhibited. The coloured boxes represent the growth of the strain lacking the MarA target gene, in the presence of the various antibiotics and relative to the wild type cells. Hence, genes important for innate resistance to antibiotics are highlighted by red coloured boxes. The figure shows that the broadest sensitivity results from deletion of *tolC*; the strain becomes sensitive to many antibiotics such as nalidixic acid, novobiocin, erythromyocin, chloramphenicol, minocycline, doxycyclines etc. However, there are gaps in the protection offered by *tolC*. In particular, *xseA* (encoding the larger subunit of exonuclease VII) and *mlaFEDCB* (involved in lipid trafficking) appear to offer protection

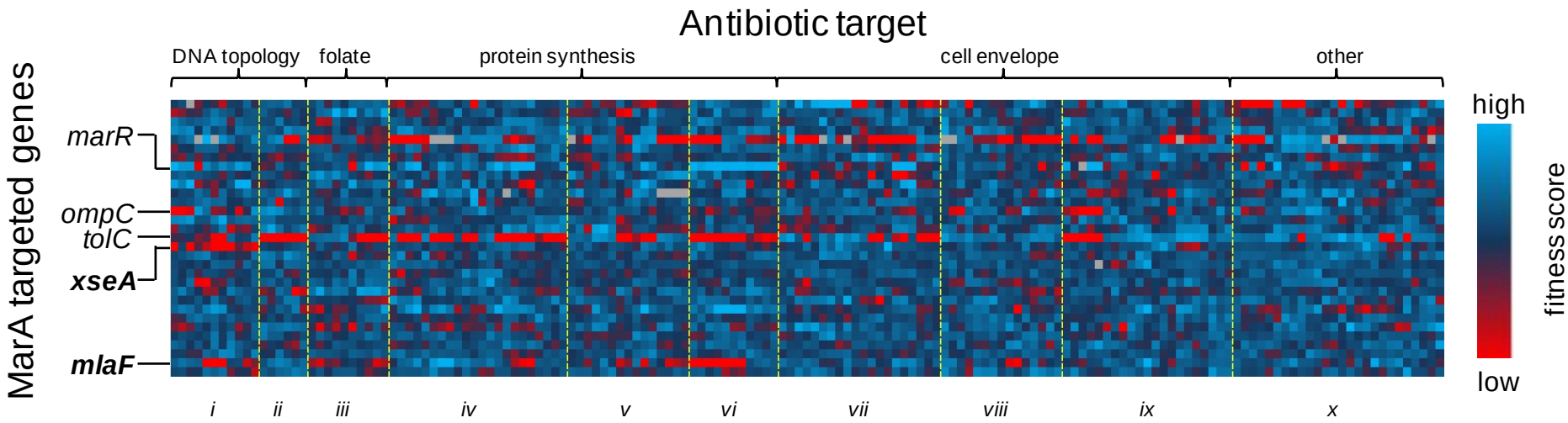


Figure 4.6: Phenotypic landscape of the MarA regulon. Heatmap generated using the growth data from Nichols *et al.,* 2011. This heatmap illustrates the fitness score in the presence of different antibiotics of the strain lacking MarA target genes compared to the parental strain. The antibiotics are bundled by their mode of action or the cellular processes they target and are labelled at the top of the heatmap. Below the heatmap the antibiotics are subdivided in different classes. The classes are as follows: *i*, quinolones; *ii*, non-quinolone topoisomerase inhibitors; *iii*, antifolates; *iv*, macrolides; *v*, aminoglycosides; *vi*, tetracyclines; *vii*, penicillins; *viii*, cephalosporins; *ix*, other cell envelope antibiotics; *x*,miscellaneous.

against the quinolone family and tetracycline family of antibiotics (Harris *et al.,* 1998, Malinverni and Silhavy, 2009). We focus on these targets in the following two chapters.

4.6 Discussion

MarA is a member of the AraC/XylS family of transcription factors found conserved in many bacteria, particularly the *Enterobacteriaceae*. MarA is a transcriptional activator inducing transcription of many genes and can enhance drug resistance. Conventionally, MarA mediated antibiotic resistance has been attributed to regulation of genes encoding the AcrAB-TolC efflux system. These are amongst only 12 well characterised targets listed in the Ecocyc database (Karp *et al.,* 2014, Keseler *et al.,* 2017). However, there are more than one hundred MarA molecules in the cells that are drug resistant (Martin *et al.,* 2008). We reasoned that MarA must control expression of more genes than is currently known.

Barbosa and Levy used *E. coli* microarrays and identified over 60 genes putatively controlled by MarA (Barbosa and Levy, 2000). From those, only 3 genes were found in common in the studies by Pomposiello and co-workers (Pomposiello *et al.,* 2001). A study by Seo *et al.* (2015) has revealed 25 SoxS targets using ChIP-exo (Seo *et al.,* 2015). Of these targets 7 are common in my MarA ChIP-seq experiment. This shows an overlap between MarA and SoxS regulons. However, this overlap is not suprising since MarA and SoxS are very similar proteins recognising the same binding site (Li and Demple, 1994). Recently, another study have shown control of expression of *E. coli* OmpF prorin by BluR, YcgZ and Lon (Duval *et al.,* 2017). *E. coli bluR* and *ycgZ* were also identified in our ChIP-seq analysis and are targets of MarA. It could be that these

genes are regulated by MarA and in turn they regulate expression of a porin, OmpF that was previously known to regulated by *micF* (Cohen *et al.,* 1988, Ziervogel and Roux, 2013).

Previously it was shown that to bind specific promoters the concentration of regulator is also important. For example, to bind promoters like *acrAB,* a high concentration of MarA is required (Martin *et al.,* 2008). It could be because of this reason our ChIP-seq on ETEC H10407 did not reveal *acrAB* as a target. Since, we carried out our experiment in a strain that has a MarA overexpressing plasmid known as pRGM9818 and may be the strain was not able to produce enough MarA that was required to bind *acrAB.*

E. coli antibiotic resistance is known to be regulated by MarA. However, other pathogenic bacteria also maintain antibiotic resistance by MarA like regulators. Southern hybridisation experiments in previous studies with an *E. coli marRAB* probe have shown that *marRAB* locus is highly conserved in many other pathogens as *Shigella flexneri, Salmonella enterica, Klebsiella pneumonia* etc (Cohen *et al.,* 1993b). Like MarA, a homologue RamA is known to be regulator of *acrAB-tolC* in *Salmonella enterica* (Abouzeed *et al.,* 2008). RamA is also an AraC/XylS family member that has a structure similar to MarA and binds DNA similarly (Ricci *et al.,* 2012). There are no signs of presence of RamA in *Shigella* or *E. coli* (Ricci and Piddock, 2009).

Even after so many studies MarA regulon was still unknown. Hence, in this work the direct MarA regulon has been defined by ChIP-seq and verified using EMSA experiments. We show that, in addition to the ArcAB-TolC efflux pump, MarA targets genes that play a role in transport, DNA damage repair, and transcription regulation. Two interesting targets are: *xseA* (encoding the larger subunit of exonuclease VII), and

mlaFEDCB (involved in lipid trafficking). However, there are many MarA targets that were revealed by our analysis that plays a role in amino acid synthesis. The role of MarA for these targets is still to be explored.

Chapter 5. MarA-dependent regulation of *xseA* is associated with DNA repair and quinolone resistance

5.1 Introduction

The ChIP-seq analysis in Chapter 4 identified MarA binding at the regulatory region of a gene known as *xseA.* This gene encodes the larger subunit of Exonuclease VII; an enzyme implicated in DNA repair (Chase and Masker, 1977, Repar *et al.,* 2013). The phenotypic analysis of MarA target genes suggested that *xseA* is a determinant for quinolone resistance. We have verified the binding of MarA to *xseA* and characterised regulation of *xseA* by MarA. Moreover, we have investigated the structure/function analysis by Poleszak *et al* (2012) and have found out the functional properties of *xseA* required for quinolone resistance.

5.2 Characterisation of P*xseA*

5.2.1 Binding of MarA to *xseA* promoter (P*xseA*)

Our ChIP-seq data suggest that, in conjunction with RNA polymerase, MarA binds a site overlapping the *xseA* promoter -35 element (Figure 5.1A). We propose that MarA binding could activate *xseA* transcription. Figure 5.1B shows the sequence of the *xseA* regulatory region, where green denotes the putative marbox and orange denotes P*xseA*. In order to verify binding of MarA to the predicted marbox at the *xseA* regulatory region, we generated two DNA fragments; *xseA1* and *xseA2* (Figure 5.1B; 5' end of fragments represented by inverted triangles). Note that, *xseA2* contains a mutation (-36C) as well as being truncated, to complete inactivation of the marbox. An EMSA experiment was done to test that MarA binding with each fragment (Figure 5.2); MarA binds *xseA1* (marbox intact) but did not bind *xseA2* (marbox inactive). To further verify MarA binding we used DNAseI footprinting; we cloned *xseA1* into a plasmid pSR and a DNA fragment was generated after digestion with *Aat*II and *Hin*dIII.

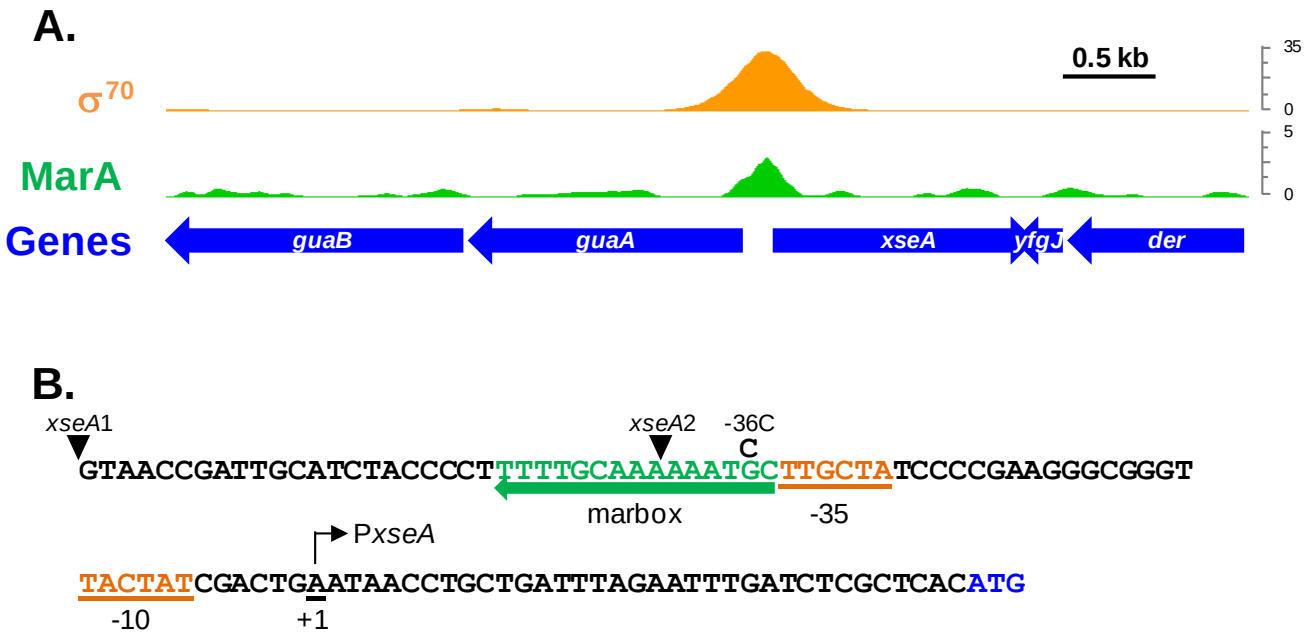


Figure 5.1: MarA and σ70 binding across the xseA and regulatory sequence of *xseA*. A. MarA binds in conjunction with RNAP and recognizes overlapping targets upstream of *xseA.* MarA binding is shown in green, σ^{70} in orange and the genes are represented in blue. B. The regulatory sequence upstream of *xseA*. The start codon is depicted in blue, marbox in green in reverse orientation represented by an arrow. Davies and Drabble (1996) have identified the transcription start site +1, underlined and shown by the bent arrow. The promoter -10 and -35 elements are shown in orange. The inverted triangles show *xseA1* and *xseA2* fragments that were used in this study; where *xseA2* has a mutation at -36C.

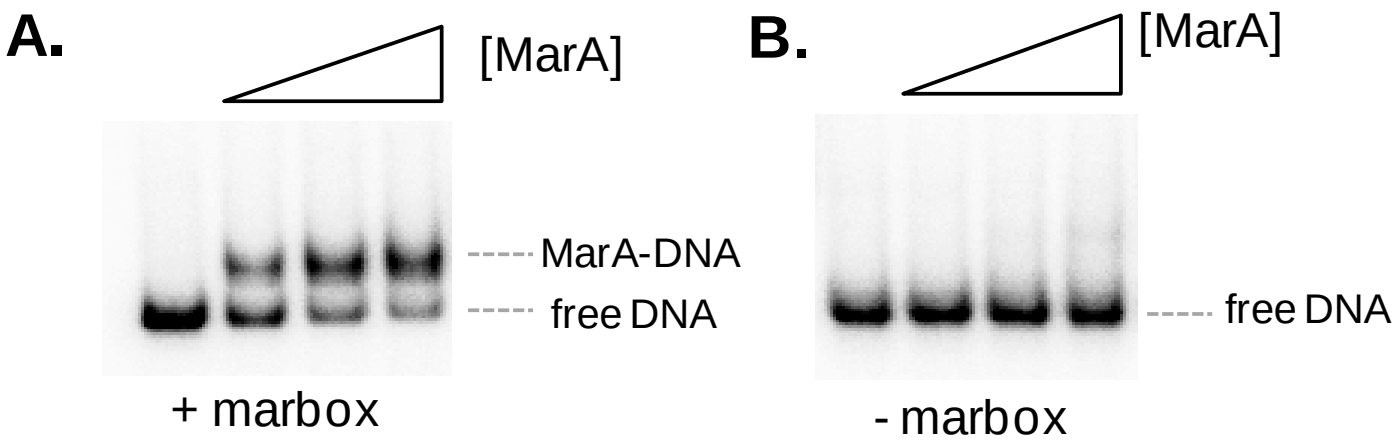


Figure 5.2: Binding of MarA to the *xseA* gene regulatory region. The panels illustrate results of EMSA. A. MarA makes a complex with DNA when the marbox was intact (*xseA1*). B. When marbox is removed, the complex of MarA with DNA is not formed (*xseA2*). MarA was used at a final concentration of 0.3, 1.0, and 1.7 µM as indicated by the triangle. Representative of n=2.

This fragment was end labelled, bound with purified MarA, and treated with DNAseI. The gel was calibrated using a Maxim/Gilbert G+A ladder. The ladder was numbered in accordance with the *xseA* start site (+1). The result is illustrated in Figure 5.3. Lane 1 shows the digested DNA fragment in the absence of MarA. Protection of the MarA binding site and DNAseI hypersensitivity on either side of the binding site is evident when MarA is added. Hence, MarA binds the marbox predicted by MEME.

5.3 P*xseA* is a MarA activated promoter

To assess whether P*xseA* was dependent on MarA; *xseA1* and *xseA2* were fused to *lacZ* in plasmid pRW50 and cells were transformed with the resulting plasmid. We determined β-galactosidase activity and the results are shown in Figure 5.4. The data show that P*xseA* is active when the marbox is present but not when the marbox was deleted. This suggests transcription activation of P*xseA* by MarA. Rob and SoxS would likely have overlapping regulatory effects and for this reason we deleted the marbox rather than deleting *marA*.

5.4 Innate resistance to ciprofloxacin requires MarA mediated *xseA* expression

We determined the minimum inhibitory concentration (MIC) of ciprofloxacin and found approximately 40 fold reduction in *xseA::kan* cells. To assess whether the ciprofloxacin hypersensitivity phenotype was dependent on *xseA* and the P*xseA* marbox we did genetic complementation experiments. Hence, derivatives of plasmid pBR322, encoding *xseA* downstream of P*xseA*, with or without the marbox, were made. In the presence of ciprofloxacin the plasmid encoding *xseA* downstream of P*xseA* was able to rescue the growth of *E. coli* strain BW25113 *xseA::kan* but pBR322 by itself was not (Figure 5.5). We also found that the P*xseA* marbox was essential for complementation; in the

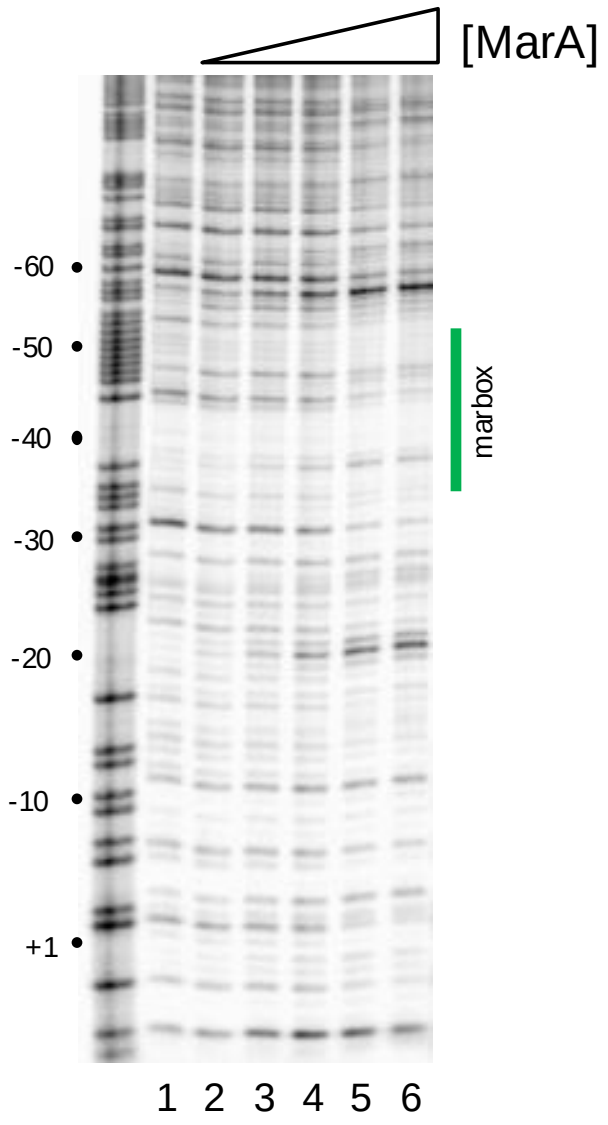


Figure 5.3: DNAse I footprint of MarA binding to the xseA gene regulatory region. For the experiment *xseA1* and purified MarA were used. The first lane had the DNAse I digested DNA without any MarA. MarA was titrated over the digested DNA from the second lane. A Maxam-Gilbert G+A sequencing ladder was used to calibrate the gel. The positions relative to the start site (+1) of *xseA* are labelled. The triangle indicates addition of MarA at concentrations of 0.3, 1.0, 1.7, 2.4, and 3.3 µM. The marbox is represented by a green line. Representative of n=2.

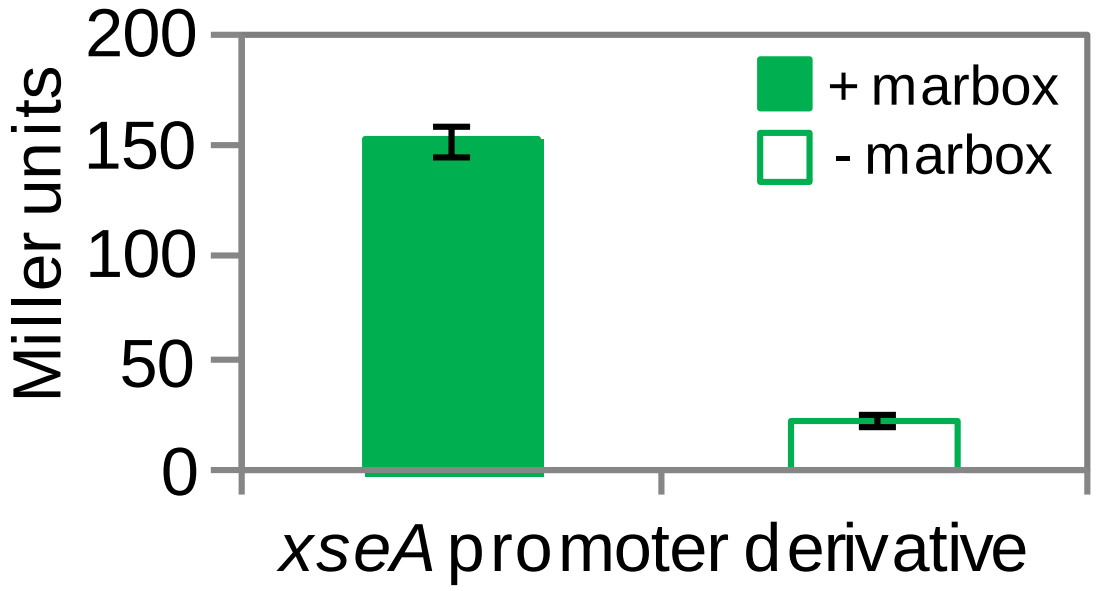


Figure 5.4: The *xseA* promoter activity requires an intact marbox *in vivo*. LacZ assay using JCB387 containing the pRW50 derivative of *xseA1* (intact marbox) and *xseA2* (marbox inactivated). Representative of n=3.

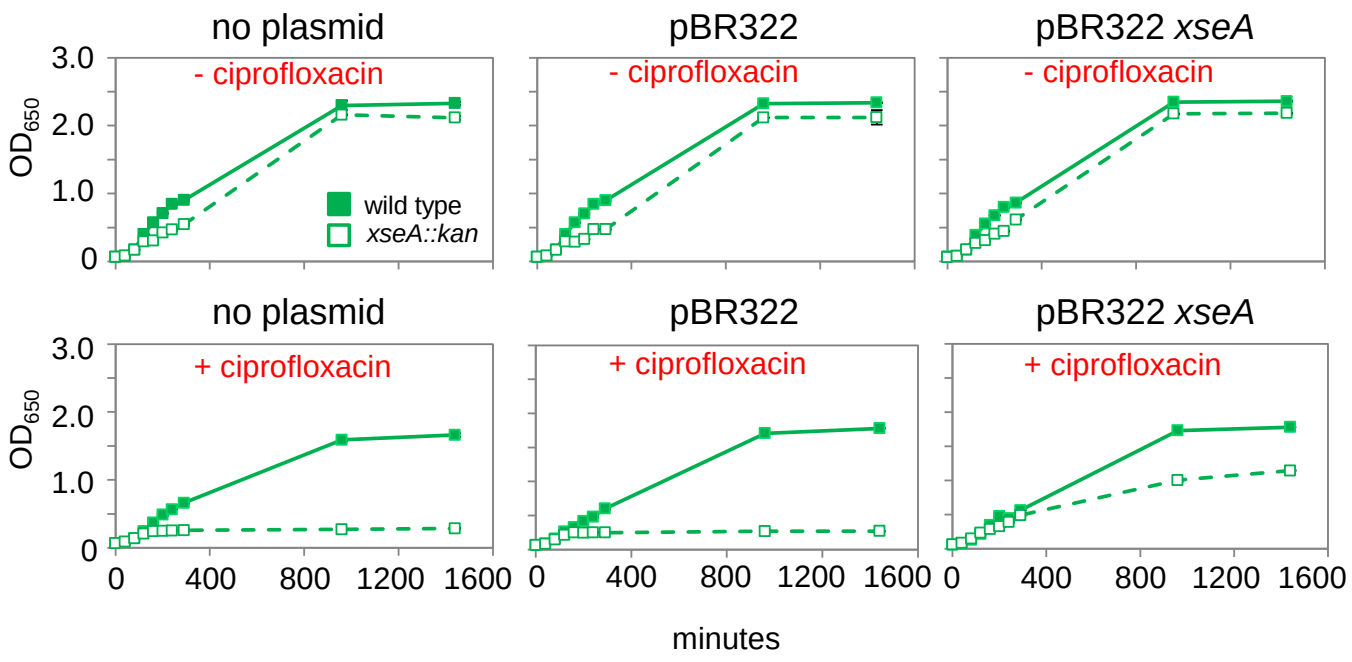


Figure 5.5: Genetic complementation of *xseA* phenotype. The graphs show the OD_{650} values of the BW25113 and *xseA::kan* grown in the absence of pBR322 (left panel), in the presence of pBR322 on its own (middle panel), and pBR322 carrying the *xseA* gene (right panel). The cells were grown in absence or presence of 0.005 µg/ml ciprofloxacin. Briefly the pBR322 on its own and a derivative of *xseA* (marbox present) were transformed in BW25113 and *xseA::kan* cells, and growth was recorded. Representative of n=3.

absence of ciprofloxacin deletion of marbox has a very little effect (Figure 5.6A), but in the presence of ciprofloxacin deleting the marbox prevented growth (Figure 5.6 B).

5.5 *xseA* determinants for DNA binding, catalysis and *XseB* interactions are required for complementation

Genes, *xseA* and *xseB* encode the large and small subunits of exonuclease VII respectively. However, *xseA* alone is targeted by MarA (Table 4.1). Hence, *xseA* might have stand alone functions. To understand the mechanism of resistance we made mutations according to the structure/function analysis of Poleszak *et al* (2012). There are 4 functional regions in *xseA*: an N-terminal OB-fold, a catalytic domain in the centre, a coiled coil domain and a shorter C-terminal domain (Poleszak *et al.,* 2012). Out of these four regions the OB folds binds DNA, and the coiled coil domain binds XseB. To determine the role of each of these activities, we introduced point mutations in *xseA* encoded by pBR322. These mutations are represented in Figure 5.7. The growth assay data is shown in Figure 5.8. Disrupting any of the *xseA* determinants that are required for proper functioning of exonuclease VII, results in hypersensitivity to ciprofloxacin.

5.6 Sub-inhibitory concentration of ciprofloxacin introduces double stranded breaks in the cells that lack MarA controlled *xseA*

Our data suggests that MarA controlled expression of *xseA* is vital for the cells treated with ciprofloxacin. We reasoned that *xseA* might help to reduce the frequency of double strand DNA breaks in cells exposed to ciprofloxacin. To test our prediction we used microscopy and visualised DNA damage in Hoescht stained cells. Note that quinolones are known to include cell filamentation and nucleoid fragmentation (Georgopapadakou and Bertasso, 1991). As expected, the *xse::kan* cells treated with a sub-inhibitory

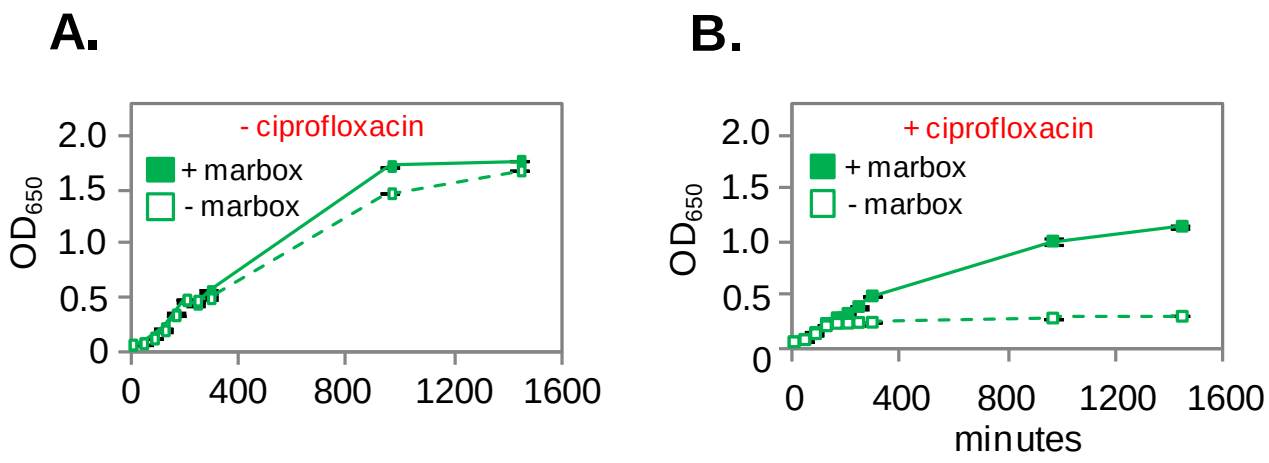


Figure 5.6: An intact marbox is required at the upstream of the promoter for genetic complementation of *xseA* phenotype. The graph shows the OD_{650} values of the BW25113 *xseA::kan* grown in: A. absence or B. presence of 0.005µg/ml ciprofloxacin. Briefly the pBR322 derivative of *xseA1* (marbox present) or *xseA2* (marbox inactive) were transformed in *xseA::kan* cells and growth was recorded. Representative of n=3.

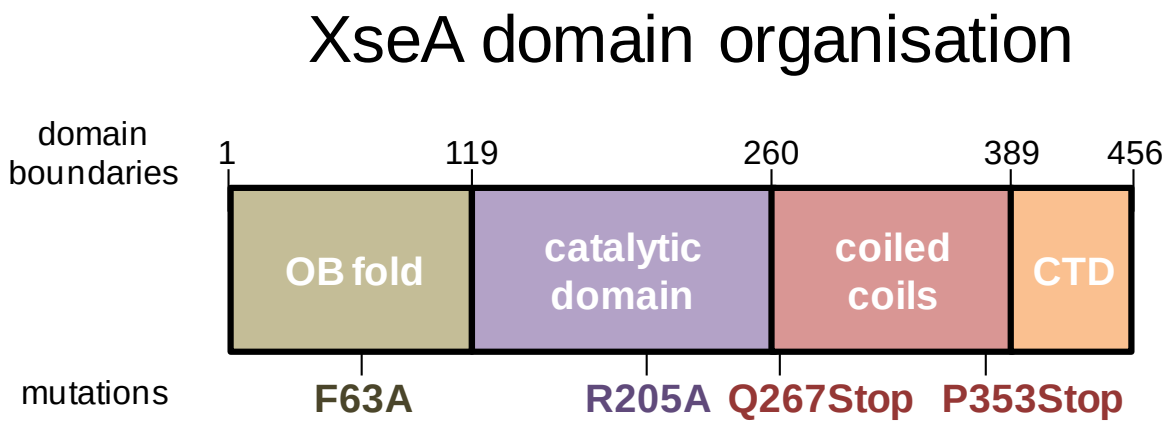


Figure 5.7: Domain organisation of XseA. The domain of XseA is organised in four regions. The OB fold shown in green binds DNA, the catalytic domain shown in purple is responsible for the catalytic function, the coiled coils domain shown in red binds XseB and the C-terminal domain (CTD) is shown in orange. The figure is based on Poleszak *et al.,* (2012).

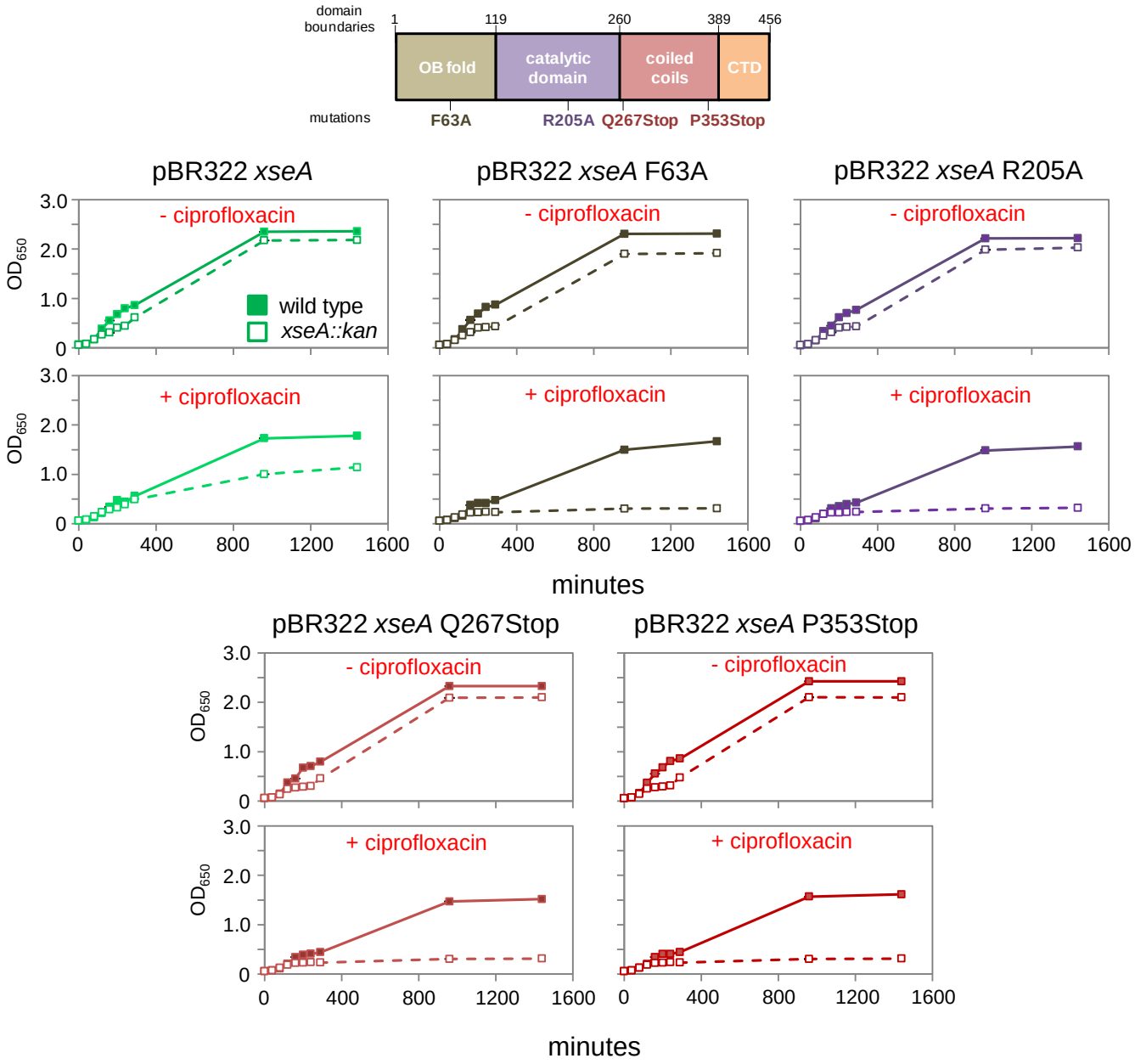


Figure 5.8: The key determinants of *xseA* are required for its proper functioning. This figure shows that *xseA* on pBR322 is able to complement. If any of the determinants, e.g., the OB fold domain that binds DNA, the catalytic activity of *xseA,* the coiled coil domain responsible for XseB interaction or another coiled coil was mutated or disrupted then *xseA* loses the complementation effect in the presence of ciprofloxacin in *xseA::kan* cells. Representative of n=3.

concentration of ciprofloxacin displayed this phenotype but wild type cells did not (Figure 5.9A, B). On complementation cells start to mimic the WT cells (Figure 5.9C). This effect is totally dependent on the marbox (Figure 5.9D). To verify that the observed changes in nucleoid conformation resulted from DNA double strand breaks, we performed pulse field gel electrophoresis. As expected, the chromosome of the untreated cells ran as tight bands and no smearing was observed, as seen in lane 1 and 2 (Figure 5.10). The chromosome of the ciprofloxacin treated cells results in DNA damage but no smear was observed in the case of wild type cells (lane 3). However, the treated cells lacking *xseA* runs as a smear in lane 4. Presence of pBR322 on its own in the treated cells lacking *xseA* had no effect (lane 5). The effect was negated in the presence of pBR322 encoding *xseA* under control of P*xseA* and marbox (lane 6). This effect was again lost when the marbox was deleted as seen in lane 7 (Figure 5.10).

5.7 Discussion

DNA damage by antibiotics in bacteria can occur by different ways. Firstly, for example, in case of bleomycin, antibiotic can directly interact with the DNA and can result in double stranded breaks. However, the mechanism of how the damage occurs is still not understood (Hecht, 2000). Second example is based on quinolones that inhibits the ligase domain in topoisomerase (Hawkey, 2003). Moreover, many antibiotics exert metabolic changes in the cells such as production of reactive oxygen species (ROS) (Dwyer *et al.,* 2014). These are mostly in the form of hydroxyl radicals that results in double stranded breaks. Also oxidised guanine starts to incorporate in the bacterial genome (Blazquez *et al.,* 2012, Foti *et al.,* 2012).

However, bacteria have mechanisms to repair the damaged DNA sometimes introduce mutation in their genome. Out of all, SOS reponse is activated on DNA damage by the activation of a protein known as RecA (Rajagopalan *et al.,* 1992). SOS genes are normally repressed by another protein known as LexA. The polymers of RecA bind single stranded DNA and when gets activated leads to the cleavage of LexA which in turn leads to thhe activation of SOS genes. These genes include DNA repair enzymes (Brent and Ptashne, 1981).

Recently, Wilkinson *et al* (2016) have shown that *E. coli* RecBCD (involved in double stranded DNA repair) is inactivated by a small phage protein Gam. This inactivity is due to the competition with RecBCD DNA bidning site. This process is found conserved in other pathogens such as *Klebsiella pneumonia*. Expression of Gam in such pathogens results in increased sensitivity to fluroquinolones (Wilkinson *et al.,* 2016b).

This chapter shows that MarA also targets processes involved in DNA repair; MarA controls *xseA* that encodes the enzyme exonuclease VII (Chase *et al.,* 1986). Cells lacking *xseA* are hypersensitive to the presence of ciprofloxacin. This is because ciprofloxacin results in double stranded breaks that are presumably more frequent when *xseA* is absent. Quinolone resistance can arise from the combined effect of Quinolone Resistance Determining Regions (QRDR) mutations (first mutation in *gyrA* and second mutation in *parC*) (Hopkins *et al.,* 2005) as well as AcrAB-tolC over-expression (Cohen *et al.,* 1989). However, the *xseA* has not been noted previously. Poleszak *et al.,* 2012 suggests that all XseA functional determinants are necessary for the activity of exo-



Figure 5.9: Ciprofloxacin induced damage of *E. coli* cells in the absence of MarA mediated *xseA* expression. All the cells were grown in a subinhibitory concentration of ciprofloxacin and were DAPI stained. A. Hoescht stained WT cells that behave as normal in the presence of ciprofloxacin. B. *xseA* cells that become sensitive in the presence of ciprofloxacin and, when DAPI stained, cells visualized under microscope have a filamentous cell like structure with fragments of chromosome inside. C. Genetic complement of *xseA*; the cells starts to behave as WT but still are longer then WT. D. The marbox is essential for complementation. The term "complement" denotes BW25113 *xseA*::kan transformed with pBR322 encoding *xseA* under control of the *xseA*1 fragment (+ marbox) or the *xseA*2 fragment (- marbox). All the above images were obtained on the same scale of 5 µm (shown as white bar). Representative of n=3.

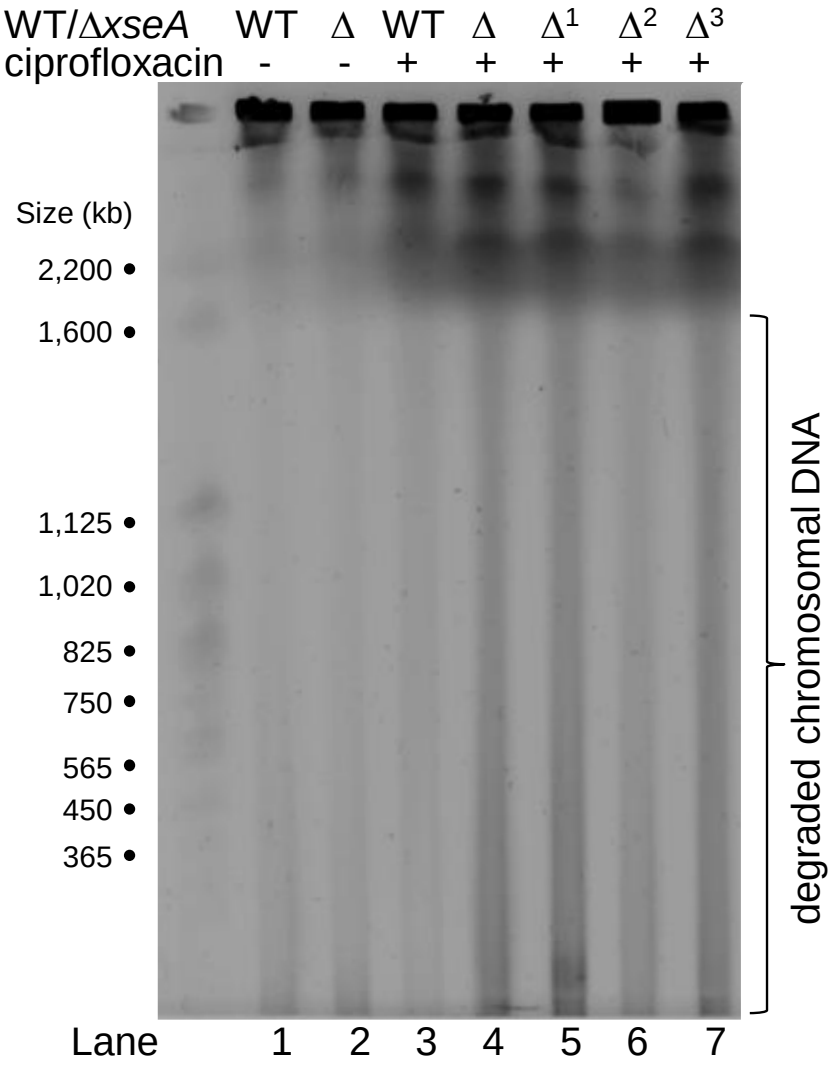


Figure 5.10: Ciprofloxacin induced DNA damage in the absence of *xseA* expression. Integrity of chromosomal DNA was analysed using PFGE. The cells used were BW25113 (WT) and *xseA*::kan derivative (Δ). The cells were transformed with a derivative containing either the plasmid on its own (empty pBR322), denoted as Δ^1; pBR322 encoding *xseA* under the control of the *xseA*1 fragment, denoted as Δ^2; or *xseA*2, denoted as Δ^3. The size markers are labelled and the absence or presence of 0.005 µg/ml ciprofloxacin is indicated on the top of the gel. Representative of n=2.

nuclease VII. This corresponded with our investigation. There are several new mechanisms, which suggest that after being induced by MarA, XseA stimulates DNA repair only in the presence of its defined catalytic activity, DNA binding ability and XseB interaction.

Chapter 6. MarA-dependent regulation of *mlaFEDCB* is associated with altered outer membrane barrier function and tetracycline resistance.

6.1 Introduction

The identified MarA binding sites in Chapter 4 included the regulatory region of an operon known as *mlaFEDCB.* Maintenance of OM lipid asymmetry (*mla*) operon encoded genes have functions associated with lipid trafficking. The phenotypic analysis of MarA targets suggests that *mlaFEDCB* is a determinant for MarA controlled resistance to tetracycline family of antibiotics. In this chapter, we have verified the binding of MarA upstream of *mlaFEDCB,* characterised nearby promoters, and investigated the regulation of *mlaFEDCB* by MarA. Moreover, we examined the functional aspects of *mlaFEDCB* associated with drug resistance.

6.2 Binding of MarA to *mlaFEDCB*

Our ChIP-seq data suggest co-binding of MarA and RNAP σ^{70} to the *mlaFEDCB* locus. (Figure 6.1A). To verify the binding of MarA, we generated two DNA fragments containing the *mlaFEDCB* regulatory region with and without the marbox respectively: *mlaF1* (135 bp) and *mlaF2* (92 bp) (Figure 6.1B) DNA fragments. These DNA fragments were radiolabelled and used for *in vitro* EMSA experiments. The data are shown in Figure 6.2. As expected, MarA bound to the fragment with the intact marbox but not the fragment without the marbox. To verify the exact site of MarA binding, we used DNAseI footprinting with the 135 bp DNA fragment. The results were consistent with the ChIP-seq analysis and EMSA, MarA binds the predicted marbox (green in Figure 6.3).

6.3 MarA activates transcription of *mlaFEDCB in vitro* and *in vivo*

To date, no promoters have been mapped at the *mlaFEDCB* regulatory region. To find the location of MarA regulated promoters, *in vitro* transcription assays were used. For

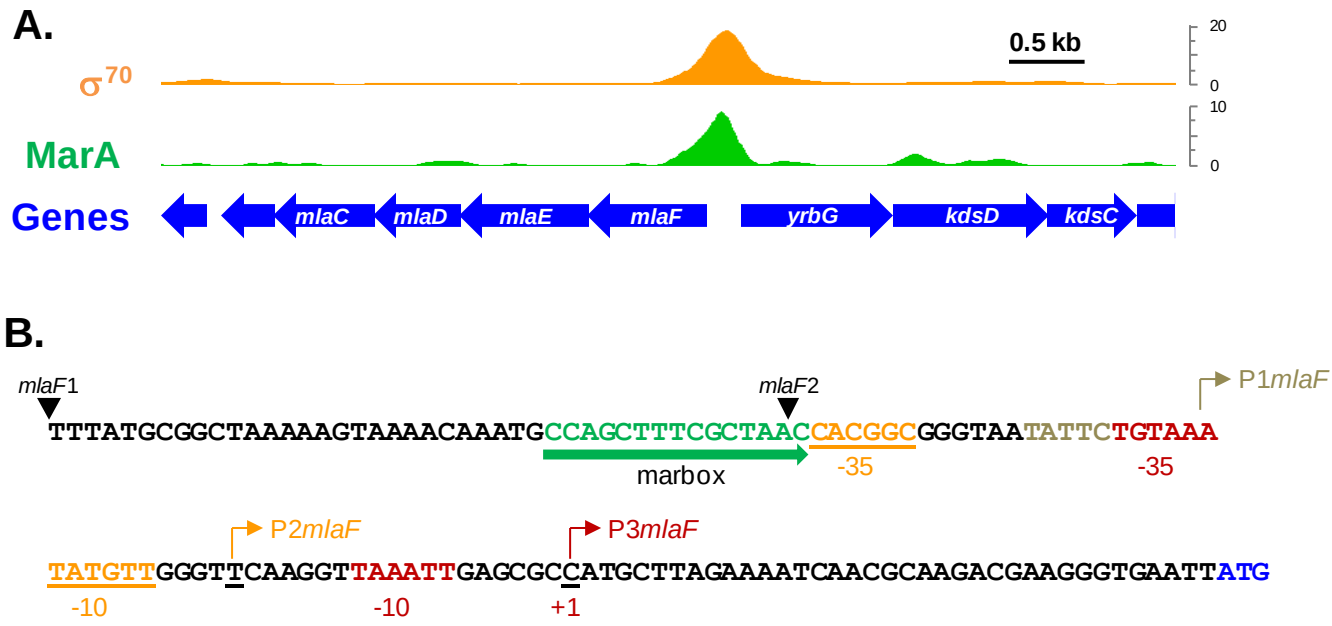


Figure 6.1: MarA and RNA polymerase bind together across the *mla* regulatory region. A. Graph showing the binding of MarA and RNAP to the *mlaFEDCB* locus. MarA binding data is shown in green, RNAP binding profile shown in orange. Genes are represented by labelled blue arrows. B. The regulatory region upstream of *mlaF.* The start codon is shown in blue, marbox in green and is in forward orientation as represented by an arrow. The promoter -10 and -35 elements are highlighted and labelled.

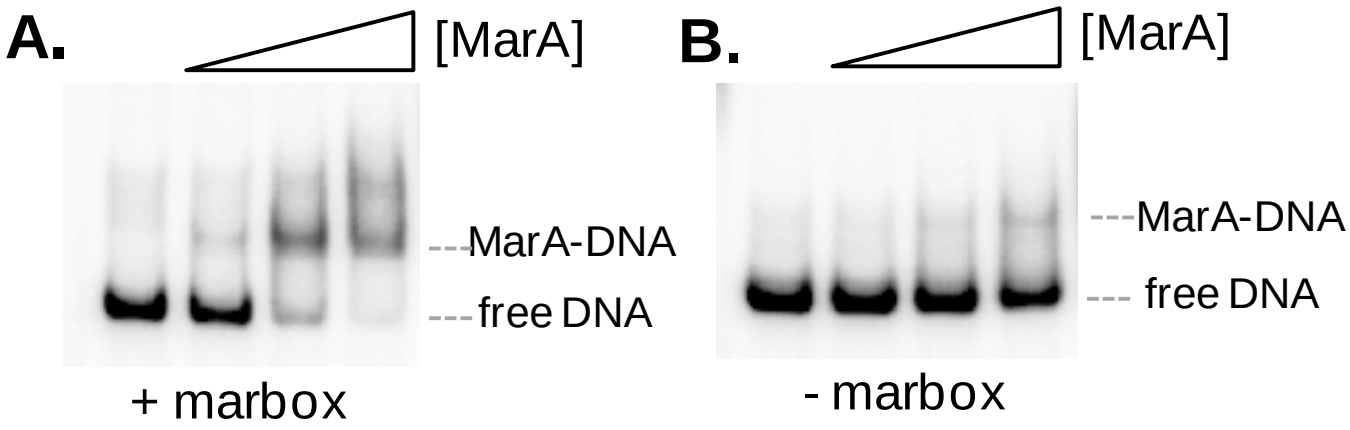


Figure 6.2: MarA binds fragments of *mlaFEDCB* in the presence of the marbox. EMSA gels, A. MarA bound with the *mlaFEDCB* fragment containing the marbox, B. MarA did not bind with the *mlaFEDCB* fragment lacking marbox. MarA was added at a final concentration of 0.3, 1.0, and 1.7 µM as indicated by the triangle. Representative of n=2.

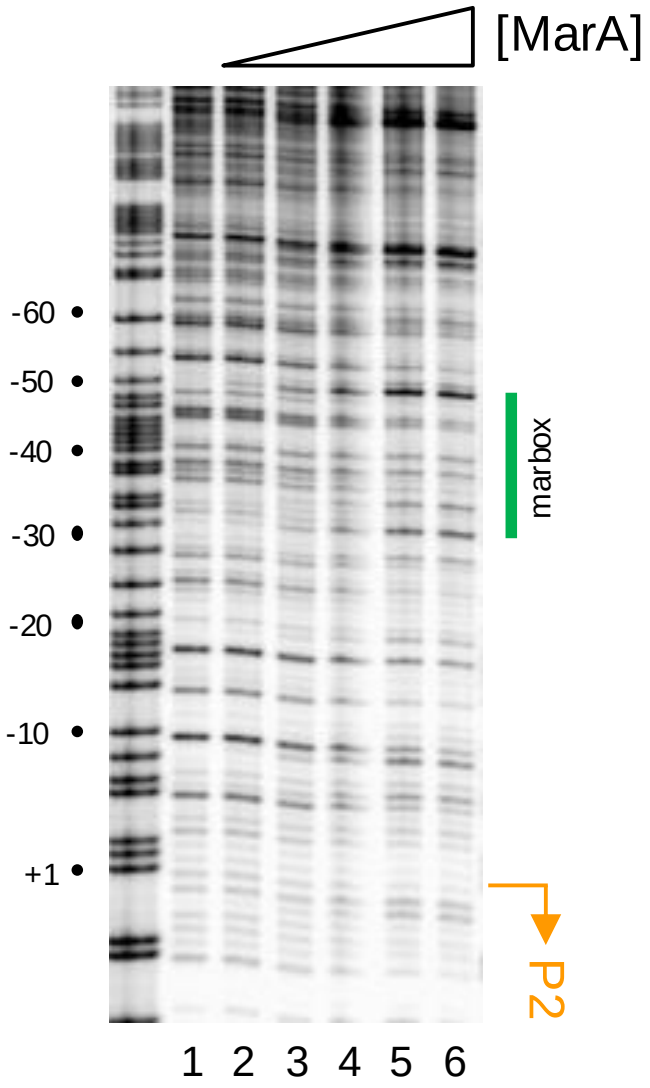


Figure 6.3: DNAse I footprint of MarA binding to the regulatory region of *mlaFEDCB*. Experiment was done using the *mlaFEDCB* regulatory region containing marbox. The first lane had the DNAse I digested DNA without MarA. MarA was titrated over the digested DNA from the second lane. A Maxam-Gilbert G+A sequencing ladder was used to calibrate the gel. The positions relative to the start site (+1) of *mlaFEDCB* are labelled. The triangle indicates addition of MarA at concentrations of 0.3, 1.0, 1.7, 2.4, and 3.3 µM. Marbox is represented by a green line. Representative of n=2.

this purpose, we cloned the 135 bp DNA fragment, containing the *mlaFEDCB* intragenic region, upstream of the *λoop* transcription terminator in pSR. The resulting construct was used as a template for *in vitro* transcription with MarA and RNAP. Figure 6.4 shows the transcripts that were generated by RNAP. Note that the pSR replication origin encodes a 108 nt transcript (RNAI) that acts as an internal control in these assays. The results show generation of three transcripts (128, 148 and 157 nt long) that must initiate within the cloned *mlaFEDCB* fragment. The transcript start sites are located between 5 and 7 bases downstream sequences similar to promoter -10 elements. We refer to the promoters as *mla*P1, *mla*P2 and *mla*P3 respectively (Figure 6.1B). The marbox is proximally adjacent to the *mla*P2 -35 element and transcription from *mla*P2 is activated on MarA binding (Figure 6.4). This shows that MarA activates *mlaFEDCB* transcription *in vitro.*

To assess transcription activation *in vivo* we generated pRW50 derivatives containing the *mlaF1* or *mlaF2* fragments with or without the marbox respectively. These constructs were then used to transform JCB387 cells and La*cZ* activities were determined by β-galactosidase assay. The results are shown in Figure 6.5 and are consistent with the *in vitro* assays; reduction in transcriptional activity was observed when the marbox was deleted. The residual activity observed after the deletion of marbox is likely from the P3 promoter.

6.4 Innate resistance to doxycycline requires MarA mediated *mlaFEDCB* expression

Figure 4.5 suggests that *mlaE::kan* cells become most sensitive to doxycycline. Hence,

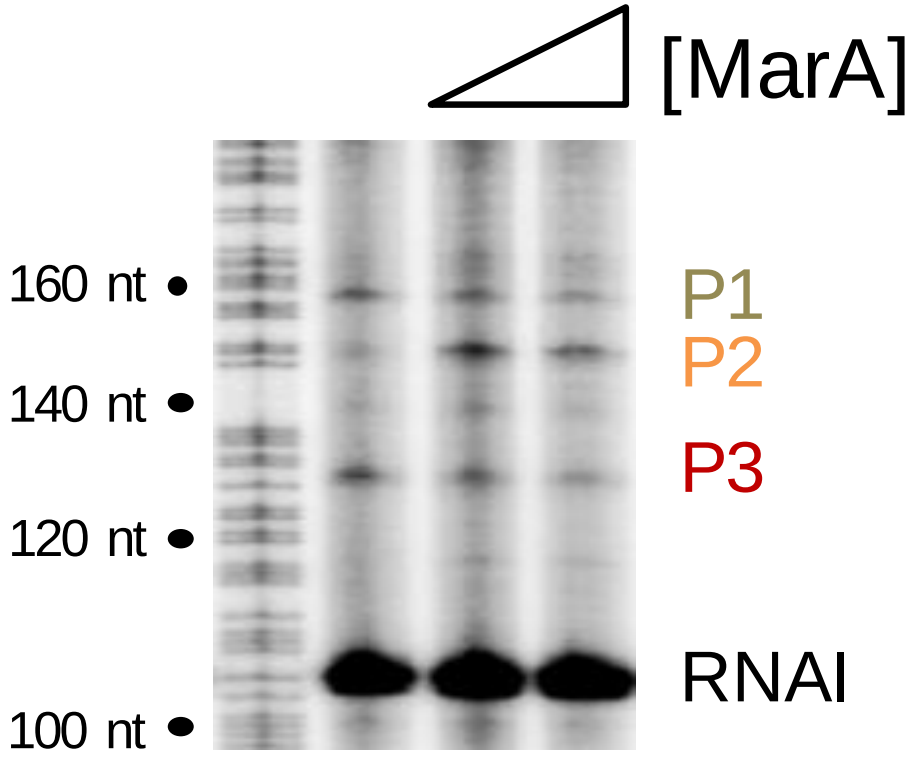


Figure 6.4: *In vitro* transcription assay revealed the presence of three promoters in the regulatory region of *mlaFEDCB*. Assays were done using the *mlaFEDCB* fragment containing the regulatory region cloned in plasmid pSR. A Maxam-Gilbert G+A sequencing ladder was used to calibrate the gel. The RNAI acts as an internal control,itis a transcript derived from the pSR replication origin. The *mlaFEDCB* fragment containing the regulatory region drives the formation of transcripts P1, P2 and P3. The triangle indicates addition of MarA at concentrations of 0.3 or 1.0 µM. Representative of n=2.

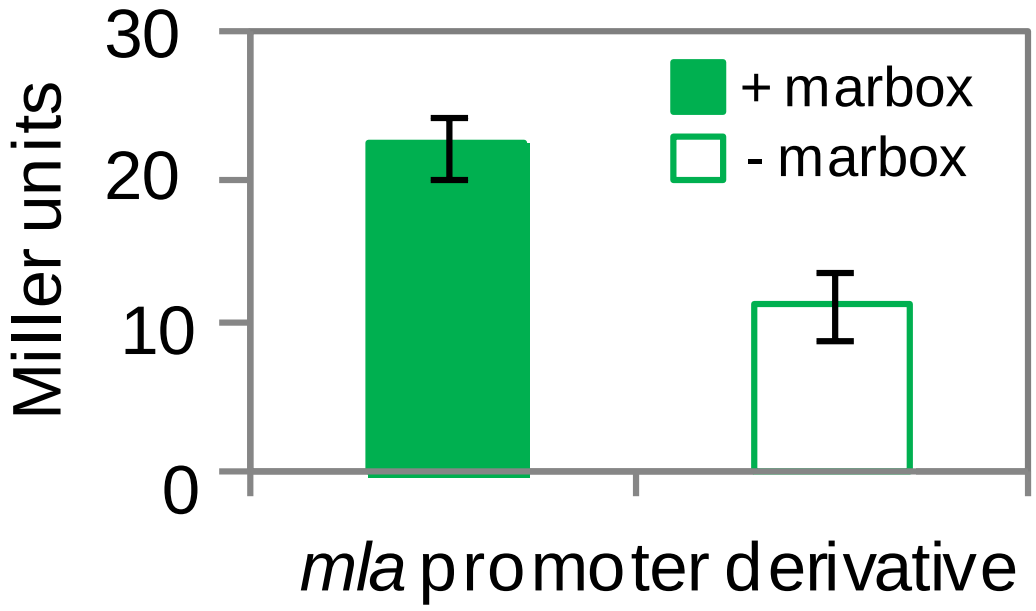


Figure 6.5: The *mlaFEDCB* activity requires an intact marbox *in vivo*. LacZ assay was done using JCB387 containing the pRW50 derivative of *mlaFEDCB* with or without marbox. Representative of n=3.

we sought to determine the MIC of doxycycline; an 8-fold MIC reduction occurred in *mlaE::kan* cells. To assess whether the doxycycline hypersensitivity phenotype was dependent on the *mla* operon and the marbox, we used genetic complementation. Hence, we generated derivatives of pBR322 encoding *mlaFEDCB,* and the upstream regulatory region, with or without the marbox. In the presence of doxycycline, the plasmid encoding *mlaFEDCB* was able to rescue the growth of *E. coli* strain BW25113 *mlaE::kan* but pBR322 by itself was not (Figure 6.6). However it was found that intact marbox was important for complementation. Since, in the presence of doxycycline, deletion of the marbox prevented growth (Figure 6.7).

6.5 Defective barrier functions but normal efflux was observed in the cells lacking *mlaE*

The barrier function of the bacterial outer membrane is dependent on its asymmetry; the outer leaflet is marked by the presence of lipopolysaccharides whereas phospholipids typify the inner leaflet. The *mlaFEDCB* operon encodes an ABC transport system that removes phospholipids from the outer leaflet (Malinverni and Silhavy, 2009). Moreover, a gene encoding MlaE from the same operon is known to form a protein channel that acts as a bridge between the cytoplasm and the periplasm. Therefore, we reasoned the *mla* locus could provide enhanced resistance to drugs via altered outer membrane barrier function or via ABC transporter activity and MlaE (Malinverni and Silhavy, 2009). To assess the possibilities two key factors, drug accumulation and drug efflux respectively, were compared in wild type and *mlaE*::kan cells. Doxycycline accumulation increased drastically when *mlaE* was disrupted (Figure 6.8) whilst ethidium bromide

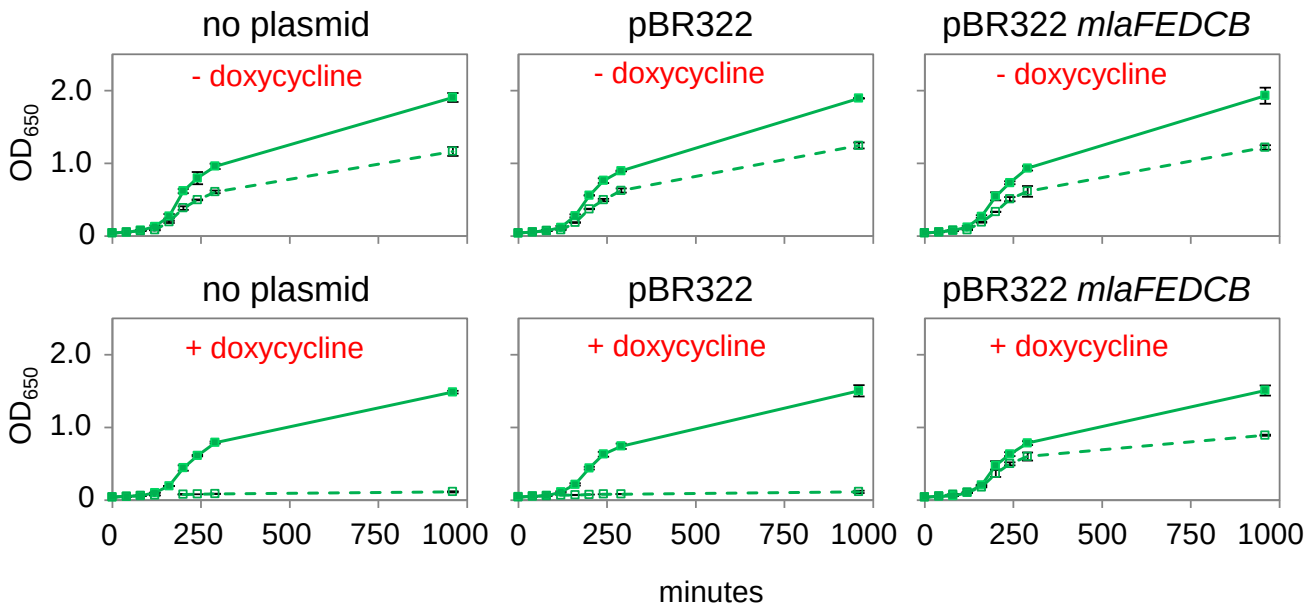


Figure 6.6: Genetic complementation of *mlaF*::kan phenotype. OD_{650} values of BW25113 and *mlaF::kan* grown in: absence of pBR322 (left panel), in the presence of pBR322 on its own (middle panel), and pBR322 carrying the *mlaFEDCB* operon (right panel). The cells were grown in absence or presence of 1 µg/ml doxycycline, top and bottom panels respectively. Briefly the pBR322 on its own and a derivative of *mlaFEDCB* (marbox present) were transformed in BW25113 and *mlaE::kan* cells and growth was recorded. Representative of n=3.

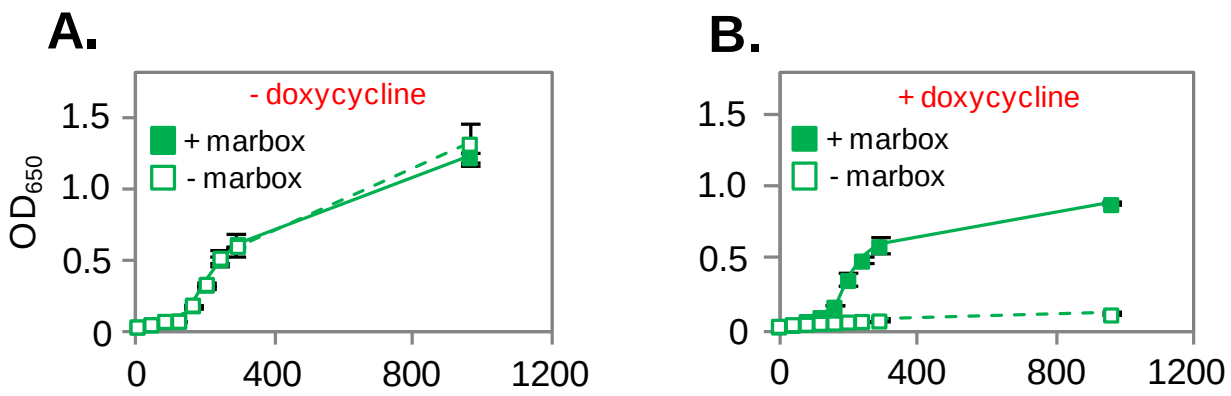


Figure 6.7: An intact marbox is required for genetic complementation of *mlaF*::kan phenotype. The graph shows the OD650 values of the BW25113 *mlaE::kan* grown in: A. absence or B. presence of 1 µg/ml doxycycline. Briefly the pBR322 derivative of *mlaFEDCB* with or without the marbox was transformed in *mlaE::kan* cells and growth was recorded. Representative of n=3.

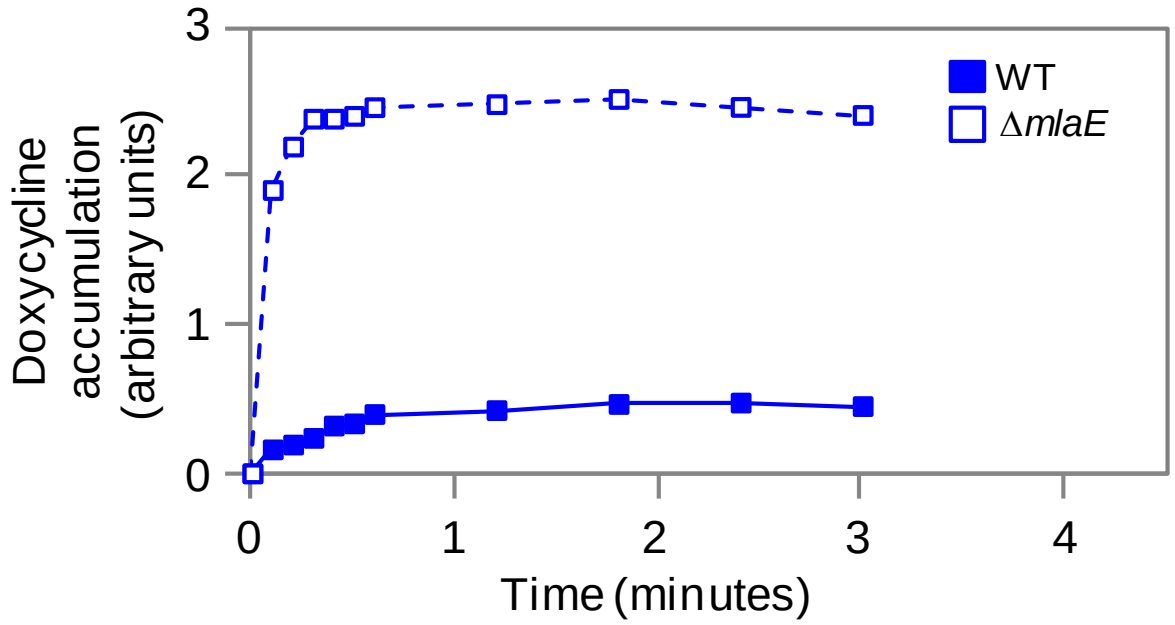


Figure 6.8: Cells lacking *mlaE* are defective for barrier function. Graph showing the accumulation of doxycycline. BW25113 cells are shown in solid blue line, the *mlaE*::kan derivative in dashed blue line. Representative of n=3.

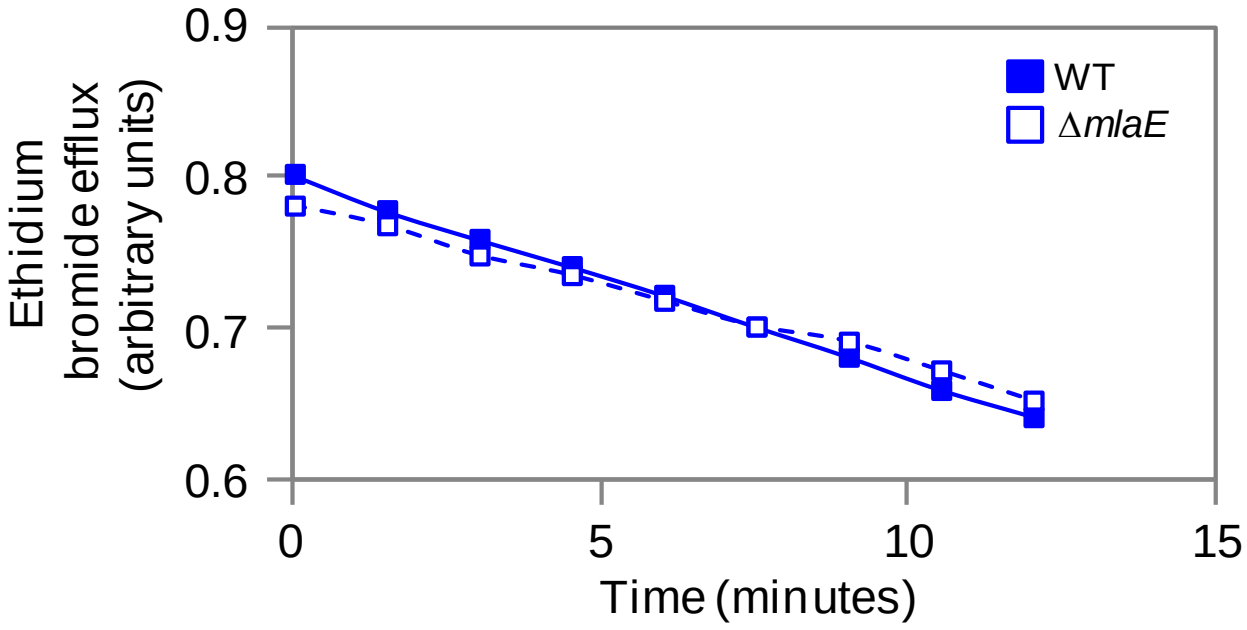


Figure 6.9: Cells lacking *mlaE* are not defective for drug efflux. Graph showing the efflux of ethidium bromide. BW25113 cells are shown in solid blue line, the *mlaE*::kan derivative in dashed blue line. Representative of n=3.

efflux remained the same in both the strains (Figure 6.9). For this reason it could be assumed that drug efflux will also remain the same in both the strains. The increased permeability suggests that MarA controls outer membrane barrier function by regulating *mlaFEDCB*.

6.6 Surface hydrophobicity increased in the cells lacking *mlaFEDCB*

A greater amount of phospholipid in a bacterium's outer leaflet should correspond to greater surface hydrophobicity. Therefore, to measure such an effect, the partitioning of cells between the organic and aqueous phase of a two solvent system (PUM buffer and p-xylene) was assessed. Cells in PUM buffer were allowed to mix with p-xylene. The organic and aqueous phases were then allowed to reach equilibrium. To detect the segregation of cells we used light absorbance. Cells with greater hydrophobicity should move towards the organic phase equating to a reduced light absorbance by aqueous phase (Rosenberg *et al.,* 1980). As expected, we found such a change; cells without *mlaFEDCB,* compared to wild type cells migrate to the organic phase more readily (Figure 6.10). We also assessed binding of cells with the hydrophobic dye crystal violet (Halder *et al*. 2015). We found that the cells lacking *mlaFEDCB* were more successful at removing crystal violet from aqueous solution (Figure 6.11).

6.7 Discussion

Cell permeability in *E. coli* depends on the assymetric distribution of lipids in the outer membrane. OmpF is a porin, which is a major component of outer membrane that regulates the entry of many small, hydrophilic molecules inside the cells

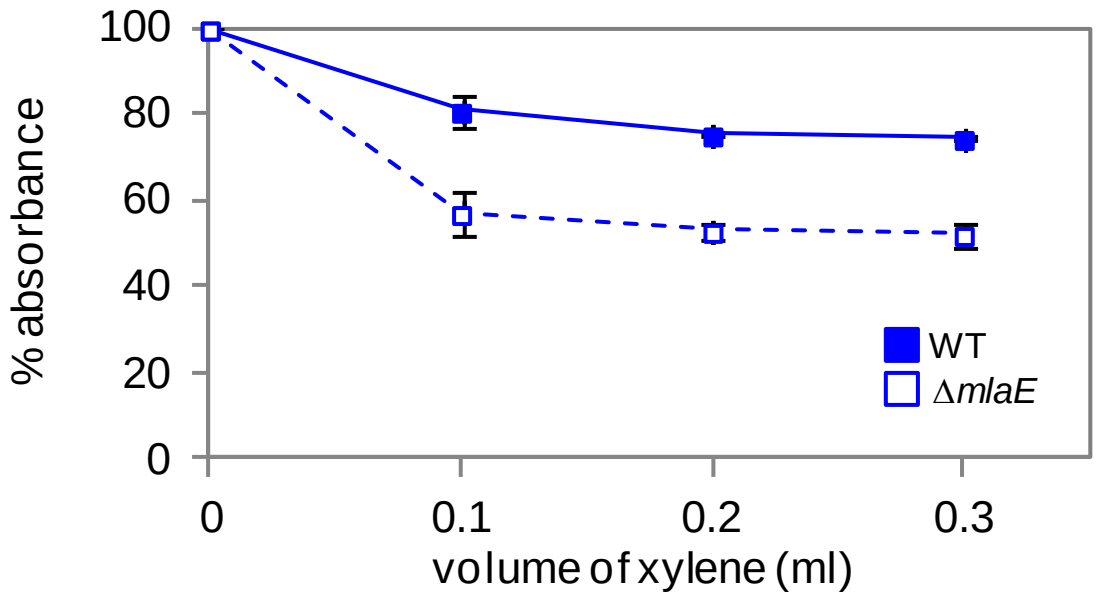


Figure 6.10: Cells that lack *mlaE* have increased surface hydrophobicity. Graph showing the percentage absorption of the aqueous layer of the mixture containing the indicated volume of p-xylene. BW5113 cells are represented by a solid blue line, *mlaE::kan* derivative by a dashed blue line. Representative of n=3.

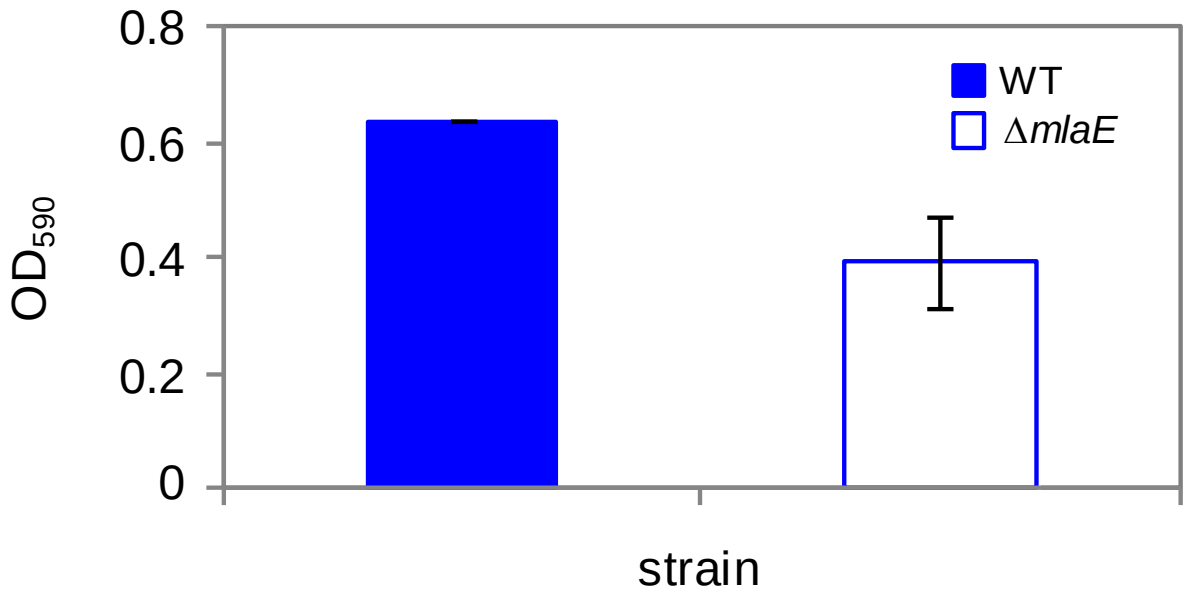


Figure 6.11: Cells that lack *mlaE* have increased surface hydrophobicity. Graph showing the OD_{590} of a crystal violet solution after mixing and incubating it with BW25113 (solid blue bar) and *mlaE::kan* derivative (open bar). Low OD_{590} value shows increased binding of crystal violet with bacterial cells. Representative of n=3.

(Chou *et al.,* 1993, Delihas and Forst, 2001, Cohen *et al.,* 1989). MarA activates a small inhibitory RNA known as *micF,* that downregulates expression of OmpF. Hence, the cells become resistance to many antibiotics by the decrease in the permeability (Cohen *et al.,* 1988).

In this chapter we studied the role of MarA in the regulation of *mlaFEDCB.* Many genes are targeted by MarA including those that have involvement with transport systems. An example is *mlaFEDCB*. The *mlaFEDCB* locus enables MarA to govern the control of lipid trafficking, hydrophobicity of cell surface and outer membrane permeability (Malinverni and Silhavy, 2009, Chng *et al.,* 2010). The tetracycline family of antibiotics have different properties; tetracycline is less hydrophobic than doxycycline (Leive *et al.,* 1984). Thus we found that the cells lacking *mlaFEDCB* become more sensitive to doxycycline than tetracycline. The data was taken and re-analysed from Nichols *et al.* (2011) (Figure 4.6). Similar to our findings others have noted a positive correlation between the *mar* resistance and compound hydrophobicity (Asako *et al.,* 1997, Ankarloo *et al.,* 2010). Interestingly, the Mla system is already predicted to have an import function; preventing phospholipid accumulation in the outer leaflet of the outer membrane. This accumulation indicates disruption in packing of the OM that reduces barrier function (Nikaido, 2005).

Until Thong et al (2016) have revealed it was unknown how MlaFEDB is organised and the role of each protein was also unknown. The authors have shown that MlaD and MlaB are required for the assembly and activation of the complex. They also show that all the proteins in the complex MlaFEDB interact tightly with each other to form a stable complex. It was shown that MlaD exits in hexameric form which makes it most stable

and is involved in binding to phospholipids (Thong *et al.,* 2016). The mechanism of working, organisation and roles are well understood. However, the consequences associated with antibiotic resistance were unknown. Hence, the regulation and function of *mlaFEDCB* associated multiple antibiotic resistance was explained and it was inferred that this aspect of the MarA system accompanies *acrAB-tolC* and the *xseA* facilitated resistance.

Chapter 7. Final conclusions and general discussions.

7.1 Final conclusions and General discussions

7.1.1 Overview

ETEC is a major cause of infection in both humans and animals (Gupta *et al.,* 2008). Since its discovery approximately 50 years ago (Sack *et al.,* 1971), ETEC and its virulence factors have been actively researched. We show that enterotoxin encoding gene expression is under the control of CRP and H-NS. We have also shown that there are many genes that can be controlled by MarA. Of particular interest were *xseA* and *mlaFEDCB.* MarA dependent regulation of these genes gives increased resistance to quinolones and tetracyclines.

Our model for regulation of enterotoxin expression has implications related to treatment of ETEC diarrahoea. Oral rehydration therapy, which is used to rehydrate patients, may have consequences for toxin gene regulation. For example, rehydration solutions contain sugar and salts; expression of ST toxin decreases when glucose is present and increases when salt is present. This is because CRP and H-NS are sensors of glucose and osmolarity respectively.

With respect to the MarA regulon, we were able to understand the regulation of two new targets. First, *xseA* encodes the larger subunit of exonuclease VII that has a role in DNA repair. Thus, cells that lack *xseA* become sensitive to fluoroquinolones. We suggest that after being activated by MarA, *xseA* helps in the repair of the DNA breaks generated by the action of fluoroquinolones. This may provide with an opportunity to acquire mutations that enhance quinolone resistance. Second is the *mlaFEDCB* operon that encodes the ABC transporter system with a role in lipid trafficking. Cells that lack any

gene of this operon become sensitive to tetracyclines. Our accumulation assay shows that after getting activated by MarA, *mlaFEDCB* helps form an enhanced barrier for tetracycline uptake. In the future, inhibitors of XseA and MlaFEDCB could increase the efficiency of quinolones and tetracycline. We suggest that MarA targets found in this study can also be targeted by Rob, SoxS, and other MarA-like regulators. Hence, a deep understanding of such regulation is necessary in order to completely understand multiple antibiotic resistance systems.

In this chapter I will discuss my data in the context of findings from other laboratories, further consider my interpretation of the data presented and highlight some future research directions.

7.1.2 Understanding why chromatin immunoprecipitation does not identify all binding sites for CRP and MarA

It is well established that chromatin immunoprecipitation, followed by either microarray deep sequencing analysis, is unable to identify all binding targets for a given gene regulatory protein (Grainger *et al.,* 2005, Grainger *et al.,* 2007, Shimada *et al.,* 2008). In some instances, it may be that such false negative signals arise because epitopes on the surface of the protein of interest are occluded by other factors. For example, an adjacently bound DNA binding protein. In other instances, known targets may have been erroneously identified. However, the most common reason for DNA targets for a given protein not being isolated by chromatin immunoprecipitation appears to be occlusion of protein binding sites by other factors. These are often nucleoid associated proteins. This is best demonstrated for FNR, where 111 of 187 targets are blocked by Fis, IHF or H-NS (Myers *et al.,* 2013). This is very likely true for all regulators, including

CRP and MarA. Hence, in Chapter 3, I show that H-NS is able to block binding of CRP to the *estA* promoters and that this has an important regulatory effect. Binding sites for CRP can also be occluded by specific transcriptional regulators. Thus, at the well characterised *lacZYA* promoter, CRP binding sites overlap with binding sites for the Lac Repressor protein (Liu *et al.,* 2004). This likely explains why CRP binding at the *lac* operon promoter is not detected using chromatin immunoprecipitation in the absence of lactose (Haycocks *et al.,* 2015).

The issue of DNA binding site occlusion is particularly pertinent to interpretation of my MarA ChIP-seq data. For instance, the marbox upstream of *acnA*, which was not detected in this work, overlaps with binding sites for CRP, FNR, and AcrA (Gruer and Guest, 1994). Indeed, examination of the Ecocyc database reveals that around half of all previously described MarA binding sites overlap with a known binding target for at least one other transcription factor (Keseler *et al.,* 2017). Of course, if one considers competition with SoxS and Rob, it is easy to see how MarA could be excluded from many targets. Consistent with this, published *in vitro* binding constants suggest that almost all known marboxes are actually preferentially bound SoxS or Rob (Table 7.1). Consistent with this, the MarA ChIP-seq presented in this work identifies known marboxes only if they bind MarA tightly *in vitro* (Table 7.1). Thus, although the acrAB operon promoter is considered a target for MarA, the locus actually has a much higher affinity for SoxS and Rob with SoxS also being more efficient at activating *acrAB* expression than MarA (Table 7.1). It is also notable that the evidence for around half of "known" MarA targets is actually quite poor (see Table 7.1 for a summary). For

example, although they are considered targets, MarA binding to *pqiA*, *inaA*, *hdeA*, *sodA*, *fpr* and *purA* cannot be detected *in vitro* (Table 7.1).

7.1.3 Potential for overlapping regulation of *mlaFEDCB* and *xseA* by SoxS and Rob

It is entirely possible that the MarA targets identified here will also be targets for SoxS and/or Rob. Indeed, this seems very likely given the known similarities in the DNA binding properties of these proteins (Martin and Rosner, 2001, Alekshun and Levy, 1997). Importantly, overlapping regulation would not alter the main conclusion that the *mlaFEDCB* and *xseA* promoters are regulated by MarA; if the genes can also be regulated by SoxS or Rob this in no way alters our interpretation. Whilst the issue of overlapping regulation has not been addressed in this thesis experiments that would tackle the problem can be proposed. For instance, *in vitro* DNA binding assays such as those presented in Figure 4.4 could be used to assess the relative binding affinities of MarA, SoxS and Rob for the different targets. Similarly, monitoring reconstituted *in vitro* transcription from the *mlaFEDCB* P2 promoter could also address the topic. It should also be possible to probe the contributions of the various transcription factors *in vitro*. Thus, ectopic overexpression of MarA, SoxS or Rob could be used to determine the relative abilities of these proteins to activate transcription. Furthermore, *in vivo* DNA binding of the different regulators could be assessed by chromatin immunoprecipitation. To this end, Seo *et al.* (2015) have already mapped genome wide SoxS binding using a technique called ChIP-exo (Seo *et al.,* 2015). Briefly, in ChIP-exo experiments, Immunoprecipitated protein DNA complexes are treated with an exonuclease to

Table 7.1: Interactions of MarA and SoxS with previously proposed targets

Gene[a]	MarA binding *in vitro*[b]	Affinity (Kd (nM^{-1})[c]			Relative activation (SoxS/MarA)[d]	Other comments[e]
		MarA	SoxS	Rob		
micF	yes	25	50	n.d.	1	
marRAB	yes	75	75	20	1.1	
ybjC	yes	320	100	35	0.8	
fumC	yes	320	75	75	5.3	
rob	yes	400*	100*	<100*	n.d.	
nfsB	yes	>500*	<200*	<200*	0.6	
acrAB	yes	800	128	35	1.8	
acnA	yes	1000	350	500	3.2	
zwf	yes	>1000	>1000	1000	2.3	
pqiA	none	n.d.	n.d.	n.d.	4	Martin *et al*. 2011 unable to detect binding of MarA or SoxS *in vitro* at concentrations tested.
inaA	none	n.d.	n.d.	n.d.	0.8	Martin *et al*. 2011 unable to detect binding of MarA or SoxS *in vitro* at concentrations tested.
hdeA	none	n.d.	n.d.	n.d.	n.d.	Schneiders *et al*. 2004 identified a potential MarA site but binding could not be detected *in vitro*.
sodA	none	n.d.	n.d.	n.d.	2.3	Martin *et al*. 2011 unable to detect binding of MarA or SoxS *in vitro* at concentrations tested.
fpr	none	n.d.	200	n.d.	23	Martin et al. 2011 detected binding of SoxS but not MarA*in vitro* at concentrations tested.
purA	yes	n.d.	n.d.	n.d.	n.d.	Schneiders *et al*. 2004 detected MarA binding *in vitro* but no comparison with SoxS was made

[a]Genes listed are those described in Ecocyc as being MarA targets on the basis of high quality evidence (usually *in vitro* DNA binding assays). Predicted targets listed by Ecocyc are not included. Underlined genes were identified by ChIP-seq as MarA binding targets in this work. Genes in bold were identified by ChIP-exo as SoxS targets.

[e]Experimental evidence for purified MarA binding to the proposed target.

[c]Affinities are reported where binding of MarA and SoxS were directly compared in the same paper. An asterisk indicates values estimated from published images of gel electrophoretic mobility shift assays rather than values directly provided by the authors. In the table, the order in which genes is based in affinity for MarA *in vitro*. If a binding constant could not be determined this is shown as n.d.

[d]Martin *et al*. 2011 directly compared the ability of MarA and SoxS to activate transcription of the respective genes. A relative activation of >1 indicates more efficient activation by SoxS. If the comparison has not been done this is shown as n.d.

[e]In some instances Ecocyc lists genes as being MarA targets on the basis of DNA binding data. On inspection, DNA binding experiments were indeed done but no binding was detected under the conditions tested.

generate short DNA fragments that precisely correspond to the DNA binding targets in the immuno precipitated protein. Using this approach, 25 binding sites for SoxS were identified in *E. coli* and 7 of these were also identified as MarA binding sites in my experiments. The overlap between the MarA and SoxS data is shown in Figure 7.1. Note that this comparison shows binding of SoxS but not MarA to the *acrAB* promoter as eluded to above.

7.1.4 The Rob paradox and its implications for this work

As noted elsewhere, members of the AraC family of transcription factors bind to DNA using a ~100 amino acid dual helix-turn-helix (HTH) motif (Rhee *et al.,* 1998). In most cases, each HTH motif engages the DNA major groove and these sites of engagement are separated by one turn of the double helix (Rhee *et al.,* 1998). The MarA and SoxS proteins are 127 and 107 amino acids in length respectively and share 42% sequence identity (66% similarity). Differences in their DNA binding determinants result in MarA and SoxS preferentially recognising distinct subsets of marboxes (Martin *et al.,* 2000). Conversely, Rob is a 289 amino acid protein that contains a dual HTH DNA binding domain and a C-terminal multimerisation domain. Comparison of the dual HTH motif in MarA, with the equivalent amino acid sequence in Rob, reveals 43% sequence identity (62% similarity). However, Rob binds the DNA major groove using only one of its two HTH motifs (Kwon *et al.,* 2000). The second HTH interacts non-specifically with the minor grove via a surface exposed arginine side chain (Kwon *et al.,* 2000). As a result, Rob binds DNA less specifically, but with a higher affinity, than MarA and SoxS (Kwon *et al.,* 2000). Hence, all marboxes appears to bind Rob with a higher affinity than SoxS

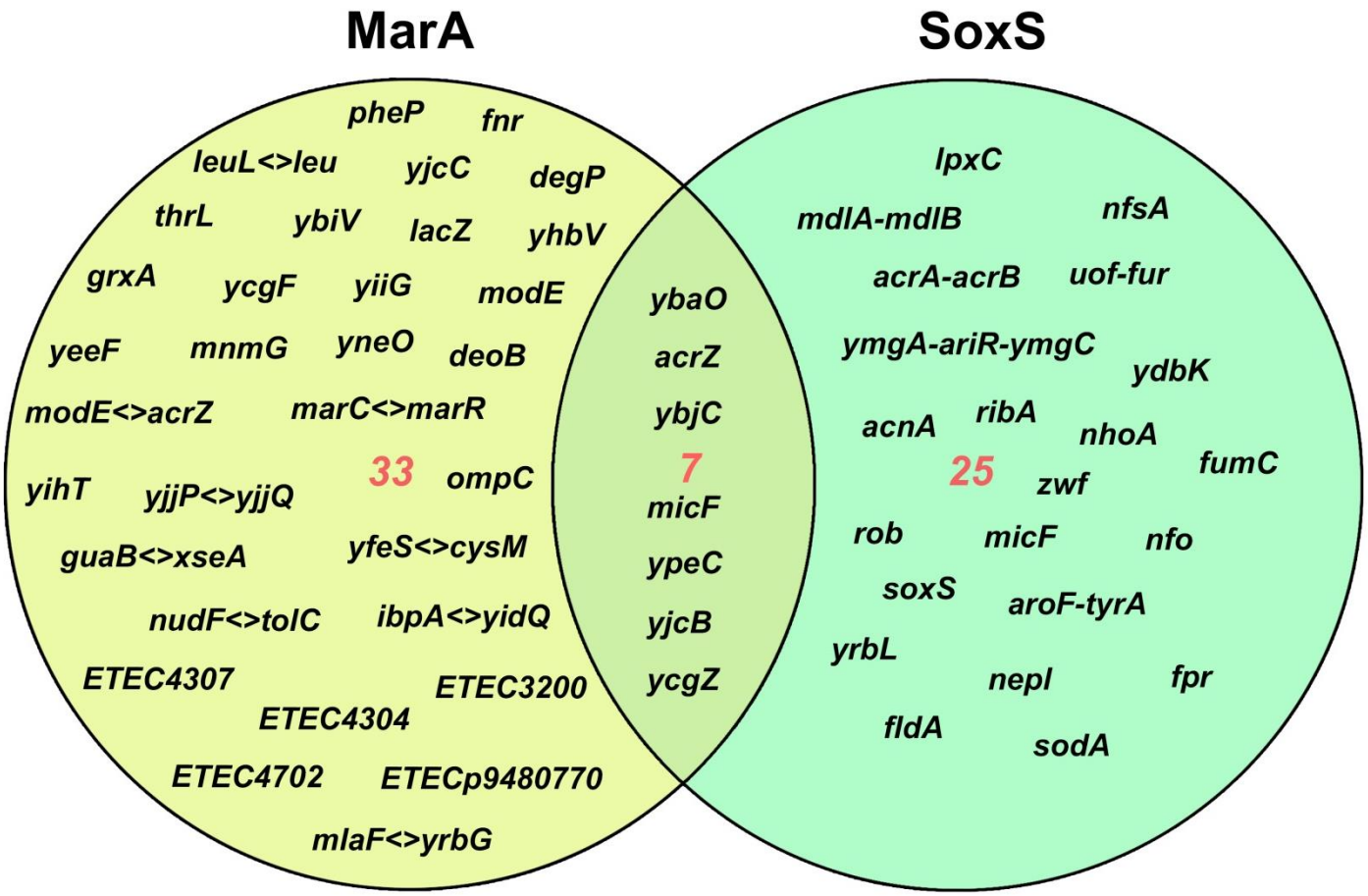


Figure 7.1: Overlap in the genome-wide DNA binding profiles of MarA in ETEC H10407 and SoxS in *E. coli* K-12. Venna diagram showing the MarA targets obtained from our analysis in light green colour on the left and the SoxS target obtained from Seo *et al.*(2015) is shown in green color on the right. The targets that are in common between MarA and SoxS are shown in the overlapping region of the two circles. The numbers in red denotes the number of targets.

or MarA ((Kwon *et al.,* 2000), Table 7.1). The three proteins also differently expressed; Rob is produced constitutively whilst MarA and SoxS are up-regulated in response to stress (Barbosa and Levy, 2000). Despite this, MarA and SoxS are able to outcompete Rob for binding many marboxes (Barbosa and Levy, 2000). This behaviour appears paradoxical but is possible because Rob is usually sequestered in foci whilst MarA and SoxS form ternary complexes with RNA polymerase that have an elevated affinity for marbox containing promoters (Barbosa and Levy, 2000, Kwon *et al.,* 2000). Thus, MarA targets that are identified in this thesis are likely to bind Rob tightly *in vitro* but this is unlikely to be indicative of specific Rob binding.

7.1.5 Why were targets such as *mlaFEDCB* and *xseA* not found in previous attempts to map the MarA regulon?

Most of the DNA targets identified here for MarA have not been identified by previous genome-wide analyses of the MarA system. Briefly, transcriptome analyses identified hundreds of genes putatively controlled by MarA (Barbosa and Levy, 2000, Pomposiello *et al.,* 2001). However, only 3 genes were common to independent studies (Martin and Rosner, 2002). Additionally, many of the targets identified by our work, including both *mlaFEDCB* and *xseA*, were not identified by either transcriptome analysis (Pomposiello *et al.,* 2001, Barbosa and Levy, 2000). In part, the failure of transcriptome analyses to identify many MarA targets may be due to complications arising from contributions to regulation made by SoxS and Rob, which can compensate for loss of MarA (Zhang *et al.,* 2008, Chubiz and Rao, 2011). Hence, MarA regulated genes cannot necessarily be identified by comparing transcription profiles generated in the presence and absence of

MarA. Interestingly, in the work of Pomposiello *et al.,* (2001) global gene expression was probed after treatment with stressors known to induce the MarA and SoxS systems and, as such, this data is difficult to interpret. An added complication, particularly for *xseA*, is that the effect of MarA on gene expression can be modest in certain conditions. Such smaller changes in gene expression are often removed from transcriptome datasets.

Although not an attempt to map the MarA regulon, Ruiz & Levy (2010) conceived a genetic screen to identify chromosomal mutations altering MarA multidrug resistance. This genetic screen also failed to identify the MarA target genes found in this thesis. In this case, there is an obvious explanation; Ruiz & Levy did not directly screen for mutations abrogating MarA mediated multidrug resistance. As Ruiz & Levy note, “we would obtain many mutants that affected MDR by mechanisms unrelated to MarA”. Instead, Ruiz & Levy screened for chromosomal mutations that prevented repression of the *hdeAB* promoter by MarA. Clearly, there is no reason to think that *mlaFEDCB* or *xseA* mutation would alter *hdeAB* repression. The same is true of most MarA regulated genes. Hence, the screen of Ruiz and Levy simply identified *hdeAB* specific regulators and mutations that alter intracellular MarA levels (Ruiz and Levy, 2010).

7.1.6 Comparison of MarA binding sites in ETEC and *Escherichia coli* K-12

A key assumption of this work is that the MarA binding sites that I have identified using ChIP-seq in ETEC strain H10407 are also MarA targets in the *E. coli* K-12 lab strain. For *xseA* (Chapter 5) and *mlaFEDCB* (Chapter 6) my subsequent analysis supports this

assumption. To better understand if other targets are shared, and to what extent, between the two strains a bioinformatics comparison was made. The *xseA*/*mlaFEDCB* genes and intergenic regions are identical in *E. coli* K-12 and ETEC. This is also the case for the majority of MarA targets we identified. Thus, the 200 base pairs upstream of each MarA target identified by ChIP-seq in ETEC are at least 99% identical at the DNA sequence level to the equivalent region in *E. coli* K-12. Similarly, the DNA sequence of the downstream gene has a minimum of 97% sequence similarity in all cases. Importantly, deletions or insertions never occur in either genes or regulatory regions when comparing MarA targets in ETEC and *E. coli* K-12. Hence, juxtaposition of regulatory elements is also conserved in all cases. It is possible that some of the apparent differences are simply mistakes in the ETEC H10407 genome sequence that is not as carefully curated and rechecked as that of *E. coli* K-12.

7.1.7 Wider roles for DNA repair systems in modulating the effects of quinolones on gram negative bacteria

This thesis demonstrates that the MarA regulated *xseA* gene, encoding the large subunit of exonuclease VII, can play an important role in modulating the effect of ciprofloxacin on *E. coli*. Unsurprisingly, because drugs like ciprofloxacin induce the formation of DNA double strand breaks, the same is true of other protein complexes involved with DNA repair. Most notably, the RecBCD complex has helicase and nuclease activity and also plays an important role. An interesting comparison is the different methods used to regulate *xseA* and *recBCD* expression (Figure 7.2). Briefly,

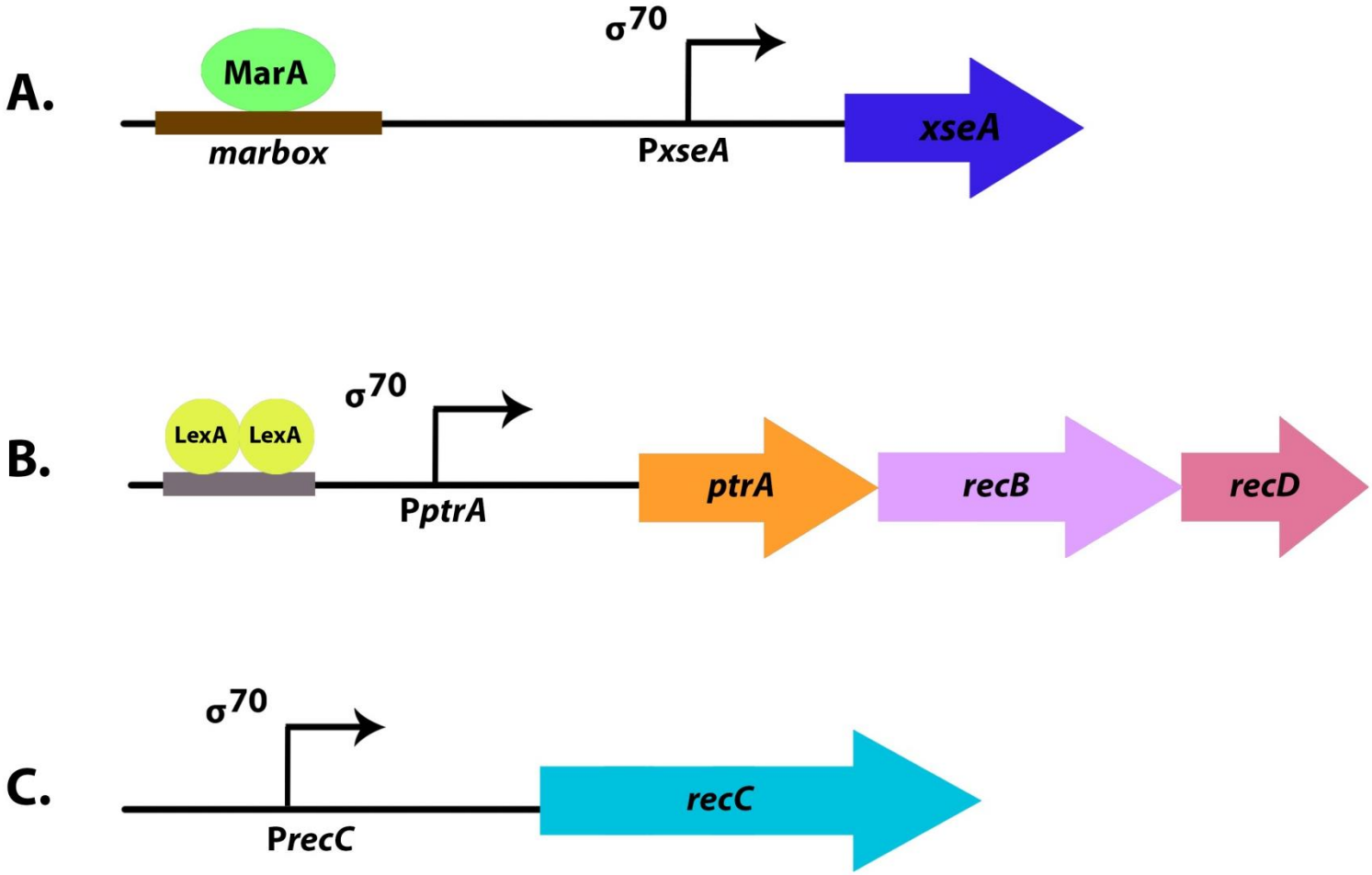


Figure 7.2: Comparison of the *xseA* and *recBCD* gene regulatory systems. All these systems are transcribed by σ^{70}. A. *xseA* regulatory system, B. *ptrA-recB-recD* operon and C. *recC* regulatory system. MarA and LexA are shown in green and yellow. The promoters are shown in black bent arrows.

recC is monocistronic and transcribed from a σ^{70} dependent promoter for which there are no identified regulatory proteins. Conversely, *recB* and *recD* reside in an operon with a third gene called *ptr* that encodes a protease. This operon is also transcribed from a σ^{70} dependent promoter that is usually repressed by the LexA protein (Claverie-Martin *et al.,* 1987, Wade *et al.,* 2005). Thus, the RecBCD complex is formed in response to stresses that induce the SOS system (Newmark *et al.,* 2005). Hence, XseA and RecBCD are produced in response to different environmental signals to contribute to DNA repair. Interestingly, the activity of RecBCD can also be regulated post-translationally by a phage encoded protein called Gam (Wilkinson *et al.,* 2016b). Briefly, Gam functions as a DNA mimic and occupies the DNA binding pocket of RecBCD to inhibit activity of the complex (Wilkinson *et al.,* 2016a). Thus, DNA repair systems can be controlled by many different signals. We note that all of the genes discussed in this section are transcribed from promoters dependent on the housekeeping RNA polymerase sigma factor. This is unlikely to be coincidental and probably represents a need for such genes to be expressed in many different growth conditions is a stress is encountered.

7.1.8 Future work

This thesis provides a platform for understanding the role of CRP and MarA in controlling gene regulation in ETEC and other bacterial strains. Whilst the CRP regulon is well-defined the set of genes controlled by MarA is much less well understood. Hence, it will be important to also investigate the genome-wide DNA binding by MarA in other important pathogens such as *Klebsiella* spp., *Salmonella* spp., *Shigella* spp. and *Yersinia* spp.. Interestingly, in these organisms, additional MarA-like proteins are found

in addition to SoxS and Rob. For instance, in *Klebsiella* spp. the RamA protein plays a key role in activating gene expression to provide protection against certain antibiotics. It is not clear why such ancillary protreins exist in bacteria other than *E. coli* and understand their presence is of key importance. This should be possible using techniques similar to those described here that should also enable a better understanding of the overlapping regulatory effects of MarA-like proteins.

References

[1] ABOUZEED, Y. M., BAUCHERON, S. & CLOECKAERT, A. 2008. ramR mutations involved in efflux-mediated multidrug resistance in Salmonella enterica serovar Typhimurium. *Antimicrob Agents Chemother,* 52**,** 2428-34.

[2] AFGAN, E., BAKER, D., VAN DEN BEEK, M., BLANKENBERG, D., BOUVIER, D., CECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., EBERHARD, C., GRUNING, B., GUERLER, A., HILLMAN-JACKSON, J., VON KUSTER, G., RASCHE, E., SORANZO, N., TURAGA, N., TAYLOR, J., NEKRUTENKO, A. & GOECKS, J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res,* 44**,** W3-W10.

[3] ALEKSHUN, M. N. & LEVY, S. B. 1997. Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. *Antimicrob Agents Chemother,* 41**,** 2067-75.

[4] ALEKSHUN, M. N. & LEVY, S. B. 1999. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol,* 7**,** 410-3.

[5] AMIT, R., OPPENHEIM, A. B. & STAVANS, J. 2003. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys J,* 84**,** 2467-73.

[6] ANANTHA, R. P., MCVEIGH, A. L., LEE, L. H., AGNEW, M. K., CASSELS, F. J., SCOTT, D. A., WHITTAM, T. S. & SAVARINO, S. J. 2004. Evolutionary and functional relationships of colonization factor antigen i and other class 5 adhesive fimbriae of enterotoxigenic Escherichia coli. *Infect Immun,* 72**,** 7190-201.

[7] ANKARLOO, J., WIKMAN, S. & NICHOLLS, I. A. 2010. Escherichia coli mar and acrAB mutants display no tolerance to simple alcohols. *Int J Mol Sci,* 11**,** 1403-12.

[8] AROLD, S. T., LEONARD, P. G., PARKINSON, G. N. & LADBURY, J. E. 2010. H-NS forms a superhelical protein scaffold for DNA condensation. *Proc Natl Acad Sci U S A,* 107**,** 15728-32.

[9] ARTSIMOVITCH, I. & LANDICK, R. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A,* 97**,** 7090-5.

[10] ASAKO, H., NAKAJIMA, H., KOBAYASHI, K., KOBAYASHI, M. & AONO, R. 1997. Organic solvent tolerance and antibiotic resistance increased by overexpression of marA in Escherichia coli. *Appl Environ Microbiol,* 63**,** 1428-33.

[11] ATLUNG, T. & INGMER, H. 1997. H-NS: a modulator of environmentally regulated gene expression. *Mol Microbiol,* 24**,** 7-17.

[12] BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., DATSENKO, K. A., TOMITA, M., WANNER, B. L. & MORI, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol,* 2**,** 2006 0008.

[13] BAE, B., FEKLISTOV, A., LASS-NAPIORKOWSKA, A., LANDICK, R. & DARST, S. A. 2015. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife,* 4.

[14] BALBAS, P., SOBERON, X., MERINO, E., ZURITA, M., LOMELI, H., VALLE, F., FLORES, N. & BOLIVAR, F. 1986. Plasmid vector pBR322 and its special-purpose derivatives--a review. *Gene,* 50**,** 3-40.

[15] BARBOSA, T. M. & LEVY, S. B. 2000. Differential expression of over 60 chromosomal genes in Escherichia coli by constitutive expression of MarA. *J Bacteriol,* 182**,** 3467-74.

[16] BARNARD, A., WOLFE, A. & BUSBY, S. 2004. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol,* 7**,** 102-8.

[17] BAXTER, K., LEE, J., MINAKHIN, L., SEVERINOV, K. & HINTON, D. M. 2006. Mutational analysis of sigma70 region 4 needed for appropriation by the bacteriophage T4 transcription factors AsiA and MotA. *J Mol Biol,* 363**,** 931-44.

[18] BELL, C. E. & LEWIS, M. 2001. Crystallographic analysis of Lac repressor bound to natural operator O1. *J Mol Biol,* 312**,** 921-6.

[19] BENOFF, B., YANG, H., LAWSON, C. L., PARKINSON, G., LIU, J., BLATTER, E., EBRIGHT, Y. W., BERMAN, H. M. & EBRIGHT, R. H. 2002. Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science,* 297**,** 1562-6.

[20] BERG, O. G. 1996. Selection intensity for codon bias and the effective population size of Escherichia coli. *Genetics,* 142**,** 1379-82.

[21] BLAKE, T., BARNARD, A., BUSBY, S. J. & GREEN, J. 2002. Transcription activation by FNR: evidence for a functional activating region 2. *J Bacteriol,* 184**,** 5855-61.

[22] BLATTNER, F. R., PLUNKETT, G., 3RD, BLOCH, C. A., PERNA, N. T., BURLAND, V., RILEY, M., COLLADO-VIDES, J., GLASNER, J. D., RODE, C. K., MAYHEW, G. F., GREGOR, J., DAVIS, N. W., KIRKPATRICK, H. A., GOEDEN, M. A., ROSE, D. J., MAU, B. & SHAO, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science,* 277**,** 1453-62.

[23] BLAZQUEZ, J., COUCE, A., RODRIGUEZ-BELTRAN, J. & RODRIGUEZ-ROJAS, A. 2012. Antimicrobials as promoters of genetic variation. *Curr Opin Microbiol,* 15**,** 561-9.

[24] BORUKHOV, S., LEE, J. & LAPTENKO, O. 2005. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol,* 55**,** 1315-24.

[25] BRENT, R. & PTASHNE, M. 1981. Mechanism of action of the lexA gene product. *Proc Natl Acad Sci U S A,* 78**,** 4204-8.

[26] BROWN, E. A. & HARDWIDGE, P. R. 2007. Biochemical characterization of the enterotoxigenic Escherichia coli LeoA protein. *Microbiology,* 153**,** 3776-84.

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

[28] BROWNING, D. F. & BUSBY, S. J. 2016. Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol,* 14**,** 638-50.

[29] BURGESS, R. R., TRAVERS, A. A., DUNN, J. J. & BAUTZ, E. K. 1969. Factor stimulating transcription by RNA polymerase. *Nature,* 221**,** 43-6.

[30] BURMANN, B. M., LUO, X., ROSCH, P., WAHL, M. C. & GOTTESMAN, M. E. 2010. Fine tuning of the *E. coli* NusB:NusE complex affinity to BoxA RNA is required for processive antitermination. *Nucleic Acids Res,* 38**,** 314-26.

[31] BURNS, C. M., RICHARDSON, L. V. & RICHARDSON, J. P. 1998. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J Mol Biol,* 278**,** 307-16.

[32] BUSBY, S. & EBRIGHT, R. H. 1999. Transcription activation by catabolite activator protein (CAP). *J Mol Biol,* 293**,** 199-213.

[33] BUSBY, S., KOTLARZ, D. & BUC, H. 1983. Deletion mutagenesis of the Escherichia coli galactose operon promoter region. *J Mol Biol,* 167**,** 259-74.

[34] CAMPBELL, E. A., MUZZIN, O., CHLENOV, M., SUN, J. L., OLSON, C. A., WEINMAN, O., TRESTER-ZEDLITZ, M. L. & DARST, S. A. 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell,* 9**,** 527-39.

[35] CARMONA, M. & MAGASANIK, B. 1996. Activation of transcription at sigma 54-dependent promoters on linear templates requires intrinsic or induced bending of the DNA. *J Mol Biol,* 261**,** 348-56.

[36] CAVANAGH, A. T. & WASSARMAN, K. M. 2014. 6S RNA, a global regulator of transcription in Escherichia coli, Bacillus subtilis, and beyond. *Annu Rev Microbiol,* 68**,** 45-60.

[37] CHAKRABORTY, A., WANG, D., EBRIGHT, Y. W., KORLANN, Y., KORTKHONJIA, E., KIM, T., CHOWDHURY, S., WIGNESHWERARAJ, S., IRSCHIK, H., JANSEN, R., NIXON, B. T., KNIGHT, J., WEISS, S. & EBRIGHT,

R. H. 2012. Opening and closing of the bacterial RNA polymerase clamp. *Science,* 337**,** 591-5.

[38] CHAO, A. C., DE SAUVAGE, F. J., DONG, Y. J., WAGNER, J. A., GOEDDEL, D. V. & GARDNER, P. 1994. Activation of intestinal CFTR Cl- channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. *EMBO J,* 13**,** 1065-72.

[39] CHASE, J. W. & MASKER, W. E. 1977. Deoxyribonucleic acid repair in Escherichia coli mutants deficient in the 5'----3' exonuclease activity of deoxyribonucleic acid polymerase I and exonuclease VII. *J Bacteriol,* 130**,** 667-75.

[40] CHASE, J. W., RABIN, B. A., MURPHY, J. B., STONE, K. L. & WILLIAMS, K. R. 1986. Escherichia coli exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (xseA). *J Biol Chem,* 261**,** 14929-35.

[41] CHATTERJI, D., OGAWA, Y., SHIMADA, T. & ISHIHAMA, A. 2007. The role of the omega subunit of RNA polymerase in expression of the relA gene in Escherichia coli. *FEMS Microbiol Lett,* 267**,** 51-5.

[42] CHEN, H., TANG, H. & EBRIGHT, R. H. 2003. Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. *Mol Cell,* 11**,** 1621-33.

[43] CHNG, S. S., RUIZ, N., CHIMALAKONDA, G., SILHAVY, T. J. & KAHNE, D. 2010. Characterization of the two-protein complex in Escherichia coli responsible for lipopolysaccharide assembly at the outer membrane. *Proc Natl Acad Sci U S A,* 107**,** 5363-8.

[44] CHO, B. K., KIM, D., KNIGHT, E. M., ZENGLER, K. & PALSSON, B. O. 2014. Genome-scale reconstruction of the sigma factor network in Escherichia coli: topology and functional states. *BMC Biol,* 12**,** 4.

[45] CHOU, J. H., GREENBERG, J. T. & DEMPLE, B. 1993. Posttranscriptional repression of Escherichia coli OmpF protein in response to redox stress: positive control of the micF antisense RNA by the soxRS locus. *J Bacteriol,* 175**,** 1026-31.

[46] CHUBIZ, L. M. & RAO, C. V. 2011. Role of the mar-sox-rob regulon in regulating outer membrane porin expression. *J Bacteriol,* 193**,** 2252-60.

[47] CLAVERIE-MARTIN, F., DIAZ-TORRES, M. R. & KUSHNER, S. R. 1987. Analysis of the regulatory region of the protease III (ptr) gene of Escherichia coli K-12. *Gene,* 54**,** 185-95.

[48] COHEN, S. P., HACHLER, H. & LEVY, S. B. 1993a. Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli. *J Bacteriol,* 175**,** 1484-92.

[49] COHEN, S. P., MCMURRY, L. M., HOOPER, D. C., WOLFSON, J. S. & LEVY, S. B. 1989. Cross-resistance to fluoroquinolones in multiple-antibiotic-resistant (Mar) Escherichia coli selected by tetracycline or chloramphenicol: decreased drug accumulation associated with membrane changes in addition to OmpF reduction. *Antimicrob Agents Chemother,* 33**,** 1318-25.

[50] COHEN, S. P., MCMURRY, L. M. & LEVY, S. B. 1988. marA locus causes decreased expression of OmpF porin in multiple-antibiotic-resistant (Mar) mutants of Escherichia coli. *J Bacteriol,* 170**,** 5416-22.

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

[52] CRAMER, P., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science,* 292**,** 1863-76.

[53] CROSSMAN, L. C., CHAUDHURI, R. R., BEATSON, S. A., WELLS, T. J., DESVAUX, M., CUNNINGHAM, A. F., PETTY, N. K., MAHON, V., BRINKLEY, C., HOBMAN, J. L., SAVARINO, S. J., TURNER, S. M., PALLEN, M. J., PENN, C. W., PARKHILL, J., TURNER, A. K., JOHNSON, T. J., THOMSON, N. R., SMITH, S. G. & HENDERSON, I. R. 2010. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. *J Bacteriol,* 192**,** 5822-31.

[54] CROXEN, M. A., LAW, R. J., SCHOLZ, R., KEENEY, K. M., WLODARSKA, M. & FINLAY, B. B. 2013. Recent advances in understanding enteric pathogenic Escherichia coli. *Clin Microbiol Rev,* 26**,** 822-80.

[55] DAME, R. T., WYMAN, C., WURM, R., WAGNER, R. & GOOSEN, N. 2002. Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1. *J Biol Chem,* 277**,** 2146-50.

[56] DARST, S. A. 2001. Bacterial RNA polymerase. *Curr Opin Struct Biol,* 11**,** 155-62.

[57] DE LAS PENAS, A., CONNOLLY, L. & GROSS, C. A. 1997. SigmaE is an essential sigma factor in Escherichia coli. *J Bacteriol,* 179**,** 6862-4.

[58] DELIHAS, N. & FORST, S. 2001. MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors. *J Mol Biol,* 313**,** 1-12.

[59] DEMPLE, B. 1996. Redox signaling and gene control in the Escherichia coli soxRS oxidative stress regulon--a review. *Gene,* 179**,** 53-7.

[60] DORMAN, C. J. 2007. H-NS, the genome sentinel. *Nat Rev Microbiol,* 5**,** 157-61.

[61] DORMAN, C. J. 2014. H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria. *Plasmid,* 75**,** 1-11.

[62] DOUZI, B., FILLOUX, A. & VOULHOUX, R. 2012. On the path to uncover the bacterial type II secretion system. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1059-72.

[63] DUVAL, V., FOSTER, K., BREWSTER, J. & LEVY, S. B. 2017. A Novel Regulatory Cascade Involving BluR, YcgZ, and Lon Controls the Expression of Escherichia coli OmpF Porin. *Front Microbiol,* 8**,** 1148.

[64] DWYER, D. J., BELENKY, P. A., YANG, J. H., MACDONALD, I. C., MARTELL, J. D., TAKAHASHI, N., CHAN, C. T., LOBRITZ, M. A., BRAFF, D., SCHWARZ, E. G., YE, J. D., PATI, M., VERCRUYSSE, M., RALIFO, P. S., ALLISON, K. R., KHALIL, A. S., TING, A. Y., WALKER, G. C. & COLLINS, J. J. 2014. Antibiotics induce redox-related physiological alterations as part of their lethality. *Proc Natl Acad Sci U S A,* 111**,** E2100-9.

[65] EGAN, S. M. 2002. Growing repertoire of AraC/XylS activators. *J Bacteriol,* 184**,** 5529-32.

[66] ERICKSON, J. W. & GROSS, C. A. 1989. Identification of the sigma E subunit of Escherichia coli RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev,* 3**,** 1462-71.

[67] ESPOSITO, D., PETROVIC, A., HARRIS, R., ONO, S., ECCLESTON, J. F., MBABAALI, A., HAQ, I., HIGGINS, C. F., HINTON, J. C., DRISCOLL, P. C. & LADBURY, J. E. 2002. H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein. *J Mol Biol,* 324**,** 841-50.

[68] ESTREM, S. T., ROSS, W., GAAL, T., CHEN, Z. W., NIU, W., EBRIGHT, R. H. & GOURSE, R. L. 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev,* 13**,** 2134-47.

[69] FERRARIS, R. P., YASHARPOUR, S., LLOYD, K. C., MIRZAYAN, R. & DIAMOND, J. M. 1990. Luminal glucose concentrations in the gut under normal conditions. *Am J Physiol,* 259**,** G822-37.

[70] FINN, R. D., ORLOVA, E. V., GOWEN, B., BUCK, M. & VAN HEEL, M. 2000. Escherichia coli RNA polymerase core and holoenzyme structures. *EMBO J,* 19**,** 6833-44.

[71] FLECKENSTEIN, J. M., HARDWIDGE, P. R., MUNSON, G. P., RASKO, D. A., SOMMERFELT, H. & STEINSLAND, H. 2010. Molecular mechanisms of enterotoxigenic Escherichia coli infection. *Microbes and Infection,* 12**,** 89-98.

[72] FLENTIE, K., GARNER, A. L. & STALLINGS, C. L. 2016. Mycobacterium tuberculosis Transcription Machinery: Ready To Respond to Host Attacks. *J Bacteriol,* 198**,** 1360-73.

[73] FOREST, K. T. 2008. The type II secretion arrowhead: the structure of GspI-GspJ-GspK. *Nat Struct Mol Biol,* 15**,** 428-30.

[74] FORTE, L. R., FREEMAN, R. H., KRAUSE, W. J. & LONDON, R. M. 1999. Guanylin peptides: cyclic GMP signaling mechanisms. *Braz J Med Biol Res,* 32**,** 1329-36.

[75] FOTI, J. J., DEVADOSS, B., WINKLER, J. A., COLLINS, J. J. & WALKER, G. C. 2012. Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science,* 336**,** 315-9.

[76] GAAL, T., ROSS, W., BLATTER, E. E., TANG, H., JIA, X., KRISHNAN, V. V., ASSA-MUNT, N., EBRIGHT, R. H. & GOURSE, R. L. 1996. DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. *Genes Dev,* 10**,** 16-26.

[77] GAASTRA, W. & SVENNERHOLM, A. M. 1996. Colonization factors of human enterotoxigenic Escherichia coli (ETEC). *Trends Microbiol,* 4**,** 444-52.

[78] GALLANT, J., BONTHUIS, P., LINDSLEY, D., CABELLON, J., GILL, G., HEATON, K., KELLEY-CLARKE, B., MACDONALD, L., MERCER, S., VU, H. & WORSLEY, A. 2004. On the role of the starved codon and the takeoff site in ribosome bypassing in Escherichia coli. *J Mol Biol,* 342**,** 713-24.

[79] GEORGOPAPADAKOU, N. H. & BERTASSO, A. 1991. Effects of quinolones on nucleoid segregation in Escherichia coli. *Antimicrob Agents Chemother,* 35**,** 2645-8.

[80] GERARD-VINCENT, M., ROBERT, V., BALL, G., BLEVES, S., MICHEL, G. P., LAZDUNSKI, A. & FILLOUX, A. 2002. Identification of XcpP domains that confer functionality and specificity to the Pseudomonas aeruginosa type II secretion apparatus. *Mol Microbiol,* 44**,** 1651-65.

[81] GORDON, B. R., LI, Y., WANG, L., SINTSOVA, A., VAN BAKEL, H., TIAN, S., NAVARRE, W. W., XIA, B. & LIU, J. 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A,* 107**,** 5154-9.

[82] GRAINGER, D. C. 2016. Structure and function of bacterial H-NS protein. *Biochem Soc Trans,* 44**,** 1561-1569.

[83] 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.

[84] GRAINGER, D. C., BELYAEVA, T. A., LEE, D. J., HYDE, E. I. & BUSBY, S. J. 2003. Binding of the Escherichia coli MelR protein to the melAB promoter: orientation of MelR subunits and investigation of MelR-DNA contacts. *Mol Microbiol,* 48**,** 335-48.

[85] GRAINGER, D. C., BELYAEVA, T. A., LEE, D. J., HYDE, E. I. & BUSBY, S. J. 2004a. Transcription activation at the Escherichia coli melAB promoter:

interactions of MelR with the C-terminal domain of the RNA polymerase alpha subunit. *Mol Microbiol,* 51**,** 1311-20.

[86] GRAINGER, D. C., HURD, D., HARRISON, M., HOLDSTOCK, J. & BUSBY, S. J. 2005. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci U S A,* 102**,** 17693-8.

[87] GRAINGER, D. C., WEBSTER, C. L., BELYAEVA, T. A., HYDE, E. I. & BUSBY, S. J. 2004b. Transcription activation at the Escherichia coli melAB promoter: interactions of MelR with its DNA target site and with domain 4 of the RNA polymerase sigma subunit. *Mol Microbiol,* 51**,** 1297-309.

[88] GRAY, M. D., BAGDASARIAN, M., HOL, W. G. & SANDKVIST, M. 2011. In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of Vibrio cholerae. *Mol Microbiol,* 79**,** 786-98.

[89] GREGORY, B. D., NICKELS, B. E., GARRITY, S. J., SEVERINOVA, E., MINAKHIN, L., URBAUER, R. J., URBAUER, J. L., HEYDUK, T., SEVERINOV, K. & HOCHSCHILD, A. 2004. A regulator that inhibits transcription by targeting an intersubunit interaction of the RNA polymerase holoenzyme. *Proc Natl Acad Sci U S A,* 101**,** 4554-9.

[90] 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.

[91] GRUBER, T. M. & GROSS, C. A. 2003. Assay of Escherichia coli RNA polymerase: sigma-core interactions. *Methods Enzymol,* 370**,** 206-12.

[92] GRUER, M. J. & GUEST, J. R. 1994. Two genetically-distinct and differentially-regulated aconitases (AcnA and AcnB) in Escherichia coli. *Microbiology,* 140 (Pt 10)**,** 2531-41.

[93] GUPTA, S. K., KECK, J., RAM, P. K., CRUMP, J. A., MILLER, M. A. & MINTZ, E. D. 2008. Part III. Analysis of data gaps pertaining to enterotoxigenic Escherichia coli infections in low and medium human development index countries, 1984-2005. *Epidemiol Infect,* 136**,** 721-38.

[94] HALDER, S., YADAV, K. K., SARKAR, R., MUKHERJEE, S., SAHA, P., HALDAR, S., KARMAKAR, S. & SEN, T. 2015. Alteration of Zeta potential and membrane permeability in bacteria: a study with cationic agents. *Springerplus,* 4**,** 672.

[95] HARRIS, R. S., ROSS, K. J., LOMBARDO, M. J. & ROSENBERG, S. M. 1998. Mismatch repair in Escherichia coli cells lacking single-strand exonucleases ExoI, ExoVII, and RecJ. *J Bacteriol,* 180**,** 989-93.

[96] HAWKEY, P. M. 2003. Mechanisms of quinolone action and microbial response. *J Antimicrob Chemother,* 51 Suppl 1**,** 29-35.

[97] HAWLEY, D. K. & MCCLURE, W. R. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res,* 11**,** 2237-55.

[98] HAYCOCKS, J. R., SHARMA, P., STRINGER, A. M., WADE, J. T. & GRAINGER, D. C. 2015. The molecular basis for control of ETEC enterotoxin expression in response to environment and host. *PLoS Pathog,* 11**,** e1004605.

[99] HEATH, J. D., PERKINS, J. D., SHARMA, B. & WEINSTOCK, G. M. 1992. NotI genomic cleavage map of Escherichia coli K-12 strain MG1655. *J Bacteriol,* 174**,** 558-67.

[100] HECHT, S. M. 2000. Bleomycin: new perspectives on the mechanism of action. *J Nat Prod,* 63**,** 158-68.

[101] HERNANDEZ, V. J., HSU, L. M. & CASHEL, M. 1996. Conserved region 3 of Escherichia coli final sigma70 is implicated in the process of abortive transcription. *J Biol Chem,* 271**,** 18775-9.

[102] HINTON, D. M. 2010. Transcriptional control in the prereplicative phase of T4 development. *Virol J,* 7**,** 289.

[103] HOPKINS, K. L., DAVIES, R. H. & THRELFALL, E. J. 2005. Mechanisms of quinolone resistance in Escherichia coli and Salmonella: recent developments. *Int J Antimicrob Agents,* 25**,** 358-73.

[104] HUBIN, E. A., TABIB-SALAZAR, A., HUMPHREY, L. J., FLACK, J. E., OLINARES, P. D., DARST, S. A., CAMPBELL, E. A. & PAGET, M. S. 2015. Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proc Natl Acad Sci U S A,* 112**,** 7171-6.

[105] ISHIHAMA, A. 1992. Role of the RNA polymerase alpha subunit in transcription activation. *Mol Microbiol,* 6**,** 3283-8.

[106] ISHIHAMA, A. 2012. Prokaryotic genome regulation: a revolutionary paradigm. *Proc Jpn Acad Ser B Phys Biol Sci,* 88**,** 485-508.

[107] ISIDEAN, S. D., RIDDLE, M. S., SAVARINO, S. J. & PORTER, C. K. 2011. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine,* 29**,** 6167-6178.

[108] JENKINS, D. E., AUGER, E. A. & MATIN, A. 1991. Role of RpoH, a heat shock regulator protein, in Escherichia coli carbon starvation protein synthesis and survival. *J Bacteriol,* 173**,** 1992-6.

[109] KANSAL, R., RASKO, D. A., SAHL, J. W., MUNSON, G. P., ROY, K., LUO, Q., SHEIKH, A., KUHNE, K. J. & FLECKENSTEIN, J. M. 2013. Transcriptional modulation of enterotoxigenic Escherichia coli virulence genes in response to epithelial cell interactions. *Infect Immun,* 81**,** 259-70.

[110] KAPER, J. B., NATARO, J. P. & MOBLEY, H. L. 2004. Pathogenic Escherichia coli. *Nat Rev Microbiol,* 2**,** 123-40.

[111] KARP, P. D., WEAVER, D., PALEY, S., FULCHER, C., KUBO, A., KOTHARI, A., KRUMMENACKER, M., SUBHRAVETI, P., WEERASINGHE, D., GAMA-CASTRO, S., HUERTA, A. M., MUNIZ-RASCADO, L., BONAVIDES-MARTINEZ, C., WEISS, V., PERALTA-GIL, M., SANTOS-ZAVALETA, A., SCHRODER, I., MACKIE, A., GUNSALUS, R., COLLADO-VIDES, J., KESELER, I. M. & PAULSEN, I. 2014. The EcoCyc Database. *EcoSal Plus,* 6.

[112] KESELER, I. M., MACKIE, A., SANTOS-ZAVALETA, A., BILLINGTON, R., BONAVIDES-MARTINEZ, C., CASPI, R., FULCHER, C., GAMA-CASTRO, S., KOTHARI, A., KRUMMENACKER, M., LATENDRESSE, M., MUNIZ-RASCADO,

L., ONG, Q., PALEY, S., PERALTA-GIL, M., SUBHRAVETI, P., VELAZQUEZ-RAMIREZ, D. A., WEAVER, D., COLLADO-VIDES, J., PAULSEN, I. & KARP, P. D. 2017. The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res,* 45**,** D543-D550.

[113] KOLB, A., BUSBY, S., BUC, H., GARGES, S. & ADHYA, S. 1993. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem,* 62**,** 749-95.

[114] 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 Res,* 23**,** 819-26.

[115] KOROTKOV, K. V., SANDKVIST, M. & HOL, W. G. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat Rev Microbiol,* 10**,** 336-51.

[116] KRUMMEL, B. & CHAMBERLIN, M. J. 1989. RNA chain initiation by Escherichia coli RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry,* 28**,** 7829-42.

[117] KWON, H. J., BENNIK, M. H., DEMPLE, B. & ELLENBERGER, T. 2000. Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. *Nat Struct Biol,* 7**,** 424-30.

[118] LAMBERT, L. J., WEI, Y., SCHIRF, V., DEMELER, B. & WERNER, M. H. 2004. T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. *EMBO J,* 23**,** 2952-62.

[119] 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. *Nat Microbiol,* 2**,** 16249.

[120] LANDIS, L., XU, J. & JOHNSON, R. C. 1999. The cAMP receptor protein CRP can function as an osmoregulator of transcription in Escherichia coli. *Genes Dev,* 13**,** 3081-91.

[121] LAWSON, C. L., SWIGON, D., MURAKAMI, K. S., DARST, S. A., BERMAN, H. M. & EBRIGHT, R. H. 2004. Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol,* 14**,** 10-20.

[122] LEE, S. Y., DE LA TORRE, A., YAN, D., KUSTU, S., NIXON, B. T. & WEMMER, D. E. 2003. Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev,* 17**,** 2552-63.

[123] LEIVE, L., TELESETSKY, S., COLEMAN, W. G., JR. & CARR, D. 1984. Tetracyclines of various hydrophobicities as a probe for permeability of Escherichia coli outer membranes. *Antimicrob Agents Chemother,* 25**,** 539-44.

[124] LEVIN, J. R., KRUMMEL, B. & CHAMBERLIN, M. J. 1987. Isolation and properties of transcribing ternary complexes of Escherichia coli RNA polymerase positioned at a single template base. *J Mol Biol,* 196**,** 85-100.

[125] LEWIS, D. E. & ADHYA, S. 2002. In vitro repression of the gal promoters by GalR and HU depends on the proper helical phasing of the two operators. *J Biol Chem,* 277**,** 2498-504.

[126] 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.

[127] LIM, C. J., WHANG, Y. R., KENNEY, L. J. & YAN, J. 2012. Gene silencing H-NS paralogue StpA forms a rigid protein filament along DNA that blocks DNA accessibility. *Nucleic Acids Res,* 40**,** 3316-28.

[128] LIU, F., YANG, X., WANG, Z., NICKLASSON, M., QADRI, F., YI, Y., ZHU, Y., LV, N., LI, J., ZHANG, R., GUO, H., ZHU, B., SJOLING, A. & HU, Y. 2015. Draft genomes of four enterotoxigenic Escherichia coli (ETEC) clinical isolates from China and Bangladesh. *Gut Pathog,* 7**,** 10.

[129] LIU, M., GARGES, S. & ADHYA, S. 2004. lacP1 promoter with an extended -10 motif. Pleiotropic effects of cyclic AMP protein at different steps of transcription initiation. *J Biol Chem,* 279**,** 54552-7.

[130] LIU, X. & MATSUMURA, P. 1996. Differential regulation of multiple overlapping promoters in flagellar class II operons in Escherichia coli. *Mol Microbiol,* 21**,** 613-20.

[131] LLAMAS, M. A., IMPERI, F., VISCA, P. & LAMONT, I. L. 2014. Cell-surface signaling in Pseudomonas: stress responses, iron transport, and pathogenicity. *FEMS Microbiol Rev,* 38**,** 569-97.

[132] 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.

[133] MADHAVAN, T. P. & SAKELLARIS, H. 2015. Colonization factors of enterotoxigenic Escherichia coli. *Adv Appl Microbiol,* 90**,** 155-97.

[134] MAHON, V., SMYTH, C. J. & SMITH, S. G. 2010. Mutagenesis of the Rns regulator of enterotoxigenic Escherichia coli reveals roles for a linker sequence and two helix-turn-helix motifs. *Microbiology,* 156**,** 2796-806.

[135] MALHOTRA, A., SEVERINOVA, E. & DARST, S. A. 1996. Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell,* 87**,** 127-36.

[136] MALINVERNI, J. C. & SILHAVY, T. J. 2009. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proc Natl Acad Sci U S A,* 106**,** 8009-14.

[137] MARTIN, R. G., BARTLETT, E. S., ROSNER, J. L. & WALL, M. E. 2008. Activation of the Escherichia coli marA/soxS/rob regulon in response to transcriptional activator concentration. *J Mol Biol,* 380**,** 278-84.

[138] 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.

[139] MARTIN, R. G., GILLETTE, W. K., RHEE, S. & ROSNER, J. L. 1999. Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in Escherichia coli: sequence, orientation and spatial relationship to the core promoter. *Mol Microbiol,* 34**,** 431-41.

[140] 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. *Mol Microbiol,* 35**,** 623-34.

[141] 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.

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

[143] 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.

[144] MATHEW, R. & CHATTERJI, D. 2006. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol,* 14**,** 450-5.

[145] MILLER, J. H. 1972. *Experiments in molecular genetics,* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

[146] MINAKHIN, L., NIEDZIELA-MAJKA, A., KUZNEDELOV, K., ADELMAN, K., URBAUER, J. L., HEYDUK, T. & SEVERINOV, K. 2003. Interaction of T4 AsiA with its target sites in the RNA polymerase sigma70 subunit leads to distinct and opposite effects on transcription. *J Mol Biol,* 326**,** 679-90.

[147] MIROPOLSKAYA, N., IGNATOV, A., BASS, I., ZHILINA, E., PUPOV, D. & KULBACHINSKIY, A. 2012. Distinct functions of regions 1.1 and 1.2 of RNA polymerase sigma subunits from Escherichia coli and Thermus aquaticus in transcription initiation. *J Biol Chem,* 287**,** 23779-89.

[148] MORIN, N., TIRLING, C., IVISON, S. M., KAUR, A. P., NATARO, J. P. & STEINER, T. S. 2010. Autoactivation of the AggR regulator of enteroaggregative Escherichia coli in vitro and in vivo. *FEMS Immunol Med Microbiol,* 58**,** 344-55.

[149] MUDRAK, B. & KUEHN, M. J. 2010. Heat-labile enterotoxin: beyond G(m1) binding. *Toxins (Basel),* 2**,** 1445-70.

[150] MUNSON, G. P., HOLCOMB, L. G. & SCOTT, J. R. 2001. Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. *Infect Immun,* 69**,** 186-93.

[151] MURAKAMI, K. S. 2015. Structural biology of bacterial RNA polymerase. *Biomolecules,* 5**,** 848-64.

[152] MURAKAMI, K. S. & DARST, S. A. 2003. Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol,* 13**,** 31-9.

[153] 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.

[154] MYERS, K. S., YAN, H., ONG, I. M., CHUNG, D., LIANG, K., TRAN, F., KELES, S., LANDICK, R. & KILEY, P. J. 2013. Genome-scale analysis of escherichia coli FNR reveals complex features of transcription factor binding. *PLoS Genet,* 9**,** e1003565.

[155] NAGARAJAVEL, V., MADHUSUDAN, S., DOLE, S., RAHMOUNI, A. R. & SCHNETZ, K. 2007. Repression by binding of H-NS within the transcription unit. *J Biol Chem,* 282**,** 23622-30.

[156] NAIR, G. B. & TAKEDA, Y. 1998. The heat-stable enterotoxins. *Microbial Pathogenesis,* 24**,** 123-131.

[157] NALIN, D. R., CASH, R. A., ISLAM, R., MOLLA, M. & PHILLIPS, R. A. 1968. Oral maintenance therapy for cholera in adults. *Lancet,* 2**,** 370-3.

[158] NEWMARK, K. G., O'REILLY, E. K., POHLHAUS, J. R. & KREUZER, K. N. 2005. Genetic analysis of the requirements for SOS induction by nalidixic acid in Escherichia coli. *Gene,* 356**,** 69-76.

[159] NICHOLS, R. J., SEN, S., CHOO, Y. J., BELTRAO, P., ZIETEK, M., CHABA, R., LEE, S., KAZMIERCZAK, K. M., LEE, K. J., WONG, A., SHALES, M., LOVETT, S., WINKLER, M. E., KROGAN, N. J., TYPAS, A. & GROSS, C. A. 2011. Phenotypic landscape of a bacterial cell. *Cell,* 144**,** 143-56.

[160] NIKAIDO, H. 2005. Restoring permeability barrier function to outer membrane. *Chem Biol,* 12**,** 507-9.

[161] ODUMOSU, O., NICHOLAS, D., YANO, H. & LANGRIDGE, W. 2010. AB toxins: a paradigm switch from deadly to desirable. *Toxins (Basel),* 2**,** 1612-45.

[162] OKUSU, H., MA, D. & NIKAIDO, H. 1996. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol,* 178**,** 306-8.

[163] PARKINSON, G., WILSON, C., GUNASEKERA, A., EBRIGHT, Y. W., EBRIGHT, R. H. & BERMAN, H. M. 1996. Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J Mol Biol,* 260**,** 395-408.

[164] PARRA, M. C. & COLLINS, C. M. 2012. Mutational analysis of the N-terminal domain of UreR, the positive transcriptional regulator of urease gene expression. *Microbiol Res,* 167**,** 433-44.

[165] PEREZ-RUEDA, E. & COLLADO-VIDES, J. 2000. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res,* 28**,** 1838-47.

[166] PERRENOUD, A. & SAUER, U. 2005. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli. *J Bacteriol,* 187**,** 3171-9.

[167] POLESZAK, K., KAMINSKA, K. H., DUNIN-HORKAWICZ, S., LUPAS, A., SKOWRONEK, K. J. & BUJNICKI, J. M. 2012. Delineation of structural domains and identification of functionally important residues in DNA repair enzyme exonuclease VII. *Nucleic Acids Res,* 40**,** 8163-74.

[168] POMPOSIELLO, P. J., BENNIK, M. H. & DEMPLE, B. 2001. Genome-wide transcriptional profiling of the Escherichia coli responses to superoxide stress and sodium salicylate. *J Bacteriol,* 183**,** 3890-902.

[169] POMPOSIELLO, P. J. & DEMPLE, B. 2001. Redox-operated genetic switches: the SoxR and OxyR transcription factors. *Trends Biotechnol,* 19**,** 109-14.

[170] PON, C. L., STELLA, S. & GUALERZI, C. O. 2005. *Repression of transcription by curved DNA and nucleoid protein h-ns: A mode of bacterial gene regulation*.

[171] POSSOT, O. M., VIGNON, G., BOMCHIL, N., EBEL, F. & PUGSLEY, A. P. 2000. Multiple interactions between pullulanase secreton components involved in

stabilization and cytoplasmic membrane association of PulE. *J Bacteriol,* 182**,** 2142-52.

[172] PRESSLER, U., STAUDENMAIER, H., ZIMMERMANN, L. & BRAUN, V. 1988. Genetics of the iron dicitrate transport system of Escherichia coli. *J Bacteriol,* 170**,** 2716-24.

[173] QADRI, F., SVENNERHOLM, A. M., FARUQUE, A. S. & SACK, R. B. 2005. Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev,* 18**,** 465-83.

[174] RAINA, S., MISSIAKAS, D. & GEORGOPOULOS, C. 1995. The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of Escherichia coli. *EMBO J,* 14**,** 1043-55.

[175] RAJAGOPALAN, M., LU, C., WOODGATE, R., O'DONNELL, M., GOODMAN, M. F. & ECHOLS, H. 1992. Activity of the purified mutagenesis proteins UmuC, UmuD', and RecA in replicative bypass of an abasic DNA lesion by DNA polymerase III. *Proc Natl Acad Sci U S A,* 89**,** 10777-81.

[176] RAO, M. C. 1985. Toxins which activate guanylate cyclase: heat-stable enterotoxins. *Ciba Found Symp,* 112**,** 74-93.

[177] REPAR, J., BRISKI, N., BULJUBASIC, M., ZAHRADKA, K. & ZAHRADKA, D. 2013. Exonuclease VII is involved in "reckless" DNA degradation in UV-irradiated Escherichia coli. *Mutat Res,* 750**,** 96-104.

[178] REVYAKIN, A., LIU, C., EBRIGHT, R. H. & STRICK, T. R. 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science,* 314**,** 1139-43.

[179] RHEE, S., MARTIN, R. G., ROSNER, J. L. & DAVIES, D. R. 1998. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc Natl Acad Sci U S A,* 95**,** 10413-8.

[180] RICCI, V., BUSBY, S. J. & PIDDOCK, L. J. 2012. Regulation of RamA by RamR in Salmonella enterica serovar Typhimurium: isolation of a RamR superrepressor. *Antimicrob Agents Chemother,* 56**,** 6037-40.

[181] RICCI, V. & PIDDOCK, L. J. 2009. Ciprofloxacin selects for multidrug resistance in Salmonella enterica serovar Typhimurium mediated by at least two different pathways. *J Antimicrob Chemother,* 63**,** 909-16.

[182] RIMSKY, S. 2004. Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Curr Opin Microbiol,* 7**,** 109-14.

[183] ROBERTS, J. W., YARNELL, W., BARTLETT, E., GUO, J., MARR, M., KO, D. C., SUN, H. & ROBERTS, C. W. 1998. Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb Symp Quant Biol,* 63**,** 319-25.

[184] RODGERS, M. E. & SCHLEIF, R. 2009. Solution structure of the DNA binding domain of AraC protein. *Proteins,* 77**,** 202-8.

[185] ROSENBERG, M. 1984. Bacterial Adherence to Hydrocarbons - a Useful Technique for Studying Cell-Surface Hydrophobicity. *Fems Microbiology Letters,* 22**,** 289-295.

[186] ROSENBERG, M., GUTNICK, D. & ROSENBERG, E. 1980. Adherence of Bacteria to Hydrocarbons - a Simple Method for Measuring Cell-Surface Hydrophobicity. *Fems Microbiology Letters,* 9**,** 29-33.

[187] ROSS, W. & GOURSE, R. L. 2005. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate Escherichia coli RNA polymerase-lacUV5 promoter association. *Proc Natl Acad Sci U S A,* 102**,** 291-6.

[188] 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. *Mol Cell,* 50**,** 420-9.

[189] RUIZ, C. & LEVY, S. B. 2010. Many chromosomal genes modulate MarA-mediated multidrug resistance in Escherichia coli. *Antimicrob Agents Chemother,* 54**,** 2125-34.

[190] RUIZ, R., MARQUES, S. & RAMOS, J. L. 2003. Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL meta-cleavage pathway, are involved in dimerization. *J Bacteriol,* 185**,** 3036-41.

[191] SACK, R. B., GORBACH, S. L., BANWELL, J. G., JACOBS, B., CHATTERJEE, B. D. & MITRA, R. C. 1971. Enterotoxigenic Escherichia coli isolated from patients with severe cholera-like disease. *J Infect Dis,* 123**,** 378-85.

[192] SAECKER, R. M., RECORD, M. T., JR. & DEHASETH, P. L. 2011. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol,* 412**,** 754-71.

[193] SAHL, J. W. & RASKO, D. A. 2012. Analysis of global transcriptional profiles of enterotoxigenic Escherichia coli isolate E24377A. *Infect Immun,* 80**,** 1232-42.

[194] SANCHEZ, J. & HOLMGREN, J. 2005. Virulence factors, pathogenesis and vaccine protection in cholera and ETEC diarrhea. *Curr Opin Immunol,* 17**,** 388-98.

[195] SCHULTZ, S. C., SHIELDS, G. C. & STEITZ, T. A. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science,* 253**,** 1001-7.

[196] SEARS, C. L. & KAPER, J. B. 1996. Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion. *Microbiol Rev,* 60**,** 167-215.

[197] SEEDORFF, J. & SCHLEIF, R. 2011. Active role of the interdomain linker of AraC. *J Bacteriol,* 193**,** 5737-46.

[198] SEMSEY, S., GEANACOPOULOS, M., LEWIS, D. E. & ADHYA, S. 2002. Operator-bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding. *EMBO J,* 21**,** 4349-56.

[199] 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.

[200] SHAH, I. M. & WOLF, R. E., JR. 2004. Novel protein--protein interaction between Escherichia coli SoxS and the DNA binding determinant of the RNA polymerase alpha subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *J Mol Biol,* 343**,** 513-32.

[201] SHAH, N., DUPONT, H. L. & RAMSEY, D. J. 2009. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. *Am J Trop Med Hyg,* 80**,** 609-14.

[202] SHARMA, U. K. & CHATTERJI, D. 2010. Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity. *FEMS Microbiol Rev,* 34**,** 646-57.

[203] SHIMADA, T., ISHIHAMA, A., BUSBY, S. J. & GRAINGER, D. C. 2008. The Escherichia coli RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res,* 36**,** 3950-5.

[204] SHIMIZU, K. 2013. Metabolic Regulation of a Bacterial Cell System with Emphasis on Escherichia coli Metabolism. *ISRN Biochem,* 2013**,** 645983.

[205] SHIN, M., KANG, S., HYUN, S. J., FUJITA, N., ISHIHAMA, A., VALENTIN-HANSEN, P. & CHOY, H. E. 2001. Repression of deoP2 in Escherichia coli by CytR: conversion of a transcription activator into a repressor. *EMBO J,* 20**,** 5392-9.

[206] SHULTZABERGER, R. K., CHIANG, D. Y., MOSES, A. M. & EISEN, M. B. 2007. Determining physical constraints in transcriptional initiation complexes using DNA sequence analysis. *PLoS One,* 2**,** e1199.

[207] 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.

[208] SIXMA, T. K., PRONK, S. E., KALK, K. H., WARTNA, E. S., VAN ZANTEN, B. A., WITHOLT, B. & HOL, W. G. 1991. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature,* 351**,** 371-7.

[209] STROZEN, T. G., LI, G. & HOWARD, S. P. 2012. YghG (GspSbeta) is a novel pilot protein required for localization of the GspSbeta type II secretion system secretin of enterotoxigenic Escherichia coli. *Infect Immun,* 80**,** 2608-22.

[210] SWINT-KRUSE, L. & MATTHEWS, K. S. 2009. Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol,* 12**,** 129-37.

[211] TAGAMI, H., INADA, T., KUNIMURA, T. & AIBA, H. 1995. Glucose lowers CRP* levels resulting in repression of the lac operon in cells lacking cAMP. *Mol Microbiol,* 17**,** 251-8.

[212] TANG, H., SEVERINOV, K., GOLDFARB, A., FENYO, D., CHAIT, B. & EBRIGHT, R. H. 1994. Location, structure, and function of the target of a transcriptional activator protein. *Genes Dev,* 8**,** 3058-67.

[213] TAUSCHEK, M., GORRELL, R. J., STRUGNELL, R. A. & ROBINS-BROWNE, R. M. 2002. Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of Escherichia coli. *Proc Natl Acad Sci U S A,* 99**,** 7066-71.

[214] TENAILLON, O., SKURNIK, D., PICARD, B. & DENAMUR, E. 2010. The population genetics of commensal Escherichia coli. *Nat Rev Microbiol,* 8**,** 207-17.

[215] THONG, S., ERCAN, B., TORTA, F., FONG, Z. Y., WONG, H. Y., WENK, M. R. & CHNG, S. S. 2016. Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. *Elife,* 5.

[216] TRAMONTI, A., VISCA, P., DE CANIO, M., FALCONI, M. & DE BIASE, D. 2002. Functional characterization and regulation of gadX, a gene encoding an AraC/XylS-like transcriptional activator of the Escherichia coli glutamic acid decarboxylase system. *J Bacteriol,* 184**,** 2603-13.

[217] TYPAS, A. & HENGGE, R. 2006. Role of the spacer between the -35 and -10 regions in sigmas promoter selectivity in Escherichia coli. *Mol Microbiol,* 59**,** 1037-51.

[218] UEDA, T., TAKAHASHI, H., UYAR, E., ISHIKAWA, S., OGASAWARA, N. & OSHIMA, T. 2013. Functions of the Hha and YdgT proteins in transcriptional silencing by the nucleoid proteins, H-NS and StpA, in Escherichia coli. *DNA Res,* 20**,** 263-71.

[219] URBAUER, J. L., SIMEONOV, M. F., URBAUER, R. J., ADELMAN, K., GILMORE, J. M. & BRODY, E. N. 2002. Solution structure and stability of the anti-sigma factor AsiA: implications for novel functions. *Proc Natl Acad Sci U S A,* 99**,** 1831-5.

[220] VALENTIN-HANSEN, P., SOGAARD-ANDERSEN, L. & PEDERSEN, H. 1996. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol,* 20**,** 461-6.

[221] VASSYLYEV, D. G., SEKINE, S., LAPTENKO, O., LEE, J., VASSYLYEVA, M. N., BORUKHOV, S. & YOKOYAMA, S. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature,* 417**,** 712-9.

[222] VELEBA, M., HIGGINS, P. G., GONZALEZ, G., SEIFERT, H. & SCHNEIDERS, T. 2012. Characterization of RarA, a novel AraC family multidrug resistance regulator in Klebsiella pneumoniae. *Antimicrob Agents Chemother,* 56**,** 4450-8.

[223] VOGEL, U. & JENSEN, K. F. 1994. The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J Bacteriol,* 176**,** 2807-13.

[224] VRENTAS, C. E., GAAL, T., ROSS, W., EBRIGHT, R. H. & GOURSE, R. L. 2005. Response of RNA polymerase to ppGpp: requirement for the omega subunit and relief of this requirement by DksA. *Genes Dev,* 19**,** 2378-87.

[225] WADE, J. T., BELYAEVA, T. A., HYDE, E. I. & BUSBY, S. J. 2001. A simple mechanism for co-dependence on two activators at an Escherichia coli promoter. *EMBO J,* 20**,** 7160-7.

[226] WADE, J. T., REPPAS, N. B., CHURCH, G. M. & STRUHL, K. 2005. Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. *Genes Dev,* 19**,** 2619-30.

[227] WANG, W., LI, G. W., CHEN, C., XIE, X. S. & ZHUANG, X. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science,* 333**,** 1445-9.

[228] WASHBURN, R. S. & GOTTESMAN, M. E. 2015. Regulation of transcription elongation and termination. *Biomolecules,* 5**,** 1063-78.

[229] WEBER, H., POLEN, T., HEUVELING, J., WENDISCH, V. F. & HENGGE, R. 2005. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol,* 187**,** 1591-603.

[230] WEDEL, A., WEISS, D. S., POPHAM, D., DROGE, P. & KUSTU, S. 1990. A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science,* 248**,** 486-90.

[231] WEIGLMEIER, P. R., ROSCH, P. & BERKNER, H. 2010. Cure and Curse: E-coli Heat-Stable Enterotoxin and Its Receptor Guanylyl Cyclase C. *Toxins,* 2**,** 2213-2229.

[232] WILKINSON, M., CHABAN, Y. & WIGLEY, D. B. 2016a. Mechanism for nuclease regulation in RecBCD. *Elife,* 5.

[233] WILKINSON, M., TROMAN, L., WAN NUR ISMAH, W. A., CHABAN, Y., AVISON, M. B., DILLINGHAM, M. S. & WIGLEY, D. B. 2016b. Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. *Elife,* 5.

[234] WILSON, C. J., ZHAN, H., SWINT-KRUSE, L. & MATTHEWS, K. S. 2007. Ligand interactions with lactose repressor protein and the repressor-operator complex: the effects of ionization and oligomerization on binding. *Biophys Chem,* 126**,** 94-105.

[235] WINARDHI, R. S., GULVADY, R., MELLIES, J. L. & YAN, J. 2014. Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic Escherichia coli competes off histone-like nucleoid-structuring protein (H-NS) through noncooperative DNA binding. *J Biol Chem,* 289**,** 13739-50.

[236] YANG, J., BALDI, D. L., TAUSCHEK, M., STRUGNELL, R. A. & ROBINS-BROWNE, R. M. 2007. Transcriptional regulation of the yghJ-pppA-yghG-

gspCDEFGHIJKLM cluster, encoding the type II secretion pathway in enterotoxigenic Escherichia coli. *J Bacteriol,* 189**,** 142-50.

[237] YANG, J., TAUSCHEK, M., STRUGNELL, R. & ROBINS-BROWNE, R. M. 2005. The H-NS protein represses transcription of the eltAB operon, which encodes heat-labile enterotoxin in enterotoxigenic Escherichia coli, by binding to regions downstream of the promoter. *Microbiology,* 151**,** 1199-208.

[238] ZHANG, A., ROSNER, J. L. & MARTIN, R. G. 2008. Transcriptional activation by MarA, SoxS and Rob of two tolC promoters using one binding site: a complex promoter configuration for tolC in Escherichia coli. *Mol Microbiol,* 69**,** 1450-5.

[239] 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.

[240] ZHANG, W., ZHANG, T., WU, Y. & JIANG, J. 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *Plant Cell,* 24**,** 2719-31.

[241] ZIERVOGEL, B. K. & ROUX, B. 2013. The binding of antibiotics in OmpF porin. *Structure,* 21**,** 76-87.

[242] ZUO, Y., WANG, Y. & STEITZ, T. A. 2013. The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol Cell,* 50**,** 430-6.

Appendix

Abbreviations

A	Adenine
A (Ala)	Alanine
APS	Ammonium persulphate
AR	Activating Region
Amp^R	Ampicillin resistance
Arg, R	Arginine
Å	Ångstrom
ATP	Adenosine triphosphate
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
ºC	Celsius (degree)
CF	Colonisation factor
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP combined with sequencing
CIAP	Calf intestine alkaline phosphatase
CRP	Cyclic-AMP receptor protein
cAMP	3'-5'-cyclic adenosine monophosphate
CTD	Carboxy-terminal domain
CPZ	Chlorpromazine
D (Asp)	Aspartic acid
dH_2O	Distilled water

DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxyribonucleoside 5'-triphosphate (N = A, C, G, T)
dsDNA	Double stranded DNA
E (Glu)	Glutamic acid
E. coli	*Escherichia coli*
EDTA	Diaminoethanetetra-acetic acid
EMSA	Electrophoretic mobility shift assays
ETEC	Enterotoxigenic *E. coli*
FNR	Fumarate and nitrate reductase
G	Guanosine
GTP	Guanosine triphosphate
H. pylori	*Helicobacter pylori*
H (His)	Histidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfphonic acid
H-NS	Histone-like Nucleoid Structuring protein
HTH	Helix-turn-helix
HU	Histone-like protein from *E. coli* strain U93
I (Ile)	Isoleucine
IHF	Integration host factor
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobases
kDa	Kilodalton

L (Leu)	Leucine
LB	Lennox Broth
LT	Heat labile toxin
Mbp	Mega base pairs
M (Met)	Methionine
MarA	Multiple antibiotic resistance activator
MarR	Multiple antibiotic resistance repressor
Mg	Magnesium
Mla	Maintenance of OM lipid asymmetry
mRNA	Messenger ribonucleic acid
NCR	Non conserved region
nt	Nucleotide
NTD	Amino-terminal domain
O	Operator
OD	Optical Density
ONPG	*o*-nitrophenyl-β-D-galactopyranoside
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
ppGpp	Guanosine tetraphosphate
Q (Gln)	Glutamine
R (Arg)	Arginine

RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
SDS	Sodium dodecyl sulphate
ST	Heat stable toxin
StpA	Suppression of td^- phenotype
T	Thymine
T. aquaticus	*Thermus aquaticus*
TEMED	N,N,N',N'-tetramethylethylene diamine
Tet^R	Tetracycline resistance
TF	Transcription factor
T_m	Melting temperature
TNSC	Transcription buffer
Tris	Tris (hydroxymethyl) aminoethane
TSS	Transcription start site
U	Uracil
UP element	Upstream promoter element
V	Volts
V. cholerae	*Vibrio cholerae*
W	Watts
WT	Wild type

OPEN ACCESS Freely available online

PLOS | PATHOGENS

The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host

James R. J. Haycocks[1], Prateek Sharma[1], Anne M. Stringer[2], Joseph T. Wade[2,3], David C. Grainger[1]*

1 Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom, **2** Wadsworth Center, New York State Department of Health, Albany, New York, United States of America, **3** Department of Biomedical Sciences, School of Public Health, University at Albany, SUNY, Albany, New York, United States of America

Abstract

Enterotoxigenic *Escherichia coli* (ETEC) cause severe diarrhoea in humans and neonatal farm animals. Annually, 380,000 human deaths, and multi-million dollar losses in the farming industry, can be attributed to ETEC infections. Illness results from the action of enterotoxins, which disrupt signalling pathways that manage water and electrolyte homeostasis in the mammalian gut. The resulting fluid loss is treated by oral rehydration. Hence, aqueous solutions of glucose and salt are ingested by the patient. Given the central role of enterotoxins in disease, we have characterised the regulatory trigger that controls toxin production. We show that, at the molecular level, the trigger is comprised of two gene regulatory proteins, CRP and H-NS. Strikingly, this renders toxin expression sensitive to both conditions encountered on host cell attachment and the components of oral rehydration therapy. For example, enterotoxin expression is induced by salt in an H-NS dependent manner. Furthermore, depending on the toxin gene, expression is activated or repressed by glucose. The precise sensitivity of the regulatory trigger to glucose differs because of variations in the regulatory setup for each toxin encoding gene.

Citation: Haycocks JRJ, Sharma P, Stringer AM, Wade JT, Grainger DC (2015) The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host. PLoS Pathog 11(1): e1004605. doi:10.1371/journal.ppat.1004605

Editor: Frank R. DeLeo, National Institutes of Health, United States of America

Received October 28, 2014; **Accepted** December 5, 2014; **Published** January 8, 2015

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are either within the paper and its Supporting Information files or held in a public repository. The ChIP-seq data are available from ArrayExpress via accession number E-MTAB-2917 (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2917/)

Funding: This work was funded by Wellcome Trust (www.wellcome.ac.uk) Research Career Development Fellowship WT085092MA and Leverhulme Trust (www.leverhulme.ac.uk) project grant RPG-2013-147 awarded to DCG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: d.grainger@bham.ac.uk

Introduction

ETEC are Gram negative bacteria that cause severe diarrhoea, known as non-*vibrio* cholera, in humans [1,2]. First isolated in 1971, ETEC are responsible for 210 million infections annually, mostly in developing countries, leading to 380,000 deaths [3]. Disease results primarily from the action of two enterotoxins. The heat-labile toxin (LT) is similar in structure and function to cholera toxin [4,5]. The heat-stable toxin (ST) mimics the human hormone guanylin [6]. Both toxins are secreted by ETEC during infection. Made up of two subunits, encoded by the *eltAB* operon, LT has the configuration AB_5 [5,7]. In the gut, LT binds to host cell GM1 gangliosides and is endocytosed [8,9]. This triggers constitutive cAMP production in the affected cell [8]. The ST toxin, encoded by the *estA* gene, also interferes with cell signalling [6]. Hence, ST binds to the guanylate cyclase C receptor and stimulates overproduction of cGMP. The combined actions of LT and ST cause loss of H_2O, and electrolytes, from epithelial cells into the gut lumen [4]. Oral Rehydration Therapy (ORT) is used to redress the resulting electrolyte imbalance and rehydrate the patient [10]. In its most simple form, ORT requires only an aqueous solution of glucose and salt. Hence, the availability of metabolites and cations are a central theme of ETEC mediated disease. The effect of ORT on human physiology is well understood: glucose and Na_2^+ are transported across the epithelial membrane, along with water, to promote rehydration [11]. Surprisingly, despite the existence of molecular mechanisms that allow bacteria to respond to these signals, the consequences for ETEC are unknown.

In *E. coli*, the transcriptional response to glucose is controlled by cAMP receptor protein (CRP) [12]. In the absence of glucose, intracellular cAMP levels increase and CRP binds DNA targets with the consensus sequence 5′-TGTGA-n_6-TCACA-3′ [13]. Subsequently, gene expression is reprogrammed to make use of alternative carbon sources [14]. Note that the gene regulatory network managed by CRP includes many indirect pathways [14,15]. Hence, CRP is also a pleiotropic regulator of transcription. Whilst indirect regulatory effects are difficult to characterise, genes that are directly controlled by CRP can be divided into distinct classes [12]. At Class II targets, CRP binds to a site overlapping the promoter -35 element and interacts directly with both the N-terminal and C-terminal domains of the RNA polymerase α subunit (αNTD and αCTD). At Class I targets, CRP binds further upstream and interacts only with αCTD. This interaction can be further stabilised by UP-elements, AT-rich DNA sequences, adjacent to the CRP site, that facilitate αCTD-DNA interactions [12]. At both classes of promoter, the various contacts enhance gene expression by stabilising the transcription

Author Summary

Diarrheagenic illness remains a major disease burden in the developing world. Enterotoxigenic *Escherichia coli* (ETEC) are the leading bacterial cause of such disease; hundreds of millions of cases occur every year. The severe watery diarrhoea associated with ETEC infections results from the action of enterotoxins. The toxins target human gut epithelial cells and trigger the loss of water and electrolytes into the gut lumen. Oral rehydration therapy can counteract this process. Hence, glucose and salt solutions promote rehydration of the patient. In this work we show that the gene regulatory mechanisms controlling toxin expression respond directly to sugar and salt. Furthermore, we describe a molecular mechanism to explain these effects. Hence, we provide a starting point for the optimisation of oral rehydration solutions to reduce toxin expression over the course of an ETEC infection.

initiation complex. Unsurprisingly, most genes regulated by CRP encode proteins involved in metabolism. However, in some bacteria, CRP has been co-opted as a virulence regulator [16].

The Histone-like Nucleoid Structuring (H-NS) factor is a component of bacterial nucleoprotein. Consequently, H-NS also influences gene expression on a global scale [17]. Briefly, H-NS targets sections of the genome with a low GC content [17]. Depending on H-NS conformation, the resulting nucleoprotein complexes can be filamentous or bridged in organisation [18]. Filamentous complexes favour gene regulation by excluding RNA polymerase, and transcriptional regulators, from their targets [19,20]. Bridged complexes favour RNA polymerase trapping [21]. In all scenarios, it is thought that H-NS acts primarily to silence transcription [22]. The conformation of H-NS, and hence the way in which it modulates DNA topology, can be controlled by divalent cations. Consequently, H-NS mediated repression can be relieved by increased osmolarity [23]. Like CRP, H-NS has been incorporated into the virulence gene regulatory networks of many bacteria [17].

In this work we define the molecular trigger that controls toxin expression in ETEC. We show that CRP and H-NS are key regulatory factors. Strikingly, this allows ETEC to integrate extracellular signals of osmolarity and metabolism to control toxin production. Hence, we propose that ETEC toxicity responds directly to osmo-metabolic flux. Interestingly, the precise regulatory settings are different for each toxin encoding gene. The differences result from i) varying promoter configurations and ii) competition between CRP and H-NS for overlapping DNA targets. This is significant since fluctuations in osmolarity, and changes in the availability of metabolites, are central to ETEC infection and its treatment.

Results

Binding of CRP and H-NS across the ETEC H10407 genome

The prototypical ETEC strain H10407 reproducibly elicits diarrhoea in human volunteers and has a well-defined genome that shares 3,766 genes with *E. coli* K-12 [1]. Pathogenicity arises from 599 ancillary genes encoded by 25 discrete chromosomal loci and 4 plasmids. The plasmids, named p948, p666, p58 and p52, encode the enterotoxins. Derivatives of the *estA* gene are found on plasmids p666 (e*stA1*) and p948 (e*stA2*). A single copy of the *eltAB* operon is encoded by plasmid p666. We used Chromatin Immunoprecipitation (ChIP) coupled with next-generation DNA sequencing (ChIP-seq) to map CRP and H-NS targets across the ETEC H10407 genome. The binding profiles are shown in Fig. 1A. In each plot genes are illustrated by blue lines (tracks 1 and 2), DNA G/C content by a cyan and pink graph (track 3), H-NS binding is in green (track 4) and CRP binding is shown in orange (track 5). As expected, H-NS binding is inversely correlated with DNA G/C content (compare tracks 3 and 4). Similarly, CRP binding occurs in expected locations; 96% of the CRP binding sites are associated with the DNA logo shown in Fig. 1B (i.e. the known CRP consensus sequence (13–15)). We identified a total of 111 high-confidence CRP targets (Table 1). Of these targets 93% were present in the genome sequences of both ETEC H10407 and *E. coli* K-12. The most common location for CRP sites was in intergenic regions (66% of targets) whilst a smaller number of targets were found within genes (34%). Consistent with expectations, CRP sites were most frequently located ~40.5 bp, or ~92.5 bp, upstream of experimentally determined transcription start sites (TSSs). Surprisingly, CRP binding was restricted to the ETEC chromosome (Fig. 1Ai). Conversely, H-NS bound to chromosomal and plasmid loci (Fig. 1Ai), including all toxin encoding genes (Fig. 1Aii).

Unoccupied high-affinity CRP binding targets on p948 and p666 are bound by H-NS

To better understand the lack of CRP binding to p948 and p666 we took a bioinformatic approach. CRP targets were aligned to generate a position weight matrix (PWM). The PWM was then used to search p948 and p666 for CRP sites. A continuum of over 100 potential CRP targets was identified. However, we recognise that the vast majority of these are likely to be false positives. Hence, we next sought to differentiate between genuine CRP sites and spurious predictions. To do this, predicted sites were scored, grouped, and ranked on the basis of their match to the PWM (Fig. 2A, S1 Table). Electrophoretic mobility shift assays (EMSA) were then used to measure binding of CRP to a target from each group so that a meaningful cut-off could be established. The result is illustrated graphically in Fig. 2B. The raw data are shown in S1A Fig. We found that predicted sites with a score$<$10 did not bind CRP. To assess the affinity of CRP for all predicted targets scoring $>$10 a second set of EMSA experiments was done (S1B Fig.). Hence, we identified a total of 5 potential CRP targets on p666 and p948. Interestingly, the *estA1* and *estA2* genes, which both encode ST, were amongst the 5 targets (Fig. 2C). Remarkably, all 5 of the plasmid borne CRP targets identified *in silico*, and bound tightly by CRP *in vitro*, were occupied by H-NS *in vivo* (Fig. 2C).

The *estA2* gene is transcribed from a Class I CRP dependent promoter

To understand if CRP could regulate ST production we focused first on *estA2*. This derivative of the toxin is more commonly associated with human disease and ETEC H10407 is somewhat unusual in also encoding *estA1* [24]. The sequence of the *estA2* regulatory region is shown in Fig. 3A. A 93 bp DNA fragment, containing the regulatory region, was cloned into the *lacZ* reporter plasmid pRW50 to generate a *lacZ* fusion (S2A Fig.). The *estA2* TSS was then determined using mRNA primer extension analysis. We detected a single extension product, of 109 nucleotides (nt) in length (Fig. 3B). The position of the TSS is labelled "+1" in Fig. 3A. Promoter -10 (5′-TTAAAT-3′) and -35 (5′-TTGCGC-3′) elements were observed at the expected positions upstream of the TSS. Throughout this work we refer to this promoter, highlighted purple in Fig. 3A, as P*estA2*. To confirm CRP binding

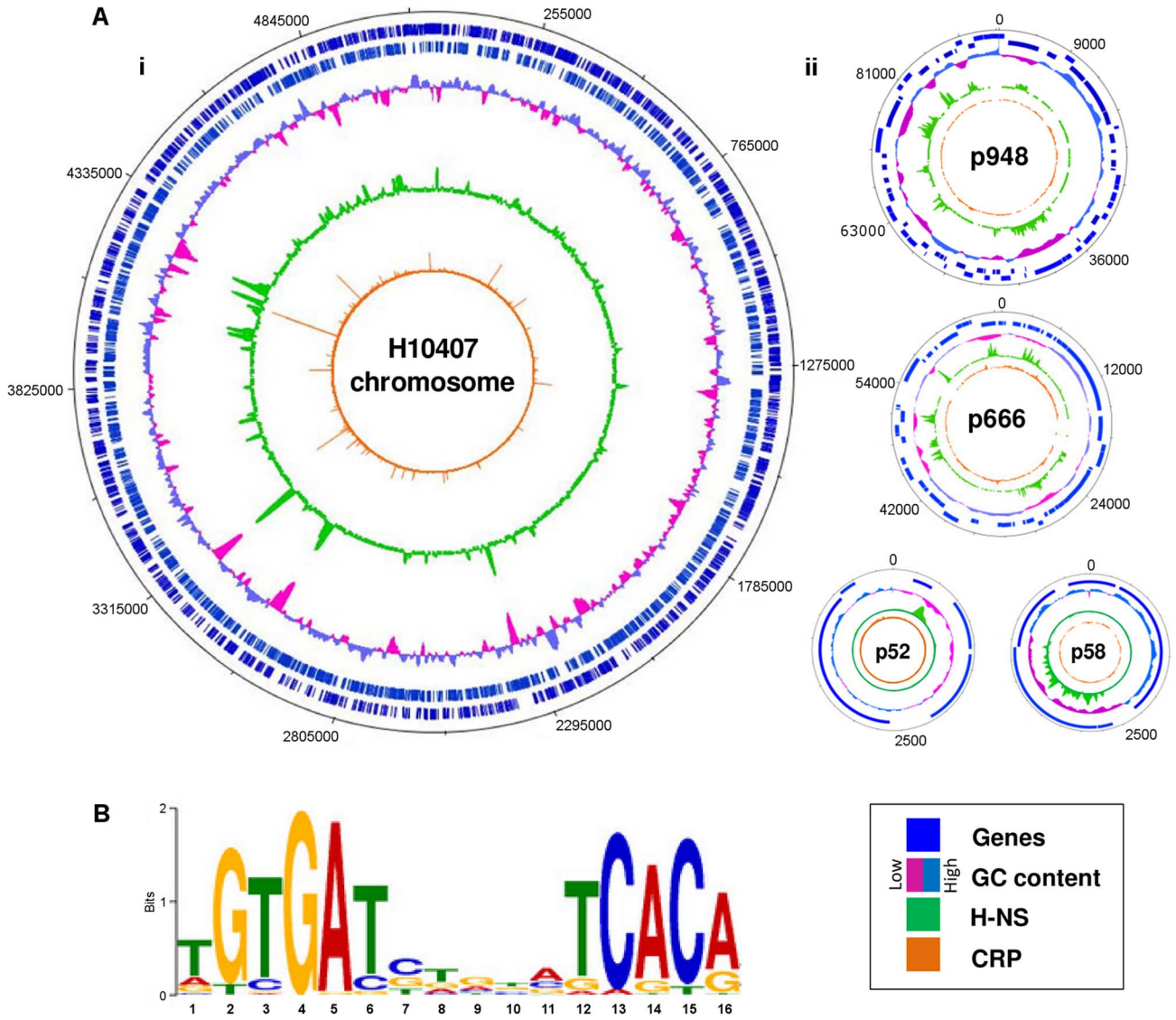


Fig. 1. Distribution of CRP and H-NS across the ETEC H10407 genome. A) The panel shows maps of the ETEC H10407 chromosome (i) and associated plasmids (ii). In each plot, tracks 1 and 2 (blue lines) show the position of genes, track 3 (purple and cyan graph) is a plot of DNA GC content, track 4 (green) is the H-NS binding profile and track 5 (orange) is the CRP binding profile. B) A DNA sequence motif generated by aligning regions of the ETEC H10407 chromosome bound by CRP.
doi:10.1371/journal.ppat.1004605.g001

at the predicted site we used DNase I footprinting (Fig. 3C). As expected, CRP protected the predicted target from digestion. Additionally, CRP induced DNase I hypersensitivity in the centre of the site. Note that the CRP site is centred 59.5 bp upstream of the TSS and adjacent to an AT-rich sequence that may be an UP element (Fig. 3A). Thus, we hypothesised that P*estA2* is a class I CRP activated promoter. To test our hypothesis we first determined whether CRP could indeed activate P*estA2*. To do this, we compared LacZ expression in M182Δ*lac* and M182Δ*lac*Δ*crp* cells carrying the P*estA2::lacZ* fusion. The data show that loss of CRP results in a 3-fold decrease in LacZ expression from P*estA2* (Fig. 3D). We next tested the ability of CRP to activate P*estA2 in vitro*. The 93 bp DNA fragment was cloned upstream of the λ*oop* terminator in plasmid pSR. In the context of this construct a 112 nt transcript is generated by RNA polymerase from P*estA2 in vitro*. The amount of transcript can then be quantified by electrophoresis. The result of the analysis, with and without CRP, is shown in Fig. 3E. As expected, an intense band corresponding to the 112 nt transcript was observed. Production of the transcript was stimulated by CRP. Note that CRP had no effect on production of the 108 nt control RNAI transcript from the plasmid replication origin. Finally, we examined the AT-rich DNA sequence (highlighted blue in Fig. 3A) located between the CRP site and the promoter -35 element. We found that increasing the GC content of the putative UP-element altered migration of the 93 bp DNA fragment on an agarose gel, consistent with a change in DNA topology (S3A Fig.). Moreover, these changes to the UP-element rendered P*estA2* insensitive to CRP *in vivo* and *in vitro* (S3B Fig.).

H-NS excludes CRP from the *estA2* promoter and represses *estA2* transcription

Promoters can be liberated from H-NS repression if separated from flanking, H-NS bound, DNA [25]. We reasoned that this

Table 1. High-confidence CRP binding sites on the ETEC H10407 chromosome identified by ChIP-seq.

Peak Centre[a]	Binding Site(s)[b]	Gene(s)[c]	K-12 Homologues[d]
45284	**TGTGA**TTGGTA**TCACA**	*ETEC_0040*	*caiT*
92014	A**GTGA**TGGATG**TCAC**G	(*ETEC_0078*)	(*cra*)
176905	A**GCG**TTCCACG**TCACA**	*(ETEC_0150)*	*(hemL)*
408885	**TGTGA**TCTCTC**TC**G**CA**	*ETEC_0385/ETEC_0386*	*yahN/yahO*
461874	**TGTG**CGCAAGA**TCACA**	*ETEC_0434*	*ddlA*
463095	TT**TG**CGCGAGG**TCACA**	*(ETEC_0436)*	*(phoA)*
468009	A**GG**G**A**TCTGCG**TCACA**	*ETEC_0443*	*aroM*
492973	ATC**GA**TTGCGT**TCAC**G	*ETEC_0464*	***tsx***
540805	**TGTGA**TCTTTA**TCACA**	*ETEC_0511*	*maa*
574230	GA**TGA**CGACGA**TCACA**	*(ETEC_0538)*	*(ybaT)*
683187	A**GTGA**TCGAGT**T**A**ACA**	*ETEC_0628*	***cstA***
697540	A**GTGA**TTTGCG**TCACA**	*ETEC_0639*	*rnk*
739223	C**GT**T**A**CCCTTG**TC**G**CA**	*ETEC_0680*	*rihA*
941002	**TGTGA**TGAGTA**TCAC**G	*ETEC_0869*	*ybiJ*
958866	**TGTGT**ACGAAA**TCACA**	*ETEC_0886/ETEC_0887*	***ybiS/ybiT***
1128472	n.d.	(*ETEC_1030*)	(*yccS*)
1205350	A**GTGA**TGTAGA**TCACA**	*ETEC_1101*	*ycgZ*
	TGA**GA**TCGAGCA**CACA**		
1263558	**TTTGA**CGGCTA**TCAC**G	*ETEC_1166*	***ptsG***
1274886	**TGTGA**TCTGGA**TCACA**	*ETEC_1176/ETEC_1177*	***ycfQ*/*bhsA***
1301786	GA**TGA**TCCGCA**TCACA**	(*ETEC_1206*)/*ETEC_1207*	ETEC-specific/ETEC-specific
1348166	AT**TGA**ACAGGA**TCACA**	*(ETEC_1259)/ETEC_1260*	*(rluE)/icd*
1376374	G**GTGA**GCTGGC**TCACA**	*ETEC_1292/ETEC_1293*	***ycgB*/*dadA***
1388620	A**GTGA**GCCAGT**TAACA**	*(ETEC_1303)*	*(dhaI)*
1541732	CG**TGA**ACCGGG**TCACA**	*ETEC_1443/ETEC_1444*	*ycjZ/mppA*
1567885	GTTA**A**GTAAAA**TCACA**	*ETEC_1462/ETEC_1463*	***paaZ/paaA***
1701402	**TGTGA**TGGATG**TCACT**	*ETEC_1568*	*ydeN*
1767726	**TGTGA**TTAACAG**CACA**	*ETEC_1628*	***mlc***
1777143	**TGTGA**TCTAGCG**C**CA**A**	*ETEC_1637*	*pntA*
1811426	CG**TGA**TCAAGA**TCACG**	*(ETEC_1668A)*	(ETEC specific)
1859265	AT**TGA**GCGGGA**TCACA**	*(ETEC_1713)*	*(sufS)*
1887513	A**GTGA**TGCGCA**TCAC**G	*ETEC_1737*	*aroH*
	TGC**GA**GGTGTG**TCACA**		
2126754	**TGTG**GCGTGCA**TCACA**	n.a.	n.a.
2201816	G**GTGA**CGCGCG**TCACA**	*ETEC_2057*	*yedP*
2210222	C**GTGA**TCTCGCG**CACA**	*ETEC_2065/ETEC_2066*	*yedR*/ETEC-specific
2458348	**TGTGA**TCTGAA**TC**T**CA**	*ETEC_2278*	***cdd***
	TGC**GA**TGCGTCG**C**G**CA**		
2492757	AT**TGA**TCGCCC**TCACA**	*ETEC_2309*	*yeiQ*
2555083	CG**TGA**CCAAAG**TCTCA**	*(ETEC_2360)*	*(yfaQ)*
2729713	TT**TGA**AGCTTG**TCACA**	*ETEC_2510/ETEC_2511*	*mntH/nupC*
2735124	A**GT**T**A**TTCATG**TCAC**G	*ETEC_2514*	*yfeC*
2795423	**TGTGA**GCCATGA**CACA**	*(ETEC_2572)/ETEC_2573*	***(aegA)/narQ***
2810983	C**GTGA**TCAAGA**TCACA**	*ETEC_2586*	***hyfA***
2887131	**TTTGA**TCTCGC**TCACA**	(*ETEC_2666*)/*ETEC_2665*	(***xseA***)/***guaB***
3012645	**TGTGA**TCCCCACA**ACA**	(*ETEC_2793*)	(***ung***)

Table 1. Cont.

Peak Centre[a]	Binding Site(s)[b]	Gene(s)[c]	K-12 Homologues[d]
3048307	TT**TGA**CGAGCA**TCACC**	*(ETEC_2822)*	*(emrB)*
3132920	G**GTGA**CCGGTT**TCACA**	*ETEC_2905/ETEC_2906*	*ascG/ascF*
3161660	**TGTGA**CCGTGG**TC**G**CA**	*(ETEC_2933)*	***(nlpD)***
3184337	CG**TGA**TGCGTG**T**A**ACA**	*(ETEC_2956)/ETEC_2955*	*(cysI)/cysH*
3196088	**TGTGA**TTACGA**TCACA**	*ETEC_2966/ETEC_2967*	*ygcW/yqcE*
3223792	A**GTGA**TCTTGA**TC**T**CA**	*ETEC_2986*	*sdaC*
	A**GT**T**A**TGTATC**T**AT**CA**		
3234980	**TG**C**GA**TCGTTA**TCACA**	*(ETEC_2994)/ETEC_2995*	*(fucU)/fucR*
3265047	**TGTGA**CCTGGG**TCAC**G	*ETEC_3017*	*rppH*
3324543	**TGTG**GGCTACG**T**A**ACA**	*(ETEC_3075)*	*(ydhD)*
3361162	n.d.	*ETEC_3105*	***serA***
3368992	**TTTGA**TGCACCG**CACA**	*(ETEC_3113)*	*(ygfI)*
3382158	**TGTGA**TCTACAA**CAC**G	*ETEC_3126*	*cmtB*
3390811	**TGTGA**TTTGCT**TCACA**	*ETEC_3133*	***galP***
3408173	**TGTGA**TGTGGA**T**A**ACA**	*ETEC_3154*	***nupG***
3442697	**TGTGA**TGATTG**TC**G**CA**	*ETEC_3186*	ETEC-specific
3558573	A**GTGA**TTTGGC**TCACA**	*ETEC_3291*	*ygiS*
3580767	A**GTGA**CTTGCA**TCACA**	*(ETEC_3318)*	*(yqiH)*
3635301	AT**TGA**TCTAAC**TCAC**G	*ETEC_3362*	***uxaC***
3642302	CT**TGA**AGTGGG**TCACA**	*(ETEC_3372)*	*(yqjG)*
3665634	**TGTGA**TCAATG**TCA**AT	*ETEC_3393/ETEC_3394*	*garP/garD*
	TGTGCTTTAGCG**C**G**CA**		
3721308	G**GTGA**TTGATG**TCAC**C	*(ETEC_3446)*	*(greA)*
3785700	CG**TG**GGTCGCA**TCACA**	*(ETEC_3510)*	*(mreC)*
3878729	G**GTGA**TTTTGA**TCAC**G	*ETEC_3614/ETEC_3615*	*ppiA/tsgA*
3908574	G**GTGA**TCGCGC**TCACA**	*(ETEC_3645)*	***(hofM)***
3918861	**TGTGA**GTGGAA**TC**G**CA**	*ETEC_3652/ETEC_3653*	*yhgE/pck*
3986400	C**GTGA**TTTTATC**CACA**	*ETEC_3707*	***rpoH***
4105040	AG**TA**A**G**GCAAG**TCC**C**T**	n.a.	*n.a.*
4111116	**TGTGA**CGGGGC**TA**A**CA**	*(ETEC_3806)*	*(wecH)*
4153055	**TGTGA**TCTGAA**TCACA**	*ETEC_3840*	***yibI***
	TGTGATCTACAG**CA**TG		
4153191	**TGTGA**TTGATA**TCACA**	*ETEC_3841*	*mtlA*
	TGTGATGAACG**TCAC**G		
4158433	n.d.	*ETEC_3846*	*lldP*
4196869	**TG**CA**A**TCGATA**TCACA**	*ETEC_3886*	*dinD*
4251326	CT**T**ACTCCTGC**TCACA**	*ETEC_3938*	ETEC specific
4266125	G**GTGA**TGGCATC**C**G**C**G	*(ETEC_3956)*	*(nepI)*
4290730	GG**TGA**GCAAAAC**CAC**G	*(ETEC_3979)*	*(yidR)*
4322430	AT**TGA**CCTGAG**TCACA**	*(ETEC_4010)*	*(yieL)*
4340544	CT**TGA**CCACGG**TCA**G**A**	*(ETEC_4025)/ETEC_4024*	*(atpA)/atpG*
4344649	**TGTGA**TCTGAAG**CAC**G	*ETEC_4030*	*atpI*
4373517	**TGT**A**A**TGCTGG**T**A**ACA**	*(ETEC_4051)*	*(ilvG)*
4402013	C**GTG**CTGCATA**TCAC**G	*(ETEC_4077)*	*(rffM)*
4412999	C**GTGA**TCAATT**TA**A**CA**	*ETEC_4085/ETEC_4085*	*hemC/cyaA*
4438352	G**GTGA**TGAGTA**TCAC**G	*ETEC_4107/ETEC_4108*	*ysgA/udp*

Table 1. Cont.

Peak Centre[a]	Binding Site(s)[b]	Gene(s)[c]	K-12 Homologues[d]
	TGTGATTTGAA**TCAC**T		
4508745	**TGTGA**TATTTG**TCACA**	(*ETEC_4165*)/*ETEC_4164*	(*fdhD*)/*fdoG*
4517442	C**GTGA**TCGCTG**TC**C**CA**	(*ETEC_4173*)	(*rhaA*)
4564670	**TG**C**GA**TCCGCC**TCA**T**A**	*ETEC_4216/ETEC_4217*	*ptsA/frwC*
4668870	**TGT**A**A**CAGAGA**TCACA**	*ETEC_4289/ETEC_4290*	***malE/malK***
4725047	**TGTG**CGGATGA**TCACA**	n.a.	n.a.
4731402	**TGTGA**TCTTGCG**CA**T**A**	(*ETEC_4365*)	(*aphA*)
4761367	C**GTGA**TGGCTG**TCAC**G	*ETEC_4389*	*fdhF*
4846352	n.d.	*ETEC_4464*	ETEC-specific
4848117	C**GTGA**GTTCTG**TCACA**	n.a.	n.a.
4863253	**TTTGA**TCAACA**TC**G**CA**	(*ETEC_4478*)	(ETEC-specific)
4873926	G**GTGA**TCTATT**TCACA**	*ETEC_4486/ETEC_4487*	*aspA/fxsA*
4930149	**TGTGA**TGAACT**TCA**A**A**	*ETEC_4545/ETEC_4546*	*yjfY/rpsF*
4940903	**TGTGA**TCACTA**TC**G**CA**	*ETEC_4557/ETEC_4558*	ETEC-specific/*ytfA*
4993073	**TGTGA**CTGGTA**TCTC**G	(*ETEC_4604*)	(*valS*)
5002854	**TGT**A**A**CCTTTG**TCACA**	*ETEC_4610/tRNA-Leu*	*yjgB/tRNA-Leu*
5030724	**TG**C**GA**TGAATG**TCACA**	*ETEC_4633/ETEC_4634*	*gntP/uxuA*
5129400	CG**T**ACCGTCGG**TCACA**	*(ETEC_4736)*	*(yjjI)*
5129944	**TGTGA**TGTATA**TC**GA**A**	*ETEC_4736/ETEC_4737*	*yjjI/deoC*

[a]Chomosome coordinate of the ChIP-seq peak in H10407. Underlined text indicates that the ChIP-seq peak maps to sequence that is not conserved in *E. coli* K-12.
[b]CRP binding site sequence predicted by MEME. "n.d." indicates that MEME did not detect a putative binding site.
[c]Genes in parentheses indicate that the ChIP-seq peak is located within that gene. Downstream genes are only listed if the annotated gene start is ≤300 bp downstream of the CRP ChIP-seq peak. "n.a." indicates that no genes starts are ≤300 bp from the CRP ChIP-seq peak.
[d]*E. coli* K-12 homologues are listed for the ETEC genes in the previous column. Genes in parentheses indicate that the ChIP-seq peak is located within that gene. "n.a." indicates that no genes starts are ≤300 bp from the CRP ChIP-seq peak. "ETEC-specific" indicates that there is no K-12 homologue. Underlined genes have been identified as CRP targets in a previous ChIP-chip study [15]. Bold genes are listed as CRP targets in the Ecocyc database.
doi:10.1371/journal.ppat.1004605.t001

might be why, when isolated on the 93 bp fragment, P*estA2* was active and dependent on CRP. To test this logic we generated a further two P*estA2::lacZ* fusions using the pRW50 plasmid system. The additional P*estA2* DNA fragments were both 460 bp in length and include the full *estA2* gene that was entirely bound by H-NS in our ChIP-seq assay (Fig. 2C). The CRP site was ablated in one of the additional fragments by introducing point mutations that are predicted to disrupt CRP binding. The sequence of the DNA fragments is shown in S2A Fig. The *lacZ* fusions are illustrated graphically in Fig. 4A. Our expectation was that the longer 460 bp fragment would bind H-NS whilst the starting 93 bp fragment would not. To test this prediction we used ChIP. Thus, we compared H-NS binding to the different P*estA2* containing fragments *in vivo*. Fig. 4B shows results of a PCR analysis to measure enrichment of the P*estA2* locus. As expected, P*estA2* was only enriched in anti-H-NS immunoprecipitates when in the context of the 460 bp fragment. Crucially, enrichment is specific because, in a set of control PCR reactions, there was no enrichment of the *yabN* locus in any immunoprecipitate.

Our ChIP analysis suggests that the 460 bp fragment containing P*estA2* is subject to regulation by H-NS. To confirm that this was the case, the various pRW50 derivatives were used to transform M182Δ*lac* and M182Δ*lac*Δ*hns* cells. We then measured LacZ activity, driven by P*estA2*, in the transformants. Consistent with our expectations the data show that P*estA2* is repressed 5-fold by H-NS only in the context of the 460 bp DNA fragment (Fig. 4C). Importantly, mutations in the CRP binding site abolish P*estA2* activity in the absence of H-NS. Hence, the measured LacZ expression must be driven by P*estA2* rather than any spurious promoters located within the *estA2* gene. Taken together our ChIP-seq and LacZ activity data show that H-NS prevents CRP from activating P*estA2*.

The *estA2* and *estA1* promoters are differently regulated by CRP but similarly regulated by H-NS

The *estA1* regulatory region, located on plasmid p666, contains a sequence similar to P*estA2* (Fig. 5A). We expected that this sequence would be the *estA1* promoter (P*estA1*). To test this expectation we created a 92 bp P*estA1::lacZ* fusion, equivalent to the 93 bp P*estA2::lacZ* fusion described above, and mapped the 5′ end of the resulting mRNA. As expected, the primer extension product was 109 nt in length (Fig. 5B). Hence, P*estA1* and P*estA2* use equivalent TSSs. However, we were surprised that the intensity of the P*estA1* primer extension product *increased* in cells lacking CRP (Fig. 5B). Closer examination of the alignment in Fig. 5A shows that, whilst P*estA1* and P*estA2* are similar, there are differences in the sequence and position of key promoter elements. To try and understand which changes result in the aberrant behaviour of P*estA1* we made a set of hybrid promoters. The hybrid constructs are derived from the CRP-activated *estA2* promoter. In each hybrid, named P*estA2.1* through P*estA2.7*, a region of P*estA2* was replaced with the equivalent region from

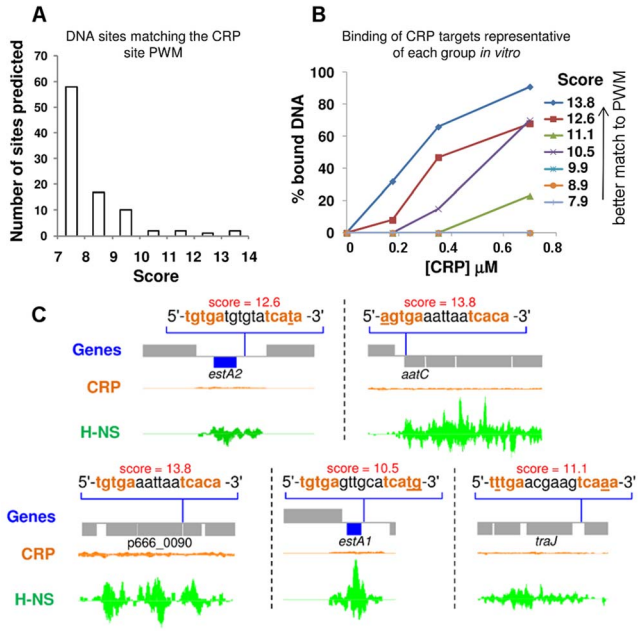


Fig. 2. Unoccupied CRP sites on p666 and p948 align with H-NS bound regions. A) A histogram showing the number of putative CRP binding sites in each of 7 discrete bins. Each bin is delineated by the "score" of the putative CRP site. A high score indicates a better match to the Position Weight Matrix that represents the consensus for CRP binding. B) The graph illustrates binding of CRP to a target from each of the bins shown in Panel A. CRP was used at concentrations of 0, 175, 350 or 700 nM. C) ChIP-seq data for CRP and H-NS binding at five regions of plasmids p666 and p948 that contain unoccupied CRP targets bound by CRP *in vitro*. The CRP and H-NS binding profiles are plots of sequence read counts at each position of the genome on both the top (above the central line) and bottom (below the central line) strand of the DNA. The y-axis scale is the same in each panel. The scale for H-NS binding is 1,785 reads on each strand and for CRP binding is 14,000 reads on each strand.
doi:10.1371/journal.ppat.1004605.g002

P*estA1* (see underlined sequences in Fig. 5C). The ability of the different hybrid promoters to drive *lacZ* expression, with and without CRP, was then tested. The results are shown in Fig. 5D. Note that, in Fig. 5D, the composition of each hybrid promoter is indicated in the grid below the graph. For example, P*estA2.1* is derived from P*estA2* but contains the P*estA1* CRP site. As expected, both P*estA1* and P*estA2* were able to drive *lacZ* expression but CRP had opposite effects. Moreover, maximal expression from P*estA1* was 3-fold lower than from P*estA2*. Only P*estA2.3* and P*estA2.5*, which both carried the same changes in the promoter -35 element, exhibited a reversed dependence on CRP. Hence, the P*estA1* -35 element must be responsible for the altered CRP dependence. All other hybrid promoters exhibited an overall reduction in activity compared to the parent P*estA2* construct. We conclude that this combination of changes results in the lower activity of P*estA1*. Note that both P*estA1* and P*estA2* were bound by H-NS in our ChIP-seq analysis (Fig. 2C). We reasoned that cloning P*estA1*, with flanking DNA, would reveal H-NS mediated repression. We generated a derivative of the P*estA1::lacZ* fusion where the downstream boundary was extended to include the entire *estA1* gene (S2B Fig., Fig. 5Ei). As expected, transcription from P*estA1* was repressed by H-NS in the presence of downstream DNA (Fig. 5Eii).

The *eltAB* operon is indirectly repressed by CRP and directly repressed by H-NS

We next turned our attention to the LT toxin promoter (P*eltAB*) [26,27]. Previously, Bodero and Munson [27] showed that transcription from this promoter was repressed by CRP. A mechanism for repression was proposed whereby CRP acted directly by binding three DNA targets overlapping P*eltAB* [27]. Even so, no CRP binding at P*eltAB* was identified by our ChIP-seq analysis (Fig. 6A). It is possible that this is because H-NS also excludes CRP from this locus (Fig. 6A). However, we also failed to identify CRP targets at P*eltAB* in our bioinformatic screen, even below the stringent cut-off (Fig. 2, S1 Table). In retrospect, this appears to be because all of three P*eltAB* CRP binding sites contain at least 4 mismatches to the consensus for CRP binding (Fig. 6A). Hence, we measured the affinity of CRP for P*eltAB* using EMSA assays. In parallel, we tested CRP binding to P*estA2* as a control. As expected, CRP bound tightly to P*estA2* at low concentrations (Fig. 6B, lanes 1–6). At high CRP concentrations further non-specific binding was observed (evidenced by a conspicuous "smear" in DNA migration in lane 7). In the equivalent experiment, with P*eltAB*, no specific binding of CRP was observed (lanes 8–13). However, non-specific CRP binding was again detectable at high protein concentrations (lane 14). Hence, CRP does not bind specifically to P*eltAB*. We hypothesised that previously observed changes in P*eltAB* activity, in cells lacking CRP, may occur indirectly. To test this, we cloned a 359 bp DNA fragment, containing P*eltAB*, into our pRW50 *lacZ* expression system. We also made a truncated 118 bp derivative of this construct where two of the three putative CRP targets were removed. A derivative of the truncated 118 bp construct, where the remaining CRP site was completely ablated by point mutations, was also made. The DNA sequences of the different constructs are shown in S2C Fig. They are illustrated graphically in Fig. 6Ci. Consistent with previous measurements, we found that transcription from P*eltAB* increased 2.5 fold in the absence of CRP. However, the response of P*eltAB* was identical when the CRP binding sites were removed (Fig. 6Cii). Hence, although CRP represses transcription from P*eltAB*, this must occur indirectly.

Given the configuration of H-NS binding at the *eltAB* locus (Fig. 6A) we reasoned that P*eltAB* would be repressed by H-NS in the presence of sufficient flanking DNA. As we had done previously for P*estA1* and P*estA2*, we compared the binding of H-NS to P*eltAB* in the presence and absence of the downstream flanking sequence. The different DNA constructs are illustrated in Fig. 7A and results of ChIP experiments to measure H-NS binding are shown in Fig. 7B. As predicted, enrichment of P*eltAB*, in immunoprecipitations with anti-H-NS, was only observed in the presence of downstream DNA. Importantly, this enrichment was specific to P*eltAB* and not observed for the control locus *yabN*. Corresponding LacZ activities, for the different DNA constructs, measured in M182 or the Δ*hns* derivative, are shown in Fig. 7C. Incorporation of flanking DNA downstream of P*eltAB* resulted in a 15-fold reduction in LacZ activity that was largely relieved in the absence of H-NS.

CRP and H-NS allow the *estA1*, *estA2* and *eltAB* promoters to respond to glucose and salt

Given the established regulatory connections between CRP and glucose, and between H-NS and salt, we next measured changes in the activity of P*estA1*, P*estA2* and P*eltAB* in response to glucose and salt. A complete description of assay conditions is provided in the Materials and Methods section. Briefly, to establish the range

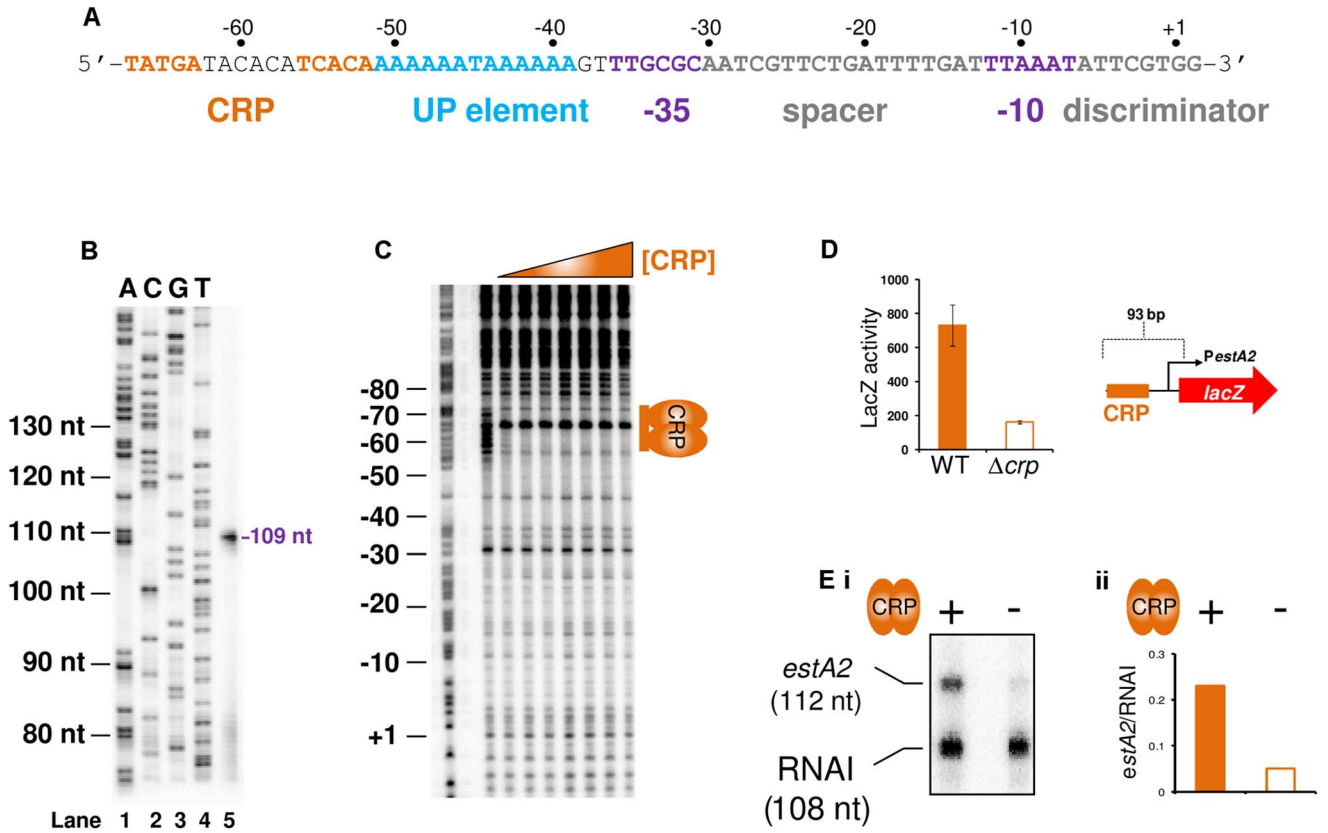


Fig. 3. The *estA2* promoter is activated by a Class I CRP dependent mechanism. A) Sequence of the *estA2* gene regulatory region. The CRP binding site is shown in orange, the UP element is blue and the promoter -10 and -35 elements are shown in purple. The different promoter positions are numbered relative to the transcription start site (+1). B) Location of the P*estA2* transcription start site. The gel shows the product of an mRNA primer extension analysis to determine the *estA2* transcription start site (Lane 5). The gel was calibrated using arbitrary size standards (A, C, G and T in Lanes 1–4). C) Binding of CRP to P*estA2*. The panel shows the result of a DNAse I footprint to monitor binding of CRP to the 93 bp P*estA2* DNA fragment. The gel is calibrated with a Maxim-Gilbert DNA sequencing reaction. CRP was added at concentrations of 0.35–2.1 μM. D) CRP is required for transcription from P*estA2 in vivo*. The panel shows a cartoon representation of the 93 bp P*estA2::lacZ* fusion and a bar chart illustrates LacZ activity in lysates of cells carrying this fusion. Assays were done in LB medium. E) i) Stimulation of P*estA2* by CRP *in vitro*. The figure shows the results of an *in vitro* transcription reaction. The 112 nt transcript initiates from P*estA2* and the 108 nt RNAI transcript is an internal control. CRP was added at a concentration of 350 nM and RNA polymerase was added at a concentration of 400 nM. ii) quantification of band intensities from the *in vitro* transcription analysis.
doi:10.1371/journal.ppat.1004605.g003

of conditions across which the promoters were able to respond, we examined the effect of titrating glucose or salt into the growth medium individually. In all experiments, we used the promoter::*lacZ* fusions that included downstream flanking DNA. This was to ensure that signals sensed by both CRP and H-NS could be integrated. As expected, the activity of P*estA1* was low. Consequently, the effects of glucose and salt were negligible (S4A Fig.). Conversely, the activity of P*estA2* was sensitive to both glucose and salt (S4B Fig.). Thus, *lacZ* expression driven by P*estA2* was repressed by glucose (orange line) and enhanced by salt (green line). As expected, P*eltAB* activity increased in the presence of both salt and glucose, but induction by salt was more prominent (S4C Fig.). We hypothesised that, for P*estA2*, the inhibitory effect of glucose should override the stimulatory effect of salt. Our reasoning was that, although H-NS can repress P*estA2*, the promoter is ultimately dependent on CRP for activity. Hence, we examined the effect of adding salt and glucose, to cells carrying the P*estA2::lacZ* fusion, separately and in combination (Fig. 8A). As predicted, the inhibitory effect of glucose was dominant (Fig. 8Ai) and was still observed in the absence of H-NS (Fig. 8Aii). Conversely, the stimulatory effect of salt required H-NS (compare green bars in Fig. 8). Importantly, in a separate experiment, we also showed that the effect of glucose on P*estA2* activity requires that the CRP site is intact (S4D Fig.). The combined effect of salt and glucose on P*eltAB* was more difficult to predict because CRP acts via an undefined, and indirect, mechanism. The result of the analysis (Fig. 8B) shows that the stimulatory effects of salt and glucose on transcription from P*eltAB* are not additive. Moreover, the stimulatory effect of glucose requires H-NS.

The response of P*eltAB* and P*estA2* to CRP and H-NS is conserved in other ETEC isolates and during host cell attachment

Examination of all sequenced ETEC genomes reveals slight variations in the sequence of the *eltAB* and *estA2* promoter sequences (recall that ETEC H10407 is somewhat anomalous in also encoding *estA1*). Thus, we next sought to understand if our model for regulation of LT and ST expression was broadly applicable. We focused our efforts on ETEC E24377A since i) the genome has been sequenced and ii) a vast array of independently generated transcriptomic data are available for this organism

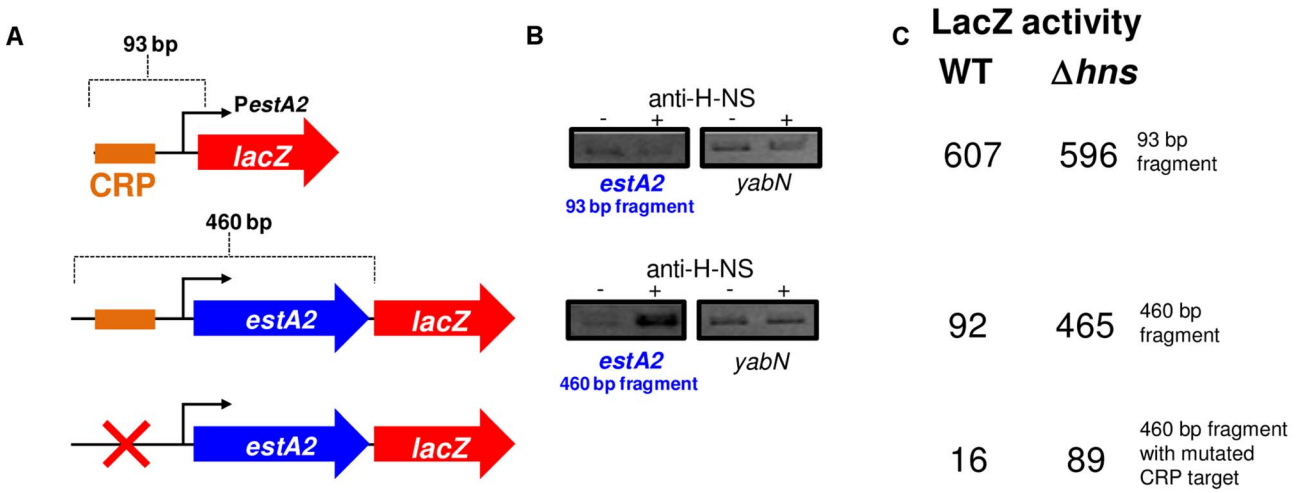


Fig. 4. The *estA2* promoter is repressed by H-NS. A) The panel shows different P*estA2*::*lacZ* fusions. The *lacZ* gene is shown as a red arrow and the *estA2* gene is shown as a blue arrow. P*estA2* is illustrated using a bent arrow and the CRP binding site is shown as an orange box. B) H-NS binds to P*estA2* only in the presence of flanking DNA. ChIP-PCR was used to measure binding of H-NS to the different P*estA2* derivatives cloned in pRW50. PCR products were generated using primers that could detect P*estA2* in the context of both the 93 bp fragment and the longer 460 bp fragment. C) The values are β-galactsidase activity values for lysates of M182, or M182Δ*hns*, carrying the different P*estA2* derivatives. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g004

[28,29]. Using ETEC E24377A DNA as a template, we generated a 460 bp P*estA2*, and 1126 bp P*eltAB* DNA fragment. The sequences are shown in S2D Fig. The DNA fragments were cloned into pRW50 and the ability of the promoters to drive *lacZ* expression in response to CRP and H-NS was measured. As expected, transcription from P*estA2* was repressed by H-NS and activated by CRP whilst P*eltAB* was repressed by H-NS (Fig. 9A). We observed no effect of CRP on transcription from *PeltAB* in the context of the 1126 bp ETEC E24377A fragment. This is not unexpected because CRP acts indirectly and these indirect CRP effects have only previously been observed in the context of short DNA fragments containing P*eltAB* that are not subject to direct repression by H-NS. We note that Sahl and Rasko previously examined the global transcriptome response of E24377A to glucose levels and bile salts [28]. In exact agreement with our model for toxin regulation, and the data in Fig. 9A, this study confirmed that i) salt induced expression of both toxins and ii) glucose inhibited expression of *estA2* [28]. Fortuitously, changes in the ETEC E24377A transcriptome, prompted by ETEC attachment to human gut epithelial cells, have also been quantified comprehensively [29]. Briefly, in these experiments, ETEC were added to sets of Caco-2 intestinal epithelial cell tissue cultures. Over a time course, ETEC that had adhered to host cells were separated from non-adhered ETEC. The transcriptomes of adhered and non-adhered ETEC were then compared. By mining these data, we next sought to determine if our model was consistent with observed changes in the transcription of *crp*, *hns*, *eltA* and *estA* during host cell attachment. Briefly, our data predict that changes in *estA* expression should be directly correlated to changes in the level of CRP and inversely correlated with changes in levels of H-NS. Conversely, levels of *eltA* expression should be inversely correlated with levels of H-NS. The result of the analysis is illustrated in Fig. 9B. The data show that the relative levels of *crp* transcription in attached and unattached cells are similar (orange line). However, levels of *hns* transcription change dramatically (green line) 60 minutes after host cell attachment. As predicted by our model, levels of *estA2* and *eltA* transcription (dashed lines) inversely track changes *hns* transcript levels. When undertaking this analysis we noticed that, although there was little change in the relative level of *crp* mRNA between attached and unattached ETEC cells, the absolute level of *crp* mRNA did fluctuate across the time course of the experiment and between biological replicates. Strikingly, when these absolute mRNA levels are compared there is a clear linear relationship between *crp* and *estA2* expression (Fig. 9C). Note that in Fig. 9C the absolute level of *hns* mRNA has been added in parenthesis for each data point. Remarkably, the only two outlying data points in this plot correspond to the two samples with increased *hns* expression. We conclude that regulation of *estA2* and *eltA* by CRP and H-NS is important during the attachment of ETEC to human intestinal epithelial cells, and that the regulatory control of ETEC toxins is conserved across different strains.

Disrupting the regulatory switch attenuates ETEC virulence

Taken together, our data suggest that CRP and H-NS form a regulatory switch that controls ETEC toxicity. We next sought to examine the effect of disabling the switch on virulence. This is not straightforward because no animal model faithfully mimics the disease caused by ETEC in humans. However, intranasal mouse models have been used as a proxy for measuring *E. coli* pathogenicity [30]. Importantly, pathogenic *E. coli* cause more severe disease in this model than non-pathogenic strains [30]. Furthermore, ETEC strains lacking genes encoding toxins and known colonisation factors are less virulent in this model [31]. We opted to disrupt the regulatory switch by removing the *crp* rather than the *hns* gene. This was a deliberate decision since *E. coli* strains lacking *hns* are severely attenuated for growth in laboratory conditions. Conversely, the *crp* null derivative of ETEC H10407 was only mildly compromised for growth in liquid culture. Hence, we compared pathogenicity of ETEC H10407, and the *crp* derivative, using the intranasal mouse model [30]. Note that the outcome of this experiment is difficult to predict since the effects of CRP on pathogenicity likely go far beyond the control of toxin expression. However, it is reasonable to assume that ETEC virulence should differ in cells lacking *crp*. The median survival of mice challenged with wild type ETEC was 53 hours and the

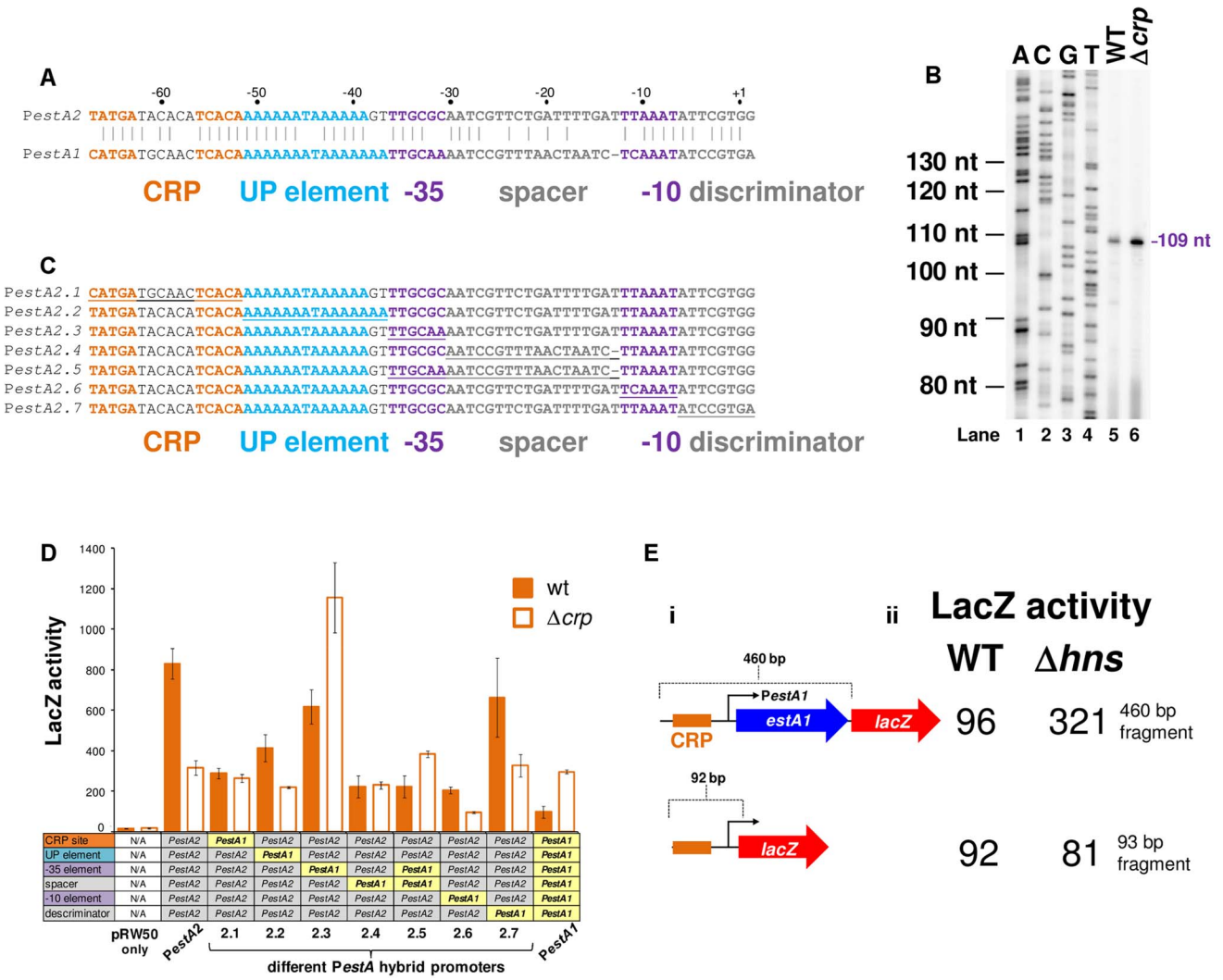


Fig. 5. Comparison of P*estA1* and P*estA2* reveals differential activity and regulation by CRP. A) Comparison of P*estA1* and P*estA2*. The panel shows the DNA sequences of P*estA1* and P*estA2*. Bases that are identical are highlighted by a solid vertical line. The CRP sites are shown in orange, the UP element in blue and the core promoter elements in purple. The sequences are numbered with respect to the transcription start site (+ 1). B) Location of the P*estA1* transcription start site. The gel shows products from an mRNA primer extension analysis (Lanes 5 and 6). The gel was calibrated using arbitrary size standards (A, C, G and T in Lanes 1–4). C) Sequences of hybrid *estA* promoters. The sequences labelled *estA2.1* through *estA2.7* are derivatives of the 93 bp P*estA2* DNA fragment where different sequence elements have been replaced with the equivalent sequence from P*estA1*. D) The bar chart shows β-galactosidase activity measurements for lysates obtained from cultures of M182, or the Δ*crp* derivative, containing the indicated hybrid promoter fragment was fused to *lacZ*. E) The panel shows different P*estA1*::*lacZ* fusions. The *lacZ* gene is shown as a red arrow and the *estA1* gene is shown as a blue arrow. P*estA1* is illustrated using a bent arrow and the CRP binding site is shown as an orange box. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g005

mortality rate was 100%. Conversely, the median survival of mice challenged with Δ*crp* ETEC was 72 h and 20% of the mice survived (Fig. 9D). Thus, whilst the full extent to which CRP co-ordinates the ETEC virulence programme remains to be determined, CRP is clearly central to the pathogenic response.

Discussion

A complex hierarchy of salt and glucose-dependent regulation controls toxin expression

We propose that toxin expression in ETEC can be controlled by osmo-metabolic flux. This is relevant to conditions in the small intestine (osmolarity equivalent to 300 mM NaCl) disease symptoms (the extrusion of cations and cAMP into the gut lumen) and treatment (the ingestion of solutions containing glucose and salt) [7–11,32]. A molecular model, describing how the different signals are integrated, is illustrated in Fig. 10. Two gene regulatory proteins, CRP and H-NS, are central to our model. Hence, H-NS directly represses the expression of *eltAB*, *estA1* and *estA2* (pathways "a" and "b" in Fig. 10). For *estA2* and *eltAB* this repression can be relieved, in an H-NS dependent manner, by increased osmolarity. At P*estA2* CRP directly activates transcription by a Class I mechanism (pathway "c"). H-NS can interfere with this process by competing with CRP for binding at P*estA2* (pathway "d"). Finally, CRP can indirectly repress expression of *eltAB* via an unknown pathway that is influenced by H-NS ("e"). Both pathways "c" and "e" are sensitive to glucose availability because of their dependence on CRP. We speculate that pathway

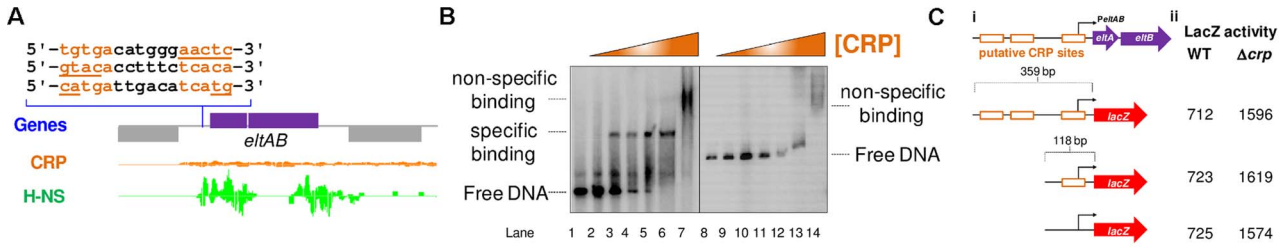


Fig. 6. The *eltAB* promoter is indirectly repressed by CRP. A) The Panel shows ChIP-seq data for CRP and H-NS binding at the *eltAB* locus. The sequence of 3 putative CRP binding sites proposed by Bodero and Munson (2009) are shown. The CRP and H-NS binding profiles are plots of sequence read counts at each position of the genome on both the top (above the central line) and bottom (below the central line) strand of the DNA. The y-axis scale for H-NS binding is 1,785 reads on each strand and for CRP binding is 14,000 reads on each strand. B) Results of an Electorphoretic Mobility Shift Assay to measure binding of CRP to the 93 bp P*estA2* fragment (Lanes 1–7) or the 359 bp P*eltAB* fragment (Lanes 8–14). Specific and non-specific binding of CRP is indicated to the left and right of the gel. CRP was added at a concentration of 0.2–7.0 μM. C) Panel (i) shows different P*eltAB::lacZ* fusions. The *lacZ* gene is shown as a red arrow and the *eltAB* operon is shown in purple. P*eltAB* is illustrated using a bent arrow and the putative CRP binding sites are shown as open orange boxes. In panel (ii) the values are β-galactsidase activity measurements taken in M182 or the Δ*crp* derivative. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g006

"e" may include H-NS since the effects of salt and sugar on *eltAB* expression were epistatic (Fig. 8). Our model for H-NS repression of *eltAB* is consistent with previous work [26]. However, our conclusion that *eltAB* is indirectly repressed by CRP disagrees with a previous study [27]. Even so, we were able to faithfully reproduce most of the observations previously described by Bodero and Munson [27]. We note that Bodero and Munson previously suggested that CRP may bind targets at P*eltAB* with a 7, rather than 6, base pair spacer between the two CRP half sites. Such CRP targets have never been described amongst hundreds of known CRP regulated promoters. Furthermore, we found no such CRP sites in our ChIP-seq analysis. Given that these DNA sequences can be deleted, without negating the effect of CRP on P*eltAB* activity, the regulatory effect of CRP must be indirect.

Oral Rehydration Therapy is likely to impact on toxin expression

Our model for regulation of ST and LT expression is pertinent to both ETEC mediated disease and its treatment. ST and LT trigger the extrusion of H_2O, cations, and cAMP (the cofactor for CRP) from the small intestine into the gut lumen [4–9]. Furthermore, solutions of salt and glucose are consumed by patients to reverse this process [10,11]. We speculate that, during infection, extrusion of electrolytes and cAMP into the gut lumen could create a positive feedback loop to drive toxin expression. Importantly, our model also suggests that ORT may provide benefits beyond stimulating rehydration of the patient. The concentration of glucose used in ORT is ~10-fold higher than is required to repress *estA2* expression. Hence, even if 90% of glucose present in ORT solutions is absorbed before reaching the site of infection, sufficient glucose should be present to down regulate toxin expression. Furthermore, even though salt is able to induce expression of *estA2* and *eltAB*, the effect is only observed at concentrations far higher than those found in ORT solutions.

Differential regulation of *estA1* and *estA2* by CRP

Our observation that *estA1* and *estA2* are oppositely regulated by CRP is intriguing given the similarities between the promoter sequences of these genes. Differential regulation is dependent on the promoter -35 element (Fig. 5). At Class I CRP regulated promoters an αCTD protomer sits between CRP and domain 4 of the RNA polymerase σ subunit, which is bound to the promoter -35 element [12]. Thus, one possible explanation is that changes in the -35 element result in subtle repositioning of σ. This could result in unproductive interactions between αCTD and σ when CRP is present.

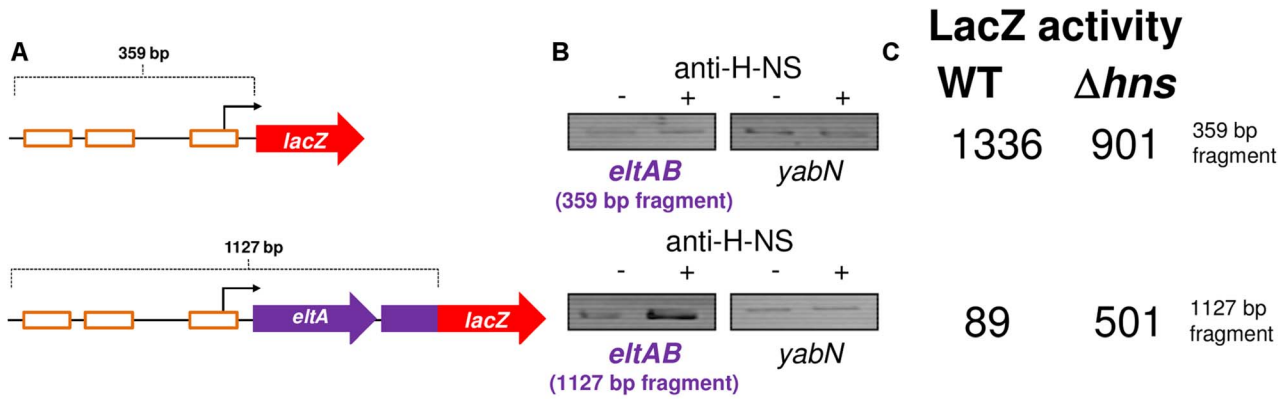


Fig. 7. The *eltAB* promoter is directly repressed by H-NS. A) The panel shows two P*eltAB::lacZ* fusions. B) The results of a ChIP-PCR analysis used to measure binding of H-NS to the two different P*eltAB* derivatives shown in Panel A. C) The values are β-galactsidase activity measurements for lysates of M182, or the Δ*hns* derivative, carrying the different P*estA2::lacZ* fusions. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g007

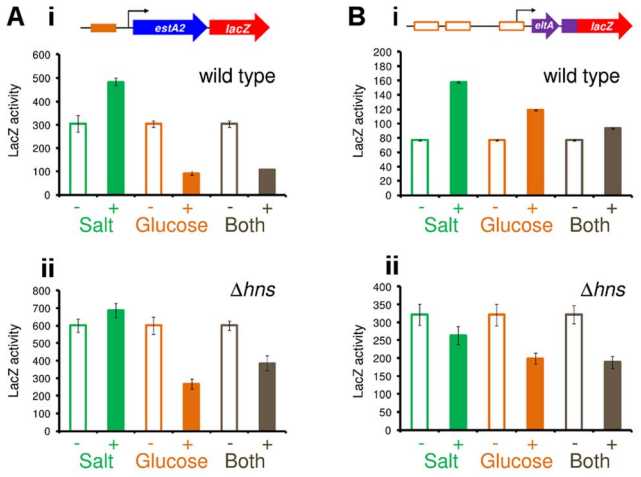


Fig. 8. H-NS and CRP integrate signals of osmolarity and metabolism to control expression of LT and ST. The figure shows β-galactosidase activity measurements for lysates obtained from cultures of M182 (i) or the Δ*hns* derivative (ii) containing A) P*estA2* or B) P*eltAB* fused to *lacZ* in plasmid pRW50. Cultures were grown in the presence and absence of 2% glucose and/or salt (60 mM NaCl and 20 mM KCl). Assays were done in M9 minimal medium so that the glucose and salt concentrations could be more accurately controlled.
doi:10.1371/journal.ppat.1004605.g008

H-NS prevents CRP regulation of select target genes

Our data indicate that several strong CRP binding sites in the H10407 genome are occluded by H-NS. This strongly suggests that the CRP regulon has evolved to incorporate additional environmental signals through the action of H-NS. The repressive effect of H-NS on transcription has been widely described [23]. H-NS represses transcription predominantly by occluding the binding of RNAP or by trapping RNAP at promoters [20]. Recently, it was shown that H-NS occludes many binding sites for the CRP homologue, FNR, in *E. coli* [21]. Thus, occlusion of transcription factor binding sites appears to be a major function of H-NS, especially for CRP family proteins. Note that, in order to exclude CRP from target promoters, sites of H-NS nucleation and CRP binding need not overlap precisely. For example, at both *estA1* and *estA2*, maximal H-NS binding is observed within the coding sequence of the gene (Fig. 2C). Despite this, H-NS oligomerisation across adjacent DNA is sufficient to prevent CRP binding.

Conclusions

In summary, our model provides a framework for better understanding ETEC mediated disease and its treatment. Moreover, our catalogue of CRP and H-NS binding targets provide a useful community resource for further studies of all *E. coli* strains. In particular, our ChIP-seq data for CRP report >50 targets not identified previously in *E. coli* K-12 and 8 ETEC-specific targets. Finally, our data show how very small changes in the organisation

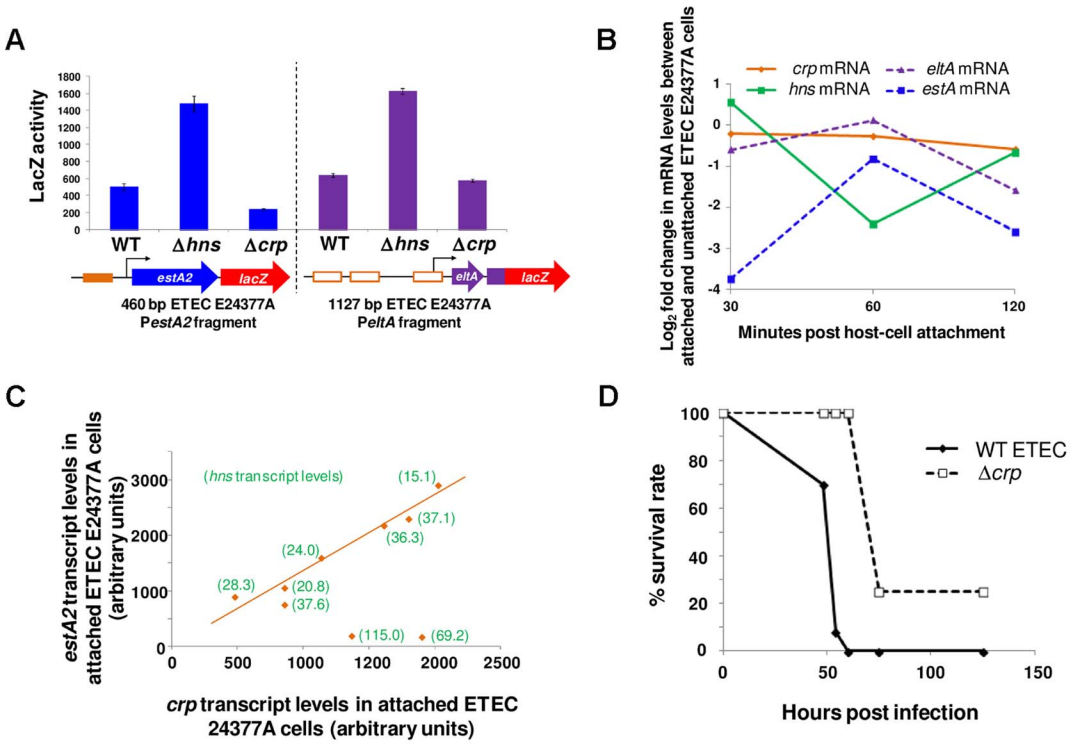


Fig. 9. Modulation of *estA2* and *eltA* transcription during attachment of ETEC E24377A to gut epithelial cells. A) The figure shows β-galactosidase activity measurements for lysates obtained from cultures of M182 or the Δ*hns* and Δ*crp* derivatives containing P*estA2* (460 bp fragment) or P*eltAB* (1126 bp fragment) from ETEC 24377A fused to *lacZ* in plasmid pRW50. B) The panel shows $\log_2$ fold changes in the transcription of *crp*, *hns*, *eltA* and *estA* in ETEC E24377A cells over a two hour incubation with a Caco-2 intestinal epithelial cell culture (29). The $\log_2$ values represent the fold change in transcription between ETEC cells attached and unattached to Caco-2 cells at each time point. C) The panel shows a scatter plot of absolute *crp* and *estA2* mRNA levels in ETEC E24377A attached to Caco-2 intestinal epithelial cells. Each data point represents a different biological replicate. For each data point the absolute level of *hns* mRNA is shown in parenthesis. D) The panel shows the survival rate of BALB/C mice (n = 30) after intranasal inoculation with wild type ETEC H10407 or the Δ*crp* derivative.
doi:10.1371/journal.ppat.1004605.g009

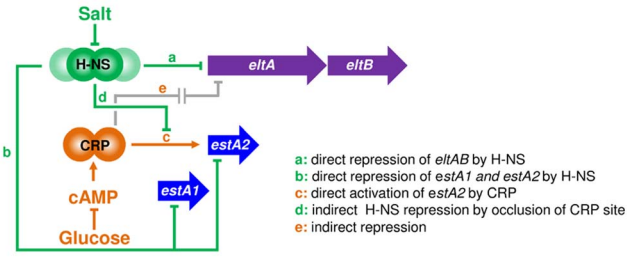


Fig. 10. An osmo-metabolic gene regulatory circuit comprised of CRP and H-NS controls expression of LT and ST. The diagram illustrates the regulatory effects of salt, cAMP and glucose on transcription from the various ST and LT promoter regions.
doi:10.1371/journal.ppat.1004605.g010

of gene regulatory regions can have major effects on gene expression, such that transcription responds differently to the same environmental cues.

Materials and Methods

Strains, plasmids and oligonucleotides

ETEC strain H10407 is described by Crossman *et al.* [1]. The C-terminal *crp*-3×FLAG tag was introduced into the H10407 chromosome using the recombineering method of Stringer *et al.* [33]. Wild type *E. coli* K-12 strains JCB387 and M182 have been described previously [34,35]. The *Δhns* M182 derivative was generated by P1 transduction of *hns*::kan from *E. coli* K12 derivative YN3144 (a gift from Ding Jin). Plasmids pRW50 and pSR are described by Lodge *et al.* [36] and Kolb *et al.* [37]. More detailed descriptions of strains and plasmids, along with the sequences of oligonucleotides, are provided in S2 Table.

ChIP-seq

Cultures were grown to mid-log phase in M9 minimal medium with 1% (w/v) fructose at 37°C. Targeted ChIP experiments (Fig. 4 and 6) were done exactly as described by Singh and Grainger [38] using P*estA2* or P*eltAB* fragments cloned in pRW50 carried in strain M182. The ChIP-seq was done as described extensively by Singh *et al.* [25] using strain H10407. Briefly, H-NS and CRP-3×FLAG were immunoprecipitated using protein A sepharose (GE Healthcare) in combination with 2 µL of anti-H-NS or 2 µl of anti-FLAG respectively. After immunoprecipitation and washing, beads were resupended in 100 µL 1× Quick Blunting Buffer (NEB) with dNTPs (as specified by the manufacturer) and 2 µL Quick Blunting Enzyme Mix, and incubated for 30 minutes at 24°C with gentle mixing. After being collected by centrifugation, the beads were again washed and the associated DNA was A-tailed by resuspension of beads in 100 µL 1× NEB buffer #2 supplemented with 2 mM dATP and 10 units of Klenow Fragment (3′→5′ exo-; NEB). Following incubation for 30 minutes at 37°C, with gentle mixing, the beads were again collected and washed. Illumina adapters (1 µl NEXTflex ChIP-seq barcoded adapters; BioO Scientific) were added to beads resuspended in 100 µL 1× Quick Ligation reaction buffer and 4 µL Quick T4 DNA Ligase (NEB), and incubated for 15 minutes at 24°C with gentle mixing. After washing the beads, the DNA was the eluted into a fresh tube by addition of 100 µL ChIP elution buffer (50 mM Tris–HCl, pH 7.5, 10 mM EDTA, 1% SDS) and incubation at 65°C for 10 minutes. The eluate was collected by centrifugation for one minute at 4000 rpm. Crosslinks were reversed by incubation for 10 minutes at 100°C. Samples were purified by phenol extraction and precipitated with ethanol, 40 µg glycogen and 8.3 mM sodium acetate. DNA was pelleted for 15 minutes at 4°C at top speed in a microcentrifuge, washed with 70% ethanol, dried and resuspended in 11 µL H_2O. After quantification by PCR each library was amplified, purified and resuspended in 20 µL H_2O. Libraries were the sequenced using a HiSeq 2000 sequencer (Illumina; University at Buffalo Next Generation Sequencing Core Facility). Sequence reads were aligned to non-repetitive sequences in the *E. coli* H10407 genome using CLC Genomics Workbench and overall coverage was determined using custom Python scripts. Sequence reads have been submitted to the EBI ArrayExpress database and can be accessed using accession number E-MTAB-2917.

Bioinformatics

ChIP-seq peaks were identified as described previously [25]. We refer to these peaks as "high stringency" peaks. A second round of peak calling was performed in which the sequence read threshold values (i.e. the minimum number of sequence reads at a given genomic position that is required for a peak to be called) was reduced by 20%. We refer to these peaks as "low stringency" peaks. MEME [39] was used to identify enriched sequence motifs in the sequences from 50 bp upstream to 50 bp downstream of the high stringency peak centres. Thus, we identified a motif closely resembling the known CRP consensus site in many of the regions surrounding high stringency ChIP-seq peaks. These CRP site sequences are included in Table 1. Those high stringency peaks for which MEME did not identify a motif were used for a second round of analysis using MEME. This also identified a motif closely resembling the known CRP consensus site. These CRP site sequences are also included in Table 1. We used MEME to identify enriched sequence motifs in the low stringency peak list. This also identified a motif closely resembling the known CRP consensus site. These CRP site sequences are also included in Table 1. "High-confidence" ChIP-seq peaks listed in Table 1 include all the high stringency peaks but only those low stringency peaks for which we identified a motif using MEME. A complete list of all peaks, including low stringency peaks for which a motif was not identified by MEME, is provided in S3 Table. In order to assess the location of CRP sites with respect to TSSs we used the targets listed in Table 1. For each target the predicted sequence from MEME was used in a BLAST search against the *E. coli* K-12 MG1655 genome. All but 11 CRP sites in ETEC had a single perfect match in the *E. coli* K-12 chromosome. For each perfect match the distance from the centre of the CRP site to all transcription start sites was calculated. Transcription start site coordinates are from Kim *et al.* [40] and Cho *et al.* [41]. Distances between −200 and +100 were selected and all other distances were discarded. Distances were then grouped in bins of 5 bp each and the most common distance bins were identified. Note that, because the position of the CRP site was transposed onto the *E. coli* K-12 genome, the distance between CRP sites and TSSs

The PWM describing CRP binding sites was generated using the PREDetector software package and our previous list of 68 CRP binding sites in the *E. coli* K-12 genome [15,42]. Subsequent bioinformatic screens of plasmids p666 and p948 were done by importing the relevant genbank files into PREDetector and running a binding site search with a cut-off of 7 using settings that did not exclude CRP sites within genes. The "score" for each site predicted by PREDetector increases if a closer match to the PWM is found. To generate the chromosome and plasmid maps shown in Fig. 1 we used DNA plotter software [43].

Data shown in Fig. 9B–C were extracted from the publically available datasets of Kansal *et al.* [29] that measure changes in the

ETEC E24377A transcriptome upon contact with Caco-2 intestinal epithelial cells. The data are hosted under the GEO accession code GSE40427. For each assay condition (planktonic and attached ETEC cells) we extracted the signal intensity for microarray probe sets A1527 (*crp*), UTI189_C1433 (*hns*), D4754 (*eltA*) and D4048 (*estA*). The average signal intensity was calculated and the fold change in transcription in attached compared to planctonic ETEC cells was determined for each time point. The data in Fig. 9C show a comparison of absolute signal intensities for probe sets A1527 (*crp*) and D4048 (*estA*) compared for each of the two replicates obtained at 30, 60 or 120 minutes after attachment to host cells. Signal intensities obtained after 30 minutes growth in LB medium (three replicates) are also included in this analysis.

Proteins

The CRP and σ^{70} purification was done exactly as described previously [44,45]. RNA polymerase core enzyme was purchased from Epicenter. RNA polymerase holoenzyme was generated by incubating the core enzyme with an equimolar concentration of σ^{70} at room temperature for 20 minutes prior to use. H-NS was overexpressed in T7 express cells from plasmid pJ414*hns*. After overexpressing H-NS, cells were collected from the culture by centrifugation and resuspended in buffer A (20 mM Tris-HCl pH 7.2, 1 mM EDTA and 10% (*v/v*) glycerol) containing 100 mg/ml PMSF. Cells were lysed by sonication and the sample was cleared by centrifugation. The supernatent was loaded directly onto a Heparin column (Amersham) pre-equilibrated with buffer A. A linear NaCl gradient was applied and H-NS was found to elute at approximately 500 mM NaCl. The peak fractions were pooled and diluted 3-fold with buffer A. The sample was then loaded onto an S-FF column (Amersham) pre-equilibrated with Buffer A. A NaCl gradient was applied and H-NS eluted at approximately 550 mM NaCl. The H-NS containing fractions were then dialysed against a buffer containing 20 mM Tris HCl (pH 7.2), 300 mM KCl and 10% Glycerol (*v/v*)for storage at −80°C.

DNAse I footprinting and Electrophoretic Mobility Shift Assays

DNA fragments for DNAse I footprinting or EMSA assays were excised from pSR by sequential digestion with *Hin*dIII and then *Aat*II. After digestion, fragments were labelled at the *Hin*dIII end using [γ-^{32}P]-ATP and T4 polynucleotide kinase. DNAse I footprints and EMSA experiments were then done as described by Grainger *et al*. [45] except that cAMP was added to reactions at a concentration of 0.2 mM. Radio-labelled DNA fragments were used at a final concentration of ~10 nM. Note that all *in vitro* DNA binding reactions contained a vast excess (12.5 μg ml^{-1}) of Herring sperm DNA as a non-specific competitor. Footprints were analysed on a 6% DNA sequencing gel (molecular dynamics). The results of all footprints and EMSA experiments were visualized using a Fuji phosphor screen and Bio-Rad Molecular Imager FX.

Primer extension assays

Transcript start sites were mapped by primer extension, as described in Lloyd *et al*. [46] using RNA purified from strains carrying the 92 bp P*estA1* or 93 bp P*estA2* fragment cloned in pRW50. The 5′ end-labelled primer D49724, which anneals downstream of the *Hin*dIII site in pRW50, was used in all experiments. Primer extension products were analysed on denaturing 6% polyacrylamide gels, calibrated with size standards, and visualized using a Fuji phosphor screen and Bio-Rad Molecular Imager FX.

In vitro transcription assays

The *in vitro* transcription experiments were performed as described previously Savery *et al*. [35] using the system of Kolb *et al*. [38]. A Qiagen maxiprep kit was used to purify supercoiled pSR plasmid carrying the different promoter inserts. This template (~16 μg ml−1) was pre-incubated with purified CRP in buffer containing 0.2 mM cAMP, 20 mM Tris pH 7.9, 5 mM $MgCl_2$, 500 μM DTT, 50 mM KCl, 100 μg ml^{-1} BSA, 200 μM ATP, 200 μM GTP, 200 μM CTP, 10 μM UTP with 5 μCi [α-32P]-UTP. The reaction was started by adding purified *E. coli* RNA polymerase. Labelled RNA products were analysed on a denaturing polyacrylamide gel.

β-galactosidase assays and addition of glucose and salt to growth medium

β-Galactosidase assays were done using the protocol of Miller [47]. All assay values are the mean of three independent experiments with a standard deviation <10% of the mean. Cells were grown aerobically at 37°C to mid-log phase in LB medium unless stated otherwise. For all experiments investigating the effects of glucose and salt M9 minimal medium was used so that the glucose and salt concentrations could be controlled more accurately. The amount of glucose is shown as percentage *w/v*. The addition of "salt" refers to a 3:1 molar ration of NaCl to KCl. We have arbitrarily described 30 mM NaCl and 10 mM KCl as being a "1%" salt solution.

Intranasal mouse infection model assays

Strains of ETEC were grown in Luria Broth (LB) to an OD_{600} of 1.0. Groups of 10 mice (8–10 week old BALB/c) were infected intranasally with approximately 1×10^9 colony forming units of bacteria in 100 μl of inoculums according to Byrd *et al*. [30]. Mice were monitored daily for 6 days post-infection for weight and morbidity.

Ethics statement

The protocol 12-02-015IBT "Oral Immunization of Mice with Enterotoxigenic: *E coli* (ETEC)" has been approved by the Noble Life Sciences IACUC committee. All animal care and use procedures adhere to the guidelines set by the Public Health Service Policy, U.S. Dept. of Agriculture (USDA) and the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health.

Supporting Information

S1 Fig Binding of CRP to predicted targets *in vitro*. A) The data show binding of CRP to a target from each of the bins shown in Fig. 2. B) CRP binding to remaining targets scoring >10. CRP was used at concentrations of 0, 175, 350 or 700 nM. The "score" describes how well the predicted target matches the PWM.
(PDF)

S2 Fig Promoter DNA fragments used in this work. A) ETEC H10407 P*estA2* containing DNA fragments. B) ETEC H10407 derived DNA fragments containing P*estA1*. C) ETEC H10407 P*eltAB* containing sequences. D) DNA fragments containing sequences upstream of the *estA2* and *eltAB* genes of ETEC E24377A.
(PDF)

S3 Fig A) Mutations in P*estA2* UP-element alter the migration of the promoter DNA on an agarose gel. The DNA sequences used are shown in part (i) and the mobility of the fragments, on an agarose gel, are shown in part (ii). Note that each sample has been loaded in duplicate. B) Mutating the UP-element renders P*estA2* uninducible by CRP. Part (i) shows LacZ activity data for the different promoter fragments cloned in pRW50. The pRW50 derivatives were used to transform M182 or the Δ*crp* derivative. Part (ii) shows the result of in vitro transcription assays using the different promoter fragments, cloned in pSR, as a template.
(PDF)

S4 Fig Activity of different promoter::*lacZ* fusions in the presence of increasing glucose and salt concentrations. The figure shows β-galactosidase activity measurements for lysates obtained from cultures of M182 carrying the A) *estA1* B) *estA2* or C) *eltAB* promoters cloned in pRW50. Panel D) shows β-galactosidase activity values for lysates of M182 and M182Δ*hns* cells, carrying the *estA2* promoter, or a derivative lacking the CRP site, cloned in pRW50. Cells were grown in the presence or absence of 2% glucose. Assays were done in M9 minimal medium so that the glucose and salt concentrations could be more accurately controlled.
(PDF)

S1 Table Putative CRP binding targets on ETEC plasmids p948 and p666 identified by PREDetector.
(DOCX)

S2 Table Strains, plasmids and oligonucleotides.
(DOCX)

S3 Table All CRP binding sites on the ETEC H10407 chromosome identified by ChIP-seq.
(DOCX)

Acknowledgments

The authors would like to thank Ian Henderson for advice and support, Doug Browning for critical comments on the manuscript, and George Munson for helpful discussions.

Author Contributions

Conceived and designed the experiments: JTW DCG. Performed the experiments: JRJH PS AMS. Analyzed the data: JRJH PS AMS JTW DCG. Contributed reagents/materials/analysis tools: AMS JTW. Wrote the paper: JTW DCG.

References

1. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, et al. (2010) A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. J Bacteriol 192:5822–5831.
2. Sack RB (2011) The discovery of cholera - like enterotoxins produced by *Escherichia coli* causing secretory diarrhoea in humans. Indian J Med Res 133: 171–80
3. Gupta SK, Keck J, Ram PK, Crump JA, Miller MA, et al. (2008) Analysis of Data Gaps Pertaining to Enterotoxigenic Escherichia coli Infections in Low and Medium Human Development Index Countries, 1984–2005. Epidemiology and Infection 136:721–738.
4. de Haan L, Hirst TR (2004) Cholera toxin: a paradigm for multi-functional engagement of cellular mechanisms. Mol Membr Biol 21:77–92.
5. Zhang RG, Scott DL, Westbrook ML, Nance S, Spangler BD, et al. (1995) The three-dimensional crystal structure of cholera toxin. J Mol Biol 251:563–73.
6. Taxt A, Aasland R, Sommerfelt H, Nataro J, Puntervoll P (2010) Heat-stable enterotoxin of enterotoxigenic *Escherichia coli* as a vaccine target. Infect Immun 78:1824–31.
7. Yamamoto T, Tamura T, Yokota T (1984) Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans. J Biol Chem 259:5037–44.
8. de Haan L, Verweij WR, Feil IK, Holtrop M, Hol WG, et al. (1998) Role of GM1 binding in the mucosal immunogenicity and adjuvant activity of the *Escherichia coli* heat-labile enterotoxin and its B subunit. Immunology 94:424–430.
9. Saslowsky DE, te Welscher YM, Chinnapen DJ, Wagner JS, Wan J, et al. (2013) Ganglioside GM1-mediated transcytosis of cholera toxin bypasses the retrograde pathway and depends on the structure of the ceramide domain. J Biol Chem 288:25804–9.
10. Nalin DR, Cash RA, Islam R, Molla M, Phillips RA (1968) Oral maintenance therapy for cholera in adults. Lancet 2:370–3.
11. Guerrant RL, Carneiro-Filho BA, Dillingham RA (2003) Cholera, diarrhea, and oral rehydration therapy: triumph and indictment. Clin Infect Dis 37:398–405.
12. Busby S, Ebright RH (1999) Transcription activation by catabolite activator protein (CAP). J Mol Biol 293:199–213.
13. Parkinson G, Wilson C, Gunasekera A, Ebright YW, Ebright RH, et al. (1996) Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. J Mol Biol 260:395–408.
14. Zheng D, Constantinidou C, Hobman JL, Minchin SD (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. Nucleic Acids Res 32:5874–93.
15. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. Proc Natl Acad Sci U S A 102:17693–8.
16. Rossiter AE, Browning DF, Leyton DL, Johnson MD, Godfrey RE, et al. (2011) Transcription of the plasmid-encoded toxin gene from enteroaggregative *Escherichia coli* is regulated by a novel co-activation mechanism involving CRP and Fis. Mol Microbiol 81:179–91.
17. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, et al. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science 313:236–8.
18. Liu Y, Chen H, Kenney LJ, Yan J (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. Genes Dev 24:339–44.
19. Lim CJ, Lee SY, Kenney LJ, Yan J (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. Sci Rep 2:509.
20. Myers KS, Yan H, Ong IM, Chung D, Liang K, et al. (2013) Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. PLoS Genet 9(6):e1003565.
21. Dame RT, Wyman C, Wurm R, Wagner R, Goosen N (2002) Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1. J Biol Chem 277:2146–50.
22. Dorman CJ (2007) H-NS, the genome sentinel. Nat Rev Microbiol 5: 157–61.
23. Atlung T, Ingmer H (1997) H-NS: a modulator of environmentally regulated gene expression. Mol Microbiol 24: 7–17.
24. Steinsland H1, Valentiner-Branth P, Perch M, Dias F, Fischer TK, Aaby P, Mølbak K, Sommerfelt H (2002) Enterotoxigenic *Escherichia coli* infections and diarrhea in a cohort of young children in Guinea-Bissau. J Infect Dis 186: 1740–7.
25. Singh SS, Singh N, Bonocora RP, Fitzgerald DM, Wade JT, et al. (2014) Widespread suppression of intragenic transcription initiation by H-NS. Genes Dev 28: 214–219.
26. Yang J, Tauschek M, Strugnell R, Robins-Browne RM (2005) The H-NS protein represses transcription of the *eltAB* operon, which encodes heat-labile enterotoxin in enterotoxigenic *Escherichia coli*, by binding to regions downstream of the promoter. Microbiology 151: 1199–1208.
27. Bodero MD, Munson GP (2009) Cyclic AMP receptor protein-dependent repression of heat-labile enterotoxin. Infect Immun 77:791–8.
28. Sahl JW, Rasko DA (2012) Analysis of global transcriptional profiles of enterotoxigenic Escherichia coli isolate E24377A. Infect Immun 80: 1232–1242.
29. Kansal R, Rasko DA, Sahl JW, Munson GP, Roy K (2013) Transcriptional modulation of enterotoxigenic *Escherichia coli* virulence genes in response to epithelial cell interactions. Infect Immun 81: 259–70.
30. Byrd W, Mog SR, Cassels FJ (2003) Pathogenicity and immune response measured in mice following intranasal challenge with enterotoxigenic *Escherichia coli* strains H10407 and B7A. Infect Immun 71: 13–21.
31. Byrd W, Boedeker EC (2013) Attenuated *Escherichia coli* strains expressing the colonization factor antigen I (CFA/I) and a detoxified heat-labile enterotoxin (LThK63) enhance clearance of ETEC from the lungs of mice and protect mice from intestinal ETEC colonization and LT-induced fluid accumulation. Vet Immunol Immunopathol 152: 57–67.
32. Gupta S, Chowdhury R (1997) Bile affects production of virulence factors and motility of *Vibrio cholerae*. Infect Immun 65: 1131–1134.
33. Stringer AM, Singh N, Yermakova A, Petrone BL, Amarasinghe JJ, et al. (2012) FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. PLoS One 7: e44841.

34. Page L, Griffiths L, Cole JA (1990) Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria. Arch Microbiol 154:349–54.
35. Busby S, Kotlarz D, Buc H (1983) Deletion mutagenesis of the Escherichia coli galactose operon promoter region. J Mol Biol 167:259–274.
36. Lodge J, Fear J, Busby S, Gunasekaran P, Kamini NR (1992) Broad host range plasmids carrying the Escherichia coli lactose and galactose operons. FEMS Microbiol Lett 74:271–6.
37. 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 Res 23:819–26.
38. Singh SS, Grainger DC (2013) H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions. PLoS Genet 9: e1003589.
39. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37(Web Server issue):W202–8.
40. Kim D, Hong JS, Qiu Y, Nagarajan H, Seo JH, et al. (2012) Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. PLoS Genet 8(8):e1002867.
41. Cho BK, Kim D, Knight EM, Zengler K, Palsson BØ (2014) Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states. BMC Biol 12:4.
42. Hiard S, Maree R, Colson S, Hoskisson PA, Titgemeyer F, et al. (2007) PREDetector: A new tool to identify regulatory elements in bacterial genomes. Biochem Biophys Res Commun 357(4):861–4.
43. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25: 119–20.
44. Savery NJ, Lloyd GS, Kainz M, Gaal T, Ross W, et al. (1998) Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase alpha subunit. EMBO J 17:3439–3447.
45. Grainger DC, Goldberg MD, Lee DJ, Busby SJ (2008) Selective repression by Fis and H-NS at the *Escherichia coli dps* promoter. Mol Microbiol 68:1366–1377.
46. Lloyd GS, Hollands K, Godfrey RE, Busby SJ (2008) Transcription initiation in the *Escherichia coli* K-12 malI-malX intergenic region and the role of the cyclic AMP receptor protein. FEMS Microbiol Lett 288: 250–7.
47. Miller J (1972) Experiments in Molecular Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.