STUDIES WITH AGGR: THE MASTER REGULATOR OF VIRULENCE IN ENTEROAGGREGATIVE *ESCHERICHIA COLI*

BY

GABRIELLE SOPHIE CHRISTIE

A THESIS

SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INSTITUTE OF MICROBIOLOGY AND INFECTION
SCHOOL OF BIOSCIENCES
COLLEGE OF LIFE AND ENVIRONMENTAL SCIENCES
UNIVERSITY OF BIRMINGHAM
SEPTEMBER 2022

Abstract

Escherichia coli exists as part of the human flora in the gastrointestinal tract. Some strains cause intestinal and extra-intestinal infections, and these bacteria are amongst the most prevalent of Gram-negative pathogens. There are many pathotypes of diarrhoeagenic *E. coli*, including enteroaggregative *E. coli* (EAEC). EAEC is a significant cause of diarrhoeal disease in low-, middle- and high-income countries. Aggregative adherence fimbriae (AAFs) mediate EAEC adherence to epithelial cells. A 'typical' EAEC strain contains the master virulence regulator, AggR, which is a transcription factor, encoded on a virulence plasmid (pAA), that has over 40 virulence genes in its regulon. Previous work demonstrated that AggR acts as a monomer as a Class II activator of transcription at its dependent promoters. The aim of this project was to investigate complex AggR-regulated promoters, focussing on a bi-directional promoter, and to examine variations in the organisation of dependent promoters between strains. A further objective was to define the key determinants for AggR activity with an aim of understanding how its activity is regulated.

The *yicS-nlpA* bi-directional promoter is known to be present on the chromosome in more than ten *E. coli* strains, though it has been identified in at least 30 more strains. The YicS protein is thought to be involved in biofilm formation in avian pathogenic *E. coli* (APEC), but YicS is not required for biofilm formation in EAEC 042. An investigation of the regulation of the *yicS* and *nlpA* promoters, from *E. coli* K-12, EAEC strain 042 and EAEC strain H92/3, by AggR identified the key determinants that permit AggR-dependent activation of the *yicS* promoter in

EAEC 042, weaker activation in EAEC H92/3, and no activation in *E. coli* K-12. In contrast, in all these strains, the *nlpA* promoter is weakly repressed by AggR.

In my study, the genome of the 'atypical' EAEC strain H149/5, which lacks a virulence plasmid, was sequenced and several AggR-activated promoters were identified. I identified the *aap* and *caf1M* genes, which demonstrates that this strain carries the dispersin (Aap) surface protein and Caf1M, a homologue of AAF proteins. An AggR-like transcription factor was identified, CfaD H149/5, that is chromosomally encoded. The AraC subgroup transcription regulators, Rns and AggR, could substitute for CfaD for activation of promoters from H149/5. My results showed that CfaD H149/5 has a different activation profile to AggR.

Through mutational analysis, key determinants for AggR activity were located throughout both the N-terminal domain (NTD) and the C-terminal domain (CTD). None of the AggR mutants were strongly *trans*-dominant over wild-type AggR and this argues that AggR might function by pre-recruitment. An investigation of how AggR activity is regulated suggested the existence of a bi-stable virulence switch. Consistent with this, I showed that expression of AggR is not dependent on any external signal. Measurement of cell-to-cell variation of AggR activity indicates that a small proportion of cells reach a critical threshold of AggR production, and I argue that this 'flips' the virulence switch to the 'on' state. Taken together, the available experimental data suggest that the regulation of AggR activity is dependent on a feedforward loop with a bi-stable switch.

Dedicated to my family,

my parents, Sarah and Andrew

my sister, Kate

and Fergus

Acknowledgements

Thank you to the BBSRC and MIBTP for funding this PhD project. Thank you to the University of Birmingham for 8 wonderful years as an undergrad and postgrad.

I would like to thank my supervisors, Professor Stephen Busby, Dr Douglas Browning and Dr Damon Huber for the support and guidance.

Steve, thank you so much for all of your help and advice with my PhD project and thesis. It's been a pleasure to be your PhD student and to have had the opportunity to work with you.

Doug, thank you for all you've taught me about being a successful research scientist both during my undergrad project and during my PhD – and thank you for taking a chance on me!

Georgina Lloyd, firstly, thank you so much for reading my thesis drafts. Your support has been invaluable – you let me vent and offered me advice, I couldn't have coped without our weekly coffee dates!

A special thank you to Rita Godfrey for your guidance and sharing your considerable expertise. And to Jo Hothersall, Munirah Alhammadi, and other lab members over the years – it's been a pleasure working with you all and thank you for everything you've done to help me, not least for your friendship and the fun that we've had! Thank you also to the Grainger group for letting us Busby's crash your lab meetings.

Jess Blair, Helen McNeil, and Sarah Element – thank you for your guidance with flow cytometry.

To those that I am lucky to call my friends and family around the world – thank you for everything.

Thank you to my dear friends: Ali, Tash, Jem, Ash, Maria, Nicole, Jess, Niah, Ailbhe, and all of the wonderful friends I've made through MIBTP.

To my wonderful family in Wellington and the Wairarapa, thank for your generosity and support. Grandma and Grandad, thank you for always reading my papers, but please don't try to read this thesis! And Grandma, thank you for your

continued concern for my “mutant babies”, you can see them all thriving in Chapter 5!
And to my family,
Luca, Vale and Giosuè – thank you for your encouragement and support, and for the funny videos and interesting articles!
Kate, Harry, and Harold – thank you for your support, formatting expertise, letting me stay for extended periods (I know I'm always the favourite dog-sitter!) and for sharing wine and whiskey with me.
Fergus, my furry companion, we lost you almost a year ago, but you’ll always be in our hearts and in our thoughts when we see your pear tree.
Last in order but not of importance, my parents for the wonderful life you've given me, for your love and support, for teaching me the important values I live by and for encouraging me even when you have no idea what I’m talking about!

Table of Contents

Table of Figures

List of Tables

List of abbreviations

A: Adenosine

AA: Aggregative adherence

AAFs: Aggregative adherence fimbriae

Aai: AggR Activated Island

Aap: Antiaggregation protein or dispersin 4

Aar: AggR-activated regulator

Amp^R: Ampicillin-resistance

AMR: Antimicrobial resistance

ANR: AraC Negative Regulator

APS: Ammonium persulfate

bp: Base pair

C: Cytosine

CIP: Calf intestinal alkaline phosphate

Cm^R: Chloramphenicol-resistance

CRE: Core recognition element

CRP: Cyclic AMP receptor protein

CTD: Carboxyl terminal domain

DMEM: Dulbecco's modified Eagle medium

DBD: DNA-binding domain

dsDNA: Double stranded DNA

DNA: Deoxyribonucleic acid

EAST-1: Enteroaggregative heat-stable toxin

EBD: Effector binding domain

FSC-H: Forward scatter height

G: Guanine

GFP: Green fluorescent protein

H-NS: Histone like-nucleoid structuring protein

HUS: Hemolytic uremic syndrome

HTH: Helix-turn-helix

Kb: Kilo base pair

Km^R: Kanamycin-resistance

LB: Luria-Bertani broth

LPS: Lipopolysaccharides

nt: Nucleotide

NTD: Amino terminal domain

NTP: Nucleoside triphosphate (N=A, C, G or T)

OD: Optical density

ONPG: Ortho-nitrophenyl-b-D-galactopyranoside

ORF: Open reading frame

pAA: Aggregation adherence plasmid

PAI-I: *pheU* pathogenicity island

PCR: Polymerase chain reaction

PMT: Photomultiplier tube

RBS: Ribosome binding sites

RNA: Ribonucleic acid

RNAP: RNA polymerase

RPitc: RNAP-promoter initial transcribing complex

RPo: Open promoter complex

ShET1: *Shigella* enterotoxin 1

Stx: Shiga toxin

ssDNA: single stranded DNA

SOC: Super Optimal broth with Catabolite repression

T: Thymine

TBE: Tris/Borate/EDTA buffer

TEC: Transcription elongation complex

Tet^R: Tetracycline resistance

TEMED: N,N,N',N'-Tetramethylethylenediamine

TSS: Transcription start site

T1SS: Type I secretion system

T6SS: Type VI secretion system

v/v: Volume per volume

w/v: Weight per volume

Amino acid standard single letter codes were used.

Chapter 1 Introduction

1.1 *Escherichia coli* – an overview

In the late 19th century, a very small microorganism called *Escherichia coli* was identified by Theodor Escherich (El-Hajj and Newman, 2015; Hacker and Blum-Oehler, 2007; Clements *et al.*, 2012; Lim *et al.*, 2010). *E. coli* is a rod-shaped and very small Gram-negative bacterium, approximately 0.5 μm wide by 1-2 μm long (El-Hajj and Newman, 2015; Davies, 1996). Most *E. coli* strains exist harmlessly as part of the normal human flora in the gastrointestinal tract (Davies, 1996; Hacker and Blum-Oehler, 2007; Lim *et al.*, 2010; Prager *et al.*, 2014). However, pathogenic *E. coli* are significant contributors to the burden of disease, including serious nosocomial infections, and one of the most prevalent Gram-negative pathogens (Davies, 1996; Hacker and Blum-Oehler, 2007; Miajlovic and Smith, 2014). Pathogenic strains are distinguished from harmless commensal isolates by their acquisition of virulence-associated genes through plasmids, transposons, pathogenicity islands, or bacteriophages (Hacker and Blum-Oehler, 2007; Lim *et al.*, 2010; Prager *et al.*, 2014). Pathogenic *E. coli* can be categorised by presence of specific virulence factors, the serotype, the mechanisms of pathogenicity, or the clinical presentation (Lim *et al.*, 2010).

There are many established pathotypes in the group of diarrhoeagenic *E. coli* (DEC) including Enteroaggregative *Escherichia coli* (EAEC), diffusely adhering *E. coli* (DAEC), Shiga toxin-producing *E. coli* (STEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), EHEC and

verocytotoxin *E. coli* (VTEC) (Jensen and Hebbelstrup Rye Rasmussen, 2017; Navarro-Garcia and Elias, 2011; Dias *et al.*, 2020; Kaper *et al.*, 2004).

1.2 EAEC prevalence and identification

The impact of EAEC on global public health is a growing concern due to the association with serious outbreaks of disease, high morbidity and mortality rates, and the broad range of susceptible patients (Andrade *et al.*, 2017; Yasir *et al.*, 2019; Navarro-Garcia and Elias, 2011; Croxen *et al.*, 2013). EAEC is a significant cause of diarrhoeal disease in adults and children, both in industrial and developing countries (Dias *et al.*, 2020; Yasir *et al.*, 2019; Jenkins *et al.*, 2005; Morin *et al.*, 2010), and is associated with diarrhoea, weight loss, malnutrition and growth retardation in children in developing countries (Bamidele *et al.*, 2019; Havt *et al.*, 2017; Ellis *et al.*, 2020). EAEC is responsible for 8-32% of acute diarrhoeal cases in infants and children, depending on geographic location (Navarro-Garcia and Elias, 2011; Bamidele *et al.*, 2019). It can also be attributed to 20-30% of persistent diarrhoeal incidences, described as lasting more than 14 days (Bamidele *et al.*, 2019), including persistent paediatric diarrhoea (Jensen and Hebbelstrup Rye Rasmussen, 2017; Nataro *et al.*, 1993; Navarro-Garcia and Elias, 2011), persistent enteric infections in HIV-infected patients, and extra-intestinal infections including urinary tract infections and septicaemia (Ellis *et al.*, 2020; Yasir *et al.*, 2019; Mellies and Barron, 2006).

EAEC is typically linked to foodborne outbreaks and is one of the leading causes of traveller's diarrhoea (Santiago *et al.*, 2017; Morin *et al.*, 2010). In 2011, a serious foodborne outbreak occurred in Germany which led to over 4000 gastroenteritis cases, 850 of which progressed to hemolytic uremic syndrome

(HUS), 900 hospitalisations and 54 deaths (Yasir *et al.*, 2019; Berger *et al.*, 2016; Lang *et al.*, 2018; Buchholz *et al.*, 2011; Boisen *et al.*, 2015; Ellis *et al.*, 2020). Globally, this was the highest occurrence of *E. coli*-related HUS and the largest foodborne outbreak in Germany (Berger *et al.*, 2016; Boisen *et al.*, 2015). The strain implicated was EAEC O104:H4, which had a combination of enterohemorrhagic *E. coli* (EHEC) and EAEC virulence determinants, including the EHEC virulence factor Shiga toxin (Stx), and carried an EAEC-like pAA plasmid (Ellis *et al.*, 2020; Berger *et al.*, 2016; Lang *et al.*, 2018; Buchholz *et al.*, 2011; Brzuszkiewicz *et al.*, 2011; Boisen *et al.*, 2015).

The paradigm EAEC strain 042 was first isolated from a child in Peru with acute diarrhoea in 1983 (Nataro *et al.*, 1985). This strain is comprised of a 100 kb aggregative adherence plasmid, pAA, and the 5.2 Mb circular chromosome (Estrada-Garcia *et al.*, 2014; Sheikh *et al.*, 2002; Yasir *et al.*, 2019). The pAA plasmid encodes AggR, which is an important regulator of many virulence factors implicated in pathogenicity, including aggregative adherence fimbriae (AAFs) (Jonsson *et al.*, 2017; Huttener *et al.*, 2018; Bamidele *et al.*, 2019; Morin *et al.*, 2010; Berger *et al.*, 2016; Yasir *et al.*, 2019; Dias *et al.*, 2020). AAFs mediate the EAEC-unique aggregative adherence to intestinal epithelial cells, that has been characterised as a “stacked-brick” pattern (Jonsson *et al.*, 2017; Jensen and Hebbelstrup Rye Rasmussen, 2017; Berger *et al.*, 2016; Huttener *et al.*, 2018; Bamidele *et al.*, 2019; Dias *et al.*, 2020; Estrada-Garcia *et al.*, 2014). This characteristic pattern of adherence has been detected in patients presenting diarrhoeal disease in many different countries and across a range of age groups

(Estrada-Garcia *et al.*, 2014), and this allows identification, despite the heterogeneity of EAEC (Ellis *et al.*, 2020; Estrada-Garcia *et al.*, 2014).

1.3 Pathogenesis of EAEC and clinical manifestation

The major stages of EAEC pathogenesis include colonisation of the host intestinal mucosa, biofilm formation, toxin secretion, cytotoxic damage and intestinal inflammation (Figure 1.1) (Morin *et al.*, 2010; Yasir *et al.*, 2019; Dias *et al.*, 2020; Navarro-Garcia and Elias, 2011; Mohamed *et al.*, 2007; Ellis *et al.*, 2020). Several virulence factors thought to be involved in pathogenesis, have been identified (Ellis *et al.*, 2020). AAFs mediate adherence to the intestinal epithelium and are believed to be central to pathogenicity in certain EAEC strains (Morin *et al.*, 2010; Jonsson *et al.*, 2017). AAFs are a family of chaperone-usher adhesins, and five major variants have been identified so far, AAF/I to AAF/V (Ellis *et al.*, 2020; Morin *et al.*, 2010; Jonsson *et al.*, 2017). Antiaggregation protein (Aap), dispersin, facilitates the dispersion of AAFs on the bacterial surface by binding to the surface of lipopolysaccharides (LPS) on the outer membrane (Asadi Karam *et al.*, 2017; Harrington *et al.*, 2006; Dias *et al.*, 2020; Berger *et al.*, 2016; Boll *et al.*, 2013). This effect has been suggested to be due to dispersin counteracting surface charge on the LPS and AAFs (Asadi Karam *et al.*, 2017). Enterotoxins and cytotoxins are then produced and secreted, causing cytotoxic damage. These include enteroaggregative heat-stable toxin (EAST-1) and *Shigella* enterotoxin 1 (ShET1), and autotransporters (e.g. SepA, SigA, Pic, Pet) (Morin *et al.*, 2010; Ellis *et al.*, 2020; Dias *et al.*, 2020; Navarro-Garcia and Elias, 2011). Following this, mucosal inflammation is triggered, and the host immune response leads to interleukin-1 and interleukin-8 secretion (Morin *et al.*, 2010;

Yasir *et al.*, 2019; Navarro-Garcia and Elias, 2011; Jensen and Hebbelstrup Rye Rasmussen, 2017). It is important to understand how EAEC causes disease, in order to appreciate the differences between asymptomatic pathogenic strains (Franca *et al.*, 2013), particularly as genotyping has not identified any robust link between any single gene (or combination of genes) and virulence (Ellis *et al.*, 2020; Boisen *et al.*, 2020).

Globally, antibiotic resistance is a growing concern. There is an increasing prevalence of antimicrobial resistance (AMR) in clinical *E. coli* strains; EAEC has previously been tested for antibiotic sensitivity and, though it is susceptible to quinolones, it is highly resistant to many antibiotics, including ampicillin and trimethoprim (Dias *et al.*, 2020; McHale *et al.*, 2018). Due to the decreasing susceptibility to antibiotics and effective immune system evasion by EAEC during infection, antibiotics are reserved for severe cases, particularly since development of new antibiotics has slowed (Jensen and Hebbelstrup Rye Rasmussen, 2017; Hebbelstrup Jensen *et al.*, 2014; Roberts *et al.*, 2018).

1.4 Virulence determinants of EAEC

The genomes of EAEC strains are heterogenic and highly mosaic (Bamidele *et al.*, 2019; Jensen and Hebbelstrup Rye Rasmussen, 2017; Estrada-Garcia *et al.*, 2014; Ellis *et al.*, 2020). Pathogenicity is orchestrated by key virulence factors (Bamidele *et al.*, 2019), and formation of a biofilm is an important pathogenic trait of EAEC, often associated with persistent infection. Genes involved include *aatA,*

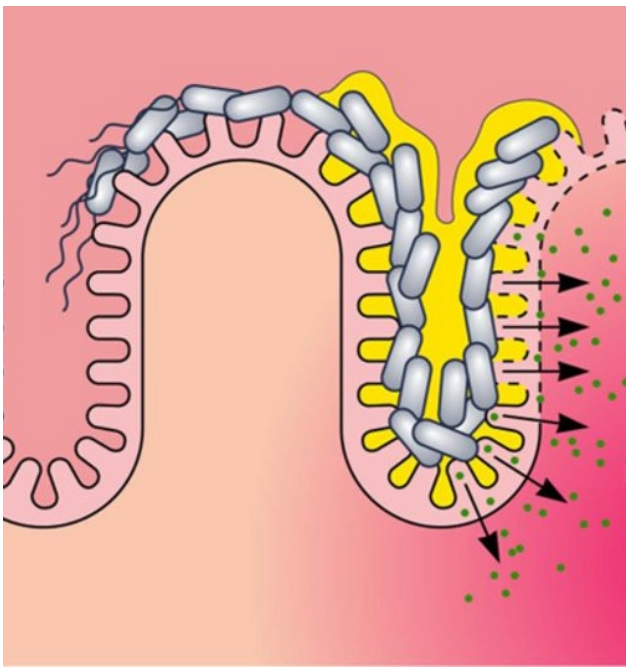


Figure 1.1 Pathogenesis of EAEC

The figure illustrates EAEC colonisation of the gut by aggregative adherence to intestinal epithelium, forming a biofilm (yellow), followed by the secretion of toxins (green dots). Adapted from Hebbelstrup Jensen *et al.* (2014).

aggR, pic, sepA, sigA and the AAFs (Jensen and Hebbelstrup Rye Rasmussen, 2017). Some of these, and many other virulence genes, are encoded on a 72-120 kb plasmid, pAA, indicating its central role in pathogenicity (Huttener *et al.*, 2018; Berger *et al.*, 2016; Jonsson *et al.*, 2017) and there is a high degree of conservation between the pAA of different EAEC strains (Savarino *et al.*, 1994). Some strains have pAA plasmids that also encode toxins, for example in EAEC strain 042 the autotransporters Pet and EAST-1, and SepA in a German O104:H4 outbreak strain (Jonsson *et al.*, 2017). In the prototypical EAEC strain 042, chromosomally encoded genes include ShET1, Pic and the type VI secretion system; the AAF genes, dispersin, EAST-1, Pet and AggR are encoded on pAA2, the pAA plasmid prototype (Morin *et al.*, 2010; Okeke *et al.*, 2000). Certain virulence genes encoded on pAA, particularly AggR-regulated AAF/I fimbriae, enhance inflammation, and allow adherence to epithelial cells (Boisen *et al.*, 2013; Huttener *et al.*, 2018; Berger *et al.*, 2016).

1.4.1 Attachment Adherence Fimbriae (AAF)

The AAF variant, AAF/I, which is expressed in EAEC strain 17-2, was found to be present in the 2011 German outbreak strain, O104:H4 (Elias *et al.*, 1999; Boisen *et al.*, 2013). In fact, the presence of pAA, which encodes the AAF genes and other important virulence factors, was shown to affect the probability of HUS occurrence; loss of pAA reduced the occurrence of HUS and high incidence of HUS was associated with pAA (Huttener *et al.*, 2018; Berger *et al.*, 2016). A major factor in this is likely to be the translocation of Shiga toxin (Stx2) across the intestinal epithelium, that promotes inflammation (Boisen *et al.*, 2013; Huttener *et al.*, 2018; Berger *et al.*, 2016).

AAFs are highly hydrophobic adhesins, required by many EAEC strains for their adherence to epithelial cells (Sheikh *et al.*, 2002; Nishi *et al.*, 2003; Savarino *et al.*, 1994), and the stacked-brick pattern allows distinction from EPEC and DAEC, which present localised adherence and diffuse adherence patterns (Jonsson *et al.*, 2017; Jensen and Hebbelstrup Rye Rasmussen, 2017; Berger *et al.*, 2016; Huttener *et al.*, 2018; Bamidele *et al.*, 2019; Dias *et al.*, 2020; Estrada-Garcia *et al.*, 2014). Five different allelic variants of AAF have been described (Harrington *et al.*, 2006).

The AAF/I and AAF/II fimbrial subunits are 25% identical but have greater homology with members of the Dr family of adhesins (Elias *et al.*, 1999; Harrington *et al.*, 2006), which is a family of fimbrial adhesins present in EPEC and DAEC (Harrington *et al.*, 2006), than with each other (Elias *et al.*, 1999). AAF variants also have different morphologies (Czeczulin *et al.*, 1997). Hence, AAF/I forms fimbriae 2-3 nm in diameter and arranged in flexible bundles (Elias *et al.*, 1999); whereas AAF/II, expressed in EAEC strain 042, forms semi-rigid bundles of filaments and that are 5 nm in diameter (Elias *et al.*, 1999; Czeczulin *et al.*, 1997). The two unlinked regions containing the AAF/I encoding genes, *agg*, are separated by 9 kb segment containing the 117-bp long EAST1 gene, *astA*, on a pAA virulence plasmid (Nataro *et al.*, 1994; Savarino *et al.*, 1994; Elias *et al.*, 1999; Menard and Dubreuil, 2002).

As there are some virulent EAEC strains that do not express any of the known AAF variants, it is expected that there are other adhesins still to be discovered (Ito *et al.*, 2014; Harrington *et al.*, 2006; Bhargava *et al.*, 2009).

1.4.2 Dispersin and the Aat secretion system

In most EAEC strains, on the virulence plasmid, there is a small open reading frame (ORF) upstream of the *aggR* gene that encodes the 10 kDa Aap protein, known as dispersin (Harrington *et al.*, 2006; Prager *et al.*, 2014; Fujiyama *et al.*, 2008). Virulence plasmids usually also encodes the ABC transporter complex, Aat, which is required for the export of dispersin (Prager *et al.*, 2014; Harrington *et al.*, 2006; Nishi *et al.*, 2003). Both dispersin and Aat are regulated by AggR (Harrington *et al.*, 2006).

Dispersin is secreted to the surface of EAEC cells, forming a capsule, and promoting dispersal of EAEC on intestinal mucosa by binding noncovalently to LPS of the outer membrane (Harrington *et al.*, 2006; Sheikh *et al.*, 2002; Nishi *et al.*, 2003). The predicted amino acid sequence of dispersin is more hydrophilic than AAF, and its presence is thought to neutralise the negative charge of LPS to allow AAF to spread and bind to distant sites (Sheikh *et al.*, 2002; Harrington *et al.*, 2006).

1.5 Bacterial transcription

The ability of many bacteria to adapt to environmental conditions is largely due to the global regulation and modulation of gene expression. Included in this response are disease-causing genes, notably virulence determinants that facilitate colonisation of the host and bacterial survival (Thomas and Wigneshweraraj, 2014).

Bacteria contain a single DNA-dependent RNA polymerase (RNAP) that transcribes DNA into RNA. The complete RNAP holoenzyme is formed by a core enzyme, with $\beta\beta'\alpha_2\omega$ subunits, interacting with a σ subunit, which provides the

key interactions for site-specific transcription initiation. The σ subunit is comprised of four independently folded domains and is responsible for promoter recognition and interaction: domain 4 interacts with the promoter -35 element and domain 2 with the -10 element (Browning and Busby, 2004; Mejia-Almonte *et al.*, 2020; Browning and Busby, 2016; Sutherland and Murakami, 2018). These elements are named according to the approximate base pair (bp) distance they occur upstream of the transcription start site (TSS). Following DNA-binding, RNAP unwinds the double helix strands to form an open complex, exposing the DNA strand to be transcribed (Browning and Busby, 2004; Browning and Busby, 2016). The frequency of transcription of a gene is positively correlated with the strength of its promoter. Strong promoters contain consensus sequence elements, TTGACA for the -35 element and TATAAT for the -10, however, most promoters have non-consensus sequences and transcription can still occur (Browning and Busby, 2004; Li and Zhang, 2014; Rhodius and Mutalik, 2010). This is often due to activators, a type of transcription factor, which improve the RNAP affinity for the promoter.

1.5.1 RNA polymerase

The discovery of RNAP, over 60 years ago, was followed ten years later by the discovery of the σ subunit, this has since expanded to a more detailed understanding of the structure and function of each subunit of the holoenzyme (Figure 1.2) (Ebright *et al.*, 2019).

The α subunit, encoded by the *rpoA* gene, is the second smallest in RNAP and its primary function is to act as a scaffold for assembly of the β and β' subunits. The subunit exists as a homodimer with two identical subunits, each composed

of two independently folded domains separated by an unstructured flexible linker (Sutherland and Murakami, 2018; Browning and Busby, 2004; Lee *et al.*, 2012), which allows flexibility in positioning RNAP at promoters (Lee *et al.*, 2012). The αNTD (α-amino terminal domain) is further separated into two parts, one of which is involved in dimerization with the other α-subunit, forming a highly conserved hydrophobic core (Sutherland and Murakami, 2018). The αCTDs (α-carboxyl terminal domain) are not required for assembly of RNAP or for basal level transcription, in fact, these domains interact with the UP element at strong promoters, adjacent to the σ subunit domain 4 interacting with the -35 hexamer (Sutherland and Murakami, 2018; Lee *et al.*, 2012).

The RNAP β and β' subunits contain the enzyme active site, responsible for transcript initiation, elongation, and termination. It is made up of β and β' determinants that contact template DNA, incoming NTPs and RNA products during transcription (Browning and Busby, 2004; Opalka *et al.*, 2000). The β and β' subunits are encoded in the *rpoB* and *rpoC* genes and make up the majority of the RNAP mass (Sutherland and Murakami, 2018; Naryshkina *et al.*, 2001). These subunits contain highly conserved regions, similar to human RNAP subunits, that are separated by non-conserved spacer regions (Dieci *et al.*, 1995; Opalka *et al.*, 2000; Naryshkina *et al.*, 2001). The crab claw shape of RNAP, creates a cleft for template DNA to enter into the active site. The cleft is formed by the β and β' subunits, each making up one of the arms of the crab claw (Sutherland and Murakami, 2018).

The RNAP β and β' subunits interact with the σ subunit. The β'-subunit has a high affinity interaction with σ domain 2 (Nickels *et al.*, 2005). The β flap domain, which

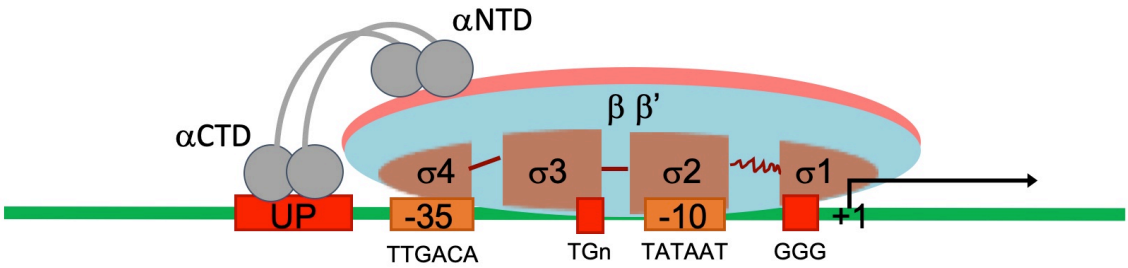


Figure 1.2 Schematic diagram of RNAP holoenzyme bound at a promoter

The figure illustrates the RNAP holoenzyme bound to a promoter, with subunits bound at promoter elements: the αCTDs contact the UP-element, σ region 1 interacts with the discriminator, σ region 2 contacts the -10 hexamer (TATAAT consensus sequence), σ region 3 contacts the extended -10 element (TGn) and σ region 4 contacts the -35 hexamer (TTGACA consensus). Adapted from Browning and Busby (2004).

blocks the RNA exit pathway from the active site, is also required to contact and position the σ-subunit domain 4 at promoter DNA for sequence-specific contacts with the -35 element (Sutherland and Murakami, 2018; Nickels *et al.*, 2005).

The ω-subunit, encoded by the *rpoZ* gene, is the smallest subunit of RNAP and is not directly involved in transcription. The ω-subunit is thought to act as a chaperone to assist in β' folding, additionally, it has a role in maintaining and protecting the catalytic core of the active site by binding to the β' domain (Sutherland and Murakami, 2018; Browning and Busby, 2004).

The holoenzyme is formed of RNAP together with a single σ factor, which is required for transcription initiation, as it carries the major determinants for the specificity promoter recognition. Though there are seven types of σ factor in *E. coli*, σ^{70} is the most abundant and has the highest affinity for core RNAP (Lee *et al.*, 2012; Sutherland and Murakami, 2018). Housekeeping genes are expressed by RNAP holoenzyme with σ^{70}, this σ^{70} factor is comprised of four independently folding domains, each with distinct functions. For example, domain 2 recognises single stranded DNA (ssDNA) of the non-template strand of the -10 element, and domain 4 recognises the -35 element as double stranded DNA (dsDNA) (Sutherland and Murakami, 2018).

1.5.2 Promoters

Promoter recognition is a major step in transcription initiation: sequence-specific interactions of the RNAP with core promoter elements ensure the open complex can be formed, and subsequent steps of transcription can then occur (Browning and Busby, 2004; Winkelman *et al.*, 2016). RNAP binding specificity relies on core sequence elements: the discriminator, -10 hexamer, extended -10, -35

hexamer and UP element (Browning and Busby, 2004; Lee *et al.*, 2012). Each promoter element contributes to this specificity to different extents, particularly as the elements differ between promoters, and they can compensate for deficiencies due to other non-consensus elements (Browning and Busby, 2004; Lee *et al.*, 2012). In fact, there is not naturally occurring "perfect" promoter, composed completely of consensus elements, as RNAP would bind too tightly (Browning and Busby, 2004). At promoters, the holoenzyme σ region 2.4 ($\sigma_{2.4}$) recognises the -10 element, which is located 10 bp upstream from the TSS, and σ region 4.2 ($\sigma_{4.2}$) recognises the -35 element, located 35 bp upstream from the TSS (Browning and Busby, 2004; Yang and Lewis, 2010). However, the remaining promoter elements are also important for promoter recognition. The discriminator element, at positions -4 to -6, is also recognised by $\sigma_{1.2}$ (Lee *et al.*, 2012; Davis *et al.*, 2017; Forquet *et al.*, 2021). Directly upstream of the -10 hexamer is a 3-4 bp motif, called the extended -10, which contacts domain 3 of the σ subunit (Browning and Busby, 2004; Lee *et al.*, 2012). The UP element, located approximately 20 bp upstream of the -35 element, is contacted by the RNAP αCTDs (Browning and Busby, 2004; Lee *et al.*, 2012).

1.5.3 Open complex formation and transcript initiation

There are three sequential steps in gene transcription: promoter recognition and transcription initiation, elongation of the RNA transcript, and transcript termination (Yang and Lewis, 2010). The first step is the most highly regulated (Kapanidis *et al.*, 2006). For promoter recognition, the RNAP holoenzyme scans DNA for specific promoter sequences (Browning and Busby, 2004; Yang and Lewis, 2010). Then, the catalytically competent open complex, RPo, is formed with an

~13 bp segment of the promoter surrounding the TSS unwound, forming an open "transcription bubble" (Kapanidis *et al.*, 2006; Winkelman *et al.*, 2016; Mazumder *et al.*, 2021; Browning and Busby, 2004). RNAP $\sigma_{2.3}$ binds to the non-template strand of the bubble, termed the core recognition element (CRE) formed of positions -4 to +2; this interaction causes the free template strand of the bubble to move into the active site for RNA synthesis, guided by the CRE (Browning and Busby, 2004; Zhang *et al.*, 2012; Petushkov *et al.*, 2015). The nucleotide of the single-stranded template-strand aligned with the active site will be the location of the transcript 5' end (Lee *et al.*, 2012; Winkelman *et al.*, 2016; Mazumder *et al.*, 2021). The TSS is generally located 7 bp downstream of the -10 element, however, this can vary to between 6-10 bp downstream (Winkelman *et al.*, 2016). The RPo is competent for transcription if the appropriate nucleoside triphosphates are present (Lee *et al.*, 2012).

Formation of the transcription bubble starts when DNA base pairs at the upstream end of the promoter -10 element are broken, and the most highly conserved base of the -10 element, A_{-11}, flips into the domain 2 σ-recognition pocket (Mazumder *et al.*, 2021; Bae *et al.*, 2015). The proposed mechanism for DNA unwinding and loading into the *E. coli* RNAP-σ^{70} holoenzyme active centre is 'bind-unwind-load-and-lock': the RNAP β and β' subunit clamp is closed when promoter DNA unwinds in the upstream-downstream direction and the clamp closes further to lock the ssDNA in the active centre (Mazumder *et al.*, 2021). The dsDNA and ssDNA junction at the upstream end of the bubble is maintained by an interaction between duplex T_{-12} and σ (Sutherland and Murakami, 2018; Bae *et al.*, 2015).

Typically, as the RNAP-promoter initial transcribing complex (RPitc) begins RNA synthesis, using nucleoside triphosphate substrates, there are several rounds of abortive synthesis of short, ~9-15 nt, RNA products (Kapanidis *et al.*, 2006; Bae *et al.*, 2015). DNA scrunching, which is proposed to be the driving force for promoter escape, is required for abortive synthesis: the upstream trailing-edge of RNAP is stationary relative to DNA and the downstream DNA is drawn into RNAP, this creates a stressed intermediate state due to the unwinding and compaction of DNA (Wade and Struhl, 2008; Winkelman *et al.*, 2016).

1.5.4 Transcription elongation and termination

After promoter recognition and initiation, the RNAP transitions into the elongation complex, and, following promoter escape, extends the RNA chain (Browning and Busby, 2004; Wade and Struhl, 2008). As soon as ~12 nt transcript has formed, the upstream exit channel is filled, and the σ region 4 and β-flap interaction is destabilised. This disruption is the first step in σ dissociation, the -35 element is released, and this allows the escape of RNAP from the promoter and the start of the elongation phase (Bae *et al.*, 2015; Yang and Lewis, 2010).

There are two steps in elongation: first, the nucleoside triphosphate (NTP) binds to the transcription elongation complex (TEC) resulting in a conformational change to the trigger loop of the RNAP catalytic core; then, the NTP is incorporated, and pyrophosphate is released. The rate of elongation is not uniform in *E. coli* as synthesis of RNA includes pauses and backtracking: pauses can be resolved spontaneously or lead to backtracking. GreA and GreB factors restart backtracked TEC by binding in the channel and cleaving the 3’-end of the

nascent transcript in the active site (Washburn and Gottesman, 2015; Ray-Soni *et al.*, 2016).

There are three modes of transcription termination, the final step of gene expression; these are intrinsic, Rho-dependent and Mfd-dependent (Roberts, 2019).

Rho-dependent and Mfd-dependent termination are reliant on the activity of ATP and the RNA translocase, Rho, or the DNA translocase, Mfd, respectively (Roberts, 2019).

For Rho-dependent termination, Rho preferentially binds to cytosine-rich sequence of the nascent transcript following RNA transcript emerging from the TEC exit channel and the RNA is fed through the central pore, which triggers translocase activity. The RNAP dissociates after all of the RNA has been translocated through Rho (Ray-Soni *et al.*, 2016). Mfd termination is triggered in response to random DNA damage and usually results in premature termination and no useful transcript. This mode of termination provides DNA repair via the transcription-coupled repair factor (TCRF, aka Mfd). Mfd recognises TECs paused due to DNA damage and induces TEC dissociation and subsequent DNA repair (Shi *et al.*, 2020; Ray-Soni *et al.*, 2016).

Intrinsic, or Rho-independent, termination depends on the interaction elements in the most recently synthesised RNA with RNAP as they pass through the exit channel (Roberts, 2019). This mode of termination is mediated by RNA hairpins, which form after transcription of palindromic guanine-cytosine rich template DNA that is followed by a run of adenines (Roberts, 2019; Ray-Soni *et al.*, 2016; Sutherland and Murakami, 2018; Washburn and Gottesman, 2015). Pausing

occurs, which allows sufficient time for the terminator hairpin to form in the RNAP exit channel (Ray-Soni *et al.*, 2016). The NusA protein contributes to the efficiency of termination by contacting both the flap tip and pause hairpin, which are required to interact for hairpin-stabilised pausing (Roberts, 2019; Ray-Soni *et al.*, 2016). At the 7^{th} or 8^{th} uridine nucleotide at the end of the released RNA, synthesis stops, and the TEC dissociates. The timing of hairpin formation is critical and must be coordinated with transcription of the termination point (Washburn and Gottesman, 2015; Roberts, 2019). The efficiency of this type of termination depends on the length of the hairpin loop and the sequence and size of the stem (Washburn and Gottesman, 2015).

1.6 Transcription regulation

Transcription factors play a big role in bacterial responses to changes in the environment, by coupling gene expression with external signals to maintain bacterial cell fitness. Each transcription factor may control just one or two genes, or a large number of genes, and the regulation of a transcription factor itself involves either its expression or its activity (Browning and Busby, 2004; Browning and Busby, 2016). Transcription factors are generally composed of two domains: one for sensing signals, through effector binding of a small protein or ligand, or due to abundance of the transcription factor due to synthesis, turnover and sequestering; and one for interacting with DNA promoter targets, most commonly a helix-turn-helix (HTH) motif (Ishihama, 2012; Martinez-Antonio and Collado-Vides, 2003; Browning and Busby, 2016).

Transcription initiation can be regulated by transcription factors, positively or negatively, with transcription factors binding to specific promoters, or groups of

promoters, to recruit or block RNAP binding to DNA respectively (Browning and Busby, 2004; Chen *et al.*, 2021). Activators function via different mechanisms to improve the ability of RNAP to recognise and bind to promoters to initiate transcription. Conversely, repressors work by blocking or disrupting promoters to prevent RNAP binding and transcription initiation. Most promoters are regulated by more than one transcription factor and the most common mechanism to confer transcription factor specificity to a promoter is dimerization, which includes homodimerization, or multimerization (Browning and Busby, 2004; Browning and Busby, 2016).

1.6.1 Simple activation

Different mechanisms have evolved to facilitate activation of transcript initiation. Class I activators bind upstream of the promoter -35 element and recruit RNAP to the promoter by binding to the αCTD (Figure 1.3A) (Browning and Busby, 2004; Browning and Busby, 2016). Class II activators have a DNA-binding site that overlaps the promoter -35 element, leading to activator-dependent recruitment of RNAP by interaction with σ region 4. Note that, sometimes, Class II activators can also contact the RNAP αCTDs (Figure 1.3B) (Browning and Busby, 2004; Browning and Busby, 2016). The *E. coli* cyclic AMP receptor protein, CRP, has been shown to function as both a Class I and Class II activator; for example, it is a Class I activator at the *lac* promoter but a Class II activator at the *galP1* promoter (Browning and Busby, 2004; Lee *et al.*, 2012; Rhodius *et al.*, 1997).

A third class of activator functions by binding at or near promoter elements, that are spaced non-optimally. Activator binding causes a conformational change in the promoter DNA that allows RNAP to access the now-correctly positioned

promoter elements (Figure 1.3C). An example of this is found with the MerR-family of transcriptional regulators (Browning and Busby, 2004; Brown *et al.*, 2003; Browning and Busby, 2016).

Some promoters have developed to be co-dependent on two activators, because the individual activators are not able to activate alone (Lee *et al.*, 2012). An example of this occurs in EAEC strain 042 where the *pet* promoter is co-dependent on both Fis and CRP. Although CRP binds to a target that overlaps the promoter -35 element, it is located sub-optimally, and the upstream binding of Fis offsets this deficiency (Lee *et al.*, 2012; Rossiter *et al.*, 2011).

For a select group of activators, another mechanism has been characterised called 'pre-recruitment' or 'DNA scanning. AraC-family transcriptional regulators SoxS and MarA have been shown to function via this method: the transcription factor interacts with the σ-subunit and then the activator-RNAP complex scans the DNA for a dependent promoter (Zafar *et al.*, 2011; Lee *et al.*, 2012; Taliaferro *et al.*, 2012).

1.6.2 Repression by transcription factors

Repressors are another type of transcription factor that interact with promoters to prevent RNAP binding, thereby inhibiting transcription initiation (Browning and Busby, 2004; Browning and Busby, 2016). There are three main mechanisms for repression, the simplest of which is steric hindrance in which the repressor binding-site overlaps with a key promoter element to block RNAP access and binding (Figure 1.4A). An example of repression by steric hindrance is the binding of LacI to the *lac* promoter (Rojo, 1999).

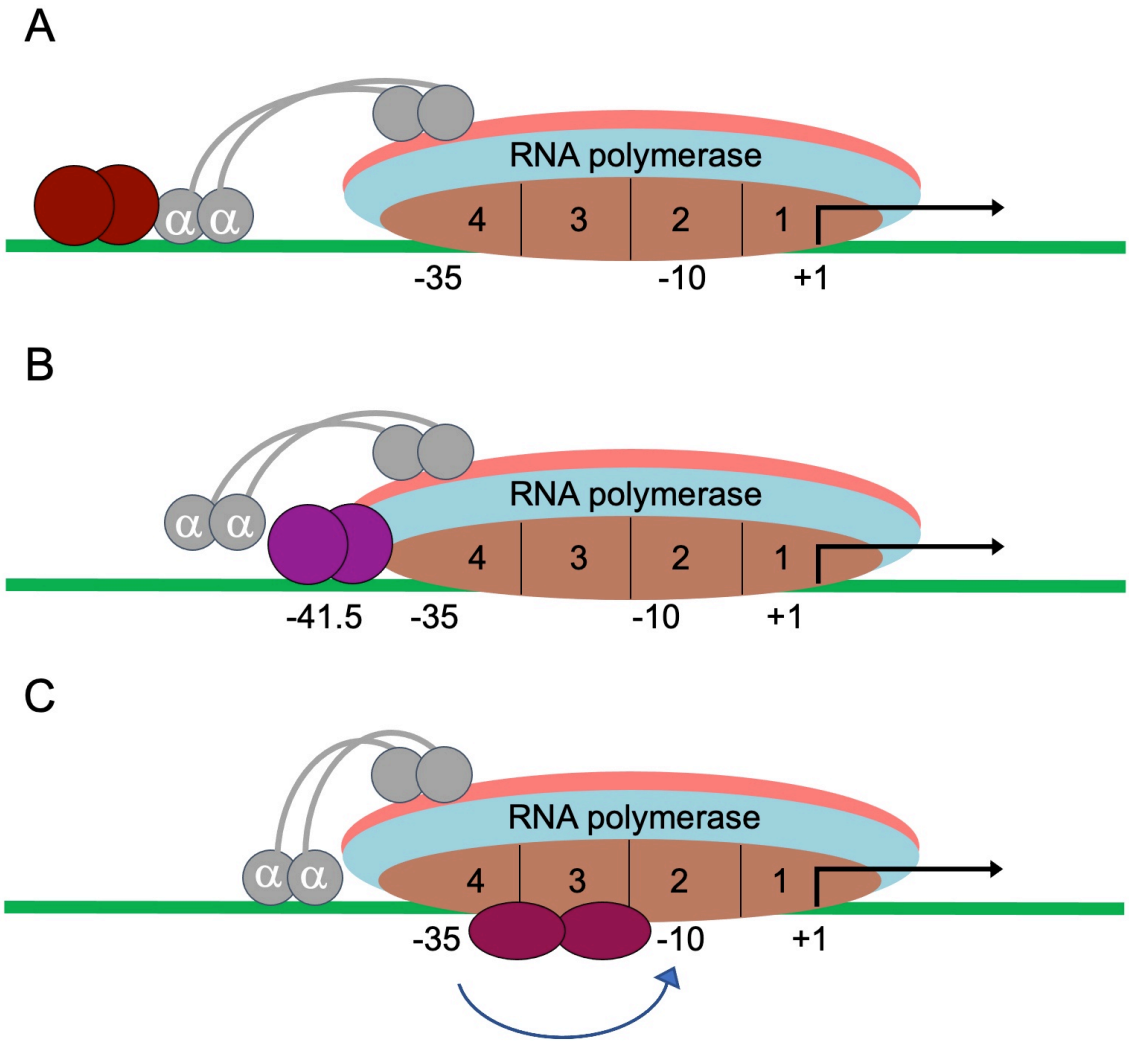


Figure 1.3 Mechanisms of activation of transcription initiation

A. Class I: The figure illustrates the binding of an activator (brown) to a site upstream of the -35 hexamer and its interaction with the αCTD of RNAP to facilitate transcription.
B. Class II: The figure shows the binding of an activator (purple) at a position overlapping the -35 hexamer of the promoter and its interaction with σ region 4 to promoter transcription.
C. Activation by conformational change: the figure shows the binding of an activator (red) to the promoter at a position between the -10 and -35 elements, which causes a conformational change to facilitate transcription.

Adapted from Browning and Busby (2004).

Repression can also occur by DNA looping where repressors bind both upstream and downstream of a promoter, which causes looping that prevents RNAP binding and transcript initiation (Figure 1.4B). An example of this is the action of GalR protein at the *galETK* promoter (Choy and Adhya, 1992).

The third mechanism involves repressors that bind to the activator to prevent interaction with RNAP (Figure 1.4C) (Browning and Busby, 2004; Browning and Busby, 2016). An example of this 'anti-activator' mechanism is the repression of some CRP-activated promoters by CytR that binds directly to CRP to prevent CRP-dependent activation (Shin *et al.*, 2001).

1.7 Overview of the AraC-family of transcription factors and AggR

AggR is the master regulator of EAEC virulence. It belongs to a subgroup of the AraC/XylS family of transcription factors, with sequence similarity to Rns, CfaD and CfaR from ETEC, and VirF from *Shigella flexneri* (Santiago *et al.*, 2014; Yasir *et al.*, 2019; Porter and Dorman, 2002; Munson *et al.*, 2001). Most bacterial transcription factors can be categorised into a family, due to sequence similarities, particularly in the DNA-binding domain (Pis Diez *et al.*, 2022); the AraC family is one of the best characterised (Browning and Busby, 2004). AraC/XylS transcription regulators can be plasmid or chromosomally encoded and the G+C content varies (Gallegos *et al.*, 1997). The AraC/XylS family of transcriptional regulators is one of the most widespread, especially amongst positive regulators (Gallegos *et al.*, 1997). The AraC/XylS transcription factors regulate a diverse range of biological processes, including metabolism, stress

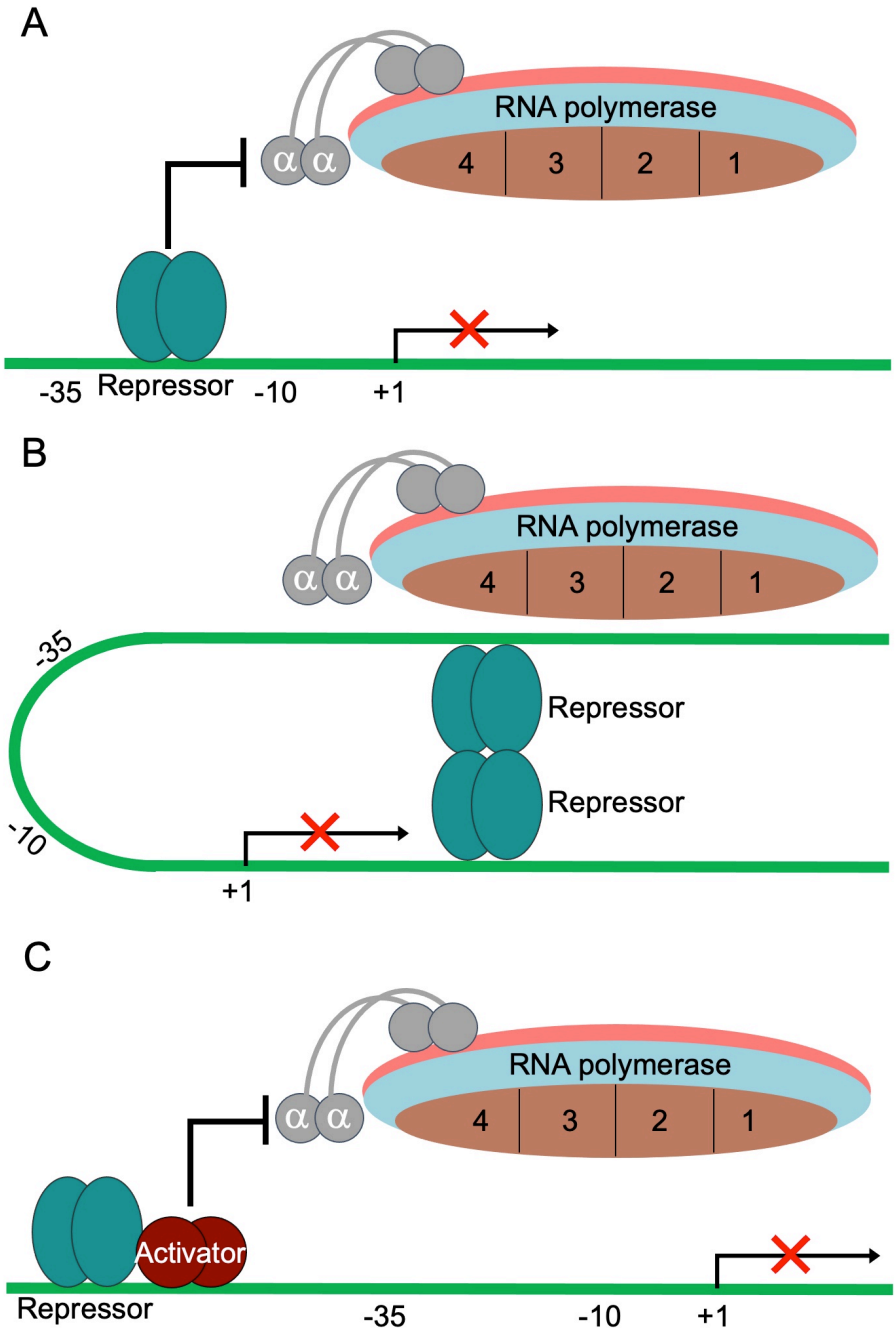


Figure 1.4 Mechanisms of repression of transcription initiation

A. Steric hindrance: a repressor (blue) binds to the promoter between promoter elements key for promoter recognition.
B. Formation of a loop: repressors (blue) binds upstream and downstream of the promoter which causes a loop to form and prevent RNAP binding.
C. Anti-activators: a repressor (blue) binds to an activator (brown) to prevent recruitment of RNAP.

Adapted from Browning and Busby (2004).

response and virulence (Ibarra *et al.*, 2008; Gallegos *et al.*, 1997; Cortes-Avalos *et al.*, 2021). Some AraC/XylS family transcription factors support bacterial colonisation of the gastrointestinal tract, urinary system or respiratory tract by stimulating synthesis of proteins that facilitate adhesion to epithelial cells, including fimbriae (e.g. AggR, Rns, CfaD) and invasins (e.g. VirF). Some transcription factors respond to stressors e.g. oxidative stress, antibiotics, and transition between growth stages. Some transcription factors cross-regulate specific genes and can only regulate targets when overproduced (e.g. SoxS, MarA and Rob) (Gallegos *et al.*, 1997).

AraC itself, the 'founder' member of the family, regulates metabolism through controlling the transport and degradation of arabinose. AraC was the first AraC/XylS member to be purified, characterised and identified as the regulator of the L-arabinose operon in *E. coli* (Gallegos *et al.*, 1997). Three gene products are required for L-arabinose catabolism, encoded by *araB, araA,* and *araD* (Schleif, 2010), and expression of these genes is controlled by AraC (Hamilton and Lee, 1988). The *araBAD* operon is under the control of the *ara* promoter p_{BAD}, which is regulated by the activators CRP and AraC (Schleif, 2010; Hamilton and Lee, 1988; Dhiman and Schleif, 2000; Ammar *et al.*, 2018). The presence of arabinose is required to induce AraC binding to the promoter site overlapping the -35 hexamer and subsequent transcript initiation (Ammar *et al.*, 2018; Hamilton and Lee, 1988; Dhiman and Schleif, 2000).

Most AraC/XylS transcription factors have 250-300 amino acids and, typically, have two domains: a C-terminal DNA-binding domain (DBD) and an N-terminal multimerization or effector binding domain (EBD). The DBD is approximately 100

residues and has a highly conserved tertiary structure: two HTH DNA binding motifs, each containing three α-helices, and connected by a single longer α-helix (Ibarra *et al.*, 2008; Schuller *et al.*, 2012; Egan, 2002). These HTH motifs interact with DNA bases displayed in two adjacent major grooves on the same face of the DNA helix. The C-terminal HTH is mostly conserved and improves DNA binding affinity, the N-terminal HTH is more variable, its function is thought to be responsible for individual binding specificities (Schuller *et al.*, 2012). In contrast, the EBD is variable in different AraC family members, consisting of 100-200 residues. Its function in many AraC family transcription factors is not known, though some use it for both dimerization and effector binding (Ibarra *et al.*, 2008; Schuller *et al.*, 2012; Egan, 2002). The NTD is required for the function of the C-terminal-DBD in many AraC-family transcription factors, presumably to stabilise the active conformation. An example of this is MelR, which is regulated by its interaction with melibiose (Kahramanoglou *et al.*, 2006). However, some AraC transcription factors are comprised of only the C-terminal-DBD, for example MarA and SoxS (Grainger *et al.*, 2004; Lee *et al.*, 2012; Schuller *et al.*, 2012).

1.7.1 An overview of the EAEC AggR regulon

AggR is considered to be a vital element of EAEC pathogenicity. AggR autoactivates its own expression and has more than 44 virulence genes in its regulon (Morin *et al.*, 2010; Morin *et al.*, 2013). Regulation of AggR activity occurs at the transcriptional level, of *aggR* gene expression, and also at the post-transcriptional level, by the action of Aar (AggR-activated regulator) (Santiago *et al.*, 2017; Huttener *et al.*, 2018; Dias *et al.*, 2020). 'Typical' EAEC strains carry this 'master' regulator and are considered more virulent than atypical strains,

though virulence genes are also present in atypical isolates (Andrade *et al.*, 2017; Yasir *et al.*, 2019; Morin *et al.*, 2013; Jensen and Hebbelstrup Rye Rasmussen, 2017; Jenkins *et al.*, 2005; Boisen *et al.*, 2020). AggR is closely related to the fimbrial activator Rns and the target DNA-binding sites are similar (Yasir *et al.*, 2019; Morin *et al.*, 2013). The consensus AggR DNA-binding site (sequence: AWWWWWWWTATC, where W=A/T) is located 21-23 bp upstream from the -10 region at target promoters, and overlaps with the -35 element, suggesting that AggR might activate via a Class II activation mechanism (Figure 1.5) (Yasir *et al.*, 2019; Abdelwahab *et al.*, 2021). Many AraC/XylS transcription factors activate transcription by assisting RNAP recruitment and interacting with domain 4 of the σ-subunit (Cortes-Avalos *et al.*, 2021).

AggR activates genes responsible for AAF/II fimbriae synthesis (Huttener *et al.*, 2018) and the *aafD* genes, which encode its chaperone protein (Yasir *et al.*, 2019). Previously, the presence of *aggR* was used to identify EAEC isolates but now *aatA* is the target (Dias *et al.*, 2020; Bamidele *et al.*, 2019). *aatA* is an important AggR-activated EAEC gene as it encodes a T1SS (type I secretion system) which secretes dispersin and is therefore vital to pathogenicity (Harrington *et al.*, 2006; Jensen and Hebbelstrup Rye Rasmussen, 2017; Bamidele *et al.*, 2019; Dias *et al.*, 2020). The Aai (AggR Activated Island) type VI secretion system (T6SS), encoded in the *pheU* pathogenicity island (PAI-I, *aaiA-Y*), is also regulated by AggR (Bamidele *et al.*, 2019; Jonsson *et al.*, 2017; Morin *et al.*, 2010; Dias *et al.*, 2020; Navarro-Garcia *et al.*, 2019). T6SS has been shown to contribute to pathogenicity and has been implicated in biofilm formation and antibacterial activity (Navarro-Garcia *et al.*, 2019).

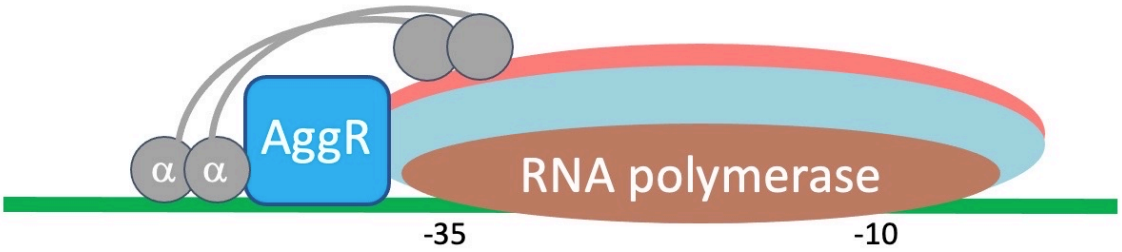


Figure 1.5 Mechanism of AggR-mediated activation

The figure shows the proposed mechanism of AggR-mediated activation, AggR binds at a target promoter to a site overlapping the -35 hexamer and interacts with the RNAP αCTD and σ subunits. AggR recruits RNAP to the promoter by contacting σ region 4.
Adapted from Browning and Busby (2004).

1.8 An overview of the Aar regulon in EAEC

AggR functions via a feed-forward loop whereby it autoactivates its own expression. However, AggR also activates the expression of its repressor, Aar. Aar is encoded by *aar*, which is one of the most upregulated genes in the AggR regulon (Santiago *et al.*, 2014; Boisen *et al.*, 2020). Aar was the first characterised member of the AraC Negative Regulator (ANR) family, which is a large family of bacterial gene regulators expressed in clinically important Gram-negative pathogens (Santiago *et al.*, 2017). ANRs act as anti-activators to modulate the activity of certain AraC family members to precisely adjust virulence gene expression levels in particular environments (Hodson *et al.*, 2017; Santiago *et al.*, 2016). ANRs can prevent DNA binding and dimerization of cognate master-regulators by binding to the central linker domain (Santiago *et al.*, 2016; Hodson *et al.*, 2017). Common structural features include a low molecular mass (<10 kDa) and highly conserved residues with a predicted secondary structure of three highly conserved α-helix motifs spanning the protein (Santiago *et al.*, 2016). In EAEC 042, Aar was identified for its ability to repress AggR activity, though a DNA-binding domain has not been identified in Aar, therefore, it is thought that Aar binds directly to AggR to inhibit its ability to activate genes in its regulon (Santiago *et al.*, 2017; Yasir *et al.*, 2019). Aar has been shown to bind with high affinity to the central linker domain of AraC-like members to prevent DNA binding activity (Santiago *et al.*, 2017; Santiago *et al.*, 2016; Hodson *et al.*, 2017). Consistent with this, EAEC strains from Mali and Brazil without *aar* were found to have increased pathogenicity (Santiago *et al.*, 2014).

Aar is thought to have another role by binding to H-NS, which is known to repress some AggR activities (Santiago *et al.*, 2017). Recall that H-NS is a nucleoid-associated protein that silences AggR expression, and the AggR regulon, by binding to AT-rich DNA sequences, effectively blocking the promoter (Santiago *et al.*, 2017; Morin *et al.*, 2010; Berger *et al.*, 2016). Hence, interaction with H-NS is an important regulatory function of ANR family members (Santiago *et al.*, 2017). It has been proposed that, when the Aar concentration is low, Aar relieves H-NS repression, which allows AggR upregulation at certain AT-rich promoters. Conversely, when Aar concentration is high, Aar binds to AggR and drastically reduces AggR upregulation, though Aar still removes the H-NS repression at high or low Aar concentration (Mickey and Nataro, 2020). Another nucleoid-associated protein, FIS, a global regulatory protein factor, may also enhance AggR expression (Huttener *et al.*, 2018; Morin *et al.*, 2010; Berger *et al.*, 2016).

The activity of AggR is regulated by its level, due to AggR-dependent activation of the *aggR* promoter, and by Aar. AggR has both the variable NTD and conserved CTD characteristic of most AraC-family transcriptional regulators, and it has been determined that AggR functions as a monomer, thus it is assumed that the NTD is a receptor for a ligand. However, in EAEC, there is no evidence of a ligand, thus, the suggestion is that AggR is regulated by a feedforward loop with a bistable switch (Midgett *et al.*, 2021; Morin *et al.*, 2010; Yasir *et al.*, 2019; Abdelwahab *et al.*, 2021). This has led us to a model (see Chapter 6 and Chapter 7) that conflicts with the Jacob and Monod principle that transcriptional control circuits involving transcription factors are deterministic (Jacob and Monod, 1961). Thus, we suggest that regulation is dependent on a stochastic switch, existing in

two states: ‘on’ and ‘off’ with switch flipping triggered by AggR production reaching a critical threshold (Moxon, 2009; Palmer *et al.*, 2013; Abdelwahab *et al.*, 2021).

1.9 Aims and objectives of the project

Muhammad Yasir investigated AggR from EAEC strain 042 and several AggR-dependent promoters, primarily from strain 042. He found that these promoters had a ‘simple’ organisation, consisting of a single DNA binding site for AggR (AWWWWWWTATC, where W=A/T) located 21-23 bp upstream of a -10 hexamer element. Thus, Muhammad concluded that AggR functions as a monomer by Class II activation (Figure 1.5) (Yasir *et al.*, 2019). However, there are more complex AggR-dependent promoters that may not follow this ‘simple’ activation mechanism. Furthermore, the *aggR* gene from the prototypical strain 042 is not present in all pathogenic strains of EAEC. Thus, my aims for this project were:

- To investigate more complex AggR-regulated promoters.
- To determine possible variations in the organisation of AggR-regulated promoters between strains.
- To investigate the function of transcription factors related to AggR.
- To define key determinants for AggR activity and understand how activity is regulated.

To do this, simple and complex promoters from different EAEC strains were studied, concentrating on the *yicS-nlpA* bi-directional promoter (Chapter 3). The results give insight into how genes have evolved to join or leave the AggR-regulon. In Chapter 4, I studied AggR-activated promoters from different EAEC

strains and discovered an AggR-like transcription factor that has a subtly different activation profile, suggesting that, through evolution, *aggR* drifts like its target promoters. In Chapter 5, I outlined a detailed mutational analysis of AggR from EAEC 042 to define key determinants for its activity, and I found that these are present throughout the NTD and CTD. A sub-objective in this study was to find a strongly *trans*-dominant *aggR* allele which might be utilised in therapy. However, this failed, suggesting pre-recruitment as a mechanism for the activation of AggR at target promoters. Finally, I investigated a possible stochastic model for the regulation of AggR activity (Chapter 6). Current experimental data suggests that the NTD is not required for ligand binding or multimerization but, in fact, it drives the CTD into an active conformation. Based on the ensemble of experimental evidence, I argue that AggR-activity is regulated stochastically by a bi-stable switch that is controlled by a feedforward loop with dampening.

Chapter 2 Materials and Methods

2.1 Suppliers

The chemicals used in this study were purchased from Fisher Scientific UK, Sigma-Aldrich, New England Biolabs (NEB) and Bioline, unless stated otherwise. Accuzyme DNA polymerase, MyTaq red DNA polymerase and dNTPs were purchased from Bioline. Q5® Site-Directed mutagenesis kit, Phusion DNA polymerase, Q5 DNA polymerase, T4 DNA ligase, calf intestinal alkaline phosphatase (CIP) and the restriction enzymes used in this study were purchased from New England Biolabs (NEB). All the deoxy-oligonucleotides were ordered from Merck Life Science UK Limited.

For solutions that required sterilising, 0.2 µm pore filters (Acrodisc®) were used for filter-sterilization and for autoclaving, the conditions were 121˚C under 15 lbin^{-2} for 15-20 minutes.

2.2 Bacterial Growth Media

2.2.1 Liquid media

All strains were grown in Luria Bertani (LB) broth with shaking at 37˚C unless stated otherwise. To prepare LB, 20 g of LB powder (Lennox) (composition: 5 g/L NaCl, 10 g/L tryptone and 5 g/L yeast extract) was dissolved in dH_20, made up to 1 litre and autoclaved. For selection of plasmid-encoded antibiotic resistances, medium was supplemented with the appropriate antibiotic. Super Optimal broth with Catabolite repression (SOC) medium was prepared to a final volume of 1 litre with distilled water and contained 10 mL KCl (250 mM), 5 mL $MgCl_2$ (2 M),

20 g tryptone, 5 g yeast extract and 0.5 g NaCl, 20 mL glucose was added after autoclaving. High glucose concentration (4.5 gl^{-1}) Dulbecco's modified Eagle medium (DMEM) was purchased from Sigma Aldrich in liquid form.

2.2.2 Solid media

E. coli strains were grown on LB agar (Sigma-Aldrich) or MacConkey agar (BD Difco) plates. For LB agar, 35 g of LB agar powder (composed of 15 g/L agar, 5 g/L NaCl, 10 g/L tryptone and 5 g/L yeast extract) was dissolved in dH_20, made to 1 litre and autoclaved. For MacConkey lactose agar, 50 g of MacConkey powder (composed of 13 g agar, 3 g peptones (meat and casein), 17 g pancreatic digest of gelatin, 5 g NaCl, 10 g lactose, 1.5 g bile salts (no. 3), 0.03 g neutral red and 0.001 g crystal violet in one litre of dH_20) was dissolved in dH_20, made up to 1 litre and autoclaved. According to manufacturer's instructions, media were prepared using distilled water and autoclaved. Appropriate antibiotics were added to autoclaved medium when they had cooled to ~55˚C.

2.2.3 Antibiotics and other supplements

Strains with plasmid-encoded antibiotic resistance genes were selected for by adding appropriate antibiotics to the media – solid and liquid – after autoclaving. Stock solutions of antibiotics were prepared by dissolving powder in the following liquids, 10 mg/mL of ampicillin in dH_2O, 50 mg/mL of kanamycin in dH_2O, and 25 mg/mL of chloramphenicol in ethanol (100%); followed by filter sterilization. The 10 mg/mL tetracycline stock solution was prepared in ethanol (50% v/v) and vortexed. Ampicillin and kanamycin were stored at 4˚C in a refrigerator, chloramphenicol and tetracycline were stored at -20°C. The final concentration of

antibiotics used was 100 µg/mL ampicillin, 25 µg/mL kanamycin, 50 µg/mL chloramphenicol, and 15 µg/mL tetracycline.

Arabinose, to induce expression of protein using the pBAD expression vector, was used at a concentration of 0.2%, with a 20% (w/v) stock solution in dH_2O, prepared, filter sterilized and stored at 4˚C.

2.3 Bacterial Strains and plasmids

2.3.1 Bacterial strains and growth conditions

E. coli K-12 strain BW25113 was used throughout this study, its genotype and the other bacterial strains used in this study are detailed in Table 2.1.

For overnight cultures, one bacterial colony was used to inoculate 5 mL of LB liquid medium supplemented with appropriate antibiotics. Cultures were incubated with shaking at 37˚C for 16-18 hours.

For sub-cultures, overnight cultures were diluted 1:100 into appropriate medium. The optical density of 600 nm (OD_{600}) was measured using a Helios Gamma Spectrophotometer (Thermo Fisher Scientific Inc.) to monitor growth of bacterial cultures.

Glycerol stocks were prepared for long-term storage of bacterial strains by mixing 500 µL of overnight culture with 500 µL of 50% glycerol stock in a cryofuge tube for storage at -80˚C.

2.3.2 Plasmids

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

Table 2.1 Strains used in this study

Strains	Genotype	Reference source
BW25113	*E. coli* K12 strain, *lacI*^{q}, *rrnBT14*, Δ*lacZ*$_{WJ16}$, *hsdR514*, Δ*araBAD*$_{AH33}$, Δ*rhaBAD*$_{LD78}$	(Baba *et al.*, 2006)
M182	*E. coli* K12 strain, Δ(*lac*IPOZY) X74, *gal*K, *gal*U, *str*A	(Casadaban and Cohen, 1980)
M182 Δ*crp*	A derivative of M182 with the *crp* gene deleted.	(Busby *et al.*, 1983)
EAEC 042	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, diarrhoeagenic in patients, expresses AAF/II, biofilm positive, carries pAA2	(Nataro *et al.*, 1995)
EAEC 042 Δ*aggR*	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, carries pAA2 with the *aggR* gene deleted.	(Dudley *et al.*, 2006)
EAEC DFB 042	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, expresses AAF/II, biofilm positive, carries pAA2.	Douglas Browning
EAEC DFB042 Δ*aggR*	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, biofilm negative, carries pAA2 with the *aggR* gene replaced by kanamycin cassette.	Douglas Browning
EAEC DFB 042 Δ*Oantigen*	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, biofilm negative, carries pAA2. An insertion in *whaC* prevents O-antigen production.	Douglas Browning
EAEC DFB 042 Δ*yicS*	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, biofilm positive, carries pAA2. The *yicS* gene has been deleted.	This study
EAEC DFB 042 Δ*yicS*	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, biofilm positive, carries pAA2. The *yicS* gene has been replaced by kanamycin resistance cassette.	This study

EAEC H149/5	Wild type, diarrhoeagenic in volunteers	(Rosa *et al.*, 1998)
EAEC H92/3	Wild type, Amp^R, Tet^R, Cm^R, diarrhoeagenic in patients, expresses AAF/II, biofilm positive, carries pAA2	(Rosa *et al.*, 1998)
EAEC I18/2	Wild type, diarrhoeagenic in patients, expresses AAF/II, biofilm positive, carries pAA2	(Rosa *et al.*, 1998)
EAEC H9/3	Wild type, diarrhoeagenic in patients, expresses AAF/II, biofilm positive, carries pAA2	(Rosa *et al.*, 1998)

2.4 Gel electrophoresis

2.4.1 Agarose gel electrophoresis

Agarose gels were used to visualise and purify DNA fragments and plasmids larger than 1 kb. A TBE buffer system was used to make and run the agarose gels; a 5x TBE stock solution was prepared by dissolving 27.5 g borate, 54 g Tris and 4.65 g EDTA in one litre of dH_2O, the stock was diluted with dH_2O to 0.5x TBE. Gels were made by mixing molten 0.8% agarose, prepared by dissolving 0.8% (w/v) agarose in 0.5x TBE, with 1 μL ethidium bromide (10 mg/mL), and then this was poured into a gel casting plate. A comb was placed into the casting tray vertically to make wells that could hold 10-40 μL of sample, depending on the well width chosen, and the gel was then left to cool and solidify before the comb was removed. The gel was placed in a horizontal electrophoresis tank and submerged in 0.5x TBE buffer. DNA samples were loaded into the wells of the gel, prepared by mixing 5 volumes of sample with loading dye (Purple (2.5% Ficoll®-400, 10 mM EDTA, 3.3 mM Tris-HCl, 0.02% Dye 1 and 0.001% Dye 2)). A molecular weight ladder (NEB or Bioline) was also prepared with loading dye and loaded into the gel to indicate the size of the DNA samples. The gels were subjected to electrophoresis at 80 V for 30-45 minutes. The gel was visualised by 300 nm UV light using a Molecular Imager (Gel DOC XR imaging system Bio-Rad Laboratories) or High Performance Ultraviolet Transilluminator (Ultra Violet Products Ltd.) and, for purification of DNA fragments, the DNA bands were excised and purified using a QIAquick gel extraction kit (Qiagen) (section 2.5.1).

Table 2.2 Plasmids used in this study

Plasmids	Description	Source
pRW50	Low copy-number *lacZ* expression vector, broad-host range, *oriV*, *lacZYA*, *tet*^{R}.	(Lodge *et al.*, 1992)
pRW50*afaB*	A derivative of pRW50 carrying *afaB* promoter fragment cloned as lacZ transcriptional fusion (EcoRI- HindIII).	Muhammad Yasir
pRW50*agg4D*	A derivative of pRW50 carrying *agg4D* promoter fragment cloned as lacZ transcriptional fusion (EcoRI- HindIII).	Muhammad Yasir
pRW50*aafD100*	A derivative of pRW50 carrying *aafD* promoter fragment cloned as lacZ transcriptional fusion (EcoRI- HindIII).	Muhammad Yasir
pRW50*yicS* K-12	A derivative of pRW50 carrying *yicS* *E.coli* K-12 promoter fragment cloned as lacZ transcriptional fusion (EcoRI-HindIII).	Douglas Browning
pRW50*yicS* 042	A derivative of pRW50 carrying *yicS* EAEC 042 promoter fragment cloned as lacZ transcriptional fusion (EcoRI-HindIII).	Douglas Browning
pRW50*yicS* H92/3	A derivative of pRW50 carrying *yicS* EAEC H92/3 promoter fragment cloned as lacZ transcriptional fusion (EcoRI- HindIII).	This study
pRW50*nlpA* K-12	A derivative of pRW50 carrying *nlpA* *yicS* *E.coli* K-12 promoter fragment cloned as lacZ transcriptional fusion (EcoRI- HindIII).	Douglas Browning
pRW50*nlpA* 042	A derivative of pRW50 carrying *nlpA* EAEC 042 promoter fragment cloned as lacZ transcriptional fusion (EcoRI-HindIII).	Douglas Browning
pRW50*nlpA* H92/3	A derivative of pRW50 carrying *nlpA* EAEC H92/3 promoter fragment	This study

	cloned as lacZ transcriptional fusion (EcoRI- HindIII).	
pRW50*yicS-10KO1*	A derivative of pRW50*yicS* substituting TAT to TCT.	This study
pRW50*yicS-10KO2*	A derivative of pRW50*yicS* substituting TAT to GCT.	This study
pRW50*yicS-AggR-BS-KO*	A derivative of pRW50*yicS* substituting TATC to CACC.	This study
pRW50*yicS-K12-10consensus*	A derivative of pRW50*yicS* substituting AGC to AAT.	This study
pRW50*yicS-P2* K-12	A derivative of pRW50*yicS* K12 downstream of the AggR-dependent promoter.	This study
pRW50*yicS-P2* 042	A derivative of pRW50*yicS* 042 downstream of the AggR-dependent promoter.	This study
pRW50*nlpA-aggRKO*	A derivative of pRW50*nlpA* substituting GATA to GGTG.	This study
pRW50*nlpA-10KO*	A derivative of pRW50*nlpA* substituting TAT to TAG.	This study
pRW50*aatP*	H149/5 *aatP* promoter fragment	Douglas Browning
pRW50*caf1M*	H149/5 *caf1M* promoter fragment	Douglas Browning
pRW50*cfaD*	H149/5 *cfaD* promoter fragment	This study
pRW50*aap*	H149/5 *aap* promoter fragment	Douglas Browning
pRW50*caf1M-down*	A derivative of pRW50*caf1M* substituting TATC to CACC.	This study
pRW50*caf1M-both*	A derivative of pRW50*caf1M* substituting AATC…TATC to CACC…CACC.	This study
pRW50*caf1M-up consensus*	A derivative of pRW50*caf1M* substituting GGAATA to TTTTTT.	This study

pRW50*caf1M-up*	A derivative of pRW50*caf1M* substituting AATC to CACC.	This study
pRW50*afaB-caf1M* hybrid	A derivative of pRW50*caf1M* substituting ACTA…CTTA to CTGT…ATGA from the *afaB* promoter.	This study
pRW50*caf1M-CCG*	A derivative of pRW50*caf1M* substituting GTG to CCG from the *afaB* promoter.	This study
pRW224	Low copy-number *lacZ* expression vector, broad-host range, *oriV*, *lacZYA*, *tet*^{R}.	(Islam *et al.*, 2011)
pRW224*bssS* 042	A derivative of pRW224 carrying 042 *bssS* promoter fragment fused to *lacZ* (EcoRI-HindIII)	This study
pRW224*bssS* K-12	A derivative of pRW224 carrying K-12 *bssS* promoter fragment fused to *lacZ* (EcoRI-HindIII)	This study
pRW400	A derivative of pRW224/U9 with *lacZ, lacY* and *lacA* genes replaced with *gfp* derived from pJB in frame downstream from promoter-cloning site.	(Bryant *et al.*, 2014; Alsharif *et al.*, 2015)
pRW400*aafD100*	A derivative of pRW400 carrying 042 *aafD100* promoter fragment cloned with EcoRI-HindIII as a transcriptional fusion to *gfp*.	Muhammad Yasir
pBAD30	A derivative of pBAD vector carrying restriction sites to clone genes under the control of the *araBAD* promoter, *amp*^{R}.	(Guzman *et al.*, 1995)
pBAD*aggR*	A derivative of pBAD30 carrying *aggR* EAEC 042 cloned using EcoRI-XbaI under the control of the *araBAD* promoter.	(Sheikh *et al.*, 2002)
pBAD*cfaD*	A derivative of pBAD30 carrying EAEC H149/5 *cfaD* cloned using	This study

	EcoRI-XbaI under the control of the *araBAD* promoter.	
pBAD*rns*	A derivative of pBAD30 carrying ETEC *rns* cloned using EcoRI-XbaI under the control of the *araBAD* promoter.	George Munson (Munson and Scott, 1999)
pBAD*aggR-I14T*	A derivative of pBAD*aggR* carrying I14T substitution.	Rita Godfrey
pBAD*aggR-N16D*	A derivative of pBAD*aggR* carrying N16D substitution.	Rita Godfrey
pBAD*aggR-Y92A*	A derivative of pBAD*aggR* carrying Y92A substitution.	Previous study (unpublished)
pBAD*aggR-Y92C*	A derivative of pBAD*aggR* carrying Y92C substitution.	Previous study (unpublished)
pBAD*aggR-Y92R*	A derivative of pBAD*aggR* carrying Y92R substitution.	This study
pBAD*aggR-Y92S*	A derivative of pBAD*aggR* carrying Y92S substitution.	This study
pBAD*aggR-N168D*	A derivative of pBAD*aggR* carrying N168D substitution.	This study
pBAD*aggR-183N*	A derivative of pBAD*aggR* carrying I183N substitution.	This study
pBAD*aggR-E191A*	A derivative of pBAD*aggR* carrying E191A substitution.	This study
pBAD*aggR-I192A*	A derivative of pBAD*aggR* carrying I192A substitution.	This study
pBAD*aggR-R195A*	A derivative of pBAD*aggR* carrying R195A substitution.	This study
pBAD*aggR-K196A*	A derivative of pBAD*aggR* carrying K196A substitution.	This study
pBAD*aggR-R197A*	A derivative of pBAD*aggR* carrying R197A substitution.	This study

pBAD*aggR-RKR195-197AAA*	A derivative of pBAD*aggR* carrying R195A-K196A-R197A substitution.	This study
pBAD*aggR-E199A*	A derivative of pBAD*aggR* carrying E199A substitution.	This study
pBAD*aggR-S200A*	A derivative of pBAD*aggR* carrying S200A substitution.	This study
pBAD*aggR-Y226A*	A derivative of pBAD*aggR* carrying Y226A substitution.	This study
pBAD*aggR-Q227A*	A derivative of pBAD*aggR* carrying Q227A substitution.	This study
pBAD*aggR-Q227STOP*	A derivative of pBAD*aggR* carrying Q227STOP substitution.	This study
pBAD*aggR-Q230A*	A derivative of pBAD*aggR* carrying Q230A substitution.	This study
pBAD*aggR-Q230L*	A derivative of pBAD*aggR* carrying Q230L substitution.	This study
pBAD*aggR-Q230M*	A derivative of pBAD*aggR* carrying Q230M substitution.	This study
pBAD*aggR-Q230R*	A derivative of pBAD*aggR* carrying Q230R substitution.	This study
pBAD*aggR-QIS230-232KIP*	A derivative of pBAD*aggR* carrying Q230K-S232P substitution.	This study
pBAD*aggR-M234A*	A derivative of pBAD*aggR* carrying M234A substitution.	This study
pBAD*aggR-T240A*	A derivative of pBAD*aggR* carrying T240A substitution.	This study
pBAD*aggR-Y242A*	A derivative of pBAD*aggR* carrying Y242A substitution.	This study
pBAD*aggR-R245A*	A derivative of pBAD*aggR* carrying R245A substitution.	This study
pBAD*aggR-R245K*	A derivative of pBAD*aggR* carrying R245K substitution.	This study

pBAD*aggR-K249A*	A derivative of pBAD*aggR* carrying K249A substitution.	This study
pBAD*aggR-K263STOP*	A derivative of pBAD*aggR* carrying K263STOP substitution.	This study
pBAD*aggR-D86N-L221S*	A derivative of pBAD*aggR* carrying D86N-L221S substitution.	Previous study (unpublished)
pBAD*aggR-Q230R-K263STOP*	A derivative of pBAD*aggR* carrying Q230R-K263STOP substitution.	Previous study (unpublished)
pBAD*aar*	A derivative of pBAD30 carrying *aar* cloned using XbaI-SphI under the control of the *araBAD* promoter.	Radwa Abdelwahab
pBAD*aar**	A derivative of pBAD*aar* with ribosome binding site substitution.	Radwa Abdelwahab
pLG339	Low copy number plasmid encodes *nrfA* under its own promoter, tet^R and kan^R.	(Stoker *et al.*, 1982)
pLG339*aggR*	A derivative of pLG339 carrying *aggR* promoter fragment and *aggR* gene cloned using EcoRI-SalI.	This study
pLG339*aggR-I14T*	A derivative of pLG339*aggR* carrying I14T substitution.	This study
pLG339*aggR-N16D*	A derivative of pLG339*aggR* carrying N16D substitution.	This study
pLG339*aggR-Y92C*	A derivative of pLG339*aggR* carrying Y92C substitution.	This study
pLG339*aggR-Y92R*	A derivative of pLG339*aggR* carrying Y92R substitution.	This study
pLG339*aggR-E191A*	A derivative of pLG339*aggR* carrying E191A substitution.	This study
pDOC/*yicS-HRs*	pDOC-K derivative containing homology regions for *yicS* gene knockout, containing the *sacB* gene and amp^R.	Gurneet Dhanoa (unpublished)

pACBSR	Gene doctoring donor plasmid, *amp*^{R}, *sacB* and *km*^{R} cassette.	Douglas Browning

2.4.2 Polyacrylamide gel electrophoresis (PAGE) of DNA

Polyacrylamide gels were used to visualise and purify DNA fragments under 1 kb in length. The 7.5% (w/v) acrylamide gel stock solution was prepared by mixing 125 mL of 30% (w/v) acrylamide (National Diagnostics) with 100 mL 5x TBE (section 2.4.1) and 20 mL glycerol (per 500 mL). To polymerise the gel, 100 μL of freshly prepared 10% (w/v) ammonium persulphate (APS) and 10 μL of N, N, N', N'-Tetramethylethylenediamine (TEMED) were added to 10 mL 7.5% polyacrylamide and the mixture was poured between two glass plates, separated by 0.15 cm thick spacers, and a comb was added to the top to make wells. After the gel had polymerised, the comb and bottom spacer were removed, and the glass plates were inserted into the vertical electrophoresis apparatus. 1x TBE was added to the upper and lower tanks and the wells were flushed with 1x TBE using a Pasteur pipette. DNA samples, prepared by mixing 5 volumes of DNA sample with one volume of loading dye, were loaded into the wells and a molecular weight ladder was also loaded to indicate DNA size. A 40 mA current was applied to the gel for 45-60 minutes. The gel was stained in 0.5 μg/mL ethidium bromide solution for 15 minutes, the DNA bands were visualised and excised as in section 2.4.1 before electroelution and purification (section 2.5.2, 2.5.3 and 2.5.4).

2.5 Extraction and purification of nucleic acids

2.5.1 Extraction of DNA fragments from agarose gels

DNA bands were excised from agarose gels (section 2.4.1) and the DNA fragments were extracted from the agarose gel slices using a QIAquick Gel

Extraction Kit (Qiagen) as specified by the manufacturer. The extracted DNA fragments were eluted from the QIAquick columns in 50 μL elution buffer.

2.5.2 Electroelution of DNA fragments from polyacrylamide gels

The DNA bands were excised from a polyacrylamide gel on a small slice (section 2.4.2). 6 mm dialysis tubing (Medicell International Ltd.) was washed thoroughly with distilled water and rinsed with 0.1x TBE. The bottom of the tubing was sealed with a dialysis clip, and the gel slice placed inside the tubing along with 200 μL of 0.1x TBE and the top was secured, ensuring there were no air bubbles present. The tubing was placed into a gel tank containing 0.1x TBE and a 30-40 mA current was applied for 20 minutes for electroelution. The solution was then transferred from the dialysis tubing to a 1.5 mL microfuge tube, the tubing was washed with 200 μL SDW and this was added to the solution in the tube and the final volume was adjusted to 400 μL for DNA extraction (sections 2.5.3 and 2.5.4).

2.5.3 Phenol/chloroform extraction of DNA

To purify DNA following electroelution (section 2.5.2) or restriction enzyme digestion (section 2.7.7), the DNA sample volume was adjusted to 400 μL and mixed with 400 μL of phenol-chloroform solution (Sigma Aldrich), vortexed and centrifuged for 2 minutes at 15,700 g. The upper aqueous layer containing DNA was transferred to a new tube, without disrupting the lower phenol-chloroform interface, and the DNA was concentrated by ethanol precipitation (section 2.5.4).

2.5.4 Ethanol precipitation of DNA

Following phenol/chloroform DNA extraction (section 2.5.3), the volume of the aqueous supernatant was made up to 400 μL with SDW, then 40 μL (1/10) 3 M sodium acetate (pH 5.2), 4 μL (1/100th volume) of 1 M $MgCl_2$ and 890 μl of ice-

cold 100% ethanol were added and the mixture was incubated for an hour at -20°C or for 15 minutes at -80˚C. This mixture was then spun for 15 minutes at 4°C in a centrifuge, the supernatant was removed and 1 ml of 70% (v/v) ethanol was added to wash the pellet. This was centrifuged for a further 10 minutes at 4°C and the supernatant was removed. The pellet was washed again with 1 ml of ice-cold 100% ethanol and spun for 10 minutes at 4°C in a centrifuge. The supernatant was then removed, and the pellet dried in a vacuum drier 30-40 minutes. The pellet was resuspended in 20-50 μL SDW.

2.5.5 Small-scale preparation of plasmid DNA using QIAprep Spin miniprep kit

To prepare plasmid DNA, overnight cultures were transferred to a 15 mL sterile Falcon tube and pelleted at room temperature in a centrifuge at 2844 g for 15 minutes. The pellet was resuspended in P1 buffer (Qiagen) and transferred to a microfuge tube. Plasmid DNA was extracted using a QIAprep Spin Miniprep Kit as detailed by the manufacturer.

2.6 Transformation of *E. coli* with plasmid DNA

2.6.1 Preparation of chemically competent cells using the method of $CaCl_2$ transformation

A single bacterial colony of the strain to be made competent was used to inoculate 5 mL LB (containing appropriate antibiotic) and incubated overnight at 37°C with shaking. The next day, the overnight culture was diluted 1:100 in 50 mL LB medium and incubated with shaking at 37°C until OD_{600} reached 0.3-0.5 (mid-exponential phase of growth). The culture was transferred to a sterile 50 mL

Falcon tube and incubated on ice for 10 minutes. The cells were spun in a centrifuge at 2844 g at 4°C for 10 minutes. After the supernatant was removed, the pellet was resuspended in 25 mL ice cold 0.1 M $CaCl_2$ and incubated on ice for 20 minutes. This was then harvested by centrifugation for 10 minutes at 2844 g at 4°C. The pellet was resuspended in 3.3 mL ice cold 0.1 M $CaCl_2$ with 15% (v/v) glycerol then incubated on ice for at least 2 hours, or overnight, before use or storage. The cells were stored in 500 μL aliquots at -80°C.

2.6.2 Transformation of plasmid DNA into chemically competent cells

To transfer plasmid DNA into competent cells, 1 μL of plasmid was added to 50-100 μL competent cells (section 2.6.1) and incubated on ice for 30-60 minutes. Cells were heat shocked at 42°C for 1.5 – 2 minutes then incubated on ice for 5 minutes. 900 μL LB medium was added to the cells and incubated at 37°C with shaking for an hour, followed by harvesting by centrifugation at 15700 g for one minute. The pellet was resuspended in 100 μL of the supernatant and plated onto an LB agar or MacConkey agar plate with appropriate antibiotic, then incubated overnight at 37°C.

2.6.3 Preparation of electrocompetent cells

An overnight culture was diluted 1:100 in 5 mL LB, supplemented with appropriate antibiotic, and incubated at 37°C with shaking until mid-exponential phase, OD_{650} between 0.4 - 0.6. The cultures were transferred to chilled sterile Falcon tubes and spun in a centrifuge for 10 minutes at 2844 g and 4°C. The supernatant was removed, and the pellet resuspended in 1 mL of ice-cold 10% (v/v) glycerol. This was transferred to a chilled microfuge tube and centrifuged for one minute at 4°C

and 15700 g. The supernatant was removed, and the pellet resuspended in 10% ice-cold glycerol, this was repeated 3 times before the pellet was resuspended in 200 µL 10% ice-cold glycerol and kept on ice to be transformed the same day (section 2.6.4).

2.6.4 Transformation of *E. coli* electrocompetent cells by plasmid DNA

For electroporation, 40 µL of electrocompetent cells (section 2.6.3) were mixed with 1-3 µL of DNA and transferred to a pre-cooled electroporation cuvette (1 mm, Geneflow cat no. E6-0050). The cuvette was transferred to an electroporator (Gene Pulser, BioRad) and pulsed twice at 2.5 kV. Immediately afterwards, 1 mL of 37°C SOC was added to the cuvette and mixed by pipetting. This was transferred to a microfuge tube and incubated at 37°C with shaking for an hour. The sample was pelleted in a microcentrifuge at 15700 g for 1 minute. 900 µL supernatant was removed and the pellet was resuspended in the remaining medium, this was plated on an LB agar or MacConkey agar plate with appropriate antibiotics and incubated at 37°C overnight.

2.7 Recombinant DNA techniques

The oligonucleotide primers used in this study are outlined in Table 2.3.

2.7.1 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify target genes and promoter fragments using Accuzyme (Bioline), Phusion™ High-Fidelity DNA polymerase (New England BioLabs® Inc.) and MyTaq red DNA polymerase (Bioline). DNA templates came from three sources: bacterial cell lysates, purified bacterial genomic DNA and purified bacterial plasmids.

Table 2.3 Oligonucleotide primers

Name	Sequence (5' to 3')	Use
D10520 (FW)	CCCTGCGGTGCCCCTCAAG	Anneals upstream of the EcoRI site in pRW50/ pRW224/ pRW225. Used for sequencing and amplification of inserts in this vector.
pRW50 REV (D78264)	GGCTGTAATGTTCTGGCATTTGG TCAGC	Anneals downstream of the HindIII site in pRW50. Used for sequencing and amplification of inserts in this vector.
AggR promoter up	GGGGGGAATTCGGAAATAAGGG GGCAAGAATACG	*aggR* upstream primer containing an EcoRI site for amplification of the *aggR* promoter fragment.
AggR promoter down	GGGGGGGGTCGACCTCAAATAA TGATATGAAACATG	*aggR* downstream primer containing a SalI site for amplification of the *aggR* promoter fragment.
042 AggR KO	CAGTTTCACTTTTTCACCCTTTTT TTAATCG	*yicS* 042 upstream primer for megaprimer PCR amplification of the *yicS042-AggRBSKO* promoter fragment.
042 -10 KO	CGTTAACTGACTCTAATGGCAAG ATCAC	*yicS* 042 upstream primer for megaprimer PCR amplification of the *yicS042-10KO1* and *yicS042-10KO2* promoter fragment.
K-12 -10 to 042	GATTACCTGACTATAATGGTAAG GTCG	*yicS* K-12 upstream primer for megaprimer PCR amplification of the *yicSK-12-10consensus* promoter fragment.
nlpA 042 -10 KO (yicS)	GTGATCTTGCCATTAGAGTCAGT TAACG	*nlpA* 042 upstream primer for megaprimer PCR amplification of the *nlpA-10KO* promoter fragment.

042 nlpA AggR KO (yicS)	CGATTAAAAAAAGGGTGAAAAAG TGAAACTG	*nlpA* 042 upstream primer for megaprimer PCR amplification of the *nlpA-AggRKO* promoter fragment.
yicS P2 primer	GGGGGGAATTCTTTCGTACAGC ATGATGAGAGCG	*yicS P2* upstream primer containing an EcoRI site for amplification of the *yicS promoter 2* promoter fragments.
nlpA F	GGGGGAATTCCTGATAATGACG CCTGTGGCGTG	*nlpA* upstream primer containing a EcoRI site for amplification of the *nlpA* promoter fragment.
nlpA R H92/3	GGGGAAGCTTCTGCCAGCAATA ATGCGGCCCCC	*nlpA* upstream primer containing a HindIII site for amplification of the *nlpA* promoter from EAEC strain H92/3 fragment.
Q5SDM_D3res olved_F	TACAAAGGCCTCTTACACATATC TGATTG	*cfaD* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *cfaD* fragment.
Q5SDM_ D3resolved_R	AGTATCCCACATAATAGAAC	*cfaD* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *cfaD* fragment.
caf1M 042 hybrid FW	TTTTAATGAGTGTGTTGCTTTTTT ATCTG	*caf1M* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *afaB-caf1M hybrid* promoter fragment.
caf1M 042 hybrid REV	TATAAAACAGTCATTAACCATTAA GGTATCAG	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *afaB-caf1M hybrid* promoter fragment.
Q5SDM_caf1M GTG_FW	TAATCTTAGTCCGTTGCTTTTTTA TCTGAC	*caf1M* upstream primer for Q5 site-directed mutagenesis PCR amplification of the

		caf1M-CCG promoter fragment.
Q5SDM_caf1MGTG_REV	TTCCATTTAGTTCATTAACC	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-CCG* promoter fragment.
Q5SDM_pcaf1MbothKO_F	GTGTTGCTTTTTCACCTGACGTTAATTCCGCACTTG	*caf1M* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-both* promoter fragment.
Q5SDM_pcaf1MbothKO_R	ACTAAGGTGATTCCATATTCCATTTAGTTCATTAACCATTAAG	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-both* promoter fragment.
Q5SDM_pcaf1MdownKO_F	GTTGCTTTTTCACCTGACGTTAATTC	*caf1M* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-down* promoter fragment.
Q5SDM_pcaf1MdownKO_R	ACACTAAGATTATTCCATTTAG	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-down* promoter fragment.
Q5SDM_pcaf1MKOboth_R	ACTAAGGTGATTCCATTTAGTTCATTAACCATTAAGGTATC	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-both* promoter fragment.
Q5SDM_pcaf1Mupconsensus_F	TTATCTTAGTGTGTTGCTTTTTTATC	*caf1M* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-upconsensus* promoter fragment.
Q5SDM_pcaf1Mupconsensus_R	AAAAATTTAGTTCATTAACCATTAAGG	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the

		caf1M-upconsensus promoter fragment.
Q5SDM_pcaf1 MupKO_F	TAAATGGAATCACCTTAGTGTGT TGC	*caf1M* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-up* promoter fragment.
Q5SDM_pcaf1 MupKO_R	GTTCATTAACCATTAAGGTATC	*caf1M* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *caf1M-up* promoter fragment.
bssSFW(EcoRI)	GGGGGGAATTCTTGACGCCCTG GCTGGCGAACTTTC	*bssS* upstream primer containing an EcoRI site for amplification of the *bssS* promoter fragment.
bssSFW(EcoRI)	GGGGGAAGCTTACCAGTCATCG GCGCTTTCCCACGTTTCC	*bssS* upstream primer containing an EcoRI site for amplification of the *bssS* promoter fragment.
pBAD24 FW	CCATAAGATTAGCGGATCCTACC	Anneals upstream of the EcoRI site in pBAD30. Used for sequencing and amplification of inserts in this vector.
pBAD24 REV	CCAGGCAAATTCTGTTTTATCAG ACC	Anneals downstream of the XbaI site in pBAD30. Used for sequencing and amplification of inserts in this vector.
AggR Y92A	CCTGATGATAATAGCNGGAATAT CAAAAGTAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Y92A* fragment.
AggR Y92D	CCTGATGATAATAGCYGGAATAT CAAAAGTAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Y92D* fragment.
AggR Y92S	CCTGATGATAATATCNGGAATAT CAAAAGTAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Y92S* fragment.

AggR I183N	CTAGCTATTAAYGCAGATGAATTTAATG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-I183N* fragment.
AggR E191A(only)	GAATTTAATGTATCAGCGATAACAATAAGG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-E191A* fragment.
AggR I192A	GTATCAGAGGCAACAATAAGGAAAAGGC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-I192A* fragment.
AggR R195A	CAGAGATAACAATAGCNAAAAGGCTTGAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-R195A* fragment.
AggR K196A	GATAACAATAAGGGCNAGGCTTGAGTCAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-K196A* fragment.
AggR R197A	CAATAAGGAAAGCNCTTGAGTCAGAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-R197A* fragment.
AggR RKR195 - 197AAA	CAGAGATAACAATAGCNGCNGCNCTTGAGTCAGAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-RKR195-197AAA* fragment.
AggR E199A	GGAAAAGGCTTGCGTCAGAGTATTTAC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-E199A* fragment.
AggR S200A	GGAAAAGGCTTGAGGCAGAGTATTTAC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-S200A* fragment.
AggR Y226A	GATAACTCAGCTCAGATATCACAAATATC	*aggR* upstream primer for megaprimer PCR

		amplification of the *aggR-Y226A* fragment.
AggR Q227A	GATAACTCATATGCGATATCACAAATATC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Q227A* fragment.
AggR Q227STOP	GATAACTCATATTAAATATCACAAATATC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Q227STOP* fragment.
AggR Q230A	CAGATATCAGCAATATCTAATATGATAGG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Q230A* fragment.
AggR Q230L	CAGATATCACTNATATCTAATATGATAGG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Q230L* fragment.
AggR Q230M	CAGATATCAATGATATCTAATATGATAGG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Q230M* fragment.
AggR Q230R	CAGATATCACGAATATCTAATATGATAGG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Q230R* fragment.
AggR M234A	CAAATATCTAATGCGATAGGATTTTCCAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-M234A* fragment.
AggR T240A	GGATTTTCCAGTGCATCATATTTC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-T240A* fragment.
AggR Y242A	CCAGTACATCAGCTTTCATTAGG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-Y242A* fragment.

AggR R245A	CATATTTCATTGCGCTTTTTGTAAAAC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-R245A* fragment.
AggR R245K	CATATTTCATTAARCTTTTTGTAAAAC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-R245K* fragment.
AggR K249A	GGCTTTTTGTAGCNCATTTTGGCATAAC	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-K249A* fragment.
AggR truncation	GACTTATTTTTAAAGCCAATAGTCTAG	*aggR* upstream primer for megaprimer PCR amplification of the *aggR-K263STOP* fragment.
I183N_Q5_SDM_ FW	TCTAGCTATTAATGCAGATGAATTTAATG	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-I183N* fragment.
I183N_Q5_SDM_ REV	GTCCATCTCTTTGATAAGTC	*aggR* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-I183N* fragment.
Q227A_Q5_SDM_FW	TAACTCATATGCGATATCACAAATATCTAATATGATAGG	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-Q227A* fragment.
Q227A_Q5_SDM_REV	TCAAGCAACAGCAATGCTG	*aggR* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-Q227A* fragment.
I14T_Q5_SDM_ FW	GATTATAAAAACCAACAATATCAGAATAC	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-I14T* fragment.
I14T_Q5_SDM_ REV	TCTTTTTCGATGTTTTGTTTTAATTTC	*aggR* downstream primer for Q5 site-directed mutagenesis

		PCR amplification of the *aggR-I14T* fragment.
N16D_Q5_SDM_ FW	AAAAATCAACGATATCAGAATACATC	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-N16D* fragment.
N16D_Q5_SDM_ REV	ATAATCTCTTTTTCGATGTTTTG	*aggR* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-N16D* fragment.
Y92C_Q5_SDM_FW	GATGATAATATGCGGAATATCAAAAGTAGATGC	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-Y92C* fragment.
Y92R_Q5_SDM_FW	GATGATAATACGCGGAATATCAAAAGTAGATGCTTGTAGTTG	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-Y92R* fragment.
Y92C/R_Q5_SDM_REV	AGGGCATCCTTTAGGCGT	*aggR* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-Y92C* and *aggR-Y92R* fragment.
E191A_Q5_SDM_FW	TAATGTATCAGCGATAACAATAAGGAAAAG	*aggR* upstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-E191A* fragment.
E191A_Q5_SDM_REV	AATTCATCTGCAATAATAGC	*aggR* downstream primer for Q5 site-directed mutagenesis PCR amplification of the *aggR-E191A* fragment.

DNA amplification was carried out in a ProFlex PCR System (Thermo Fisher Scientific Inc.), PCR cycling conditions are outlined in Table 2.4 and the annealing temperature used was dependent on the oligonucleotide primers utilised. Following DNA amplification, PCR products were visualised as detailed in section 2.4. MyTaq red DNA polymerase was used for error-prone PCR, Phusion DNA polymerase and Accuzyme were used for PCR, colony PCR and megaprimer PCR.

For use of Phusion DNA polymerase, typical reaction mixtures were composed of 10 μL of 2.5 μM forward primer, 10 μL of 2.5 μM reverse primer, 10 μL 5x buffer, 1 μL dNTPs, 0.5 μL Phusion, 5 μL 1:40 template DNA and 13 μL SDW, the reaction mix was prepared on ice and mixed gently.

2.7.2 Colony PCR

Colony PCR was used to amplify genes from pathogenic EAEC strains and to confirm the presence of a segment of DNA in a bacterial genome – cell lysates of candidate colonies were used as template DNA. To prepare DNA template, a single colony from an agar plate was mixed with 50 μL SDW in a microfuge tube and heated for 2 minutes at 100˚C. The samples were vortexed and spun in a table-top centrifuge for 2 minutes at 15,700 g and the supernatant was retained as the cell lysate. 5 μL of this sample was then used as template in the reaction and mixed with 5 μL Accuzyme reaction buffer, 0.5 μL dNTPs, 10 μL 2.5 μM forward primer, 10 μL 2.5 μM reverse primer, 18.5 μL SDW, 1 μL Accuzyme DNA polymerase. The samples were transferred to a ProFlex PCR System (Thermo Fisher Scientific Inc.), the PCR conditions are detailed in Table 2.4; the annealing

temperature depended on the primers used and the extension time depended on the insert size.

2.7.3 Error-prone PCR

For random mutagenesis of *aggR*, error-prone PCR was utilised to generate *aggR* PCR product. The PCR mixture was prepared on ice and contained: 10 μL of 2.5 μM pBAD24 FW primer, 10 μL of 2.5 μM pBAD24 REV primer, 25 μL of MyTaq RED and 5 μL of 1:40 pBAD*aggR*. The PCR cycle is outlined in Table 2.4.

2.7.4 Megaprimer PCR

Megaprimer PCR was used to introduce site-directed mutations into recombinant DNA fragments. The procedure required two separate PCR cycles, with purification steps, and three primers, including two primers that annealed to the ends of the DNA fragment and a final primer, containing the intended mutation, that annealed internally. The first PCR used the primer annealing upstream of the DNA fragment and the internal primer, with the mutation, the product was then run on a gel (section 2.4) and purified (section 2.5). Then for the second PCR reaction, to generate the full-length fragment, the purified PCR product was used as a primer with the primer annealing downstream of the fragment. PCR cycles were carried out as outlined in section 2.7.1.

2.7.5 Q5® site-directed mutagenesis (SDM) PCR

Q5® Site-Directed mutagenesis (SDM) was utilised to introduce mutations in promoter and gene fragments as outlined by the manufacturer (New England BioLabs Inc.). Q5® SDM PCR mixtures were composed of 12.5 μL Q5 Hot Start

Table 2.4 PCR conditions

Cycles	PCR step	Temperature	Time
1	Initial denaturation	98˚C	5 min
35	Denaturation	95˚C	30 sec
	Annealing	Variable	30 sec
	Extension	72˚C	~1 min/kb
1	Final extension	72˚C	10 min
1	Hold	4˚C	∞

High-Fidelity 2X Master Mix, 1.25 μL 10 μM forward primer, 1.25 μL 10 μM reverse primer, 1 μL template DNA (diluted 1:40) and 9 μL SDW, with a total volume of 25 μL. The reagents were mixed by pipetting and transferred to a ProFlex PCR System (Thermo Fisher Scientific Inc), the conditions for the PCR can be found in Table 2.5. The annealing temperature was determined according to the primers used, and the extension time was 30 seconds/kb of plasmid with insert.

2.7.6 Q5® SDM Kinase, ligase and DpnI (KLD) treatment

1 μL of Q5 SDM PCR product (section 2.7.6) was mixed with 5 μL 2x KLD reaction buffer, 1 μL 10X KLD enzyme mix and 3 μL SDW, the reagents were mixed by pipetting and incubated at room temperature for 30 minutes before transformation as in section 2.6.2.

2.7.7 Restriction digestion of DNA

Restriction enzymes were used to prepare plasmid DNA and purified PCR products for DNA cloning. For high copy number plasmids, 12-15 μL of plasmid DNA miniprep was digested with 2.5 μL of each restriction enzyme (New England BioLabs® Inc.) and 6 μL 10X NEBuffer 4 or CutSmart buffer (New England BioLabs® Inc.), with a final volume of 60 μL with SDW. The digests were incubated at 37˚C for 2 hours, then 1 μL of calf intestinal alkaline phosphatase (CIP) (New England BioLabs® Inc.) was added and the reaction incubated for one further hour to dephosphorylate the ends of the DNA to prevent re-ligation. The digested plasmid was run on a 0.8% (w/v) agarose gel (2.4.1) and DNA bands were purified as in section 2.5.1.

Table 2.5 Q5 SDM PCR conditions

Cycles	PCR Step	Temperature	Time
1	Initial denaturation	98˚C	30 seconds
25	Denaturation	98˚C	10 seconds
	Annealing	Variable	30 seconds
	Extension	72˚C	~30 seconds/kb
1	Final extension	72˚C	2 minutes
1	Hold	4˚C	∞

For low copy number plasmids (e.g. pRW50), seven DNA minipreps were combined with a final volume of 300 µL and digested with 40 µL 10X NEBuffer 4 or CutSmart buffer, 10 µL of each restriction enzyme and made up to a final volume of 400 µL with SDW. The digest was incubated for 3 hours at 37˚C before 4 µL of CIP was added and then incubated for a further hour. The digested vector was purified using phenol/chloroform purification and ethanol precipitation (sections 2.5.3 and 2.5.4).

For the insert DNA, 48 µl of purified PCR product was digested with 3 µL of each restriction enzyme and 6 µL 10X NEBuffer 4 or CutSmart buffer, made to a final volume of 60 µL with SDW. The digests were incubated at 37°C for 2 hours and run on an agarose (section 2.4.1) or polyacrylamide gel (section 2.4.2) and purified (section 2.5).

2.7.8 Ligation

DNA ligation was utilised to insert genes or promoter fragments into appropriate vectors. For the ligations, 5-12 µL restricted insert DNA, 1-5 µL restricted vector DNA, 1 µL T4 DNA ligase (New England BioLabs® Inc.), and 2 µL 10x T4 DNA ligase buffer were combined and made to a final volume of 20 µL with SDW. The ligation mixes were kept on ice for 10 minutes and then incubated overnight at 16˚C. The mixture was placed on ice before addition of 100 µL $CaCl_2$ competent cells and transformations were carried out as in section 2.6.2. To select for transformants, cells were plated on MacConkey agar with lactose and supplemented with the appropriate antibiotics. To confirm the presence of insert, colonies were grown overnight, plasmid DNA was prepared and sequenced.

2.7.9 DNA sequencing

Purified plasmid DNA was plasmid-to-profile sequenced at The Functional Genomics, Proteomics and Metabolomics Facility, University of Birmingham, UK. Sequence reactions contained 3-5 µL miniprep DNA and 3 µL of primer (1 µM), made up to a final volume of 10 µL adjusted with SDW. The sequences were then analysed using SnapGene Viewer and the amino acid sequences aligned using ClustalW (Kyoto University Bioinformatics Centre, http://www.genome.jp/tools-bin/clustalw, no date; NCBI, https://www.ncbi.nlm.nih.gov/orffinder/, no date).

2.8 β-galactosidase assays

In this study, promoter activity was measured by utilising β-galactosidase assays and the promoter fragments were cloned upstream of *lacZ* in the expression vectors: pRW50, pRW224 and pRW400.

2.8.1 β-galactosidase assays during exponential growth and stationary phase

Cells were grown overnight in LB media containing the appropriate antibiotics. Overnight cultures were diluted (1:100) in 5 ml of fresh LB containing appropriate antibiotics. Each overnight culture was diluted into two subcultures, one containing arabinose at a final concentration of 0.2%. Subcultures were incubated with shaking at 37˚C until mid-logarithmic phase (OD_{650} = 0.4 – 0.6). To lyse the cultures, 2-3 drops of each of 1% (w/v) sodium deoxycholate and toluene were added to each flask and vortexed, then incubated at 37˚C with shaking for 20 minutes to evaporate the toluene. The β-galactosidase activity of cell lysates measured in Z-buffer.

Z-buffer was prepared to a final volume of 1 litre with distilled water containing: 0.75 g potassium chloride, 0.25 g magnesium sulphate heptahydrate ($MgSO_4 \cdot 7H_20$), 8.53 g disodium hydrogen orthophosphate (Na_2HPO_4) and 4.87 g sodium hydrogen orthophosphate dihydrate ($NaH_2PO_4 \cdot 2H_20$). 1M sodium carbonate (Na_2CO_3) was prepared by dissolving 10.59 g in 100 mL distilled water. Prior to use, 270 μL β-mercaptoethanol was added to 100 mL Z-buffer and 0.0392 g ortho-nitrophenol-β-D-galactopyranoside (ONPG) was added to 10 mL Z-buffer (final concentration 13 mM).

The β-galactosidase assays were carried out at 30˚C, 100 μL lysate was added to 2mL Z-buffer and the reaction was started by adding 500 μL of the ONPG solution. The reaction proceeded until the sample turned yellow and 1 mL 1M Na_2CO_3 was added to stop the reaction. The absorbance at OD_{420} was recorded and the β-galactosidase unit was calculated using:

$$\beta galactosidase\ activity = \frac{1000 \times 2.5 \times 3.6 \times OD_{420\,nm}}{OD_{650\,nm} \times 4.5 \times t \times v}\ nmol/min/mg\ of\ bacterial\ mass$$

Where,

1000/4.5 = factor for conversion of OD_{420} into nmol ONPG, based on 1 nmol ml^{-1} ONPG having an OD_{420} of 0.0045.

2.5 = factor for conversion of OD_{650} into bacterial mass, based on OD_{650} of 1 being equivalent to 0.4 $mgml^{-1}$ bacteria (dry weight).

3.6 = final assay volume (mL)

t = incubation time (minutes)

v = volume of lysate added (mL)

The β-galactosidase activity of cells containing plasmids carrying promoter fragments were measured using three independent determinations and the mean β-galactosidase level and standard deviation were calculated for each promoter fragment. In all assays, negative controls were included with cells containing empty vector.

2.9 Gene Doctoring

The "gene doctoring" technique (Lee *et al.*, 2009) was utilised in this study to delete the *yicS* gene from the EAEC strain DFB042 chromosome (Table 2.2). The pDOC/*yicS-HRs* donor plasmid used in this study contains DNA fragments homologous to the start and end of the *yicS* gene.

2.9.1 An overview of the technique

The λ-Red recombinase system is often employed to alter the *E. coli* chromosome, introducing insertions, deletions or mutations. There are three genes, *gam, exo* and *bet*, in this system that mediate the homologous recombination of a linear DNA fragment. For pathogenic strains, a method was developed to delete chromosomal genes; pDOC plasmids carry the *sacB* gene that counterselects cells containing the donor plasmid (Lee *et al.*, 2009).

2.9.2 Gene doctoring methodology

The donor plasmid, pDOC/*yicS-HRs*, and the pACBSR plasmid were co-transformed into EAEC strain DFB042 by electroporation and grown overnight on an LB agar plate with chloramphenicol and ampicillin. The co-transformant colonies were restreaked on the same plate type and then patched onto nutrient agar plates with 5% (w/v) sucrose and kanamycin (25 μg/mL). A single colony was used to inoculate 0.5 mL LB, with ampicillin (100 μg/mL) and

chloramphenicol (50 µg/mL), in a conical flask and shaken for 2-4 hours at 37˚C until turbid. The cells were harvested by centrifugation, the supernatant was discarded, and the pellet was resuspended in 0.1 X LB. This process was repeated three times to remove residual antibiotics. The cells were resuspended in 0.5 mL of 0.1 X LB, with 0.3% arabinose to induce expression of λ-Red genes and SceI meganuclease from pACBSR and transferred to a conical flask for incubation at 37˚C for 2-3 hours. 125 µL of culture was spread onto each of four LB agar plates (5% sucrose and 50 µg/mL kanamycin), and the plates were incubated at 30˚C until colonies are visible. Colonies were restreaked onto LB agar plates (5% (w/v) sucrose and 25 µg/mL kanamycin) to give single colonies, these were recombinant candidates. Colony PCR was carried out to check the size of the target region in the chromosome and the PCR products showing chromosomal modification were confirmed by DNA sequencing (sections 0 and 2.7.9). These candidates were also checked for loss of the pACBSR plasmid, by streaking onto LB agar plates with chloramphenicol. Glycerol stocks were made from chloramphenicol and ampicillin sensitive candidates, which contained the chromosomal modification.

2.9.3 Removal of the kanamycin resistance cassette and detection of gene doctored cells

The kanamycin cassette was removed by transforming the modified strain with plasmid pCP20 and incubated overnight on LB agar with ampicillin (100 µg/mL) at 30˚C. Transformants were restreaked onto LB agar plates, without antibiotic, and incubated at 37˚C overnight. Single colonies were patched onto LB agar plates in the order of 5% (w/v) sucrose and 25 µg/mL kanamycin, 100 µg/mL

ampicillin, and non-selective. The candidates were then checked by colony PCR and sequencing (sections 0 and 2.7.9).

2.10 Biofilm assay

Overnight cultures were inoculated from single colonies of EAEC in 5 mL of LB medium. To prepare subcultures, overnight cultures were diluted 1:100 in 5 mL of DMEM medium with appropriate antibiotics and incubated for 1 hour at 37°C with shaking. From the diluted culture, 150 µL was transferred into three wells of a 96 well microtitre plate. Negative controls of the DMEM solutions without inoculum were also included. The plate was sealed with Parafilm (TM) to prevent evaporation and incubated at 37°C overnight. After ~16 hours the media was removed by smacking the plate onto a paper towel. 150 µL of 0.1% crystal violet was added to each well and incubated at 4°C for 30 minutes. Again, the plate was smacked on paper towels to remove crystal violet from the wells and submerged in deionised water to remove excess crystal violet. All traces of water were removed by smacking repeatedly on a paper towel. 150 µL of ethanol/acetone solution (80 mL ethanol and 20 mL acetone) was added to each well and the microtitre plate was left to shake slowly for 45 minutes. The absorbance was measured at OD_{595} using a Labsystems Multiskan MS plate reader (Thermo Fisher Scientific Inc).

2.11 Flow cytometry

2.11.1 Set up of the flow cytometer

The Attune® NxT Acoustic Focusing Cytometer (Invitrogen) was utilised for all experiments. The manufacturer's instructions were followed to calibrate the

machine, including the use of Attune Focusing Fluid (448862, Fisher) and Attune ® Performance Tracking Beads (4449754, Fisher). GFP has excitation/emission maxima of 482 nm / 510 nm (Telford *et al.*, 2012), thus, the Blue laser was set at 488 nm: channel BL1, 530/30 filter. For fluorophore SYTO-84 Orange Fluorescent Nucleic Acid Stain (SYTO84) (Invitrogen), which has excitation/emission maxima of 567 nm / 582 nm, the Yellow laser was set at 561 nm; channel YL1, 585/16 nm filter. The machine was cleaned with 4% bleach between samples using the Sip Sanitize function, and at the end of each experiment a Deep Clean was performed.

2.11.2 Measurement of Syto84 with GFP fluorescence

Overnight cultures were used to inoculate fresh LB, 200 µL in 5 mL, and grown with shaking at 37˚C until the OD_{600} was between 0.5 and 0.6. Then, 200 µL subculture was spun in a centrifuge for 5 minutes and resuspended in 100 µL 1 x HEPES buffered saline (HBS) (Alfa Aesar). To the bacterial suspension, 500 µL HBS and SYTO84 (final concentration 10 µM) were added and vortexed.

2.11.3 Gating strategy

Using the gating strategy in Figure 2.1, GFP fluorescence was measured. To identify when no GFP was expressed, a negative control strain was used as this allowed measurement of GFP fluorescence in cells with GFP expressed. The plots in Figure 2.1A and Figure 2.1B show the unstained control, pLG339 empty vector, and in Figure 2.1C and Figure 2.1D the stained pLG339 control; both controls show the separation of $SYTO84^+$ stained cells from background debris. A polygon gate was added to select for $SYTO84^+$ populations (Figure 2.1A and

Figure 2.1C) which were then used to analyse GFP fluorescence (Figure 2.1B and Figure 2.1D).

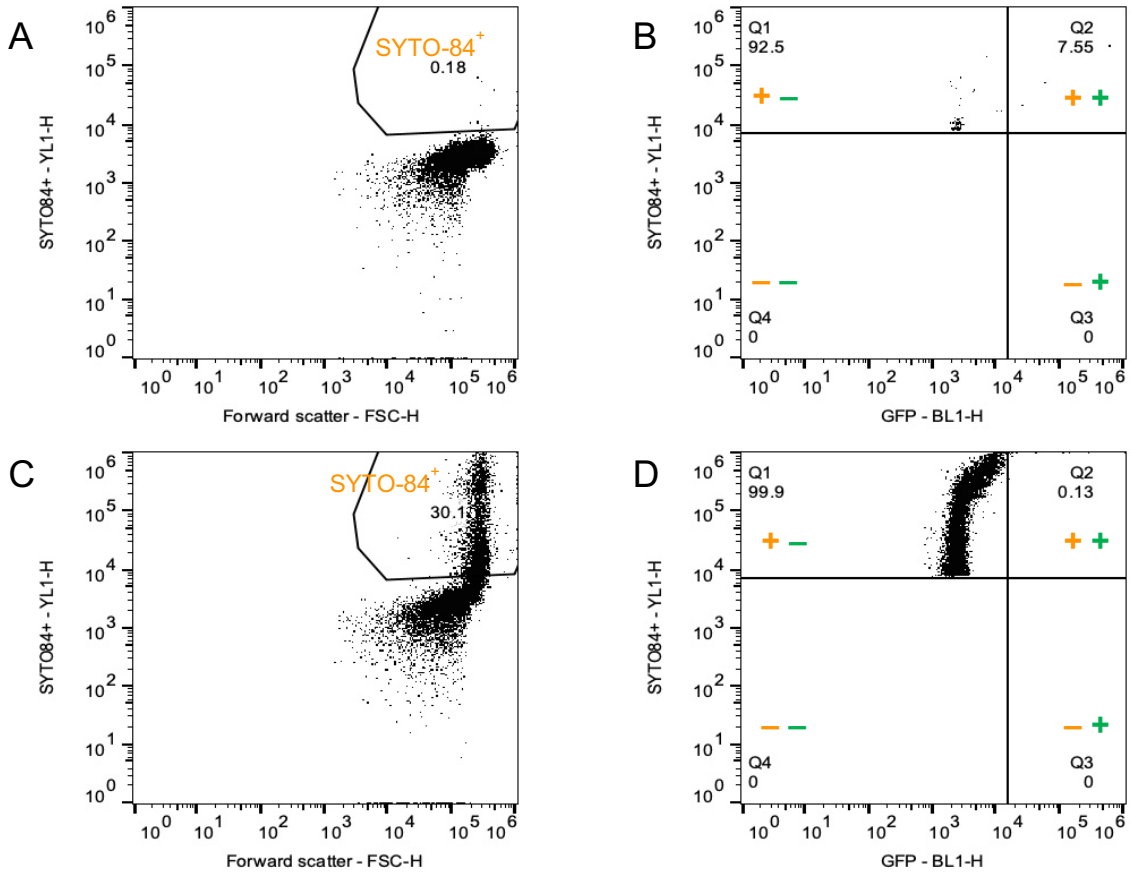


Figure 2.1 Gating strategy of empty vector control samples

The gating strategy used for analysis of flow cytometry data in this study, the plots show the negative control unstained BW25113 cells containing pRW400*aafD100* and pLG339 empty vector. In (A) and (C), cells were separated from background using channels YL1-H against FSC-H, highlighted by a polygon, and labelled SYTO84⁺ in orange. In (B) and (D), cells in the SYTO84⁺ gate were analysed using channels YL1-H against BL1-H and a quadrat gate separated the SYTO84 (labelled in orange) and GFP signals (labelled in green): Q1 for SYTO-84⁺ GFP⁻ (+ –); Q2 for SYTO-84⁺ GFP⁺ (+ +); Q3 for SYTO-84⁻ GFP⁺ (– +); and Q4 for SYTO-84⁻ GFP⁻ (– –). Data was analysed and plots were made using FlowJo.

A. One unstained sample of BW25113 cells containing pRW400*aafD100* and pLG339. The percentage of events in the SYTO84⁺ polygon gate is indicated.
B. One unstained sample of BW25113 cells containing pRW400*aafD100* and pLG339 from the SYTO84⁺ gate (A). The percentage of events in each of the quadrat gates is indicated.
C. One stained sample of BW25113 cells containing pRW400*aafD100* and pLG339. A SYTO84⁺ population of over 10,000 events was gated, highlighted by a polygon, and labelled SYTO84⁺ in orange. The percentage of events in the polygon gate of SYTO84⁺ is indicated.
D. One stained sample of BW25113 cells containing pRW400*aafD100* and pLG339 from the SYTO84⁺ gate (C). The percentage of events in each of the quadrat gates is indicated.

Chapter 3 The Action of AggR at Target Promoters

3.1 Introduction

EAEC colonises host intestinal mucosa and secretes toxins (Morin *et al.*, 2010). Some EAEC strains carry a virulence plasmid (Jonsson *et al.*, 2017) that encodes virulence genes involved in pathogenicity (Hüttener *et al.*, 2018; Berger *et al.*, 2016). An important component of EAEC pathogenicity is the master regulator of virulence, AggR, a transcription factor with more than 40 virulence genes thought to be in its regulon (Dias *et al.*, 2020; Santiago *et al.*, 2017; Morin *et al.*, 2013). To confirm this, Chris Icke, working with Ian Henderson, analysed RNA-seq data, obtained from comparing differential gene expression in the EAEC prototypical strain 042 and its Δ*aggR* derivative (Figure 3.1). The results indicated that AggR-dependent promoters are located both on the chromosome (Figure 3.1A) and the virulence plasmid, pAA2 (Figure 3.1B) (Hüttener *et al.*, 2018; Morin *et al.*, 2013; Yasir *et al.*, 2019). There are several genes highlighted in the figure, indicating some of the target promoters used in this study. Many of the AggR-dependent genes have unknown functions and this study investigated the function of the *yicS* gene expressed from the EC042_4006 locus (Section 3.3).

The organisation of a dozen different AggR-dependent promoters was investigated by Muhammed Yasir, and he showed that the AggR promoters analysed all had a similar organisation – the DNA binding site for AggR is at a

Class II position, overlapping the -35 element, with 20-22 bp between the single AggR binding site and the -10 element. Additionally, by transplanting a single DNA site for AggR into the *E. coli* CRP-dependent *melR* promoter, 22 bp upstream of the -10 element, AggR-dependence was conferred on this promoter (Yasir *et al.*, 2019). This argued that AggR functions as a monomer.

Most research has been conducted with the prototypical EAEC 042 strain (Hüttener *et al.*, 2018; Ellis *et al.*, 2019), but in recent years, many more EAEC strains have been sequenced and analysis is ongoing. Interestingly, there is enormous diversity in naturally occurring EAEC strains (Boisen *et al.*, 2020), and the pattern of virulence and presence of certain virulence genes does not follow the same pattern as in EAEC 042 or *E. coli* K-12 (which has frequently served as a non-virulent non-EAEC comparison). Therefore, my study has focussed on whether the rules for AggR-dependence follow the same patterns in different strains. I wanted to see what might be learned about AggR-dependence from looking beyond strain 042, with a focus on clinically relevant strains that were isolated from patients in Brazil (Rosa *et al.*, 1998). Hence, for comparison with EAEC 042, I looked at a promoter from the Danish EAEC strain, C1010-00 (Yasir *et al.*, 2019; Boisen *et al.*, 2008).

3.2 Analysis of simple AggR-dependent promoters

AggR activates the expression of genes important for pathogenicity including secretion systems implicated in biofilm formation and antibacterial activity (Bamidele *et al.*, 2019; Dias *et al.*, 2020; Navarro-Garcia *et al.*, 2019), including

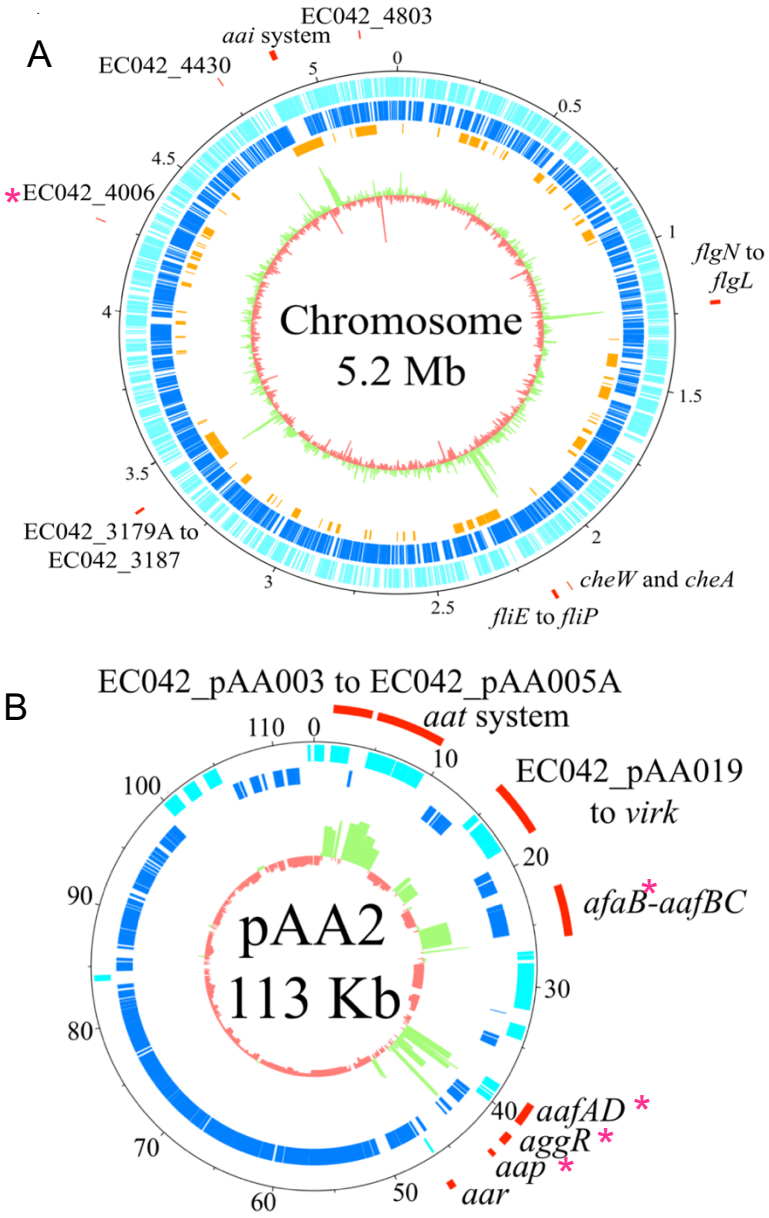


Figure 3.1 AggR transcriptomics on the EAEC 042 chromosome and plasmid

The figure shows differential RNA levels in EAEC 042 compared to an EAEC 042 Δ*aggR* strain, according to RNA-seq, on its chromosome (A) and pAA2 virulence plasmid (B). Differentially expressed genes are presented as green peaks, these identify the AggR gene targets. The light blue boxes indicate the genes in the forward orientation and the dark blue boxes are in the reverse. The central ring shows the $\log_2$ fold difference between EAEC 042 and the Δ*aggR* strain for each gene. The pink stars highlight the gene targets investigated in my study. Adapted from Yasir *et al.* (2019).

A. EAEC 042 chromosome. The orange boxes show the regions of difference in the EAEC 042 chromosome (Chaudhuri *et al.*, 2010). The base coordinates are labelled in Mb.
B. EAEC pAA2 virulence plasmid. The base coordinates are labelled in Kb.

the expression of the EAEC fimbrial antigen AAF/II (Elias *et al.*, 1999; Hüttener *et al.*, 2018). It has previously been shown that simply overexpressing AggR is sufficient for activation at target promoters (Yasir *et al.*, 2019).

3.2.1 Activation

Intergenic regulatory regions, likely to contain AggR-dependent promoters, were identified in several EAEC strains. To study these promoters, I exploited the pRW50 low copy number *lac* expression vector (Lodge *et al.*, 1992). In this plasmid, there is no promoter for the *lac* operon, so inserting promoters here drives *lacZ* expression, and subsequent β-galactosidase production, thereby facilitating measurement of the activity of the cloned promoters (Figure 3.2). PCR was used to amplify selected putative intergenic regulatory regions. To facilitate cloning into pRW50, primers that introduced an EcoRI restriction site upstream of the putative promoter and a HindIII restriction site downstream were used.

The selected AggR-target promoters looked at in my study include the promoter of strain 042 *afaB*, that encodes a pseudogene involved in fimbrial assembly (Figure 3.3A), the *agg4D* promoter, the first gene in the AAF/IV fimbrial operon of EAEC strain C1010-00 (Figure 3.3B), and the strain 042 *aafD* promoter, which controls expression of a chaperone gene in the fimbrial operon (Figure 3.3C) (Elias *et al.*, 1999; Yasir *et al.*, 2019; Ellis *et al.*, 2019). These promoters are located on virulence plasmids, *afaB* and *aafD* are found on the strain 042 pAA2 plasmid in strain 042 and *agg4D* is located on the C1010-00 pAA plasmid (Yasir *et al.*, 2019). Base sequences of the promoter fragments that were cloned into pRW50 are shown in Figure 3.3.

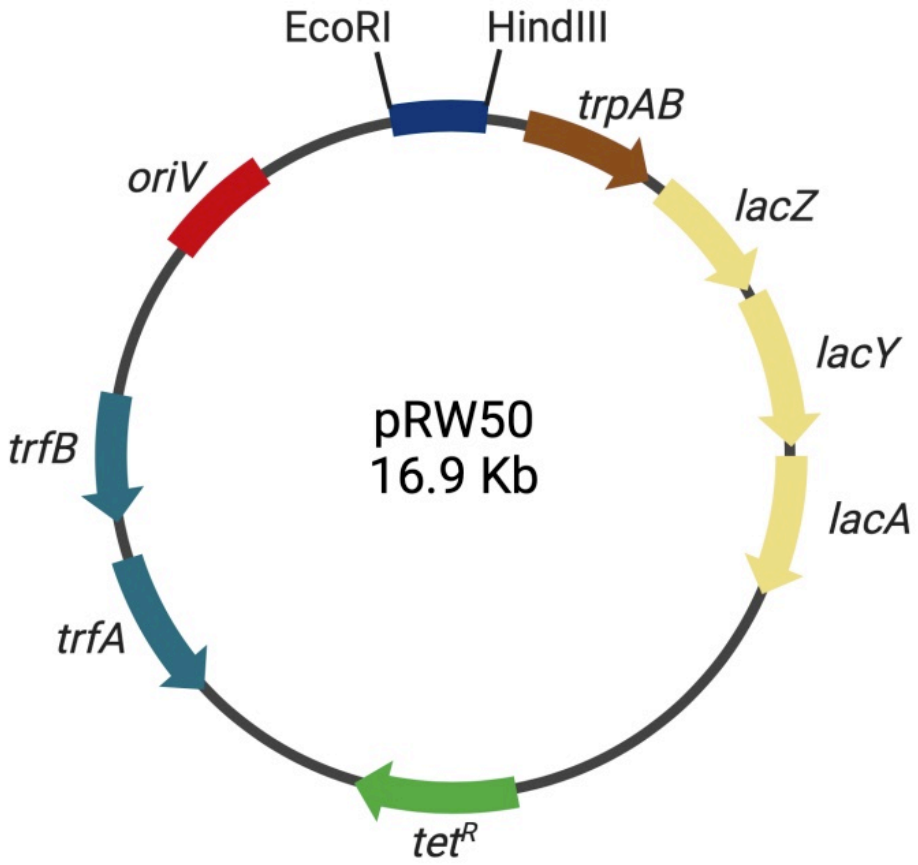


Figure 3.2 The pRW50 low copy number *lac* expression vector

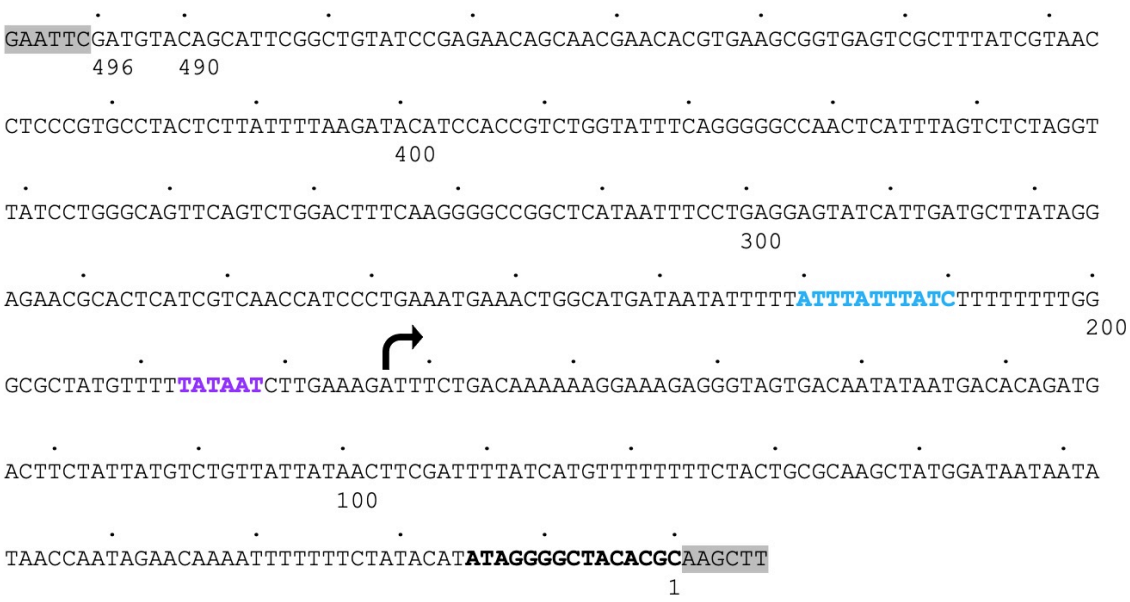
The figure shows a plasmid map of the pRW50 *lacZ* expression vector, located upstream of the *lacZYA* operon are the EcoRI and HindIII restriction sites that were used to clone promoter fragments. The plasmid carries a tetracycline resistance determinant (*tet*^{R}). Created with BioRender.com.

A

042 *afaB* regulatory region

B

C1010-00 *agg4D* regulatory region



C

042 *aafD100* regulatory region

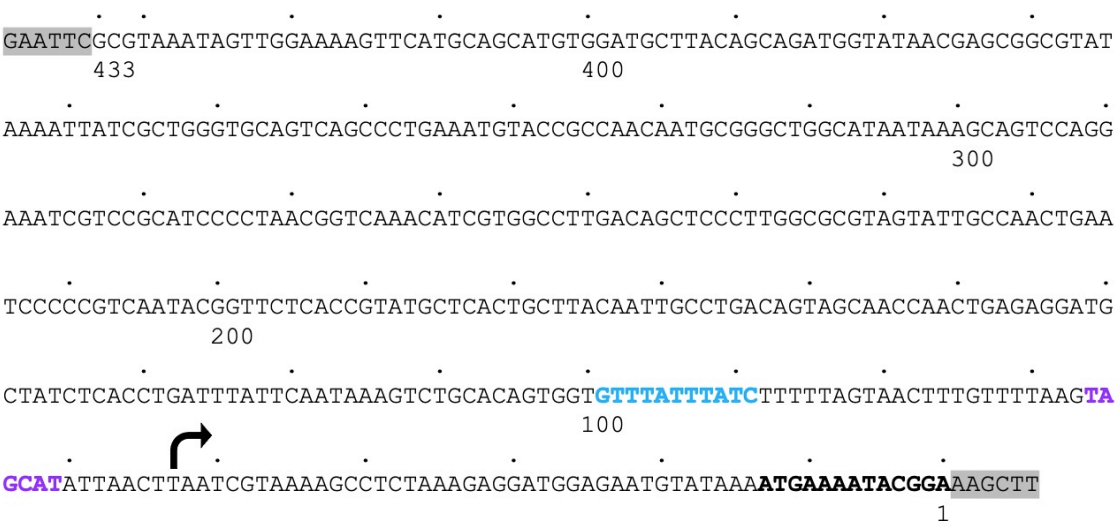


Figure 3.3 DNA base sequence upstream of genes cloned into pRW50

The figure shows the DNA sequence for promoter fragments cloned into pRW50 utilising the EcoRI and HindIII restriction sites, highlighted in grey. The -10 element is highlighted in purple, the DNA binding site for AggR is highlighted in blue and the coding sequence is in bold. The bent arrows indicate the start sites and the direction of transcription.

A. 413 bp sequence of EAEC 042 *afaB* promoter fragment
B. 496 bp sequence of EAEC C1010-00 *agg4D* promoter fragment
C. 433 bp sequence of EAEC 042 *aafD100* promoter fragment.

To measure AggR-dependent activation at target promoters, I exploited the pBAD30 vector (Figure 3.4) (Guzman *et al.*, 1995). The pBAD*aggR* recombinant plasmid contains the *aggR* gene under the control of the *araBAD* promoter (Figure 3.5A). The *aggR* gene was amplified by PCR and primers were utilised that introduced an EcoRI site upstream of the *aggR* gene and XbaI downstream (Figure 3.5B). The *araBAD* promoter can be induced by arabinose, which binds to AraC, resulting in the production of AggR (Guzman *et al.*, 1995).

In the assay system, utilised throughout this research, arabinose induces the production of AggR, which then binds at the target promoter on the pRW50 plasmid, this activates expression of *lacZ*, which produces β-galactosidase (Figure 3.6). The pRW50*afaB* recombinant plasmid was transformed into *E. coli* K-12 strain BW25113 Δ*lac* carrying pBAD30 empty vector or pBAD*aggR* (Figure *3.7*A). Cultures were grown in LB medium, with a duplicate of each sample containing 0.2% (*w/v*) arabinose, until the OD_{650} reached between 0.4 and 0.6. The β-galactosidase activity of the lysed cells was measured to determine the promoter activity. This was repeated for recombinant plasmids pRW50*agg4D* (Figure 3.7B) and pRW50*aafD100* (Figure 3.7C). The β-galactosidase activities can be seen in Figure 3.7, each panel shows the comparison of the recombinant plasmids carrying target promoters in the presence and absence of AggR. In this arabinose-induced system, expression from these promoters was activated by AggR. The apparent strength of each promoter differed, with the highest β-galactosidase activity reported when assaying *aafD100*, with over 16-fold increase over basal levels with pBAD30 empty vector (19121 and 539 units, respectively), and the lowest with *afaB*, with over 13-fold increase over basal

levels (660 and 8 units, respectively). β-galactosidase activity increased almost 13-fold over base levels with *agg4D* (1512 and 38 units, respectively). The results indicate that the *afaB, agg4D* and *aafD100* promoters are activated by AggR. The AggR-activated target promoters have a single functional DNA binding site for AggR, and the activity levels differ for each promoter.

The results also indicate that the β-galactosidase levels in the absence of AggR are higher than the basal levels in cells carrying pBAD30 empty vector. This indicates that expression of AggR from pBAD*aggR* is 'leaky'. While the *araBAD* promoter is considered to be tightly regulated (Rosano and Ceccarelli, 2014), the significant β-galactosidase levels measured in the absence of arabinose, seen in Figure 3.7, must be due to arabinose-independent expression of AggR. This arabinose-independent β-galactosidase activity can differ from promoter to promoter, however, in Figure 3.7 the background levels are comparable. I suggest that AggR binds more tightly to certain promoters, which would result in higher background levels at these promoters than at promoters with weaker AggR binding.

3.2.2 The role of CRP

The current view of a typical AggR-dependent promoter, as proposed in Yasir *et al.* (2019), is quite simple. However, the activity of many bacterial promoters is influenced by more than one transcription factor. Of all the transcription factors in *E. coli*, the cyclic AMP receptor protein, CRP, has been found to make the most interactions (Martinez-Antonio and Collado-Vides, 2003). To investigate whether CRP interacts with AggR at AggR-dependent target promoters, the activity of

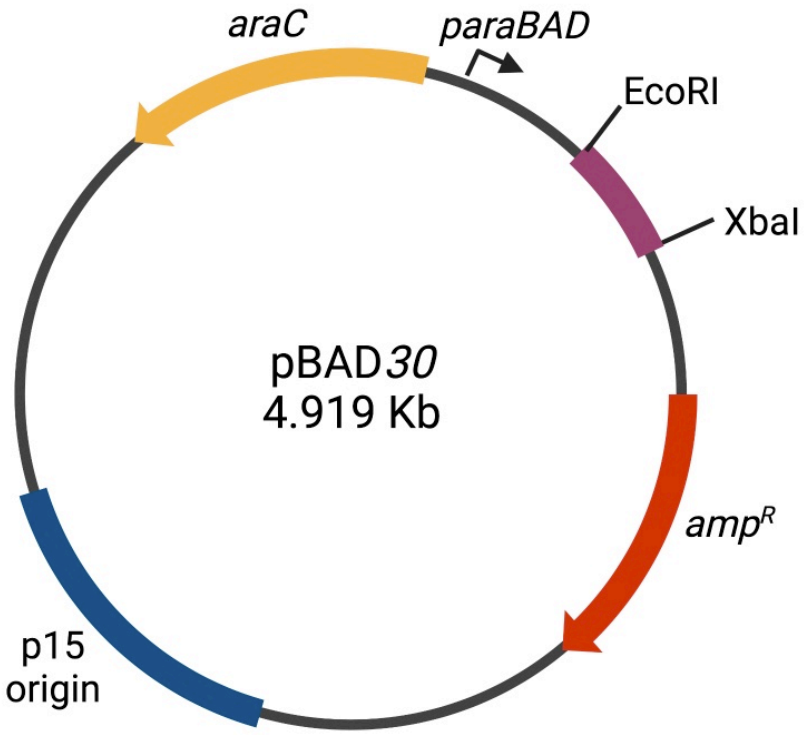
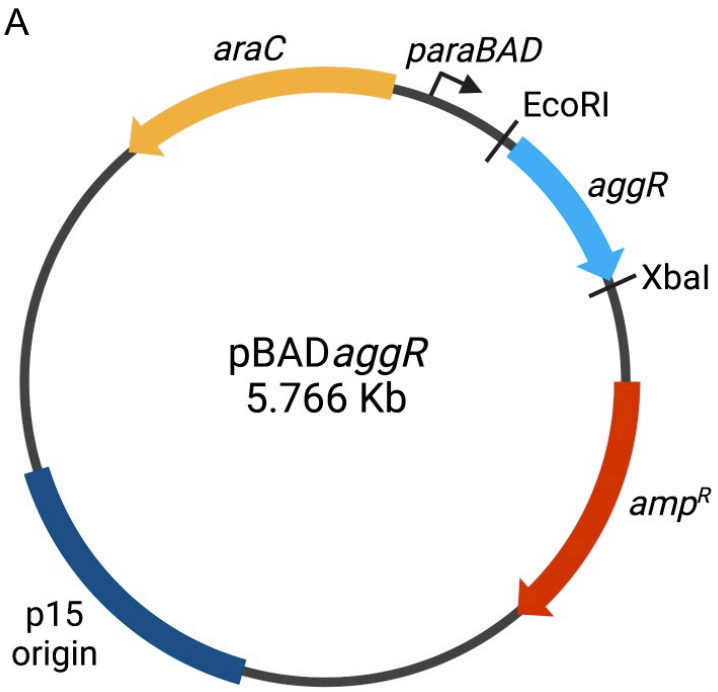


Figure 3.4 pBAD30 vector

The figure shows a plasmid map of the pBAD30 vector, with the gene encoding the arabinose-triggered transcription factor, *araC*, responsible for activation of the *paraBAD* promoter. Downstream of the promoter are the EcoRI and XbaI restriction sites, which are used to clone in gene constructs. The plasmid carries ampicillin resistance (*amp*^{R}). Created with BioRender.com.

A


araC
paraBAD
EcoRI
aggR
XbaI
pBADaggR
5.766 Kb
ampR
p15
origin

B

pBAD*aggR*

Figure 3.5 pBAD*aggR* and the DNA sequence of *aggR* gene cloned into pBAD30

A. The figure illustrates the *aggR* gene under the control of the *paraBAD* promoter, which is induced by arabinose. pBAD*aggR* contains the ampicillin resistance gene, *amp*^{R}. Created with BioRender.com.
B. The DNA sequence of the *aggR* gene cloned into pBAD30 utilising EcoRI and XbaI restriction sites, highlighted in grey. The translation start and stop codons are highlighted in pink.

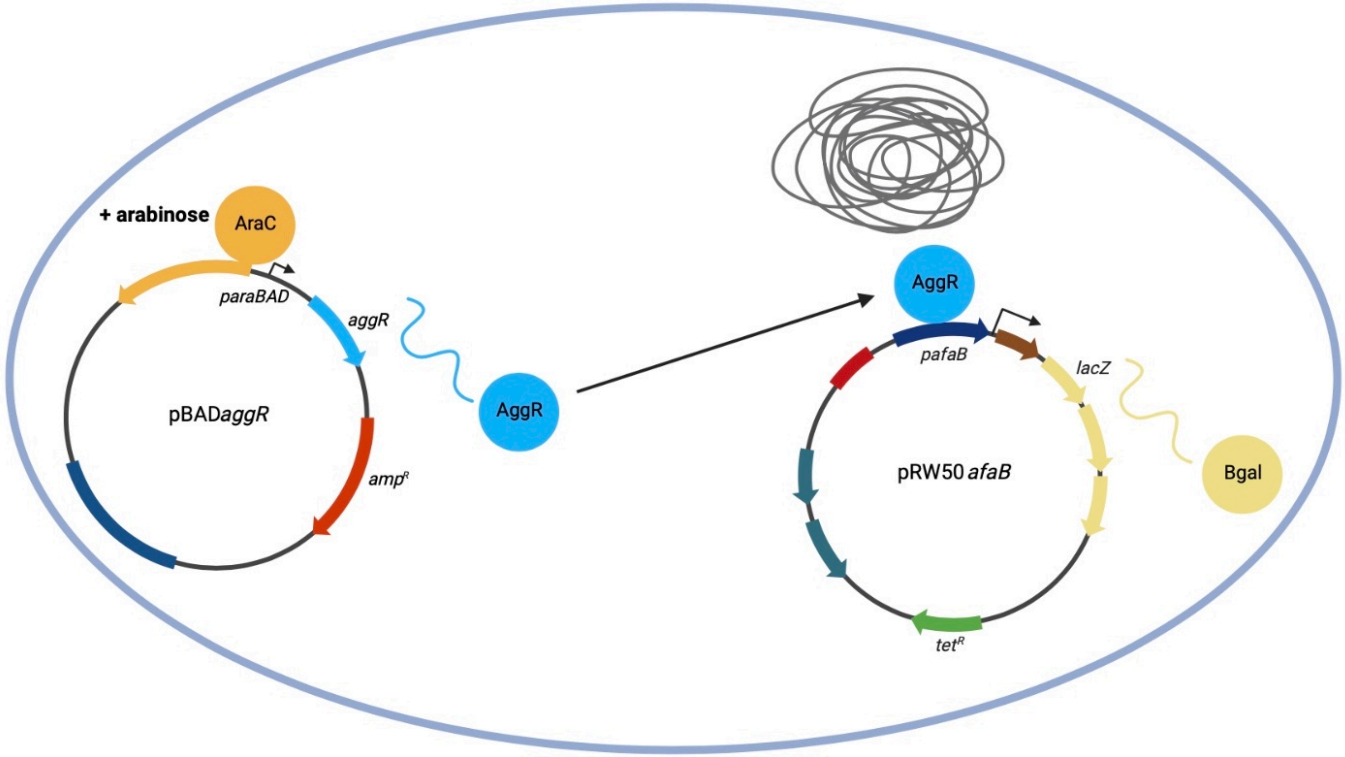
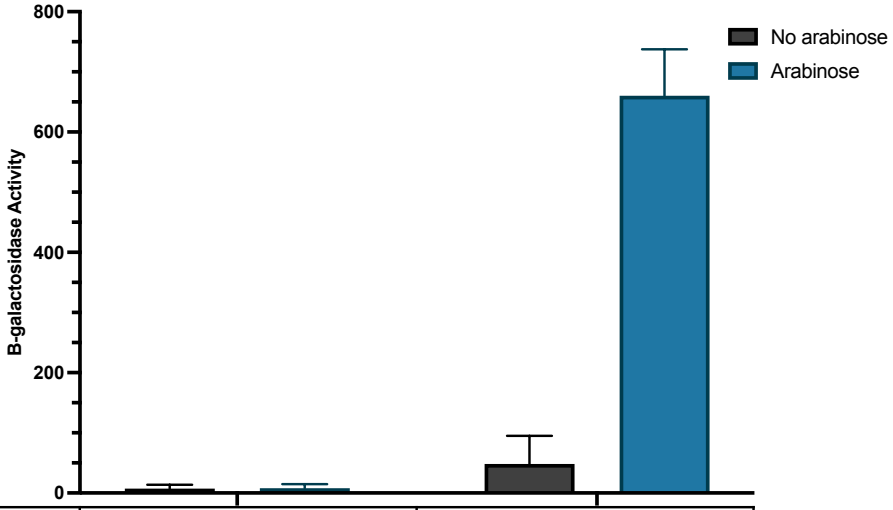


Figure 3.6 A simple assay for AggR-dependent activation at target promoters

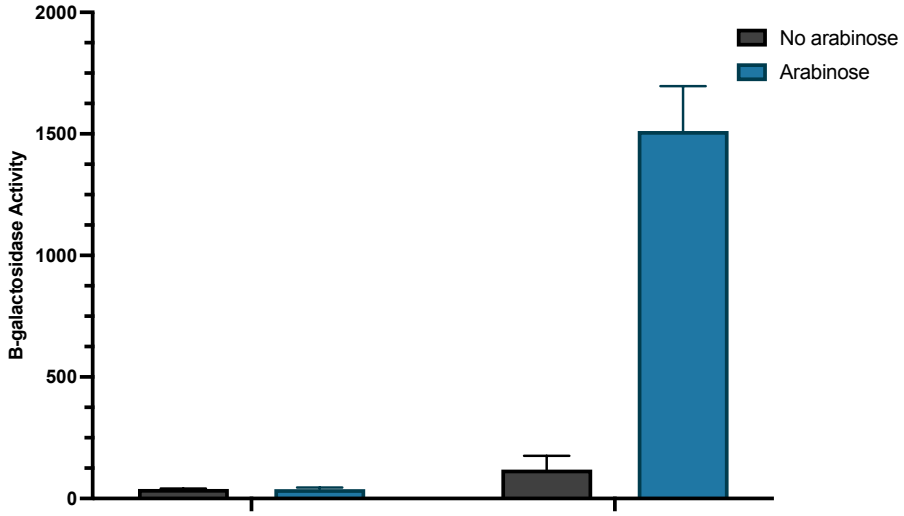
The figure illustrates the induction of the pBAD*aggR* vector by arabinose. Arabinose binds to the *araC* gene, which activates the *paraBAD* promoter and subsequent expression of the *aggR* gene. AggR then binds to the target promoter on the pRW50 vector, activating the expression of *lacZ*, producing β-galactosidase. Created with BioRender.com.

A



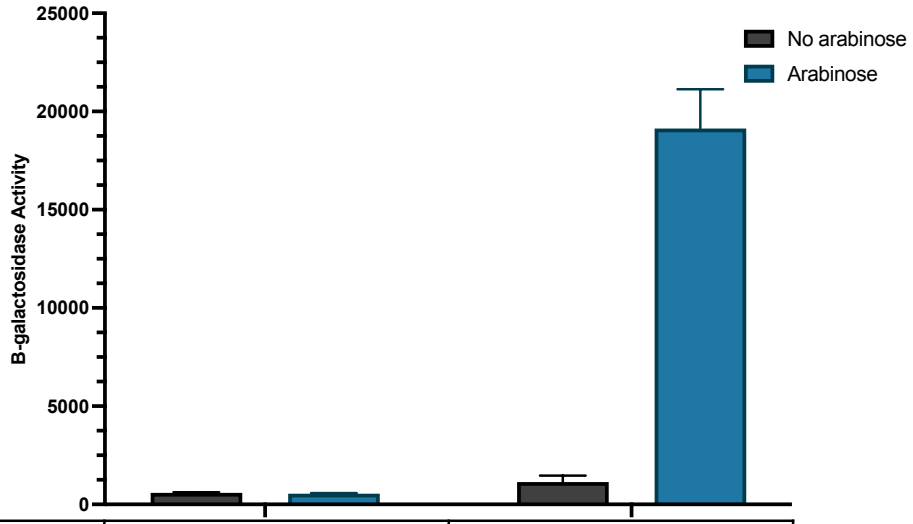
promoter in pRW50	*afaB 042*	*afaB 042*
activator in pBAD	–	AggR 042

B



promoter in pRW50	*agg4D C1010-00*	*agg4D C1010-00*
activator in pBAD	–	AggR 042

C



promoter in pRW50	*aafD100 042*	*aafD100 042*
activator in pBAD	–	AggR 042

Figure 3.7 Measurement of *afaB, agg4D, aafD100* promoter activities

The β-galactosidase activities measured in *E. coli* K-12 BW25113 Δ*lac* with the recombinant plasmids pRW50*afaB* (A), pRW50*agg4D* (B) and pRW50*aafD100* carrying either pBAD30 (empty vector) or pBAD*aggR* 042. The cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. EAEC 042 *afaB*
B. EAEC C1010-00 *agg4D*
C. EAEC 042 *aafD100.*

promoter targets was compared in *crp*+ and Δ*crp* cells. The previous measurements of AggR-dependent activation used the *araBAD* promoter, which, itself, is regulated by CRP. Therefore, here, I used a different AggR-expression system in which *aggR*, together with its own promoter was cloned into the plasmid pLG339 (Figure 3.8A and B) (Stoker *et al.*, 1982). Thus, the AggR regulatory region and *aggR* gene were amplified by PCR from EAEC 042, using primers that introduced an EcoRI restriction site upstream of the amplified sequence and a SalI site downstream (Figure 3.9).

The pLG339*aggR* recombinant plasmid was transformed into *E. coli* K-12 strain M182 Δ*lac* and into *E. coli* K-12 strain M182 Δ*lac* Δ*crp* (Busby *et al.*, 1983), containing either pRW50*afaB* or pRW50*agg4D.* Cells were grown in LB to mid-exponential phase and the β-galactosidase activity was measured from the cell lysates. Figure 3.10 shows that there was no significant difference in induction levels for the *agg4D* promoter following deletion of *crp* (one-way ANOVA, $p<0.0001$; Tukey test, $p>0.05$), however, deletion of *crp* significantly increased induction levels for the *afaB* promoter (Tukey test, $p<0.05$). From this, I conclude that AggR-dependent activation does not require the presence of CRP.

3.3 A bi-directional promoter regulated by AggR: activation and repression

One of the AggR-dependent targets identified through the RNA-seq analysis (Yasir *et al.*, 2019), called EC042_4006 (or *yicS*), is located on the EAEC 042 chromosome and drives expression of the gene for a predicted exporter protein dubbed YicS. YicS has been implicated in biofilm formation as it is located

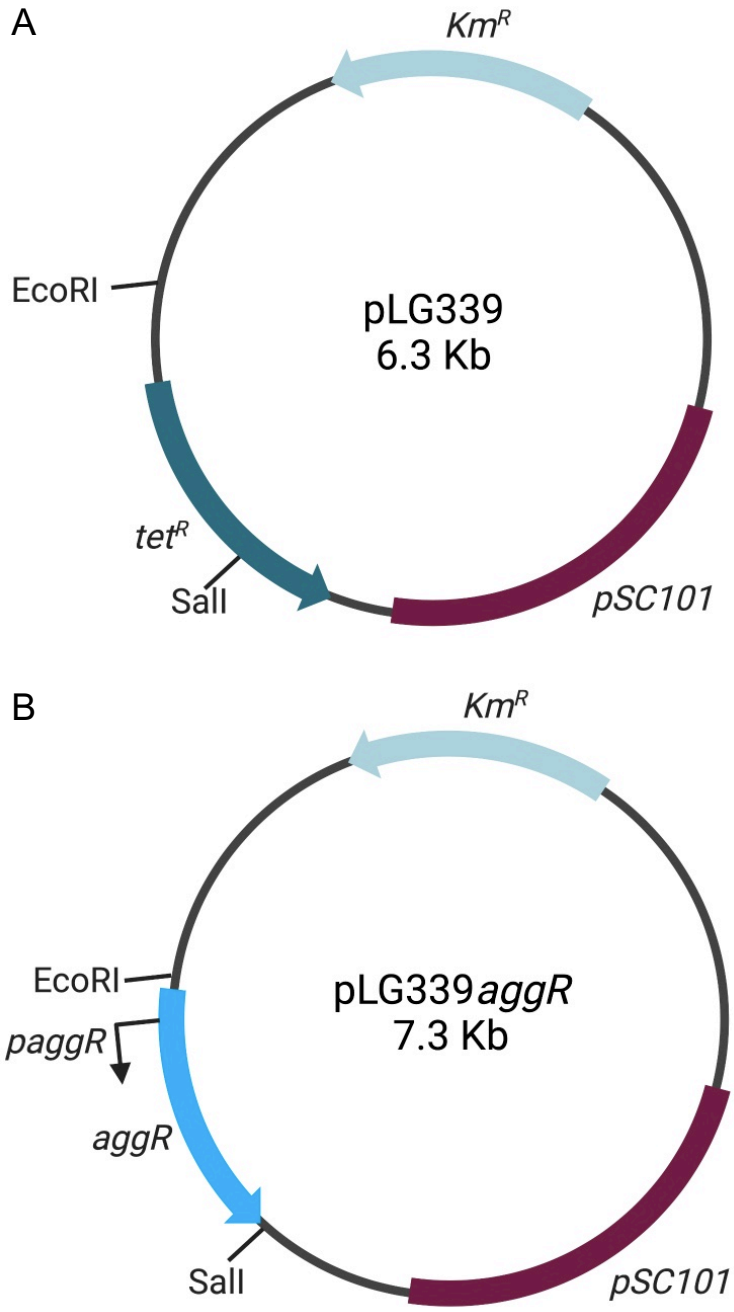


Figure 3.8 pLG339 used as a vector for *aggR*

The figure shows the plasmid maps of the pLG339 vector (A) and pLG339*aggR* (B), the plasmid carries the kanamycin resistance gene, Km^R. Created with BioRender.com.

A. pLG339
B. pLG339*aggR* with the *aggR* regulatory region and gene cloned in using EcoRI and SalI restriction sites.

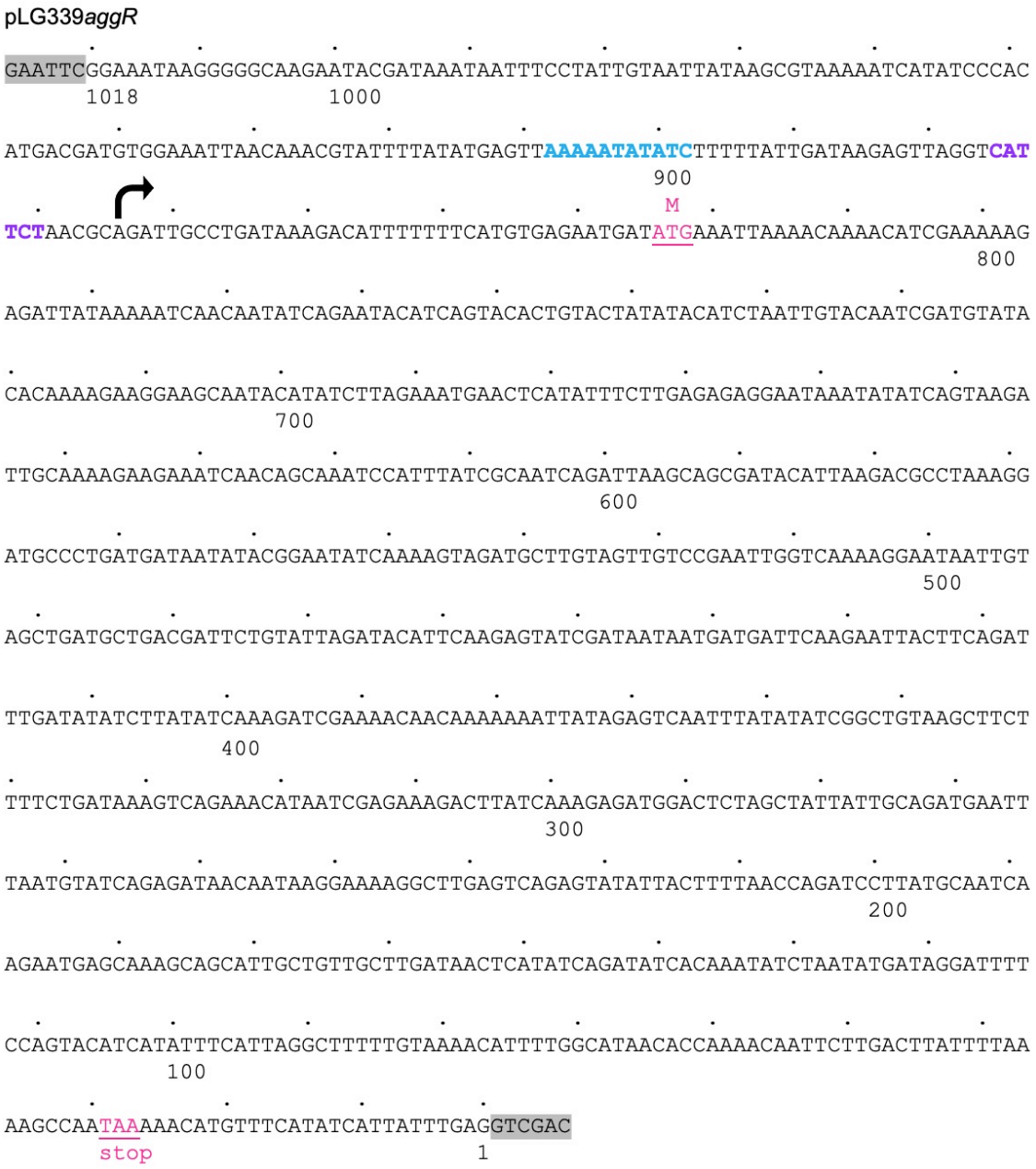
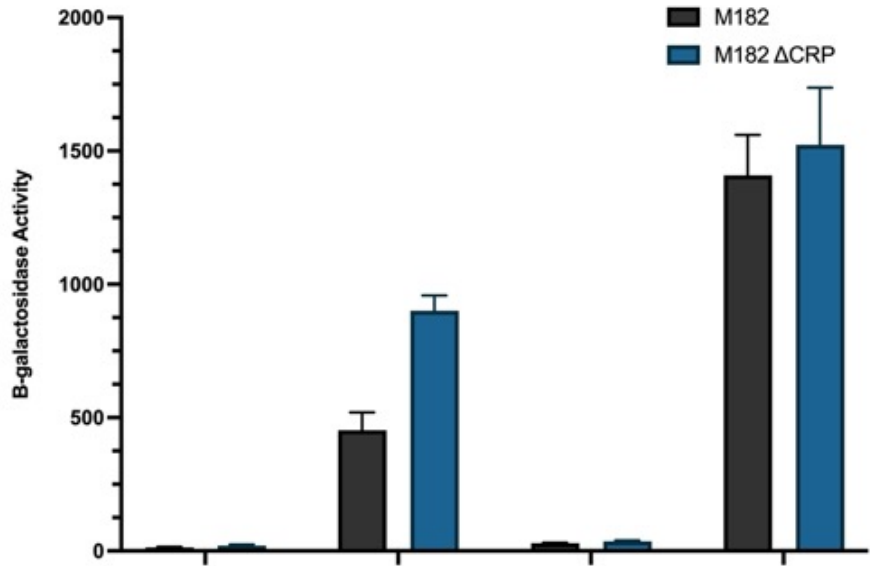


Figure 3.9 The DNA sequence of the *aggR* regulatory region and gene in pLG339

The figure shows the DNA sequence for the *aggR* regulatory region and gene cloned into pLG339 using the EcoRI and SalI restriction sites, highlighted in grey. The DNA binding site for AggR is blue and the -10 element is purple. The start and stop codons are pink.



promoter in pRW50	*afaB*	*afaB*	*agg4D*	*agg4D*
activator in pLG339*	–	AggR 042	–	AggR 042

* activator and regulatory region in pLG339

Figure 3.10 Effects of CRP on AggR-dependent activation of a target promoter

The β-galactosidase activities measured in *E. coli* K-12 M182 Δ*lac* and M182 Δ*lac* Δ*crp* with the recombinant plasmids pRW50*afaB* and pRW50*agg4D* carrying either pLG339 (empty vector) or pLG339*aggR* 042, with the *aggR* gene under the control of its own promoter. The cells were grown in LB, the black bars show the M182 cells, and the blue bars show the M182 Δ*crp* cells. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(7, 40) = 260.2$). A post-hoc Tukey's HSD test showed that the *afaB* promoter activities in M182 and M182 ΔCRP cells carrying pLG339*aggR* 042 were significantly different from each other ($p<0.05$), there was no significance for the *agg4D* promoter activities ($p>0.05$).

adjacent to the gene *csgD*, which has an important role in biofilm matrix production, though the function of YicS is not known (Verma *et al.*, 2018). The *yicS* promoter is found to overlap with the regulatory region for the AggR-regulated *nlpA* promoter, and this arrangement is present in different strains and types of *E. coli* (Verma *et al.*, 2018; Bodero *et al.*, 2007). A single DNA site for AggR has been identified in the intergenic region, which contains overlapping -10 elements. Such an arrangement indicates that the expression of the *yicS* and *nlpA* transcripts are driven by a bi-directional promoter, as illustrated in Figure 3.11 and Figure 3.12. Such promoter regions usually contain an AT-rich segment and overlapping -10 elements. Hence RNAP can bind in either direction to the duplex DNA due to the inherent symmetry, with the -10 element as the driving force (Warman *et al.*, 2021). The bi-directional *yicS-nlpA* promoter in EAEC 042 was also identified in EAEC strain H92/3, which had been isolated from children with acute diarrhoea in Rio de Janeiro, and whole genome sequenced (Rosa *et al.*, 1998). For comparison, the *yicS-nlpA* bi-directional promoter in EAEC H92/3 was isolated and assayed alongside EAEC 042 and *E. coli* K-12.

3.3.1 The *yicS* promoter: activation by AggR and essential elements

To investigate whether AggR-dependent activation at a bi-directional promoter is affected by divergent transcription, the *yicS-nlpA* intergenic regions from EAEC 042, *E. coli* K-12 and EAEC H92/3 were first compared and analysed. Figure 3.12 compares the overlapping bi-directional promoters for *yicS-nlpA* from EAEC 042, H92/3 and *E. coli* K-12. Figure 3.13 shows the alignment of the *yicS* 042, H92/3 and *E. coli* K-12 regulatory regions. Regarding the *yicS* promoter, both EAEC strains have a consensus -10 hexamer (TATAAT), however, the *E. coli* K-12

strain does not (TATAGC). Each of the strains have putative DNA binding sites for AggR that follow the organisation of the consensus AggR site identified previously (Yasir *et al.*, 2019). However, *yicS E. coli* K-12 most closely matches the consensus sequence AWWWWWWTATC (W = A or T), whereas EAEC 042 and H92/3 *yicS* have cytosine at position 2 and 1, respectively, which does not match the consensus of either adenine or thymine at these positions.

To measure *yicS* promoter activity, the *yicS-nlpA* regulatory regions from EAEC 042, EAEC H92/3 and *E. coli* K-12 were amplified by PCR (Figure 3.14), using primers with upstream EcoRI and downstream HindIII restriction sites, and then cloned into pRW50 to generate promoter *lacZ* transcriptional fusions. Recombinant plasmids were transformed into BW25113 cells, containing pBAD30 or pBAD*aggR*. Cells were grown in LB in the presence and absence of arabinose, and β-galactosidase activities were measured. Each pRW50*yicS* plasmid was assayed in the presence of AggR and with an empty vector control. Data in Figure 3.15 show high basal levels when AggR is not present (*yicS* 042, 405 units; *yicS* K-12, 314 units; and *yicS* H92/3, 366 units), this is compared to the relatively low basal levels seen with the control plasmid pRW50*afaB* (13 units). β-galactosidase activity increases in the system containing *yicS* 042 in the presence of AggR, a 7-fold increase (2708 units compared to basal pBAD30 levels, 378 units), indicating that this promoter is activated by AggR. However, in contrast, *yicS* K-12 is not activated by AggR, as there is no increase when AggR production is induced by arabinose (207 units compared to empty vector, 287 units) (Figure 3.15A). When comparing *yicS* 042 and *yicS* H92/3, there is a slight increase in activity when AggR is present, though it is minimal, roughly 1.5-fold

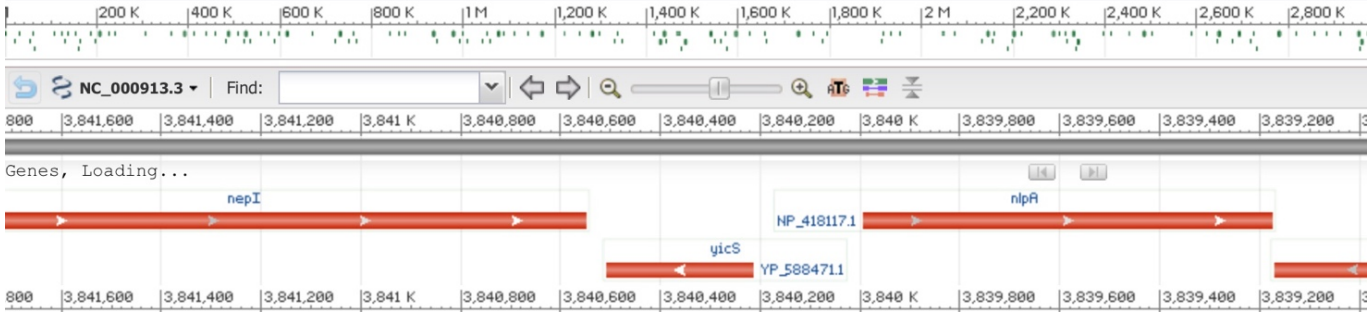


Figure 3.11 *yicS* and *nlpA* bi-directional promoter in *E. coli* K-12

The figure shows a graphic of the divergent genes (red) *yicS* and *nlpA* in *E. coli* K-12. Adapted from Rangwala *et al.* (2021).

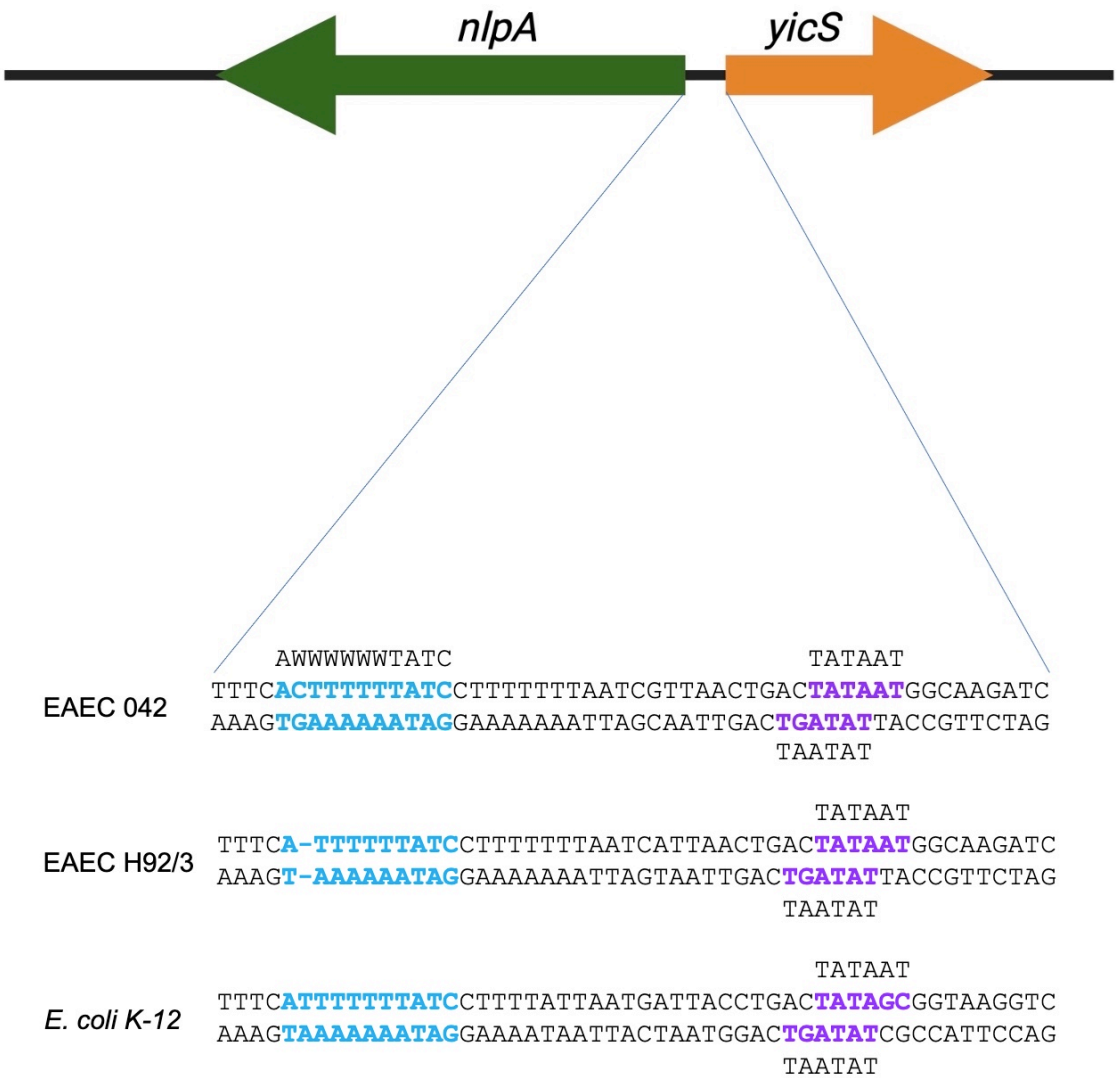


Figure 3.12 *yicS-nlpA* bi-directional promoter in EAEC 042, H92/3 and *E. coli* K-12

The figure shows the base sequence of the overlapping *yicS* and *nlpA* promoters from EAEC 042, EAEC H92/3 and *E. coli* K-12, with the *yicS* promoter elements highlighted on the top strand and *nlpA* elements on the bottom strand. The -10 elements are purple and the DNA binding sites for AggR are blue, the consensus sequences are included (W=A/T). Created with BioRender.com.

Figure 3.13 Alignment of the *yicS* regulatory regions

The figure shows the alignment of the regulatory regions from *yicS* 042 (EAEC strain 042), *yicS* H92/3 (EAEC strain H92/3) and *yicS* K-12 (*E. coli* K-12). The AggR binding sites are highlighted in blue and the -10 hexamers in purple.

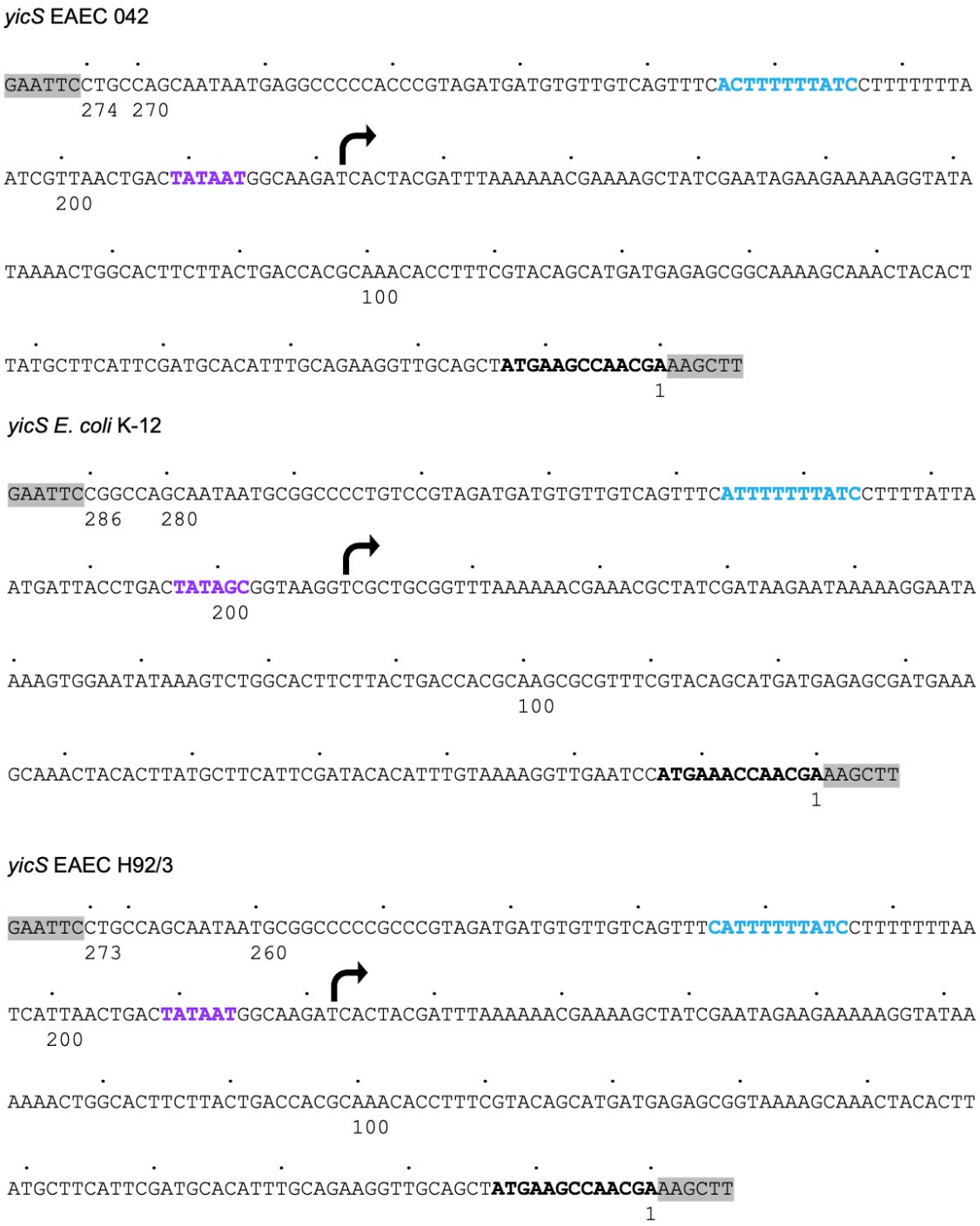
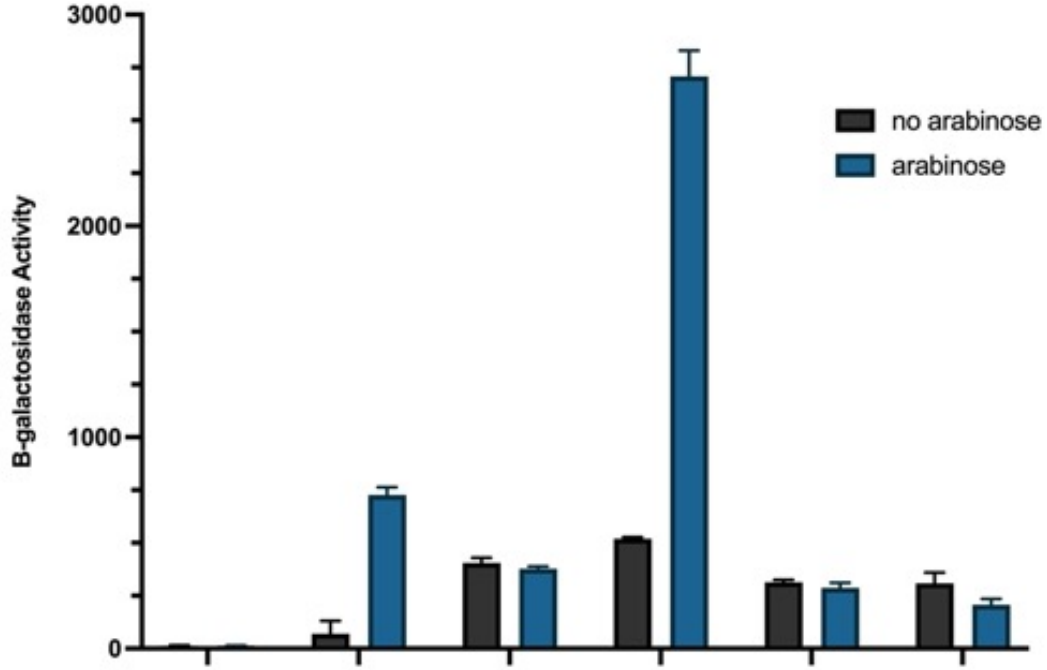


Figure 3.14 *yicS* regulatory regions cloned in pRW50

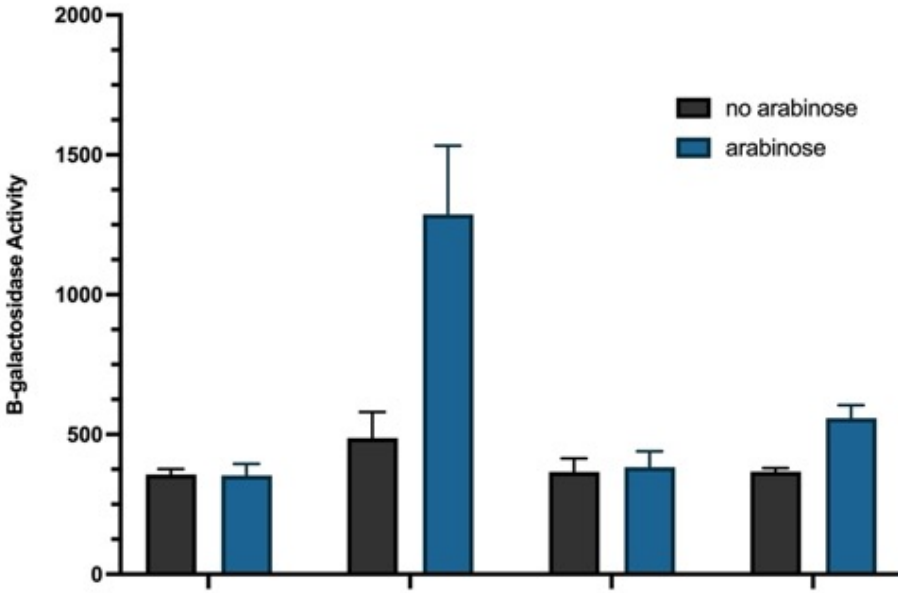
The DNA sequences for the *yicS* promoter fragments from EAEC 042, EAEC H92/3 and *E. coli* K-12. The fragments contain an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The DNA binding site for AggR is blue, the -10 hexamer is purple, and the coding sequence is in bold. The bent arrows indicate the start sites and the direction of transcription.

A



promoter in pRW50	*afaB 042*	*afaB 042*	*yicS 042*	*yicS 042*	*yicS K-12*	*yicS K-12*
activator in pBAD	_	AggR 042	_	AggR 042	_	AggR 042

B



promoter in pRW50	*yicS 042*	*yicS 042*	*yicS H92/3*	*yicS H92/3*
activator in pBAD	_	AggR 042	_	AggR 042

Figure 3.15 Measurement of promoter activities: *yicS* promoters from *E. coli* K-12, EAEC 042 and EAEC H92/3

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* containing pRW50*afaB* 042, pRW50*yicS* 042 or pRW50*yicS* K-12 (A), and pRW50*yicS* 042 or pRW50*yicS* H92/3 (B), carrying either pBAD30 (empty vector) or pBAD*aggR* 042. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. The cells containing pRW50*afaB* EAEC strain 042, pRW50*yicS* EAEC strain 042 or pRW50*yicS* *E. coli* K-12 carrying either pBAD30 (empty vector) or pBAD*aggR* 042.
B. The cells containing pRW50*yicS* EAEC strain 042 or pRW50*yicS* EAEC strain H92/3 carrying either pBAD30 (empty vector) or pBAD*aggR* 042.

(559 units compared to empty vector, 383 units), particularly when considering the high background levels of activity (Figure 3.15B). From these data, I conclude that AggR activates the *yicS* 042 promoter but there is minimal activation for the H92/3 *yicS* promoter and none for the promoter from K-12.To investigate the importance of different elements in the cloned *yicS* promoters, a series of mutations was made (Figure 3.16). Firstly, the 042 *yicS* promoter, which is activated by AggR, was subjected to mutations in the proposed DNA binding site for AggR and the -10 element. Thus, first, these were mutated to be different from consensus. Next, I noted that the K-12 *yicS* promoter, which failed to be activated by AggR, carries a consensus DNA site for AggR but a suboptimal -10 hexamer element (TATAGC rather than TATAAT). Hence, the K-12 *yicS* -10 hexamer was mutated to match consensus.

Constructs were created using the Megaprimer method of PCR amplification, briefly, the first PCR included the forward primer carrying the mutation and the pRW50 reverse primer, with pRW50*yicS* as the template. The purified product was used in the next PCR as a megaprimer, this time pRW50 forward primer was used. The purified product was cut with EcoRI and HindIII restriction enzymes and ligated in pRW50. Each mutated promoter sequence was confirmed by sequencing.

The mutant pRW50 constructs carrying the mutated promoters were transformed into BW25113 cells with pBAD*aggR* or pBAD30, and β-galactosidase activities were measured as before. Figure 3.17A shows a comparison of measured activities of the 042 *yicS* promoter constructs. The mutant 042 *yicS* promoters have similar expression levels in the presence and absence of AggR. Thus,

knocking out either the -10 region or the AggR DNA binding site negatively affects expression levels, and these elements are individually important for expression. Again, high levels of background expression in the absence of AggR were found, suggesting the presence of a promoter that is not AggR-dependent.

Remarkably, for the K-12 *yicS* promoter mutant the consensus -10 element conferred AggR-induction as expression levels increased, almost 12-fold, in the presence of arabinose (13227 units compared to empty vector basal levels, 1119 units) (Figure 3.17B). This indicates the importance of the -10 element for AggR-dependent activation. The total expression levels with the *yicS* K-12 mutant were also significantly higher than *yicS* 042 wild-type in Figure 3.17A (13227 and 2474 units, respectively; Tukey test, $p<0.05$), indicating that the *yicS* K-12 mutant promoter was more strongly activated by AggR. This is likely due to *yicS* K-12 containing a better AggR binding site (ATTTTTTTATC) than *yicS* 042, which does not match the consensus AggR DNA binding site, with cytosine at the second position as opposed to adenine or thymine. The AggR-independent basal levels with the *yicS* K-12 promoter (280 units) were comparable to basal levels in Figure 3.15 (approximately 287 units). For the mutant *yicS* K-12 promoter, the background levels were approximately 8% of the AggR-induced expression levels, whereas for *yicS* 042, the background levels were approximately 14%, which further suggests the presence of a downstream promoter.

3.3.2 A potential downstream *yicS* promoter

To determine whether there is an AggR-independent promoter downstream from the AggR-dependent *yicS* promoter, the downstream sequences were cloned into pRW50, in order to quantify any activity. This is illustrated in Figure 3.18, where

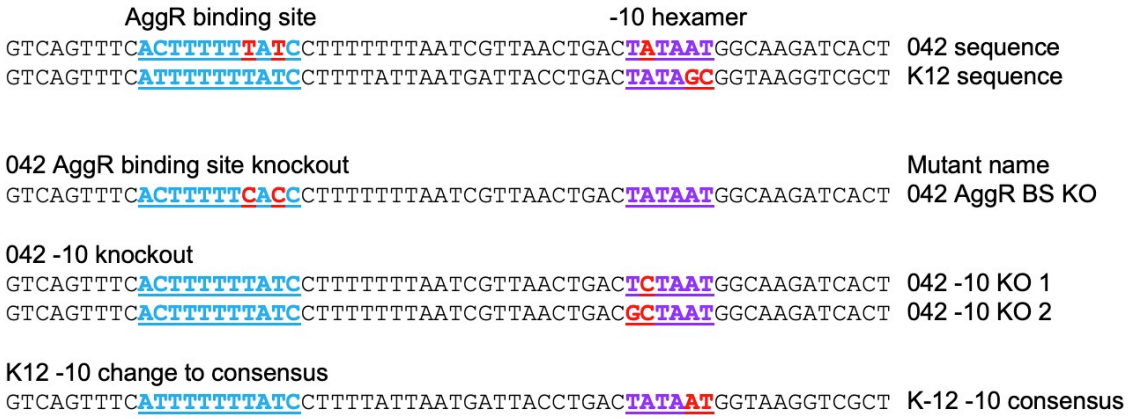
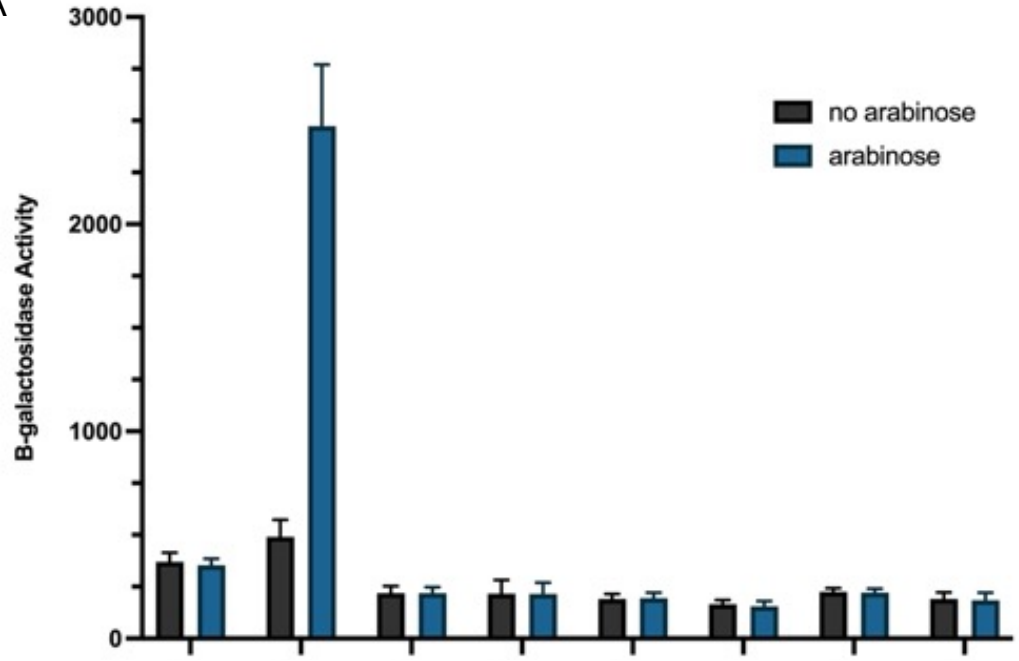


Figure 3.16 *yicS* promoter mutations

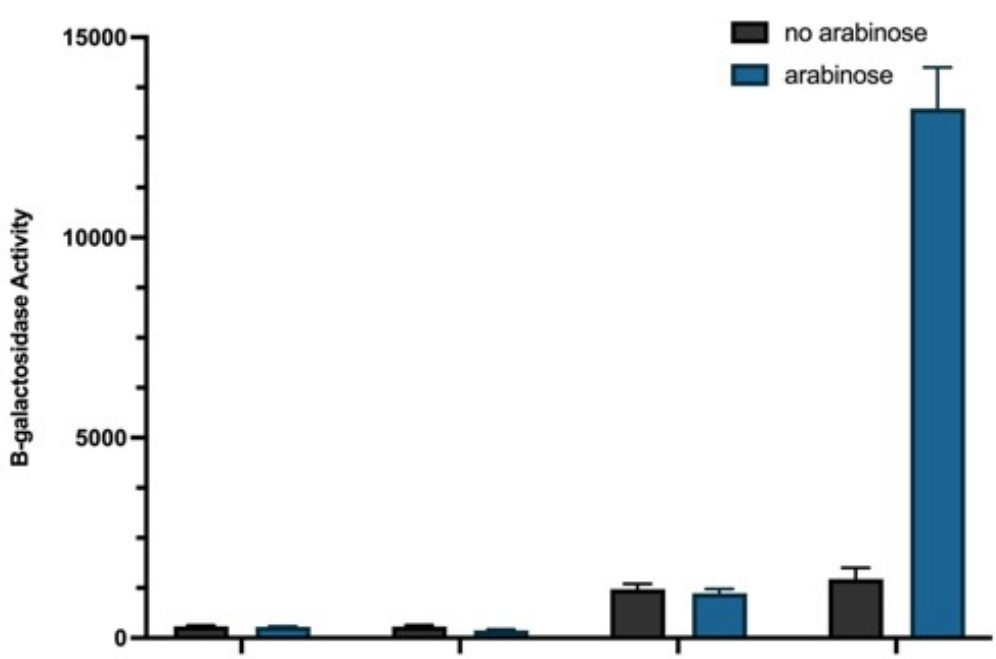
The figure shows the DNA sequence of the *yicS* promoter sequence from EAEC 042 and *E. coli* K-12. DNA sites for AggR and promoter -10 elements are highlighted in blue and purple respectively. Bases where mutations were made are highlighted in red and mutant names are included alongside the corresponding sequence.

A



activator in pBAD	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
promoter in pRW50	*yicS 042*	*yicS 042*	*-10 KO 1*	*-10 KO 1*	*-10 KO 2*	*-10 KO 2*	*AggR BS KO*	*AggR BS KO*

B



activator in pBAD	–	AggR 042	–	AggR 042
promoter in pRW50	*yicS K-12*	*yicS K-12*	*K-12 -10 consensus*	*K-12 -10 consensus*

Figure 3.17 Mutational analysis of the *yicS* promoter from EAEC 042 and *E. coli* K-12

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* containing pRW50*yicS* 042 or pRW50*yicS* 042 mutants (A), and pRW50*yicS* K-12 or pRW50*yicS* K-12 mutants (B), carrying either pBAD30 (empty vector) or pBAD*aggR* 042. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. The cells containing pRW50*yicS* EAEC strain 042, pRW50*yicS-10KO1*, pRW50*yicS-10KO2* or pRW50*yicS-AggR-BS-KO* carrying either pBAD30 (empty vector) or pBAD*aggR* 042. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 80) = 272.3$). A post-hoc Tukey's HSD test showed that the mutant promoter activities were significantly different from the wild-type *yicS* 042 promoter activity in the presence of arabinose ($p<0.05$).
B. The cells containing pRW50*yicS E. coli* K-12 or pRW50*yicS-K12-10consensus* carrying either pBAD30 (empty vector) or pBAD*aggR* 042. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(7, 40) = 821.2$). A post-hoc Tukey's HSD test showed that the mutant promoter activity was significantly different from the wild-type *yicS* K-12 promoter activity in the presence of arabinose ($p<0.05$).

the potential -10 element is highlighted. Data in Figure 3.19 compares promoter activities measured with the full-length promoter fragment containing the AggR-dependent *yicS* promoter (labelled *yicS*) and the shorter fragment carrying only the downstream sequence (labelled P2). Taking just the downstream region of pRW50*yicS* 042 removes AggR-dependence, and activity levels are similar for the downstream promoter and background levels of 042 *yicS* promoter activity (Figure 3.19A).

For *yicS* K-12, expression levels were higher when the downstream promoter was assayed without the upstream AggR-independent promoter, the activity was much higher than the background levels for *yicS* K-12 in the presence and absence of AggR (Figure 3.19B). Finally, the β-galactosidase activity was measured in the absence of *aggR*. For *yicS* 042, the activity of the downstream promoter (41-fold increase, 275 units, compared to empty vector, 7 units) was not greatly affected by upstream sequences carrying the AggR-dependent promoter (44-fold increase, 376 units). In contrast, for *yicS* K-12, activity of the downstream promoter (90-fold increase, 598 units) was reduced by the inclusion of the upstream sequences (44-fold increase, 290 units) (Figure 3.19C). This is most likely due to subtle base difference between the K12 and 042 *yicS-nlpA* intergenic regions. Additionally, the results indicate that the downstream promoters do not interfere with expression of the *yicS-nlpA* bi-directional promoter. In any case, my results indicate the presence of an AggR-independent promoter in this region which explains the high background levels of expression that are independent of AggR.

A

yicS(042)

CTGCCAGCAATAATGAGGCCCCCACCCGTAGATGATGTGTTGTCAGTTTC**ACTTTTTTATC**CTTTTTTTAATCGTT

AACTGAC**TATAAT**GGCAAGATCACTACGATTTAAAAAACGAAAAGCTATCGAATAGAAGAAAAAGGTATATAAAAC

TGGCACTTCTTACTGACCACGCAAACACCTTTCGTACAGCATGATGAGAGCGGCAAAAGCAAACTACACTTATGCT

TCATTCGATGCACATTTGCAGAAGGTTGCAGCT**ATGAAGCCAACGA**

yicS(K-12)

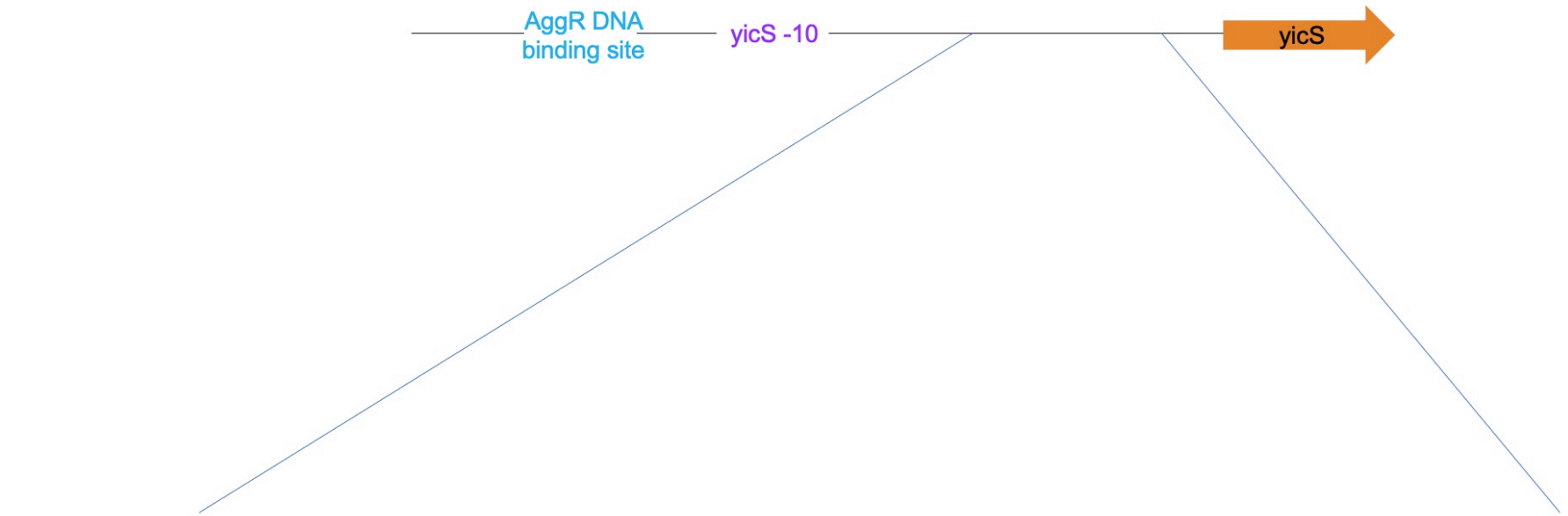
CGGCCAGCAATAATGCGGCCCCTGTCCGTAGATGATGTGTTGTCAGTTTC**ATTTTTTTATC**CTTTTATTAATGATT

ACCTGAC**TATAGC**GGTAAGGTCGCTGCGGTTTAAAAAACGAAACGCTATCGATAAGAATAAAAAGGAATAAAAGTG

GAATATAAAGTCTGGCACTTCTTACTGACCACGCAAGCGCGTTTCGTACAGCATGATGAGAGCGATGAAAGCAAAC

TACACTTATGCTTCATTCGATACACATTTGTAAAAGGTTGAATCC**ATGAAACCAACGA**

B


AggR DNA
binding site
yicS -10
yicS
yicS P2 042 TTTCGTACAGCATGATGAGAGCGGCAAAAGCAAACTACACTTATGCTTCATTCGATGCACATTTGCAGAAGGTTGCAGCTATGAAGCCAACGA
yicS P2 K12 TTTCGTACAGCATGATGAGAGCGATGAAAGCAAACTACACTTATGCTTCATTCGATACACATTTGTAAAAGGTTGAATCCATGAAACCAACGA

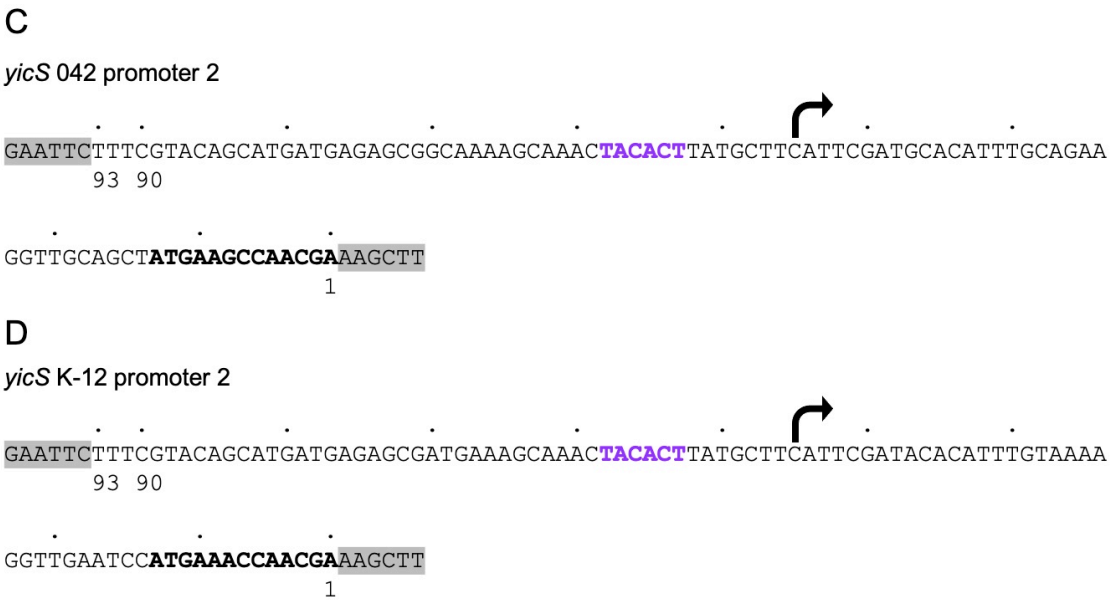
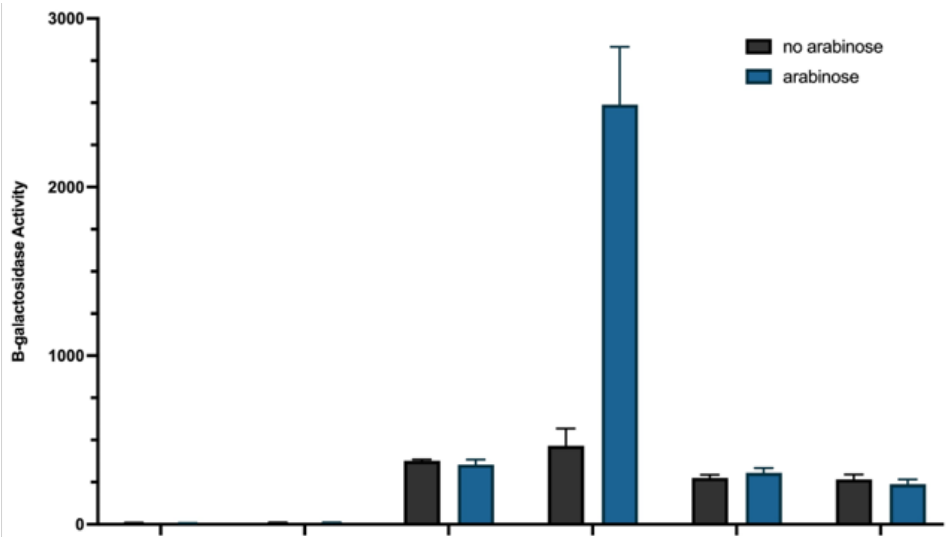


Figure 3.18 DNA sequence of the *yicS* downstream regulatory region

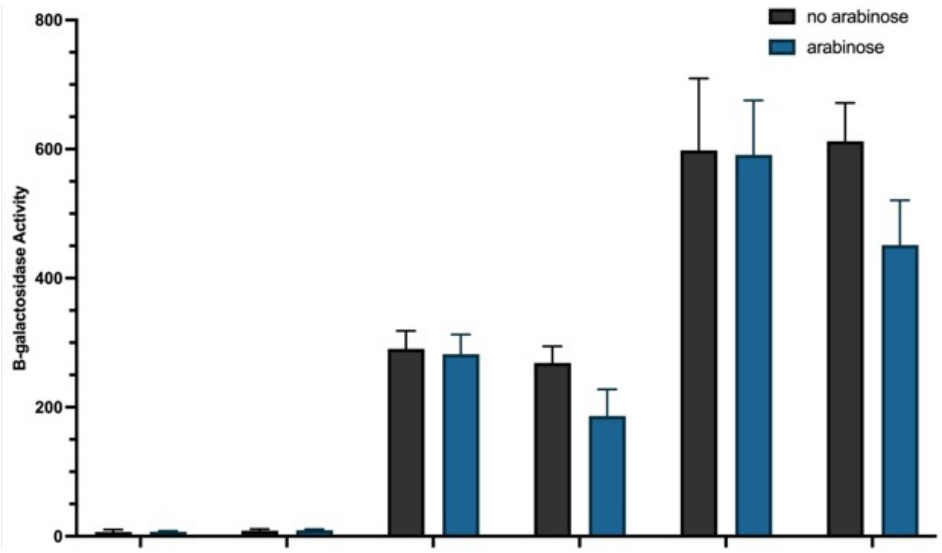
A. The DNA sequence of the *yicS* 042 and *yicS* K-12 promoter fragments that were previously cloned into pRW50, the DNA sequence downstream of the AggR-dependent promoter to be investigated is underlined (adapted from Figure 3.14).
B. The panel illustrates the downstream region of the *yicS* promoter, excluding the AggR-dependent promoter, from *E. coli* K-12 and EAEC 042.
C. The DNA sequence of the *yicS* EAEC 042, *yicS*(042) promoter 2, that was cloned into pRW50. The EcoRI and HindIII sites are highlighted in grey. The predicted -10 element is highlighted in purple, the coding sequence is in bold and the bent arrow indicates the start site and the direction of transcription.
D. The DNA sequence of the *yicS E. coli* K-12, *yicS*(K-12) promoter 2, that was cloned into pRW50. The EcoRI and HindIII sites are highlighted in grey. The predicted -10 element is highlighted in purple, the coding sequence is in bold and the bent arrow indicates the start site and the direction of transcription.

A



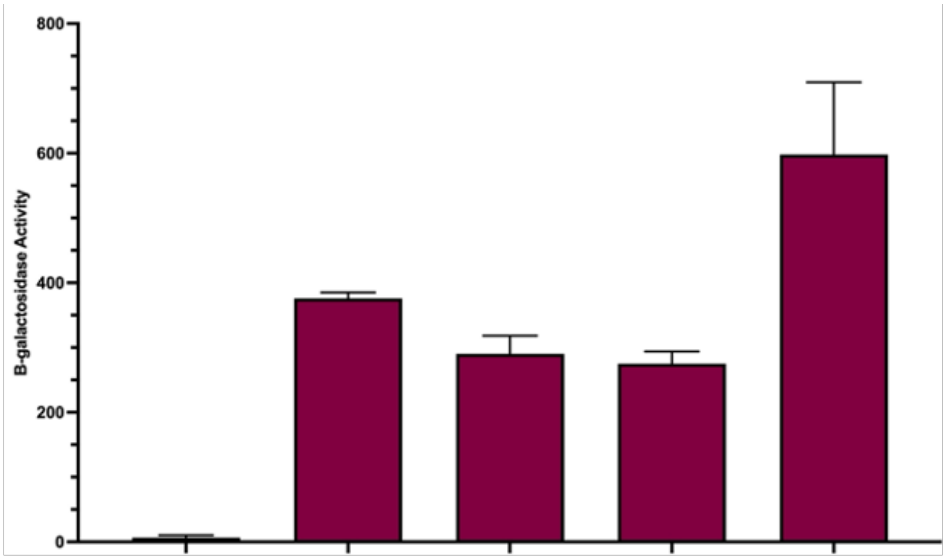
promoter in pRW50	_	_	*yicS 042*	*yicS 042*	*P2 042*	*P2 042*
activator in pBAD	_	AggR 042	_	AggR 042	_	AggR 042

B



promoter in pRW50	_	_	*yicS K-12*	*yicS K-12*	*P2 K-12*	*P2 K-12*
activator in pBAD	_	AggR 042	_	AggR 042	_	AggR 042

C



promoter in pRW50	–	*yicS 042*	*yicS K-12*	*P2 042*	*P2 K-12*

Figure 3.19 Identifying an AggR-independent promoter

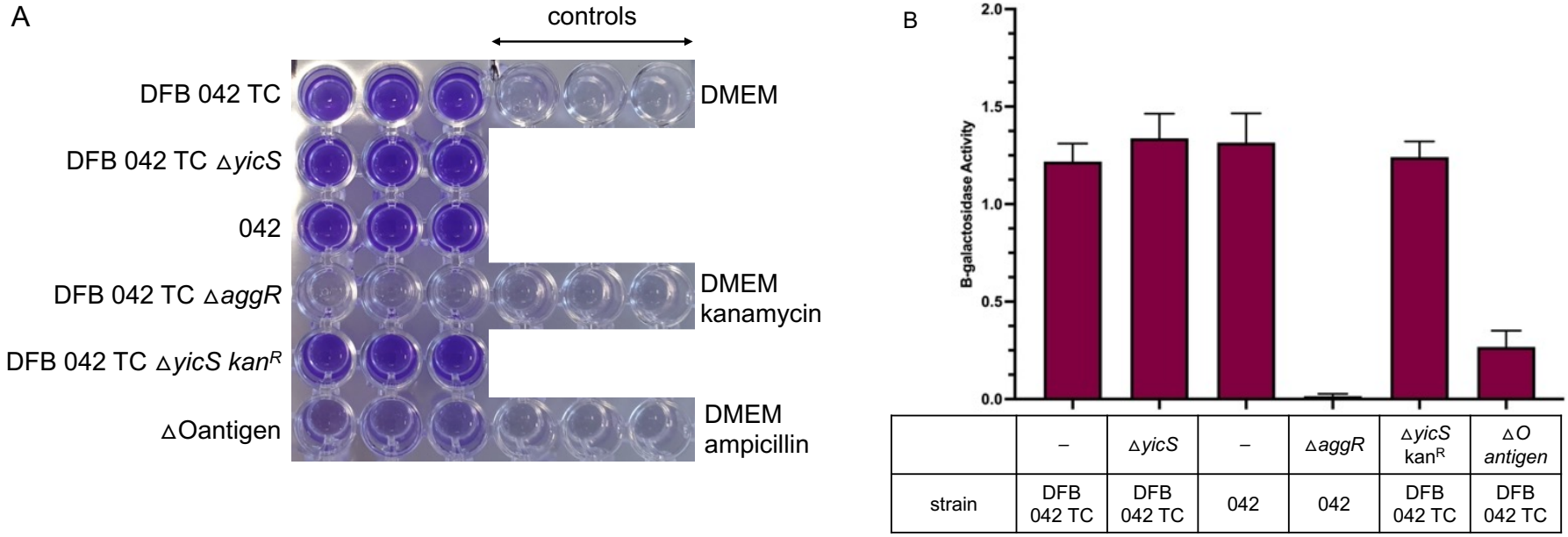
The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* containing either pRW50*yicS* 042 or pRW50*yicS* 042 promoter 2 carrying either pBAD30 (empty vector) or pBAD*aggR* 042 (A), pRW50*yicS* K-12 or pRW50*yicS* K-12 promoter 2 carrying either pBAD30 (empty vector) or pBAD*aggR* 042 (B), and pRW50*yicS* 042, pRW50*yicS* 042 promoter 2, pRW50*yicS* K-12 or pRW50*yicS* K-12 promoter 2 (C). Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. The cells containing either pRW50 (empty vector), pRW50*yicS* EAEC strain 042 or pRW50*yicS promoter 2* (P2) EAEC strain 042 carrying either pBAD30 (empty vector) or pBAD*aggR* 042.
B. The cells containing either pRW50 (empty vector), pRW50*yicS E. coli* K-12 or pRW50*yicS promoter 2* (P2) *E. coli* K-12 carrying either pBAD30 (empty vector) or pBAD*aggR* 042.
C. The cells containing either pRW50 (empty vector), pRW50*yicS* EAEC strain 042, pRW50*yicS promoter 2* (P2) EAEC strain 042, pRW50*yicS E. coli* K-12 or pRW50*yicS promoter 2* (P2) *E. coli* K-12.

3.3.3 Investigating the function of YicS in biofilms

As the function of the predicted exporter protein, YicS, is not known, I wanted to determine whether deletion of the *yicS* gene in EAEC would affect biofilm formation. This was done by gene doctoring (Lee *et al.*, 2009) using the pDOC-Kplasmid, which carries a kanamycin resistant cassette flanked by Flp recombinase restriction sites and two cloning regions on either side that are homologous to regions upstream and downstream of the *yicS* gene. Briefly, the kanamycin cassette was transferred onto the chromosome, as recombination occurred between homologous regions on the DNA substrate and the chromosome, thereby replacing the *yicS* gene and conferring kanamycin resistance. The kanamycin cassette was then excised using the Flp recombinase sites using plasmid pCP20.

Crystal violet biofilm assays were carried out, and the absorbance at OD_{595} was recorded, in order to compare biofilm formation by my YicS-deletion strain with the original EAEC 042. In Figure 3.20A, the microtitre plate is shown with the purple colour in the well indicating biofilm formation, and the OD_{595} is shown graphically in Figure 3.20B. Both the Δ*yicS* strain, and the Δ*yicS* strain with the kanamycin cassette remaining, were assayed, and the results showed that removing the *yicS* gene does not affect the ability of EAEC 042 to produce biofilm. This indicates that YicS does not appear to have a role in biofilm formation. In contrast, removing the *aggR* gene drastically reduces the level of biofilm formed, an 84-fold reduction, and removing O-antigen had a similar effect though to a lesser extent, over 4-fold reduction. O-antigen is typically found in pathogenic and



	–	Δ*yicS*	–	Δ*aggR*	Δ*yicS* kan^R	Δ*O antigen*
strain	DFB 042 TC	DFB 042 TC	042	042	DFB 042 TC	DFB 042 TC

Figure 3.20 Biofilm assay analysis of *ΔyicS*

The crystal violet biofilm assays compare the effects of deleting genes from EAEC 042 on the production of biofilm. The controls were EAEC strain 042 and DFB 042 TC. Δ indicates the removal of a gene, Δ*yicS* and Δ*yicS* kan^R for yicS removal and with the kanamycin cassette respectively, and Δ*aggR* and Δ*O antigen* for the deletion of the genes for *aggR* and the *O antigen* respectively. Each sample was included with biological triplicates.

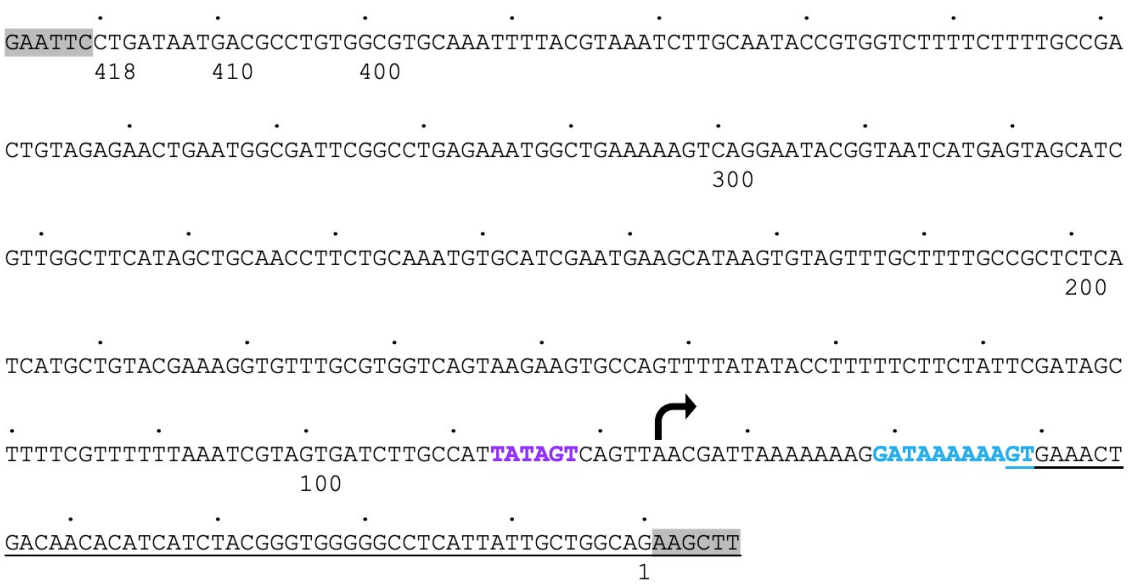
A. A photo of the 96-well plate, with biofilm production indicated by the intensity of the purple from the crystal violet stain. The buffer and antibiotic compositions for the different samples are included on the right, with DMEM media only for the first 3 rows of samples, DMEM with kanamycin for the next two rows and DMEM with ampicillin for the final row of samples.
B. The absorbance at OD_{595} was measured for each well of the 96-well plate, the results are calculated from the means of three independent samples and error bars represent the standard deviation.

lab *E. coli* strains and has a role in pathogenesis, shown to be involved in biofilm formation (Kumar *et al.*, 2015).

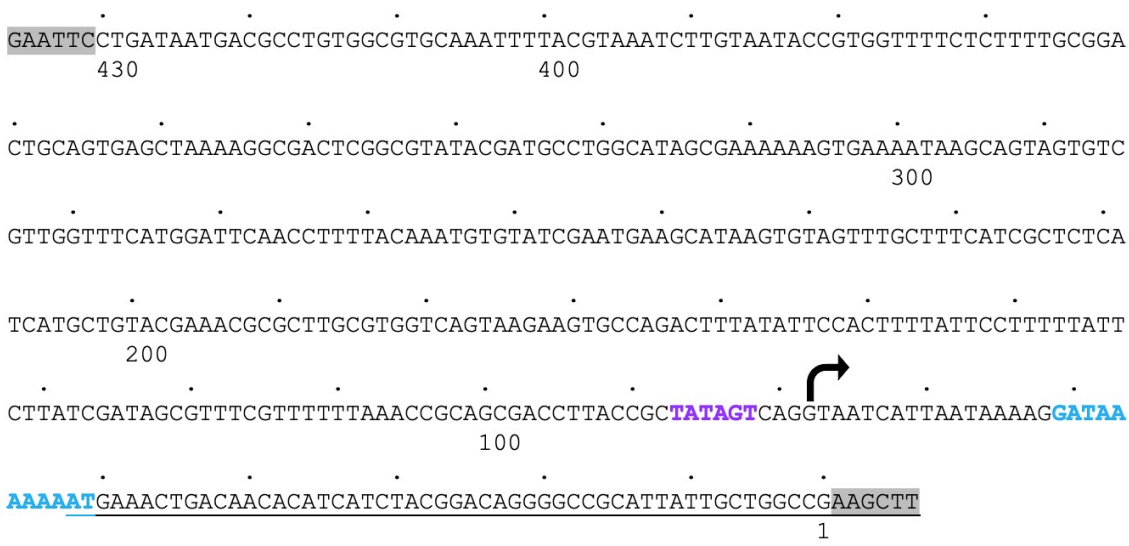
3.3.4 Repression and the nlpA promoter

The *nlpA* gene encodes an inner membrane protein, NlpA, which has a role in the biogenesis of outer membrane vesicles in ETEC (Bodero *et al.*, 2007). The *nlpA* promoter is divergent from the *yicS* promoter, with a shared AggR binding site (Figure 3.11 and Figure 3.12). Moreover, AggR is expected to repress expression from the *nlpA* promoter, as it binds downstream from the transcription start site, and AggR homologs, Rns and CfaD from ETEC, have been reported to repress expression from *nlpA* (Bodero *et al.*, 2007). To investigate this, DNA fragments, flanked by EcoRI and HindIII sites, covering the *yicS-nlpA* regulatory region from EAEC strain 042, the Brazilian EAEC strain H92/3, and *E. coli* K-12, were cloned into pRW50, such that the *nlpA* promoter drives *lacZ* expression (Figure 3.21). The recombinant plasmids were transformed into BW25113 with either pBAD30 or pBAD*aggR* and β-galactosidase activities were measured to indicate promoter activity (Figure 3.22). Results in Figure 3.22A show that cells containing pRW50*nlpA* 042 showed reduced levels of β-galactosidase expression, almost 2-fold, when pBAD*aggR* was present (44 units), compared to cells grown in the absence of arabinose (80 units). Cells containing pRW50*nlpA* K-12 showed a comparable reduction in β-galactosidase levels when expression of AggR was induced (56 units compared to 150 units), although the reductions were less than expected. Cells containing pRW50*nlpA* H92/3 showed only slight decrease in β-galactosidase levels when expression of AggR was induced (201

nlpA EAEC 042



nlpA E. coli K-12



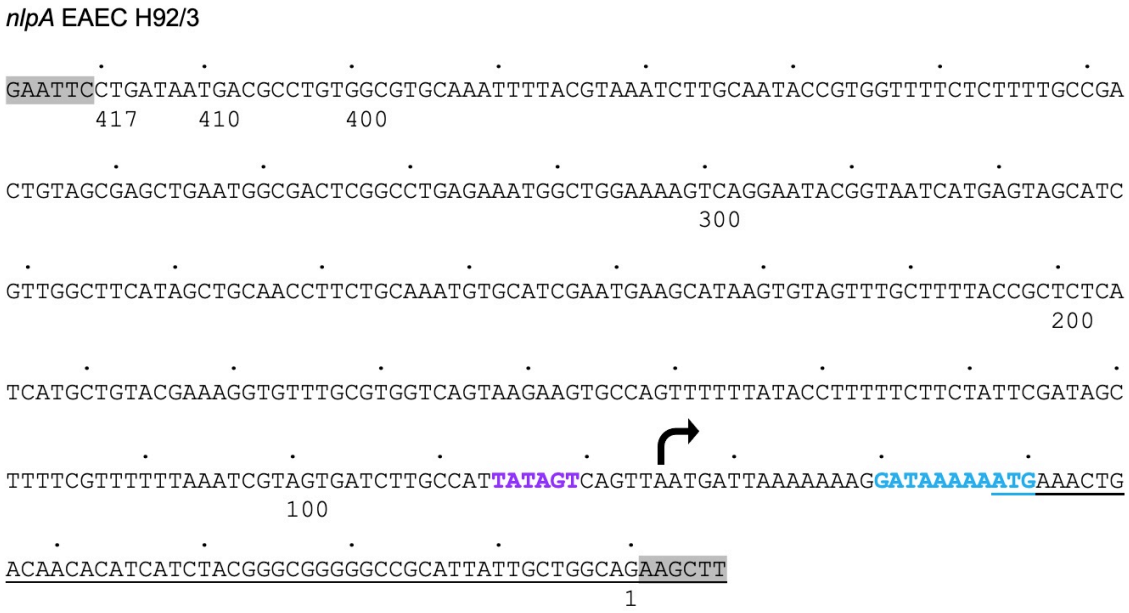
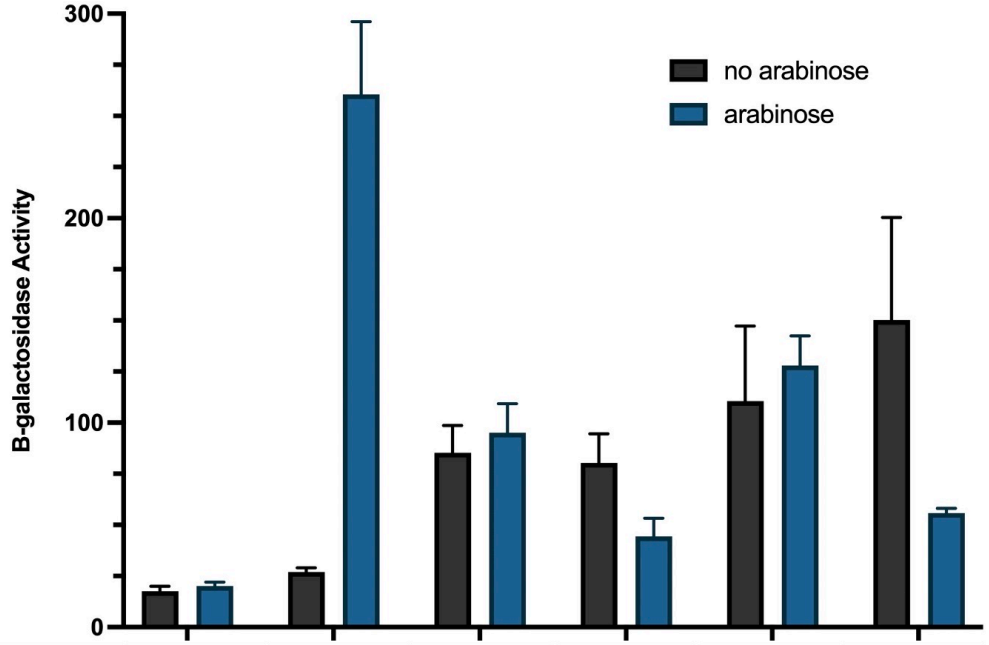


Figure 3.21 *nlpA* promoter fragments cloned in pRW50

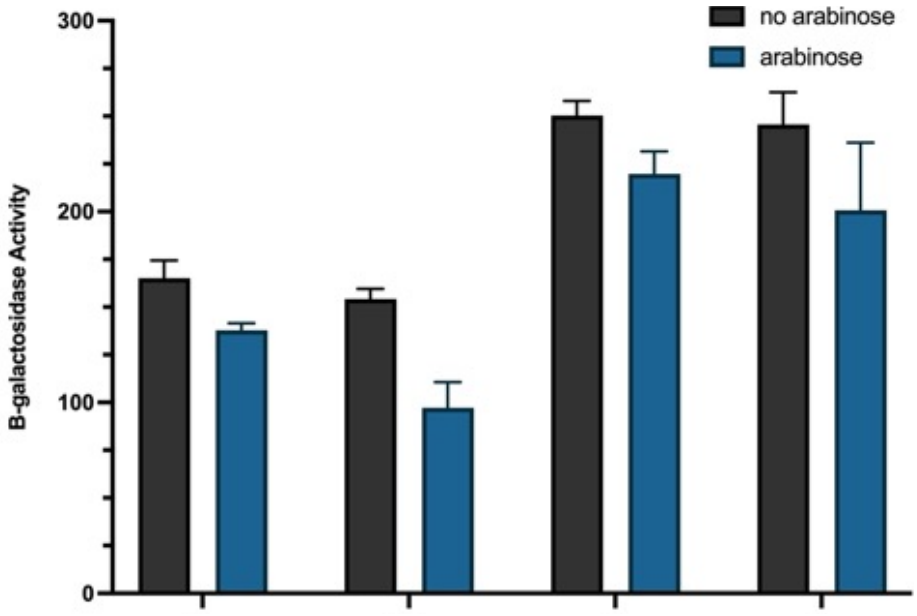
The DNA sequences covering the *nlpA* promoter from EAEC 042, EAEC H92/3 and *E. coli* K-12 were cloned into pRW50. The fragments contain an EcoRI site upstream and a HindIII site downstream, highlighted in grey, and transcription start sites are indicated by bent arrows. The -10 element is highlighted in purple, the DNA binding site for AggR is highlighted in blue and the coding sequence is underlined. The bent arrows indicate the start sites and the direction of transcription.

A



promoter in pRW50	*afaB 042*	*afaB 042*	*nlpA 042*	*nlpA 042*	*nlpA K-12*	*nlpA K-12*
activator in pBAD	–	AggR 042	–	AggR 042	–	AggR 042

B



activator in pBAD	–	AggR 042	–	AggR 042
promoter in pRW50	*nlpA 042*	*nlpA 042*	*nlpA H92/3*	*nlpA H92/3*

Figure 3.22 Measurement of promoter activities: *nlpA* promoters from *E. coli* K-12, EAEC 042 and EAEC H92/3

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* containing pRW50*afaB* 042, pRW50*nlpA* 042 or pRW50*nlpA* K-12 (A), and pRW50*nlpA* 042 or pRW50*nlpA* H92/3 (B), carrying either pBAD30 (empty vector) or pBAD*aggR* 042. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. The cells containing pRW50*afaB* EAEC strain 042, pRW50*nlpA* EAEC strain 042 or pRW50*nlpA* *E. coli* K-12 carrying either pBAD30 (empty vector) or pBAD*aggR* 042. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 24) = 28.34$). A post-hoc Tukey's HSD test showed that, in cells containing pBAD*aggR,* the promoter activities were significantly different in the presence and absence of arabinose ($p<0.05$).
B. The cells containing pRW50*nlpA* EAEC strain 042 or pRW50*nlpA* EAEC strain H92/3 carrying either pBAD30 (empty vector) or pBAD*aggR* 042. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(7, 40) = 69.40$). A post-hoc Tukey's HSD test showed that, in cells containing pBAD*aggR,* the *nlpA* promoter activities were significantly different in the presence and absence of arabinose ($p<0.05$).

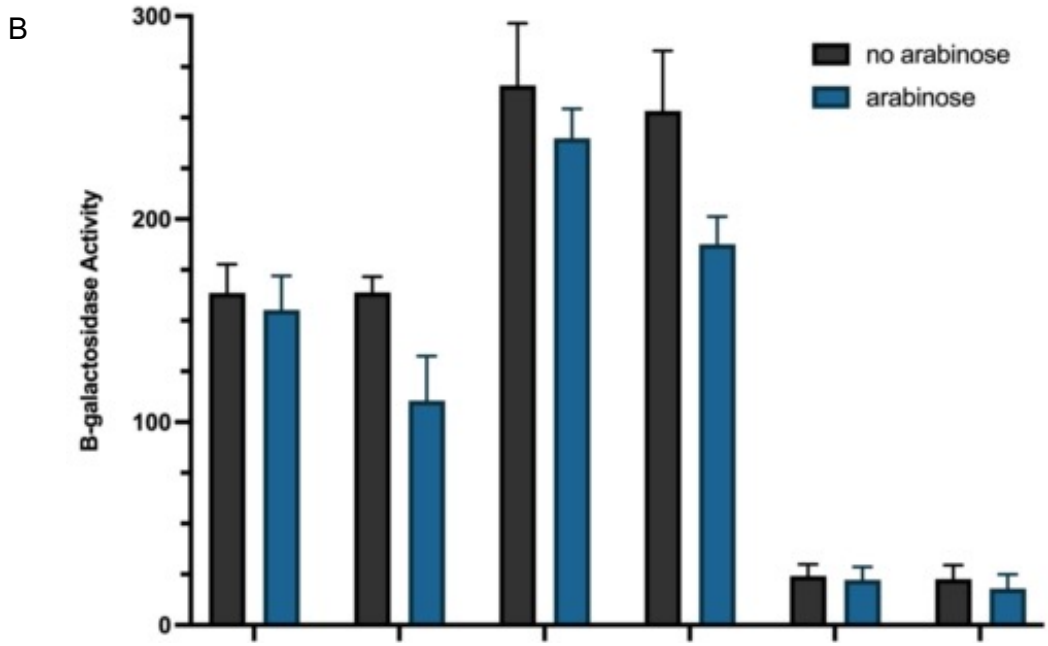
units compared to 246 units) (Figure 3.22B). Taken together, the data show that AggR-dependent repression of the 042, H92/3 and *E. coli* K-12 *nlpA* promoters is quite feeble, with the lowest reduction in cells containing pRW50*nlpA* H92/3. I had previously introduced mutations into the AggR-controlled EAEC strain 042 *yicS* promoter. Hence, I introduced some of these mutations into the EAEC strain 042 *nlpA* promoter fragments in order to alter the -10 hexamer (-10 KO) and the AggR binding site (AggR KO) (Figure 3.23A).

Constructs were made using Megaprimer PCR, as before, and primers were used that maintained an EcoRI site upstream and a HindIII site downstream of the *nlpA* promoter. The mutant pRW50 *nlpA* constructs were transformed in BW25113 cells with either pBAD30 empty vector or pBAD*aggR* and the β-galactosidase levels were measured to indicate promoter activity. The results in Figure 3.23B show that the β-galactosidase levels in cells containing pRW50*nlpA-10KO* were low both in the presence and absence of AggR (18 units and 23 units, respectively), and these levels were comparable to the basal levels in cells containing pBAD30 empty vector (22 units). Thus, altering the *yicS* -10 element, which overlaps with the *nlpA* -10 element, by changing from the consensus sequence, prevents expression from the *nlpA* promoter. In contrast, mutation of the DNA site for AggR, which destroyed AggR-dependent activation of the *yicS* promoter, does not have any great effect on the AggR-dependent repression compared to the wild-type (240 to 188 units and 155 to 111 units, respectively).

A

```
                                  -10 hexamer                    AggR binding site
nlpA042        AGTGATCTTGCCATTATAGTCAGTTAACGATTAAAAAAAGGATAAAAAAGTGAAACTGAC
AggR KO        AGTGATCTTGCCATTATAGTCAGTTAACGATTAAAAAAAGGGTGAAAAAGTGAAACTGAC
-10 KO         AGTGATCTTGCCATTAGAGTCAGTTAACGATTAAAAAAAGGATAAAAAAGTGAAACTGAC
```

B



activator in pBAD	_	AggR 042	_	AggR 042	_	AggR 042
promoter in pRW50	*nlpA 042*	*nlpA 042*	*AggR KO*	*AggR KO*	*-10 KO*	*-10 KO*

Figure 3.23 Mutational analysis of the *nlpA* promoter guided by the *yicS* promoter

A. The DNA sequences of *nlpA* EAEC 042, the DNA binding site for AggR is blue and the -10 element is purple. The bases mutated are highlighted in red and the mutant name is included alongside the corresponding sequences.
B. Measured β-galactosidase activities in *E. coli* K-12 BW25113 cells with either pBAD30 (empty vector) or pBAD*aggR*, carrying pRW50*nlpA* 042, pRW50*nlpA-aggRKO* (AggR KO) or pRW50*nlpA-10KO* (042 -10 KO). Cells were grown in the presence (blue bars) or absence (black bars) of 0.2% (*w/v*) arabinose. The b-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

This indicates that a mutation that stopped AggR-activation of the *yicS* 042 promoter did not greatly alter the reduction in β-galactosidase level in cells containing pRW50*nlpA* 042, therefore showing that AggR-dependence of the *nlpA* 042 promoter is weak and minimal. This suggests that AggR very weakly represses the *nlpA* 042 promoter, regardless of changes to the AggR-DNA binding site.

3.4 Discussion

The AggR-regulon includes genes encoding over 40 virulence factors, and many promoters were shown to have a similar organisation (Yasir *et al.*, 2019). The results in Figure 3.7 show that AggR-dependent promoters drive different levels of gene expression and, crucially, I could find no evidence for the involvement of the master regulator, CRP, in AggR-dependent activation (Figure 3.10).

There are, however, some AggR-dependent promoters that do not follow the simple structure previously characterised (Yasir *et al.*, 2019). One such case is a bi-directional promoter with overlapping -10 elements and a single AggR binding site. This bi-directional promoter falls in a regulatory region that is conserved between different strains, with small sequence differences in the promoter elements conferring or stopping AggR dependence. This is shown by the data in Figure 3.15 with the comparison of EAEC H92/3, EAEC 042 and *E. coli* K-12 *yicS* promoter activities, and it is clear that a DNA site for AggR is not necessarily sufficient to confer AggR-dependence when the promoter -10 element is unfavourable. Similarly, an AggR DNA binding site that did not follow the consensus structure reduces AggR-dependent activation. The AggR-DNA binding site consensus sequence is AWWWWWWTATC, where W = A or T, and

it was determined, utilising the *aafD* promoter, that expression levels were greatest when adenine was positioned at the start of the AggR DNA binding site (Yasir, 2017). Figure 3.15B shows that AggR-dependent activation at the 042 *yicS* promoter is greater than at the corresponding strain H92/3 promoter, which is remarkable because both promoters have consensus -10 elements (TATAAT) with slightly different AggR-DNA binding sites, ACTTTTTTATC and CATTTTTTATC in 042 and H92/3, respectively (Figure 3.12). This indicates that AggR will tolerate a non-consensus DNA binding site if the first base is adenine and not cytosine, however, the second position is not as important. Indeed, the mutational analysis shown in Figure 3.17, altering the AggR DNA binding site and the -10 elements, supports the relative importance of both promoter features. By mutating either element from consensus, AggR-dependence is stopped, and having a consensus -10 element with a consensus AggR-binding site confers AggR-dependence at a previously non-functional promoter. This shows how EAEC strains have evolved, with relative ease, to remove or add *yicS-nlpA* to the AggR-regulon.

Many promoter -10 elements show a degree of 2-fold symmetry and it is thought that this accounts for a large number of bi-directional promoters in bacterial genomes (Warman *et al.*, 2021). Such bi-directional promoters can have a 1-5 bp overlap of their -10 elements on the opposite strands, the most common organisation has the +1 and -18 positions aligned, with a 5 bp overlap, and this helps stabilise RNAP binding (Warman *et al.*, 2021). In the case of *yicS-nlpA*, the spacing is -19 bp, which is a 4 bp overlap (Figure 3.12). This means that the 2 bp overhang can be altered in either direction without affecting divergent

transcription from the other promoter. This is evidenced by *yicS* K-12, which is AggR-independent, though AggR-dependence can be conferred by altering this overhang to match the consensus -10 sequence (Figure 3.17). However, as these 2 bp do not overlap with *nlpA* -10 element, AggR repression of *nlpA* K-12 is similar to 042, even though *yicS* K-12 is AggR-independent.

There is some contention over whether bi-directional promoters' function independently. Clearly any bi-directional promoters cannot transcribe simultaneously in both directions, simply because RNAP cannot bind in both directions. Hence, one model postulates that bi-directional promoters function by blocking and interfering with each other (Warman *et al.*, 2021). However, this does not appear to be the case for *yicS-nlpA* bi-directional promoters that appear to function independently, and do not interfere with each other. A CRP-dependent bi-directional promoter was shown to activate sequentially, RNAP occupies the promoter for a brief time and infrequently in either direction, without interfering (El-Robh and Busby, 2002). This sequential activation is likely applicable to *yicS-nlpA* as the promoters are not strong, β-galactosidase levels seen in Figure 3.22 are very low which indicates that the *nlpA* promoter is very weak. Therefore, RNAP binding in either direction will not block activation in the other direction.

The levels of repression of the *nlpA* promoter by AggR, reported in Figure 3.22, are very low. However, it has previously been reported that AggR strongly represses *nlpA* (Bodero *et al.*, 2007), though it consistently has not been possible to replicate this result. Therefore, I would conclude that AggR is a very weak repressor of *nlpA* 042, K-12 and H92/3. Furthermore, the weak repression seen is independent of AggR-dependent activation *yicS*. There is a possible

explanation for this weak repression related to the mechanisms of activation and repression of transcription initiation. There are two distinguishable mechanisms of RNAP recruitment: 'normal' recruitment involves the formation of a complex between promoter DNA and the activator, which then recruits RNAP to bind; in 'DNA scanning' or pre-recruitment, an activator-RNAP complex forms and then scans the DNA to identify and bind to target promoters (Martin *et al.*, 2002). Data in this thesis strongly argue that AggR functions by pre-recruitment, which will further support this explanation for weak repression. Essentially, a repressor binds to the DNA overlapping promoter elements and blocks RNAP from binding (Browning and Busby, 2004; Browning and Busby, 2016). As AggR functions by pre-recruitment, AggR binds to the *yicS-nlpA* promoter to block RNAP, however, AggR binds as a complex with RNAP which does not effectively block transcription initiation because, according to the El-Robh and Busby 2002 argument, the bi-directional promoters do not interfere with each other.

The comparison and mutational analysis of the *yicS-nlpA* bi-directional promoter indicated that the organisation of the promoters in the different strains were relatively similar but the activity at each promoter was vastly different, and altering key positions in the -10 hexamer and DNA binding site for AggR could confer or stop AggR-dependence (Figure 3.12 and Figure 3.15). Therefore, this led to a theory that this promoter must have some significance in the EAEC virulence profile, as *E. coli* K-12 is not pathogenic. However, upon deletion of this *yicS* gene in EAEC 042, Section 3.3.3 showed that biofilm formation was not prevented, thus indicating that the function of YicS is not related to the formation of biofilm. This does not, however, indicate that YicS is not involved in pathogenicity in some

way, though more work would be needed to ascertain its function and importance in pathogenicity.

Perhaps most interestingly, divergent transcription at a bi-directional promoter, as is the case with the *yicS-nlpA* promoter, has been shown to often be a property of DNA that has been newly acquired (Warman *et al.*, 2021). As the organisation of AggR-dependent promoters and the degree of AggR-dependence varies between strains, it could be hypothesised that this *yicS-nlpA* sequence has been acquired and subsequently mutated in different strains as its irrelevance for pathogenicity became evident and therefore its biological use for particular *E. coli* strains was redundant. In fact, the alignments in Figure 3.12 and Figure 3.13 indicate how straightforward it is for a promoter to drift in and out of the AggR operon with simple mutations in the important promoter elements, the DNA binding site for AggR and -10 element.

The work detailed in this Chapter has showed that an AggR-dependent promoter requires a functional -10 element and a functional DNA binding site for AggR with proper spacing, 20-22 bp, between the promoter elements. Further investigation of other EAEC strains would offer new insights into the mechanism of AggR regulation.

Chapter 4 AggR in diverse EAEC strains

4.1 Introduction

EAEC strains are often identified by the presence of the *aggR* gene, and most work has been focussed on the 'proto-typical' 042 strain, in which AggR is encoded on a virulence plasmid, pAA2 (Morin *et al.*, 2013; Dias *et al.*, 2020). This 042 strain displays a typical aggregative adherence (AA) pattern of biofilm formation when colonising host intestinal mucosa (Table 4.1) (Braga *et al.*, 2017; Alves *et al.*, 2010; Franca *et al.*, 2013; Pereira *et al.*, 2008).

Some of my experiments, described in Chapter 3, underscore the value of comparing EAEC strain 042 with other strains, and this theme is expanded here. In Chapter 3, I explored several promoters from EAEC strain 042, a promoter from EAEC strain C1010-00, and a bi-directional promoter from EAEC strain H92/3. Strain 042 was isolated in 1983 from a child in Peru with acute diarrhoea (Nataro *et al.*, 1985). EAEC strain C1010-00 was isolated from clinical cases in Danish children under 5 years of age (Olesen *et al.*, 2005) and also carries a pAA virulence plasmid, encoding the *aggR* gene (Boisen *et al.*, 2008; Boisen *et al.*, 2020). Strain H92/3, isolated from children with acute diarrhoea in Brazil, contains a pAA virulence plasmid, with the *aggR* gene identified, and this strain also follows a typical pattern of adherence (Rosa *et al.*, 1998; Braga *et al.*, 2017; Alves *et al.*, 2010; Franca *et al.*, 2013; Pereira *et al.*, 2008). Three other strains, also isolated from clinical cases in Brazil, from children under 2 years of age with acute

diarrhoea (Rosa *et al.*, 1998), were subjected to whole-genome sequencing and analysed. Two of these strains, I18/2 and H9/3, display a typical pattern of adherence, and carry a pAA virulence plasmid with an *aggR* gene. The 'atypical' strain, H149/5, displays a discrete distinct AA pattern, and lacks a pAA virulence plasmid, encoding an *aggR* gene homologue (Morin *et al.*, 2013; Braga *et al.*, 2017; Alves *et al.*, 2010; Franca *et al.*, 2013; Pereira *et al.*, 2008).

AggR is a member of the large AraC-family of transcription factors. Most AraC family members consist of a variable N-terminal regulatory domain and a conserved C-terminal DNA-binding domain (Schuller *et al.*, 2012; Ibarra *et al.*, 2008). The 'hallmark' of the C-terminal domain is a stretch of 99 highly conserved amino acids (Gallegos *et al.*, 1997). AggR belongs to a subgroup of the AraC family (Munson *et al.*, 2001), including Rns from ETEC, a positive regulator of fimbrial expression (Nataro *et al.*, 1994), which shares strong sequence similarity to AggR (Santiago *et al.*, 2014). Rns-dependent promoters also have a DNA target for Rns in a Class II position, though this site is sometimes used for activation in conjunction with additional upstream sites (Midgett *et al.*, 2021). In contrast, AggR has previously been shown to function as a monomer with a single DNA-target for AggR, at a Class II position, sufficient to confer AggR-dependence (Yasir *et al.*, 2019). However, with results in Chapter 3 showing that AggR-regulated promoters are not always uniform, the question of whether multiple AggR-binding sites are required at some promoters arises.

Table 4.1 Some Enteroaggregative *E. coli* (EAEC) strains and their characteristics

Strain	**Phylo-genetic group**	**Serotype**	**AA pattern**	**pAA virulence plasmid**	**Virulence markers (*aggR, aap*)**	**Provenance**
042	D	O44:H18	Typical	Yes	*aggR, aap*	Peru (Nataro *et al.*, 1985)
C1010-00	B1	Orough:H1	Typical	Yes	*aggR, aap*	Denmark (Olesen *et al.*, 2005)
I18/2	D	O86:H11	Typical	Yes	*aggR, aap*	Brazil (Rosa *et al.*, 1998)
H9/3	A	O113:H12	Typical	Yes	*aggR, aap*	Brazil (Rosa *et al.*, 1998)
H92/3	D	O86:H18	Typical	Yes	*aggR, aap*	Brazil (Rosa *et al.*, 1998)
H149/5	B2	O5:H11	Discrete	No	N/A	Brazil (Rosa *et al.*, 1998)

4.2 Comparison of AggR with related transcriptional activators

4.2.1 Aligning transcription factors

All AraC family members share a 99 amino acid region of homology in their DNA-binding domain (Gallegos *et al.*, 1997). AggR belongs to a subgroup of AraC family members with the greatest sequence homology with Rns, and these regulators can be used interchangeably (Munson and Scott, 1999; Munson *et al.*, 2001). Hence, I decided to look at transcription factors in other EAEC strains, with a focus on pathogenic and 'atypical' strains.

Three strains from Brazil, H149/5, I18/2 and H9/3, were whole-genome sequenced and annotated by MicrobesNG, and an AggR-like transcription factor was identified on each chromosome and automatic genome annotation, using Prokka (Seemann, 2014), labelled these as 'CfaD' (Figure 4.1) (Rutherford *et al.*, 2000; Berriman and Rutherford, 2003; Carver *et al.*, 2008; Carver *et al.*, 2012). The amino acid sequences were compared with the master regulator from ETEC, Rns, and with AggR from EAEC strain 042 (Figure 4.2A). AggR has 66% sequence identity to Rns from ETEC (Morin *et al.*, 2013), and many of the regulatory features of AggR are similar in Rns (Munson and Scott, 1999). The C-terminal domain is highlighted to indicate the 99 amino acid region of homology in AraC members (Gallegos *et al.*, 1997), and the alignments in Figure 4.2A show that this stretch of amino acids is quite similar. Figure 4.2B shows the comparison of the amino acid sequences in AggR 042, CfaD H9/3 and CfaD I18/2, there are some differences, including one residue in the homologous region, indicating that these annotated transcription factors from the Brazilian strains are AggR-like

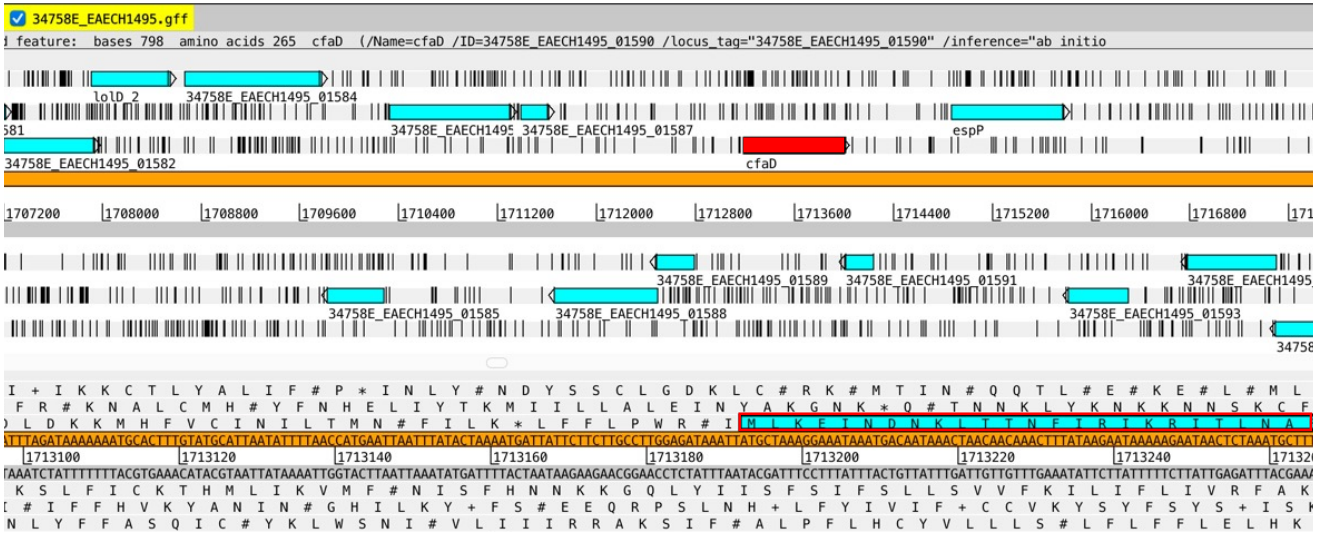
(Figure 4.2B). CfaD from H149/5, however, is very different to 042 AggR and Rns ETEC, though the C-terminal domain is homologous (Figure 4.2C). In fact, Rns ETEC and AggR 042 show greater amino acid identity (66%) (Figure 4.2D) than AggR 042 and CfaD H149/5 (33%) (Figure 4.2E). The alignments indicate that CfaD from H9/3 and I18/2 are AggR-like and Rns ETEC has sequence similarity to AggR. However, CfaD H149/5 is very different to AggR, and the other transcription factors. For this reason, the regulatory functions of CfaD H149/5, AggR 042 and Rns ETEC were compared experimentally.

4.2.2 Activation of H149/5 promoters by AggR homologues

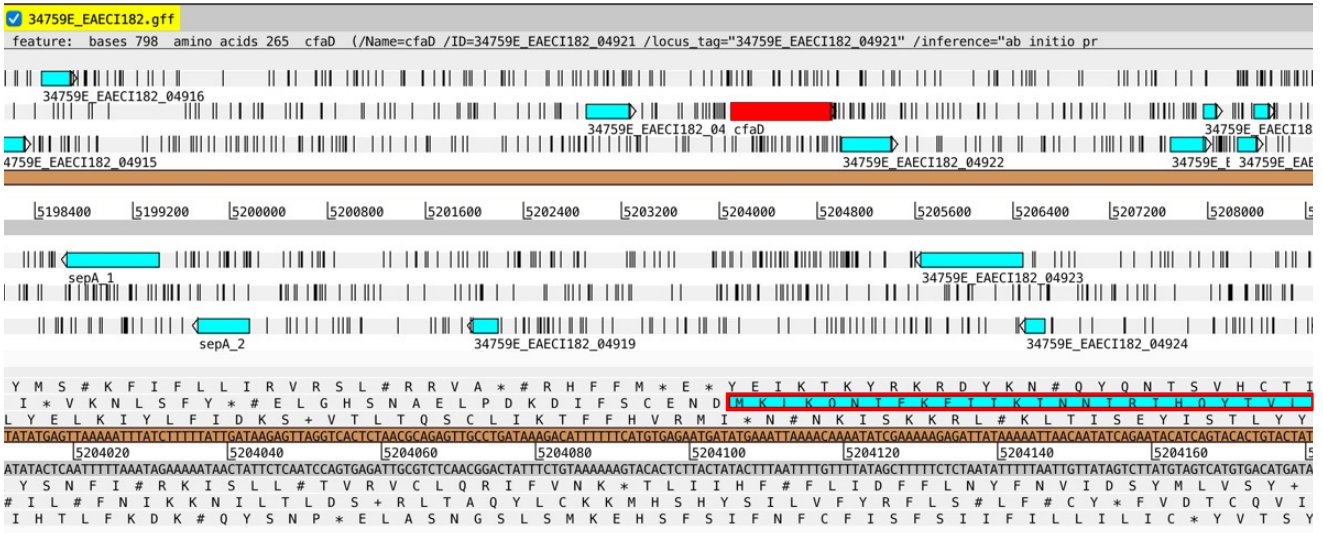
As I found that the 'atypical' EAEC strain, H149/5, encoded a transcription factor with lower sequence homology to strain 042 AggR, I decided to compare its regulatory function with 042 AggR. Additionally, I wanted to determine whether Rns from ETEC would regulate other AggR-dependent promoters. Several putative AggR binding sites were identified in the EAEC H149/5 genome, and I wanted to assay associated promoters to ascertain whether they were functional and how activity was regulated by AggR, CfaD H149/5 or Rns ETEC. Hence, using PCR, I amplified the gene encoding CfaD from EAEC H149/5 (Figure 4.3A), and the gene encoding Rns from ETEC (Figure 4.3B), with primers that introduced an EcoRI restriction site upstream and a XbaI site downstream. Resulting DNA fragments were cloned into pBAD30, placing each gene under the control of the arabinose-inducible *araBAD* promoter.

Several promoters from strain H149/5, with putative DNA binding sites for AggR, were located and then investigated. The *aatP* gene, which encodes the inner membrane permease, AatP (Nishi *et al.*, 2003), was identified in H149/5. Note

A


34758E_EAECH1495.gff
feature: bases 798 amino acids 265 cfaD (/Name=cfaD /ID=34758E_EAECH1495_01590 /locus_tag="34758E_EAECH1495_01590" /inference="ab initio
lolD_2
34758E_EAECH1495_01584
34758E_EAECH1495_01587
espP
34758E_EAECH1495_01582
cfaD
34758E_EAECH1495_01589
34758E_EAECH1495_01591
34758E_EAECH1495_01585
34758E_EAECH1495_01588
34758E_EAECH1495_01593

B


34759E_EAECI182.gff
feature: bases 798 amino acids 265 cfaD (/Name=cfaD /ID=34759E_EAECI182_04921 /locus_tag="34759E_EAECI182_04921" /inference="ab initio pr
34759E_EAECI182_04916
34759E_EAECI182_04 cfaD
34759E_EAECI182_04915
34759E_EAECI182_04922
sepA_1
34759E_EAECI182_04923
sepA_2
34759E_EAECI182_04919
34759E_EAECI182_04924

C

Figure 4.1 An AggR-like transcription factor in newly sequenced Brazilian strains

The figure shows screenshots of the main editor window of Artemis showing an annotated section of the EAEC H149/5 (A), EAEC I18/2 (B) and EAEC H9/3 (C) genome sequences. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed-in view of the amino acid sequence, following the same organisation as above. The red gene box and the corresponding highlighted amino acid sequence indicate the *cfaD* gene. Adapted from Carver *et al.* (2008).

A. The genome sequence of EAEC strain H149/5 with the *cfaD* gene highlighted in red.
B. The genome sequence of EAEC strain I18/2 with the *cfaD* gene highlighted in red.
C. The genome sequence of EAEC strain H9/3 with the *cfaD* gene highlighted in red.

A

```
CfaDH93     ELGHSNAELPDKDIFSCENDMKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGS
CfaDI182    ELGHSNAELPDKDIFSCENDMKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGS
AggR042     --------------------MKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGS
RnsETEC     --------------------MDFKYTEEKETIKINNIMIHKYTVLYTSNCIMDIYSEEEK
CfaDH149    ----------------IMLKEINDNKLTTNFIRIKRITLNAFTIMHTDNCFITLKNERDC
                              . .  .: *:*:.* :: :*:::*.** : : .:.

CfaDH93     NTYLRNELIFLERGINISVRLQKKKSTVNPFIAIRLSSDTLRRLKDALMIIYGISKVDAC
CfaDI182    NTYLRNELIFLERGINISVRLQKKKSTVNPFIAIRLSSDTLRRLKDALMIIYGISKVDAC
AggR042     NTYLRNELIFLERGINISVRLQKKKSTANPFIAIRLSSDTLRRLKDALMIIYGISKVDAC
RnsETEC     ITCFSNRLVFLERGVNISVRMQKQILSEKPYVAFRLNGDMLRHLKDALMIIYGMSKIDTN
CfaDH149    ITCTKDNFLFLEKNMTFSCEIVKVNEKLPPFSIVSFDRKSQILLKDILKEIYPLFMG---
             *   :.::***:.:.:* .: *   .  *:  . :. .    *** *  ** :

CfaDH93     SCPNWSKGIIVADADDSVLDTFKSID-HNDDSRITSDLIYLISKIENNRKIIESIYISAV
CfaDI182    SCPNWSKGIIVADADDSVLDTFKSID-HNDDSRITSDLIYLISKIENNRKIIESIYISAV
AggR042     SCPNWSKGIIVADADDSVLDTFKSID-NNDDSRITSDLIYLISKIENNKKIIESIYISAV
RnsETEC     ACRSMSRKIMTTEVNKTLLDELKNIN-SHDNSAFISSLIYLISKLENNEKIIESIYISSV
CfaDH149    GCDIQREKIITENSIAPQYQLFELITRTNDLKLKVIKSAYLIATMKNHAKIISSLYASCG
            .*    . *:. :   .  : :: *   :* .    .  ***:.::*: ***.*:* *.

CfaDH93     SFFSDKVRNTIEKDLSKRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAAL
CfaDI182    SFFSDKVRNTIEKDLSKRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAAL
AggR042     SFFSDKVRNIIEKDLSKRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAAL
RnsETEC     SFFSDKVRNLIEKDLSRKWTLGIIADAFNASEITIRKRLESENTNFNQILMQLRMSKAAL
CfaDH149    ITFTDKVKNELRKDLSKNWKISMIADKFNISEVSVRKRLSSEKTSFSQLLLQARMDRALQ
              *:***:* :.****:.*.:.:*** ** **:::****.**  .*.*:*:* **.:*

CfaDH93     LLLDNSYQISQISNMIGFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
CfaDI182    LLLDNSYQISQISNMIGFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
AggR042     LLLDNSYQISQISNMIGFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
RnsETEC     LLLENSYQISQISNMIGISSASYFIRIFNKHYGVTPKQFFTYFKGG
CfaDH149    LILDNELPLSSVSESIGISSMPYFIRVFKYFFGITPKQFSIYFRE-
            *:*:*.  :*.:*: **:** .****:*  .:*:*****  **:
```

B

```
CfaDH93     ELGHSNAELPDKDIFSCENDMKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGS
CfaDI182    ELGHSNAELPDKDIFSCENDMKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGS
AggR042     --------------------MKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGS
                                 ************************************

CfaDH93     NTYLRNELIFLERGINISVRLQKKKSTVNPFIAIRLSSDTLRRLKDALMIIYGISKVDAC
CfaDI182    NTYLRNELIFLERGINISVRLQKKKSTVNPFIAIRLSSDTLRRLKDALMIIYGISKVDAC
AggR042     NTYLRNELIFLERGINISVRLQKKKSTANPFIAIRLSSDTLRRLKDALMIIYGISKVDAC
            **************************.*******************************

CfaDH93     SCPNWSKGIIVADADDSVLDTFKSIDHNDDSRITSDLIYLISKIENNRKIIESIYISAVS
CfaDI182    SCPNWSKGIIVADADDSVLDTFKSIDHNDDSRITSDLIYLISKIENNRKIIESIYISAVS
AggR042     SCPNWSKGIIVADADDSVLDTFKSIDNNDDSRITSDLIYLISKIENNKKIIESIYISAVS
            **************************:********************:************

CfaDH93     FFSDKVRNTIEKDLSKRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAALL
CfaDI182    FFSDKVRNTIEKDLSKRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAALL
AggR042     FFSDKVRNIIEKDLSKRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAALL
            ******** **************************************************

CfaDH93     LLDNSYQISQISNMIGFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
CfaDI182    LLDNSYQISQISNMIGFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
AggR042     LLDNSYQISQISNMIGFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
            *********************************************
```

C

```
RnsETEC     MDFKYTEEK----ETIKINNIMIHKYTVLYTSNCIMDIYSEEEKITCFSNRLVFLERGVN
AggR042     MKLKQNIEK----EIIKINNIRIHQYTVLYTSNCTIDVYTKEGSNTYLRNELIFLERGIN
CfaDH149    IMLKEINDNKLTTNFIRIKRITLNAFTIMHTDNCFITLKNERDCITCTKDNFLFLEKNMT
            : :*   ::    : *:*:.* :: :*:::*.** : : .:.   *   :.::***:.:.

RnsETEC     ISVRMQKQILSEKPYVAFRLNGDMLRHLKDALMIIYGMSKIDTNACRSMSRKIMTTEVNK
AggR042     ISVRLQKKKSTANPFIAIRLSSDTLRRLKDALMIIYGISKVDACSCPNWSKGIIVADADD
CfaDH149    FSCEIVKVNEKLPPFSIVSFDRKSQILLKDILKEIYPLFMG---GCDIQREKIITENSIA
            :* .: *   .  *:  . :. .    *** *  ** :      .*    . *:. :

RnsETEC     TLLDELKNIN-SHDNSAFISSLIYLISKLENNEKIIESIYISSVSFFSDKVRNLIEKDLS
AggR042     SVLDTFKSID-NNDDSRITSDLIYLISKIENNKKIIESIYISAVSFFSDKVRNIIEKDLS
CfaDH149    PQYQLFELITRTNDLKLKVIKSAYLIATMKNHAKIISSLYASCGITFTDKVKNELRKDLS
            .  : :: *  .:* .    .  ***:.::*: ***.*:* *.   *:***:* :.****

RnsETEC     RKWTLGIIADAFNASEITIRKRLESENTNFNQILMQLRMSKAALLLLENSYQISQISNMI
AggR042     KRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAALLLLDNSYQISQISNMI
CfaDH149    KNWKISMIADKFNISEVSVRKRLSSEKTSFSQLLLQARMDRALQLILDNELPLSSVSESI
            :.*.:.:*** ** **:::****.**  .*.*:*:* **.:*  *:*:*.  :*.:*: *

RnsETEC     GISSASYFIRIFNKHYGVTPKQFFTYFKGG
AggR042     GFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
CfaDH149    GISSMPYFIRVFKYFFGITPKQFSIYFRE-
            *:** .****:*  .:*:*****  **:
```

D

```
RnsETEC     MDFKYTEEKETIKINNIMIHKYTVLYTSNCIMDIYSEEEKITCFSNRLVFLERGVNISVR
AggR042     MKLKQNIEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGSNTYLRNELIFLERGINISVR
            *.:* . *** ****** **:********* :*:*::* . * : *.*:*****:*****

RnsETEC     MQKQILSEKPYVAFRLNGDMLRHLKDALMIIYGMSKIDTNACRSMSRKIMTTEVNKTLLD
AggR042     LQKKKSTANPFIAIRLSSDTLRRLKDALMIIYGISKVDACSCPNWSKGIIVADADDSVLD
            :**:  : :*::*:**..* **:**********:**:*: :* . *: *:.::.:.::**

RnsETEC     ELKNINSHDNSAFISSLIYLISKLENNEKIIESIYISSVSFFSDKVRNLIEKDLSRKWTL
AggR042     TFKSIDNNDDSRITSDLIYLISKIENNKKIIESIYISAVSFFSDKVRNIIEKDLSKRWTL
             :*.*:.:*:* : *.*******:***:*********:**********:******::***

RnsETEC     GIIADAFNASEITIRKRLESENTNFNQILMQLRMSKAALLLLENSYQISQISNMIGISSA
AggR042     AIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAALLLLDNSYQISQISNMIGFSST
            .**** **.***********  .******* **********:*************:**:

RnsETEC     SYFIRIFNKHYGVTPKQFFTYFKGG
AggR042     SYFIRLFVKHFGITPKQFLTYFKSQ
            *****:* **:*:*****:****.
```

E

```
CfaDH149    IMLKEINDNKLTTNFIRIKRITLNAFTIMHTDNCFITLKNERDCITCTKDNFLFLEKNMT
AggR042     MKLKQN----IEKEIIKINNIRIHQYTVLYTSNCTIDVYTKEGSNTYLRNELIFLERGIN
            : **:     : .::*:*:.* :: :*:::*.** * : .:... *  :::::***:.:.

CfaDH149    FSCEIVKVNEKLPPFSIVSFDRKSQILLKDILKEIYPLFMG---GCDIQREKIITENSIA
AggR042     ISVRLQKKKSTANPFIAIRLSSDTLRRLKDALMIIYGISKVDACSCPNWSKGIIVADADD
            :* .: * :..  **  : :. .:   *** *  ** :      .*    : **. ::

CfaDH149    PQYQLFELITRTNDLKLKVIKSAYLIATMKNHAKIISSLYASCGITFTDKVKNELRKDLS
AggR042     SVLDTFKSID-NNDDSRITSDLIYLISKIENNKKIIESIYISAVSFFSDKVRNIIEKDLS
            .  : *: *  .** .  . .  ***:.::*: ***.*:* *.   *:***:* :.****

CfaDH149    KNWKISMIADKFNISEVSVRKRLSSEKTSFSQLLLQARMDRALQLILDNELPLSSVSESI
AggR042     KRWTLAIIADEFNVSEITIRKRLESEYITFNQILMQSRMSKAALLLLDNSYQISQISNMI
            *.*.:::***:**:**:::****.**  :*.*:*:*:**.:*  *:***.  :*.:*: *

CfaDH149    GISSMPYFIRVFKYFFGITPKQFSIYFRE-
AggR042     GFSSTSYFIRLFVKHFGITPKQFLTYFKSQ
            *:** .****:*  .********  **:.
```

Figure 4.2 Amino acid sequence alignments of transcription factors like AggR 042

A. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042), Rns from ETEC, CfaD H9/3 (EAEC strain H9/3), CfaD I18/2 (EAEC strain I18/2) and CfaD H149/5 (EAEC strain H149/5). The 99

amino acid sequence highlighted in orange indicates the homologous region in the AraC family.

B. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042), CfaD H9/3 (EAEC strain H9/3) and CfaD I18/2 (EAEC strain I18/2). The 99 amino acid sequence highlighted in orange indicates the homologous region in the AraC family.
C. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042), Rns from ETEC and CfaD H149/5 (EAEC strain H149/5). The 99 amino acid sequence highlighted in orange indicates the homologous region in the AraC family.
D. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042) and Rns from ETEC. The 99 amino acid sequence highlighted in orange indicates the homologous region in the AraC family.
E. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042) and CfaD H149/5 (EAEC strain H149/5). The 99 amino acid sequence highlighted in orange indicates the homologous region in the AraC family.

that the *aatP* promoter had previously been characterised from the prototypical strain 042, as well as from two Egyptian strains, E36 and E42, and shown to be activated by AggR (Abdelwahab *et al.*, 2021). Figure 4.4A shows the regulatory region upstream of *aatP* that was amplified (Figure 4.4B), utilising primers that introduced an EcoRI site upstream and a HindIII site downstream, and cloned into pRW50, upstream of the *lacZ* gene (Figure 4.4C). The pRW50*aatP* recombinant plasmid was transformed into *E. coli* strain BW25113 Δ*lac* carrying either pBAD30 empty vector, pBAD*aggR*, pBAD*rns* or pBAD*cfaD*. Cells were grown in the presence or absence of 0.2% (*w/v*) arabinose. β-galactosidase activity was measured in cell extracts, to indicate target promoter activity and responses to arabinose-induced expression of AggR 042, Rns ETEC or CfaD H149/5. Results in Figure 4.4D show that cells containing pRW50*aatP* showed significant levels of β-galactosidase expression when pBAD*aggR* (27-fold, 4944 units), pBAD*rns* (23-fold, 4147 units) or pBAD*cfaD* (19-fold, 3506 units) were present, compared to very low levels with the empty pBAD30 vector control (178 units). Furthermore, when pBAD*aggR*, pBAD*rns* or pBAD*cfaD* were present, levels were increased in response to arabinose, though the induction level differs from strain to strain. The background levels are very high in cells, in the absence of arabinose, containing pBAD*aggR,* approximately 31% of the total β-galactosidase activity. As discussed in Chapter 3 (Section 3.4), the pBAD recombinant plasmids are ‘leaky’, giving some AggR expression in the absence of arabinose, and the level of background activity must indicate how tightly the regulator binds to the promoter. The results in Figure 4.4D indicate that AggR binds very tightly at the *aatP* promoter. In conditions of full induction, 042 AggR

A

pBAD*rns*


M
GAATTCATGGACTTTAAATACACTGAAGAAAAAGAAACAATAAAAATTAATAATATTATGATTCATAAATACACTG
798
790
TATTATATACATCAAATTGTATTATGGATATATATTCGGAAGAAGAGAAAATTACATGTTTTAGTAACAGACTTGT
700
ATTTCTTGAAAGAGGGGTAAATATATCTGTAAGAATGCAGAAGCAAATTTTATCAGAAAAGCCGTATGTTGCATTC
600
AGATTGAACGGAGATATGCTAAGGCATTTAAAGGATGCATTGATGATAATATATGGCATGTCAAAAATAGATACCA
ATGCTTGTAGAAGCATGTCAAGAAAAATAATGACAACAGAGGTGAATAAAACCTTGTTGGATGAATTAAAAAATAT
500
AAACAGCCATGATAACTCAGCTTTTATATCTAGCTTGATATATTTGATTTCAAAACTTGAGAATAATGAAAAAATA
400
ATAGAATCAATTTATATATCATCTGTGAGTTTTTTTTCTGACAAGGTCAGAAATCTTATCGAAAAAGATCTATCCA
300
GAAAATGGACGCTGGGTATTATTGCAGATGCGTTTAATGCATCAGAAATAACCATCAGAAAAAGACTAGAATCTGA
200
GAATACTAATTTTAATCAGATATTAATGCAATTGAGAATGAGTAAGGCTGCGTTATTACTACTTGAAAATTCATAC
CAGATATCTCAGATATCAAATATGATTGGAATTTCCAGTGCATCTTATTTTATTAGGATTTTTAATAAACATTATG
100
GTGTTACACCAAAGCAATTTTTCACTTATTTTAAAGGTGGATAATCTAGA
1

B

pBAD*cfaD*

Figure 4.3 Cloned DNA fragments in pBAD*rns* and pBAD*cfaD*

A. The base sequence of the *rns* gene cloned into pBAD30 utilising EcoRI and XbaI restriction sites, highlighted in grey. The translation start and stop codons are highlighted pink.
B. The base sequence of the *cfaD* gene cloned into pBAD30 utilising EcoRI and XbaI restriction sites, highlighted in grey. The translation start and stop codons are highlighted pink.

A


34758E_EAECH1495.gff
ected bases on forward strand: 1704431..1704843
34758E_EAECH1495_01575
lolD_2
34758E
34758E_EAECH1495_01581
aggB
34758E_EAECH1495_01576
34758E_EAECH1495_01578
34758E_EAECH1495_01580
34758E_EAECH1495_01582
1698400
1699200
1700000
1700800
1701600
1702400
1703200
1704000
1704800
1705600
1706400
1707200
1708000
34758E_EAECH1495_01577
1704340
1704360
1704380
1704400
1704420
1704440
1704460
1704480

B

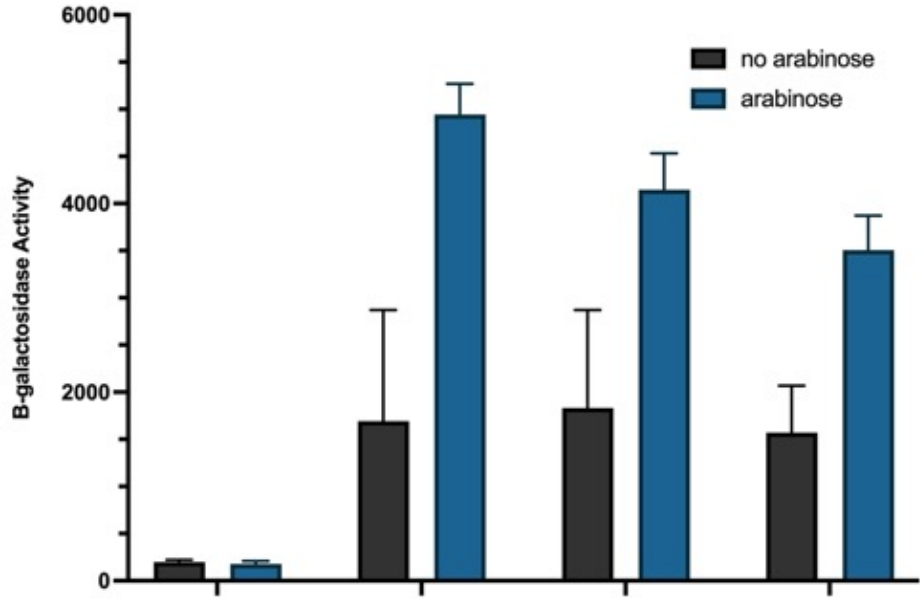

34758E_EAECH1495.gff
feature: bases 1143 amino acids 380 34758E_EAECH1495_01580 (/ID=34758E_EAECH1495_01580 /locus_tag="34758E_EAECH1495_01580" /inference="ab
34758E_EAECH1495_01575
lolD_2
34758E_EAECH1495_01581
aggB
34758E_EAECH1495_01576
34758E_EAECH1495_01580
34758E_EAECH1495_01582
1699200
1700000
1700800
1701600
1702400
1703200
1704000
1704800
1705600
1706400
1707200
1708000
1708800
34758E_EAECH1495_01577
1704740
1704760
1704780
1704800
1704820
1704840
1704860
1704880

C

aatP(H149/5)

```
GAATTCAAAAGCCAGAACGGTACCGCTAATGCAGTCGTTGTATGAGTGGCTCCAGGGGCAGATGAGCACGCTGTCG
      413          400

CGTCACTCGGATACGGCGAAAGCGTTCACCTATCTGCTGAAGCAATGGGACGCCCTGAACGAATACTGCAGTAATG
                                                 300

GCTGGGTGGAGATCGACAATAACCTGTGTGAAAACGCCCTCCGGGTAATTGCACTGGGGCGGCACTCTATCTGTAA
                                                                      200

ACTAATTTTTACCCCACGTCCTTCACGGTGCGCTTACCATTATTACGCACTGTAACCATCTCTTCTGATTAATATC

ATAAAGTGATAATAATTTATCATAAACAAACCCAGCACTAGTATATTCTGTACCAGAAATAGAGAACAATACAAAA
              100

TGTATTTATTATAGAAGGCTTTAGTCTTGTCATGTAAAAAAGCTT
                                  1
```

D



promoter in pRW50	*aatP* *H149/5*	*aatP* *H149/5*	*aatP* *H149/5*	*aatP* *H149/5*
activator in pBAD	–	AggR 042	Rns ETEC	CfaD H149/5

Figure 4.4 Analysing the aatP regulatory region from H149/5

A. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below

these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The pink highlighted section in the forward open reading frame second row indicates the *aatP* regulatory region. Adapted from Carver *et al.* (2008).

B. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The red gene box and the corresponding highlighted amino acid sequence indicate the *aatP* gene. Adapted from Carver *et al.* (2008).

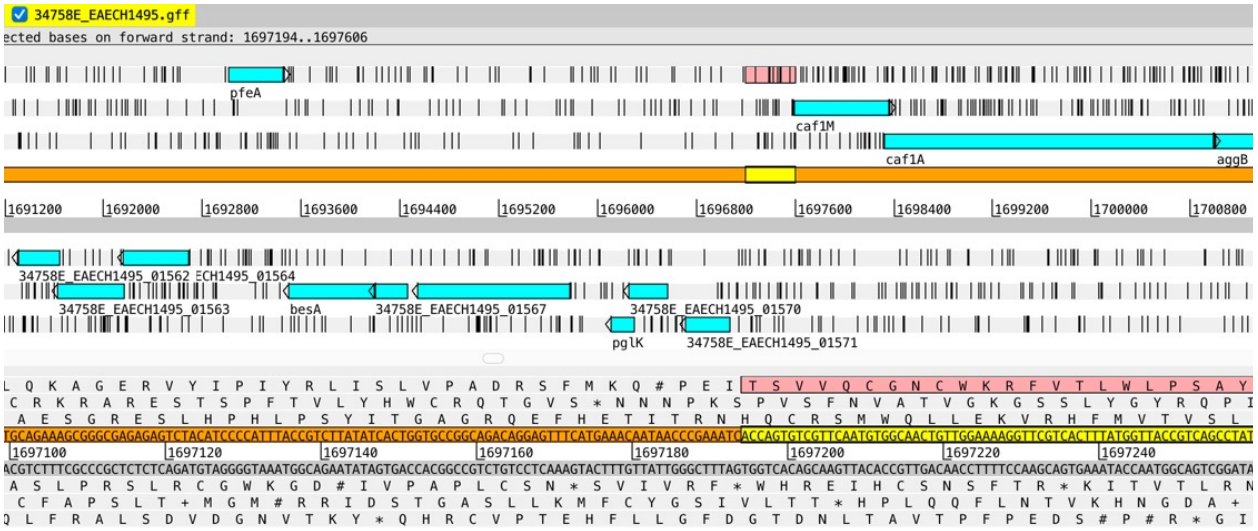
C. The DNA sequence covering the *aatP* promoter from EAEC H149/5 was cloned into pRW50. The fragment contains an EcoRI site upstream and HindIII downstream, highlighted in grey. The -10 element is highlighted in purple, the DNA binding site for AggR is highlighted in blue and the coding sequence is in bold.

D. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*aatP*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

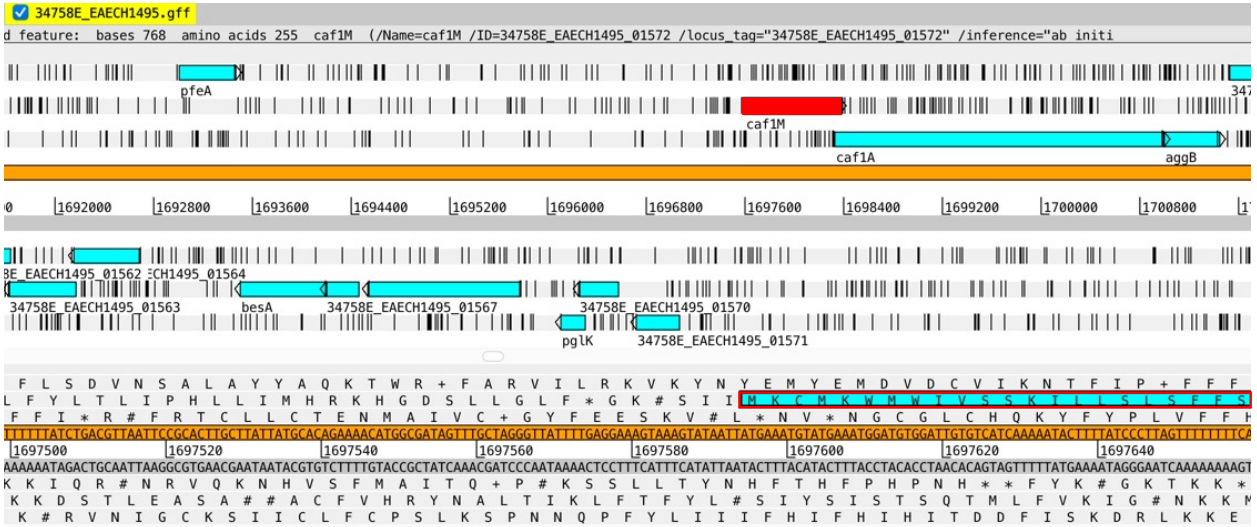
activates the *aatP* promoter to higher levels, compared to Rns and CfaD. The next EAEC H149/5 regulatory region that I investigated is adjacent to the *caf1M* gene (Figure 4.5A and Figure 4.5B). Bioinformatic analyses of the genome sequence using the Basic local Alignment Search Tool (BLAST) at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) showed significant sequence alignment of *caf1M* to fimbrial chaperone genes (Figure 4.5C) (States and Gish, 1994). This is particularly interesting because strain H149/5 has neither a pAA virulence plasmid, nor previously identified fimbrial genes, and its pattern of adherence is not typical of an EAEC strain (Franca *et al.*, 2013; Pereira *et al.*, 2008). Hence the *caf1M* regulatory region was amplified from strain H149/5, using primers introducing an EcoRI site upstream and a HindIII site downstream, and then cloned into pRW50 (Figure 4.5D). As before, β-galactosidase activity was measured in transformants, reflecting *caf1M* promoter activity, and arabinose was used to induce the expression of activators, encoded by pBAD plasmids.

The results, illustrated in Figure 4.5E, show that the β-galactosidase activity increased when CfaD-expression was induced, and this was also the case with both AggR and Rns. The increase in β-galactosidase activity was 19-fold, 22-fold and 17-fold in the presence of AggR, Rns and CfaD, respectively, compared to the basal levels with pBAD30 empty vector (22061, 25451 and 19122 units respectively compared to 1119 units). The relative background levels of expression without induction by arabinose are much lower in cells containing pRW50*caf1M* compared to cells containing pRW50*aatP* (Figure 4.4). The background levels were just over 1% of the total β-galactosidase activity in the presence of AggR, which indicates that AggR binds much less tightly to the *caf1M*

A


34758E_EAECH1495.gff
ected bases on forward strand: 1697194..1697606
pfeA
caf1M
caf1A
aggB
1691200
1692000
1692800
1693600
1694400
1695200
1696000
1696800
1697600
1698400
1699200
1700000
1700800
34758E_EAECH1495_01562
ECH1495_01564
34758E_EAECH1495_01563
besA
34758E_EAECH1495_01567
34758E_EAECH1495_01570
pglK
34758E_EAECH1495_01571
1697100
1697120
1697140
1697160
1697180
1697200
1697220
1697240

B


34758E_EAECH1495.gff
d feature: bases 768 amino acids 255 caf1M (/Name=caf1M /ID=34758E_EAECH1495_01572 /locus_tag="34758E_EAECH1495_01572" /inference="ab initi
pfeA
caf1M
caf1A
aggB
1692000
1692800
1693600
1694400
1695200
1696000
1696800
1697600
1698400
1699200
1700000
1700800
34758E_EAECH1495_01562
ECH1495_01564
34758E_EAECH1495_01563
besA
34758E_EAECH1495_01567
34758E_EAECH1495_01570
pglK
34758E_EAECH1495_01571
1697500
1697520
1697540
1697560
1697580
1697600
1697620
1697640

C

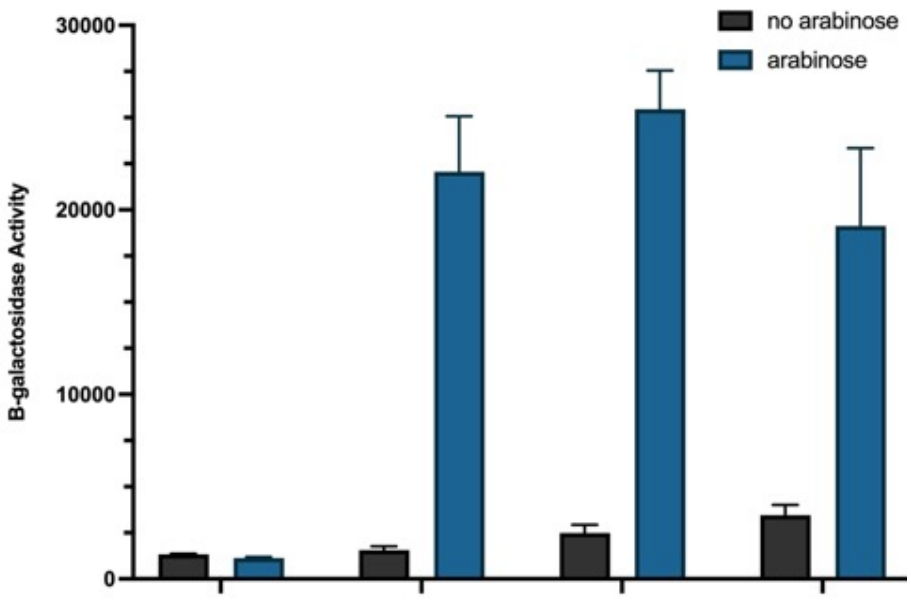
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chaperone protein Agg3D [Escherichia coli]	Escherichia coli	506	506	100%	0.0	99.61%	255	STP51891.1
hypothetical protein G854_00904 [Escherichia coli HVH 202 (4-3163997)]	Escherichia coli HVH 202 (4-3163997)	506	506	100%	0.0	100.00%	255	EQU40395.1
hypothetical protein BH689_24870 [Escherichia coli]	Escherichia coli	506	506	100%	1e-180	99.61%	255	PCM20519.1
hypothetical protein G736_01042 [Escherichia coli HVH 70 (4-2963531)]	Escherichia coli HVH 70 (4-2963531)	502	502	100%	5e-179	99.22%	255	EQP57726.1
Chaperone protein Agg3D [Escherichia coli]	Escherichia coli	501	501	100%	1e-178	99.22%	255	STL63766.1
Chaperone protein Agg3D [Escherichia coli]	Escherichia coli	501	501	100%	2e-178	98.82%	255	STF94826.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	499	499	98%	8e-178	100.00%	252	WP_032145366.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	498	498	98%	1e-177	99.60%	252	EFE7119915.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	498	498	98%	1e-177	99.60%	252	WP_032204496.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	498	498	98%	1e-177	99.60%	252	WP_096778122.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	497	497	98%	4e-177	99.60%	252	MBJ0252462.1
TPA: fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	496	496	98%	7e-177	99.21%	252	HAV7599143.1
molecular chaperone [Escherichia coli]	Escherichia coli	496	496	98%	1e-176	99.60%	252	TZB36653.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	494	494	98%	6e-176	99.21%	252	WP_032148436.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	494	494	98%	9e-176	99.21%	252	WP_087597375.1
TPA: fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	494	494	98%	9e-176	98.81%	252	HAG9594444.1
fimbria/pilus periplasmic chaperone [Escherichia coli]	Escherichia coli	493	493	98%	1e-175	98.81%	252	WP_115187621.1

D

caf1M(H149/5)

```
       .  .          .          .          .          .          .          .
GAATTCACCAGTGTCGTTCAATGTGGCAACTGTTGGAAAAGGTTCGTCACTTTATGGTTACCGTCAGCCTATTCCC
      413          400

   .          .          .          .          .          .          .          .
CGGTGGAATGCATGGGCTAGCTAAAGTGTAGCAGTATTAGGATCTGTTATTTAGATTCGCAGGTGGGGGAGTTGGT
                                              300

       .          .          .          .          .          .          .
AACGGATAATTCTACCGTAGTTAAACACAATCCGATAAATAAATAGCCCTATTGAGTTGTGGGTGGATGTAGTGAT
                                                                        200

  .          .          .          .          .          .          .          .
AATTGATGGATAAAAATATTATTTTTTTTGATACTGATACCTTAATGGTTAATGAACTAAATGGAATAATCTTAGT

     .          .          .          .          .          .          .          .
GTGTTGCTTTTTTATCTGACGTTAATTCCGCACTTGCTTATTATGCACAGAAAACATGGCGATAGTTTGCTAGGGT
               100

          .          .          .          .
TATTTTGAGGAAAGTAAAGTATAATTATGAAATGTATGAAAGCTT
                                    1
```

E



promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*
activator in pBAD	–	AggR 042	Rns ETEC	CfaD H149/5

Figure 4.5 Analysing the *caf1M* regulatory region from H149/5

A. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the

vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The pink highlighted section in the forward open reading frame second row indicates the *caf1M* regulatory region. Adapted from Carver *et al.* (2008).

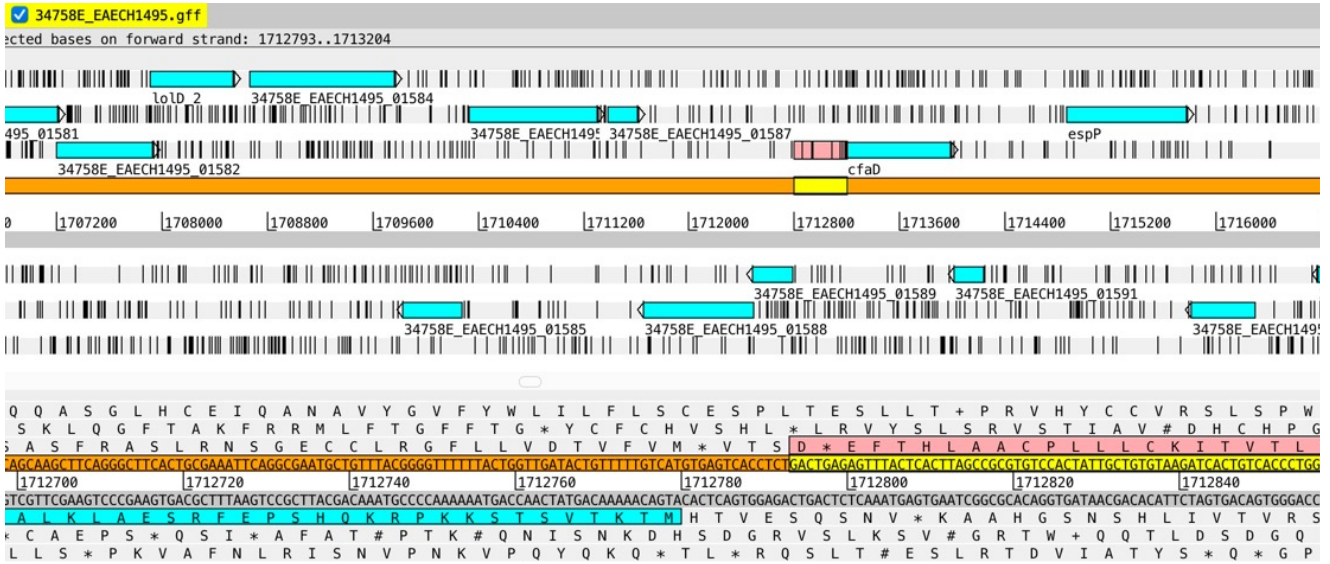
B. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The red gene box and the corresponding highlighted amino acid sequence indicate the *caf1M* gene. Adapted from Carver *et al.* (2008).
C. A screenshot of the protein BLAST search results using the FASTA sequence of the Caf1M protein from EAEC strain H149/5. The search was performed on 08/07/22. Adapted from States and Gish (1994).
D. The DNA sequence covering the *caf1M* promoter from EAEC H149/5 was cloned into pRW50. The fragment contains an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The -10 element is highlighted in purple, the DNA binding site for AggR is highlighted in blue and the coding sequence is in bold.
E. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*caf1M*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

promoter compared to *aatP*. The β-galactosidase levels measured were highest in cells containing pRW50*caf1M* in the presence of Rns, followed by AggR and CfaD respectively.

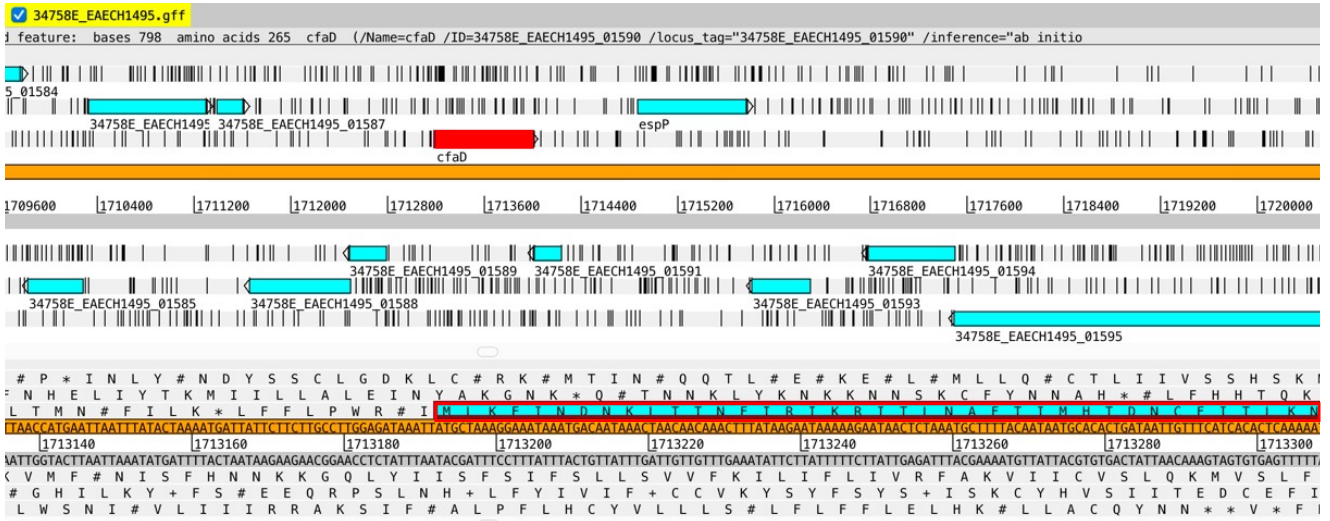
My next set of experiments with promoters from strain H149/5 involved the *cfaD* gene regulatory region. Recall that transcription of the gene encoded 042 AggR is autoactivated (see Chapter 6) (Shimada *et al.*, 2021; Morin *et al.*, 2010) and, indeed, the expression of many bacterial transcription factors is autoregulated (Martinez-Antonio and Collado-Vides, 2003; Tierrafria *et al.*, 2022), therefore, I investigated whether CfaD expression is autoactivated. A DNA fragment covering the *cfaD* promoter (Figure 4.6A and Figure 4.6B) was amplified and cloned into pRW50, using EcoRI and HindIII restriction sites (Figure 4.6C). Assays of promoter activity were run exactly as before: cultures were grown in the presence or absence of arabinose and β-galactosidase activity was measured to determine promoter activity. The results in Figure 4.6D show that the H149/5 *cfaD* gene promoter is autoactivated by CfaD: β-galactosidase activity increases in cells containing pRW50*cfaD* in the presence of CfaD H149/5 (almost 3-fold induction over pBAD30 empty vector basal levels, 10402 and 3647 units, respectively). The β-galactosidase levels measured in cells containing pRW50*cfaD* in the presence of AggR were high, with a 3-fold increase from basal levels with pBAD30 (11584 and 3647 units, respectively), though the AggR-dependent background levels were relatively low (6.5%) in cells containing pBAD*aggR* in the absence of arabinose (Figure 4.6D).

The results indicate that the *cfaD* promoter is activated by AggR and CfaD. As expected, there are substantial levels of activator-independent activity, section

A


34758E_EAECH1495.gff
ected bases on forward strand: 1712793..1713204
lolD_2
34758E_EAECH1495_01584
34758E_EAECH1495_01581
34758E_EAECH1495_01587
espP
34758E_EAECH1495_01582
cfaD
1707200
1708000
1708800
1709600
1710400
1711200
1712000
1712800
1713600
1714400
1715200
1716000
34758E_EAECH1495_01589
34758E_EAECH1495_01591
34758E_EAECH1495_01585
34758E_EAECH1495_01588
1712700
1712720
1712740
1712760
1712780
1712800
1712820
1712840

B

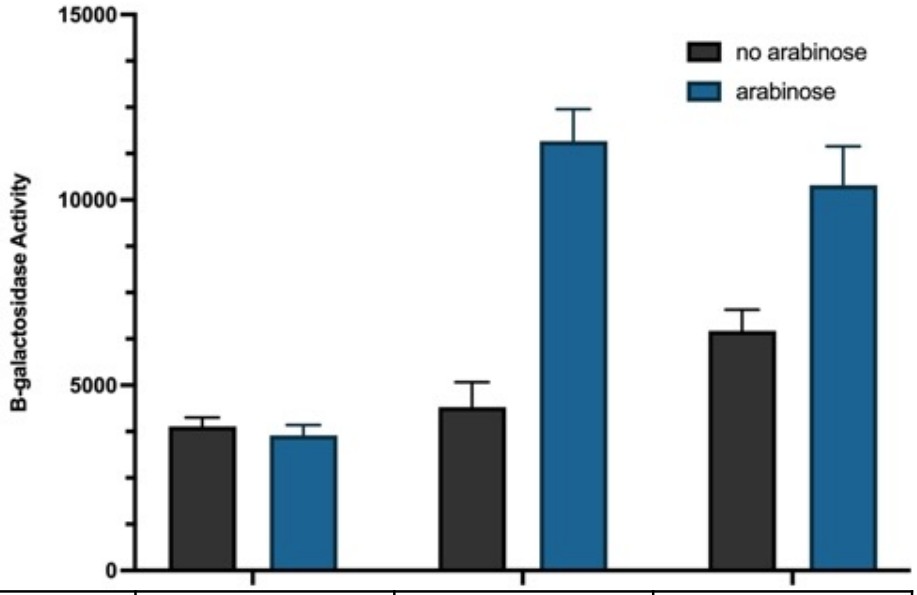

34758E_EAECH1495.gff
feature: bases 798 amino acids 265 cfaD (/Name=cfaD /ID=34758E_EAECH1495_01590 /locus_tag="34758E_EAECH1495_01590" /inference="ab initio
34758E_EAECH1495_01584
34758E_EAECH1495_01587
espP
cfaD
1709600
1710400
1711200
1712000
1712800
1713600
1714400
1715200
1716000
1716800
1717600
1718400
1719200
1720000
34758E_EAECH1495_01589
34758E_EAECH1495_01591
34758E_EAECH1495_01594
34758E_EAECH1495_01585
34758E_EAECH1495_01588
34758E_EAECH1495_01593
34758E_EAECH1495_01595
1713140
1713160
1713180
1713200
1713220
1713240
1713260
1713280
1713300

C

cfaD(H149/5)

```
GAATTCTGACTGAGAGTTTACTCACTTAGCCGCGTGTCCACTATTGCTGTGTAAGATCACTGTCACCCTGGAAAGT
      413          400

TGATCTGAGCTCTCAGTAAGTATCAATACAGTCCTCGCGAGCCGCTTACGAATGAGTATGCCGGAAGATCTCTAAA
                                          300

TGTGAATGGTTCTATTATGTGGGATACTTACAAAGGCCTCTTACACATATCTGATTGAGGTTTTTTATTATCGAAT
                                                          200

TTTGTGTAAAATTCATTTCCTATACATAAATGTAAAAATAACATCACCCATTGGCTCAAAAAGTGTCCATTAAAAA

AGAGATGATTTAGATAAAAAAATGCACTTTGTATGCATTAATATTTTAACCATGAATTAATTTATACTAAAATGAT
               100

TATTCTTCTTGCCTTGGAGATAAATTATGCTAAAGGAAAAAGCTT
                                    1
```

D



promoter in pRW50	*cfaD H149/5*	*cfaD H149/5*	*cfaD H149/5*
activator in pBAD	_	AggR 042	CfaD H149/5

Figure 4.6 Analysing the *cfaD* regulatory region from H149/5

A. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the

vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The pink highlighted section in the forward open reading frame second row indicates the *cfaD* regulatory region. Adapted from Carver *et al.* (2008).

B. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The red gene box and the corresponding highlighted amino acid sequence indicate the *cfaD* gene. Adapted from Carver *et al.* (2008).

C. The DNA sequence covering the *cfaD* promoter from EAEC H149/5 was cloned into pRW50. The fragment contains an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The -10 element is highlighted in purple, the DNA binding site for AggR is highlighted in blue and the coding sequence is in bold.

D. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*cfaD*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042 or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

6.4. The induction by AggR and CfaD, independent of arabinose, is lower compared to *aatP*; this indicates that AggR and CfaD bind more tightly to the *aatP* regulatory region than to *pcfaD* or *pcaf1M*. In summary, the three H149/5 promoters identified and assayed here have functional -10 elements and an AggR binding site located 20-22 bp upstream. The *caf1M* and *aatP* promoters are activated by AggR EAEC 042, CfaD EAEC H149/5 and Rns ETEC; AggR and CfaD also activate *pcfaD*. The proportion of AggR-dependent induction in the absence of arabinose indicates that AggR binds most tightly at the *aatP* promoter.

4.2.3 Activation at promoters from EAEC strain 042 and strain C1010-00

Since I had shown that different promoters from strain H149/5 could be activated by AggR from 042 and Rns from ETEC, as well as CfaD from H149/5, I wanted to determine whether promoter regions from strains such as 042 and C1010-00 could similarly be regulated by Rns and CfaD. Hence, plasmids pRW50*agg4D*, pRW50*afaB*, pRW50*yicS* and pRW50*nlpA*, carrying the strain C1010-00 *agg4D*, strain 042 *afaB, yicS* and *nlpA* promoters, respectively, were transformed into BW25113 cells containing either pBAD30 empty vector, pBAD*aggR*, pBAD*rns* or pBAD*cfaD*. Cultures were grown in the presence or absence of 0.2% (*w/v*) arabinose and promoter activities were determined by measuring β-galactosidase levels.

Data illustrated in Figure 4.7A and Figure 4.7B show that both AggR and Rns induce expression from the *agg4D* and *afaB* promoters, with Rns inducing higher levels than AggR. Thus, cells containing the recombinant plasmids pRW50*agg4D* or pRW50*afaB* had increased levels of β-galactosidase activity following induction of Rns expression (2288 units and 1210 units, respectively, compared

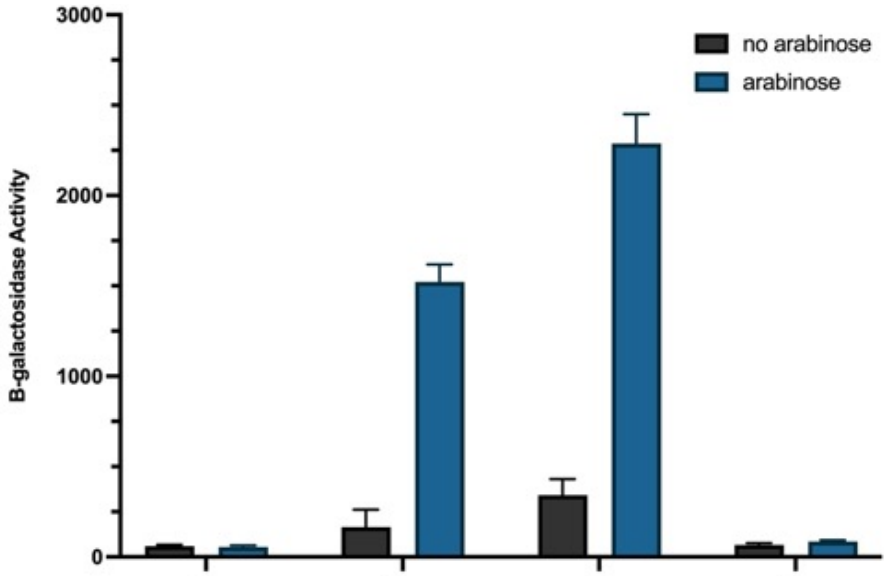
to basal levels of 56 and 25 units, respectively, with empty pBAD30) or AggR expression (27-fold induction to 1522 units, and 25-fold induction to 636 units, respectively). Unexpectedly, in cells containing pRW50*agg4D* or pRW50*afaB,* no increase in β-galactosidase activity was seen when the expression of CfaD H149/5 was induced (85 and 41 units, respectively).

Similarly, cells containing pRW50*yicS* 042 showed high β-galactosidase levels in the presence of Rns (2504 units) and in the presence of AggR (2179 units) compared to empty vector (339 units) (Figure 4.7C). In contrast, there was only a modest increase in β-galactosidase levels in cells containing pRW50*yicS* 042 when CfaD expression was induced (2-fold induction to 732 units compared to background levels of 339 units).

Cells containing pRW50*nlpA* showed reduced levels of β-galactosidase upon the induction of AggR expression (around 2-fold reduction compared to pBAD30, 87 and 157 units, respectively, Figure 4.7D), Rns expression (2-fold reduction, 113 and 157 units, respectively, Figure 4.7D) or CfaD expression (1.4-fold reduction, 33 and 66 units, respectively, Figure 4.7E). The modest repression of *nlpA* and weak activation of *yicS* seen with CfaD is consistent with the idea that RNAP binds to bi-directional promoters at different times (El-Robh and Busby, 2002); essentially, the occupying time is so brief that it does not interfere with the other promoter.

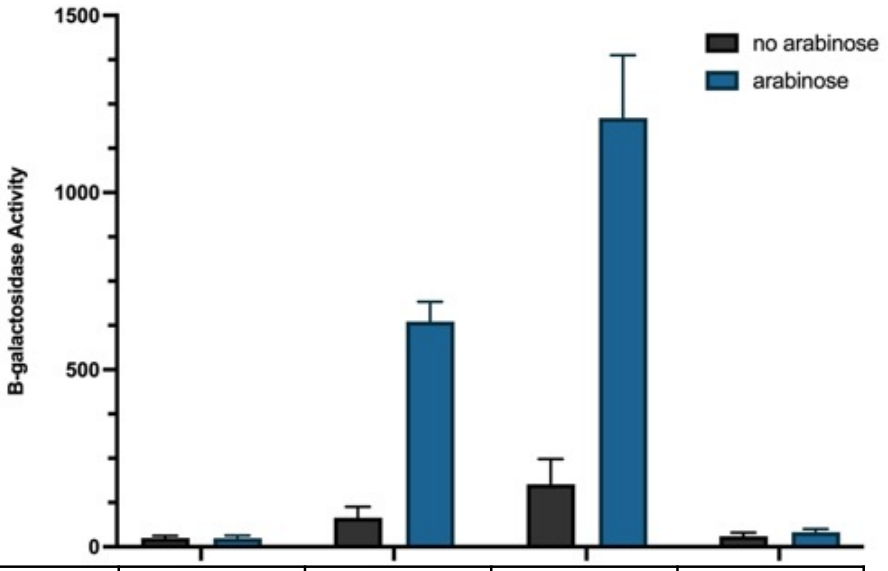
The data in Figure 4.7 show that Rns from ETEC can activate expression from AggR-dependent promoters. Most remarkably, the H149/5 promoters were activated by AggR and CfaD, however, *afaB* 042 and *agg4D* C1010-00 were not

A



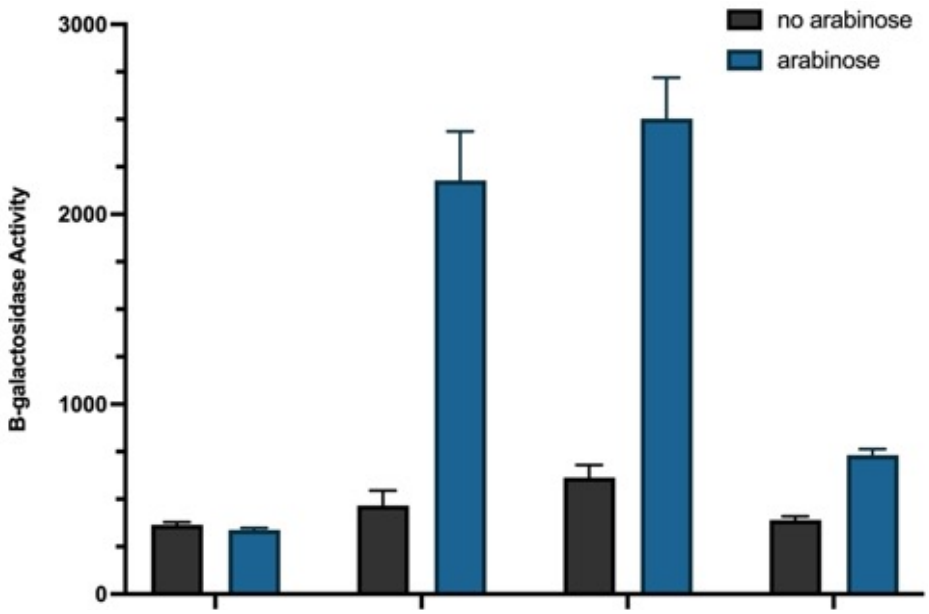
promoter in pRW50	*agg4D C1010-00*	*agg4D C1010-00*	*agg4D C1010-00*	*agg4D C1010-00*
activator in pBAD	–	AggR 042	Rns ETEC	CfaD H149/5

B



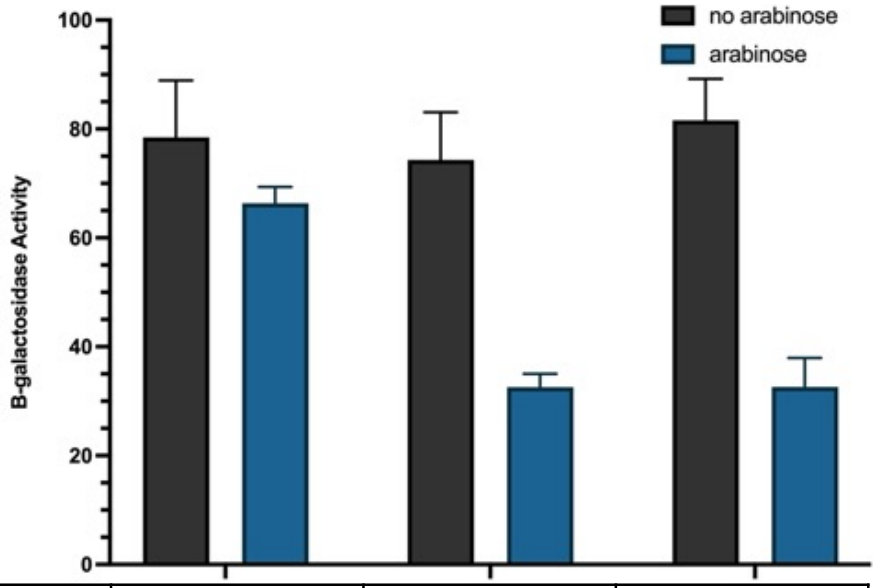
promoter in pRW50	*afaB 042*	*afaB 042*	*afaB 042*	*afaB 042*
activator in pBAD	–	AggR 042	Rns ETEC	CfaD H149/5

C



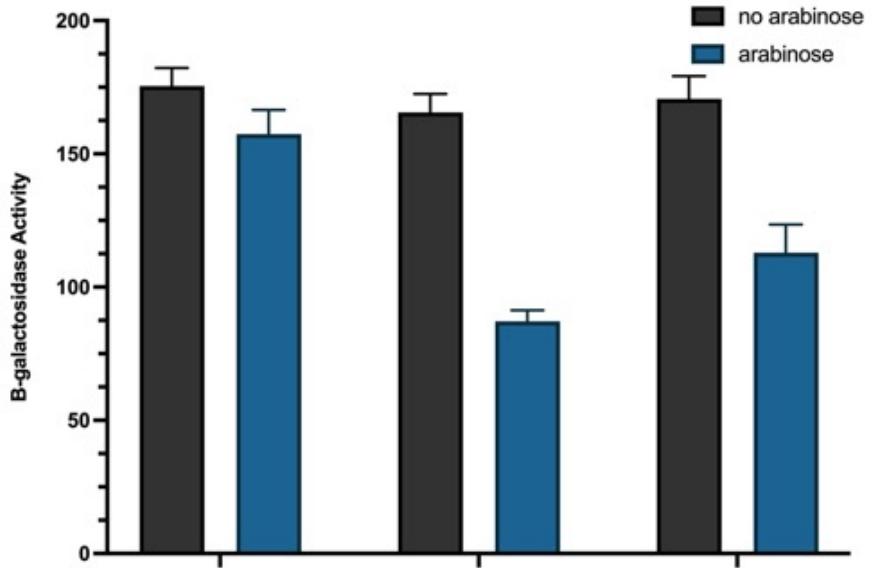
promoter in pRW50	*yicS 042*	*yicS 042*	*yicS 042*	*yicS 042*
activator in pBAD	_	AggR 042	Rns ETEC	CfaD H149/5

D



promoter in pRW50	*nlpA 042*	*nlpA 042*	*nlpA 042*
activator in pBAD	_	AggR 042	Rns ETEC

E



promoter in pRW50	*nlpA* *042*	*nlpA* *042*	*nlpA* *042*
activator in pBAD	_	AggR 042	CfaD H149/5

Figure 4.7 Measurement of promoter activities: promoters from EAEC C1010-00 and 042

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* containing pRW50*afaB* (A), pRW50*agg4D* (B), pRW50*yicS* (C), or pRW50*nlpA* (D, E). Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. The cells containing pRW50*agg4D* EAEC strain C1010-00 carrying either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5.
B. The cells containing pRW50*afaB* EAEC strain 042 carrying either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5.
C. The cells containing pRW50*yicS* EAEC strain 042 carrying either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5.
D. The cells containing pRW50*nlpA* EAEC strain 042 carrying either pBAD30 (empty vector), pBAD*aggR* 042 or pBAD*rns* ETEC.
E. The cells containing pRW50*nlpA* EAEC strain 042 carrying either pBAD30 (empty vector), pBAD*aggR* 042 or pBAD*cfaD* H149/5.

activated by CfaD, and the *yicS* and *nlpA* 042 promoters were only weakly regulated by CfaD. Therefore, I wanted to compare the base sequence of the CfaD-activated promoters with the base sequence of promoters where CfaD was unable to activate.

4.2.4 Potential upstream activator binding sites

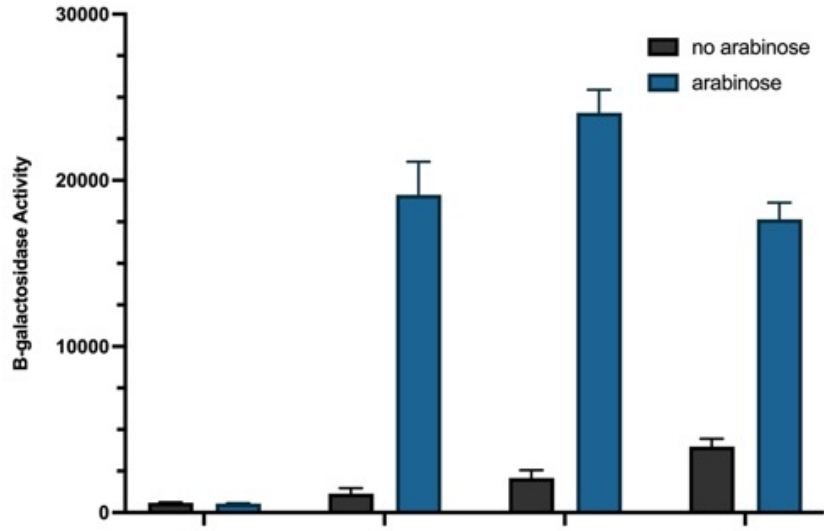
To investigate whether other EAEC 042 promoters could be activated by CfaD from EAEC strain H149/5, the pRW50*aafD100* recombinant plasmid was transformed into BW25113 carrying empty pBAD30 vector, pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5. Cultures were grown in the presence or absence of 0.2% arabinose, and β-galactosidase activity was measured to indicate promoter activity. Results in Figure 4.8A show that cells containing pRW50*aafD100* had increased β-galactosidase levels in the presence of Rns (a 44-fold increase compared to empty vector, 24060 and 539 units, respectively), these levels were higher compared to cells with AggR expression induced (35-fold increase, 19121 and 539 units, respectively). Interestingly, cells containing pRW50*aafD100* in the presence of CfaD also had increased β-galactosidase levels compared to the basal level, with a 32-fold increase compared to pBAD30 (17642 and 539 units, respectively). The background levels in the absence of arabinose were low in cells containing pRW50*aafD100* and pBAD*aggR*, with 3% of the induced β-galactosidase level when AggR expression was not induced.

As the *aafD100* EAEC 042 promoter was strongly activated by CfaD, DNA sequences from the different promoter regions were aligned (Figure 4.8B). Promoters activated by CfaD H149/5 were separated from the promoters that were not activated by CfaD. The sequence alignments show the juxtaposition of

the DNA binding site for AggR and the -10 elements, as expected, in each promoter there is an AggR DNA binding site located at a Class II position, 20-22 bp upstream of the -10 hexamer. However, I also identified a possible second AggR DNA binding site upstream from this site at some promoters (Figure 4.8B). For the CfaD-activated promoters, the putative upstream DNA binding site for AggR follows a similar organisation, whereas the promoters that show weak or no activation by CfaD have a different organisation. The putative upstream site follows a similar pattern to the AggR consensus: A/T rich and ending with TATC (*aatP* and *cfaD* H149/5), AATC (*caf1M* H149/5) or GTCT (*aafD100* 042).

This analysis led to the suggestion that the H149/5 promoters, and the 042 *aafD100* promoter, might have an upstream operator which is required to permit CfaD-dependent activation. To test this, working with the *caf1M* promoter, the suggested DNA sites for AggR were mutated, to knock out each site, either together or separately (Figure 4.9A). The putative upstream site is not a consensus AggR binding site, and so, in a parallel construction, it was made consensus. Hence, point mutations were introduced into pRW50*caf1M* using Q5 site-directed mutagenesis. The recombinant plasmids containing the pRW50*caf1M* mutant promoters, 'both', 'down', 'up' and 'up consensus', were transformed into BW25113 with either pBAD30 empty vector, pBAD*aggR* or pBAD*cfaD*. Cultures were grown in the presence and absence of 0.2% (*w/v*) arabinose and the β-galactosidase activity was measured to determine the promoter activity, with each pRW50*caf1M-mutant* being assayed alongside the wild-type pRW50*caf1M*. The data in Figure 4.9 show that cells containing the recombinant plasmid pRW50*caf1M-down* had lower β-galactosidase levels in the

A



promoter in pRW50	*aafD100 042*	*aafD100 042*	*aafD100 042*	*aafD100 042*
activator in pBAD	_	AggR 042	Rns ETEC	CfaD H149/5

B

```
                                AWWWWWWTATC          AWWWWWWTATC                                  TATAAT
aatP H149/5          CTCTTCTGATTAATATCATAAAGTGATAATAATTTATCATAAACAAACCCAGCACTAGTATATTCTGTACCA
caf1M H149/5         AACTAAATGGAATAATCTTAGTGTGTTGCTTTTTTATCTGACGTTAATTCCGCACTTGCTTATTATGCACAG
cfaD H149/5          GGCCTCTTACACATATCTGATTGAGGTTTTTTATTATCGAATTTTGTGTAAAATTCATTTCCTATACATAAA

aafD100 042          TTTATTCAATAAAGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAGTAGCATATTAAC

agg4D C1010          AAATGAAACTGGCATGATAATATTTTTATTTATTTATCTTTTTTTTGGGCGCTATGTTTTTATAATCTTGAA
afaB 042             CCTGTTTTATATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAG
yicS 042             ACCCGTAGATGATGTGTTGTCAGTTTCACTTTTTTATCCTTTTTTTAATCGTTAACTGACTATAATGGCAAG
```

Figure 4.8 Measurement of the *aafD100* promoter activity and alignments of the promoter sequences

A. The figure illustrates measured b-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*aafD100*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.
B. The figure shows an alignment of the regulatory regions from pRW50*aatP* H149/5 (EAEC strain H149/5), *caf1M* H149/5 (EAEC strain H149/5), *cfaD* H149/5 (EAEC strain H149/5), *aafD100* 042 (EAEC strain 042), *agg4D* C1010 (EAEC strain C1010-00), *afaB* 042 (EAEC strain 042), *yicS* 042 (EAEC strain 042). The promoters activated by CfaD H149/5 are highlighted in green. The DNA binding sites for AggR are highlighted in blue and the -10 hexamers are in purple.

presence of AggR or CfaD (2272 and 1178 units, respectively) than cells containing pRW50*caf1M* wild-type (16912 and 16716 units, respectively) (Figure 4.9B). Similarly, the β-galactosidase levels in cells containing pRW50*caf1M-both* were lower in the presence of AggR or CfaD (2334 and 1263 units, respectively) (Figure 4.9C). The β-galactosidase levels in cells containing either pRW50*caf1M-down* or pRW50*caf1M-both* with the expression of AggR or CfaD induced were comparable to basal levels. Thus, knocking out the downstream AggR-binding site is sufficient to stop activation by AggR or CfaD.

The β-galactosidase levels were high in cells carrying pRW50*caf1M-up consensus* in the presence of AggR or CfaD, with 28-fold (28004 units) and 27-fold (27009 units) increase, respectively, compared to empty vector basal levels (1392 units) (Figure 4.9D). Most significantly, targeting the final four bases in the potential upstream AggR-binding site in the *caf1M* regulatory region does not negatively impact AggR or CfaD activation at this promoter (Figure 4.9E). Mutating this upstream site to match the consensus DNA binding site for AggR has little or no effect on activation by AggR or CfaD.

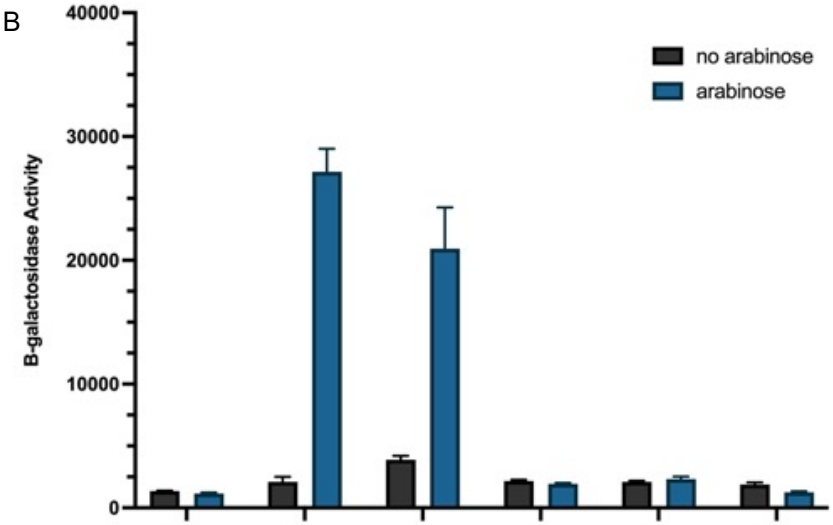
From these results, taken together, I conclude that the potential upstream site for CfaD is not required to confer CfaD- or AggR-mediated activation on this H149/5 promoter and that this has no impact on why the EAEC strain 042 and strain C1010-00 promoters are CfaD-independent.

4.2.5 A new working hypothesis

The experiments presented in the previous section show that the suggested upstream activator-binding site plays little or no role in promoter activation by CfaD. Hence, I collated the sequences of promoters that are activated by AggR

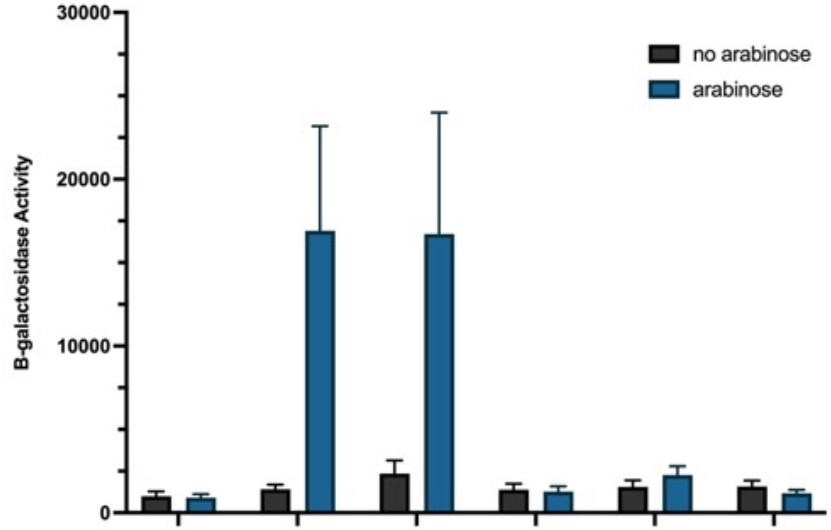
A

B



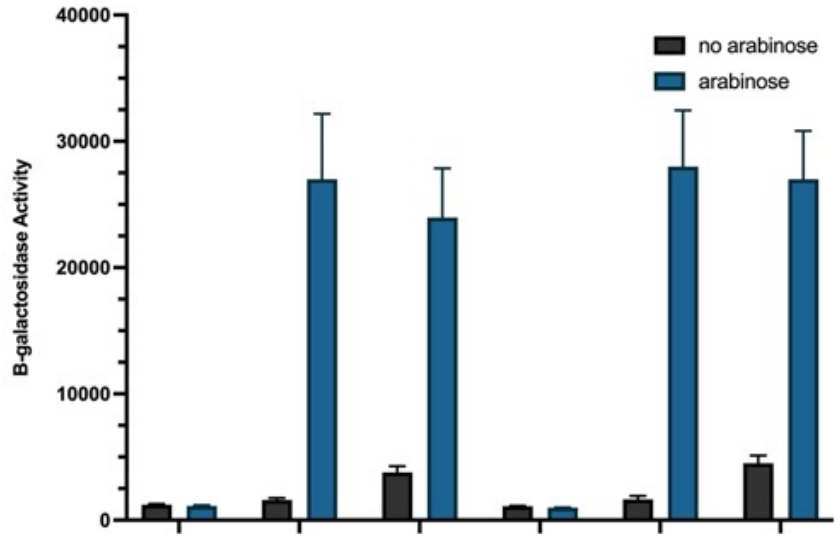
promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*caf1M both*	*caf1M both*	*caf1M both*
activator in pBAD	_	AggR 042	CfaD H149/5	_	AggR 042	CfaD H149/5

C



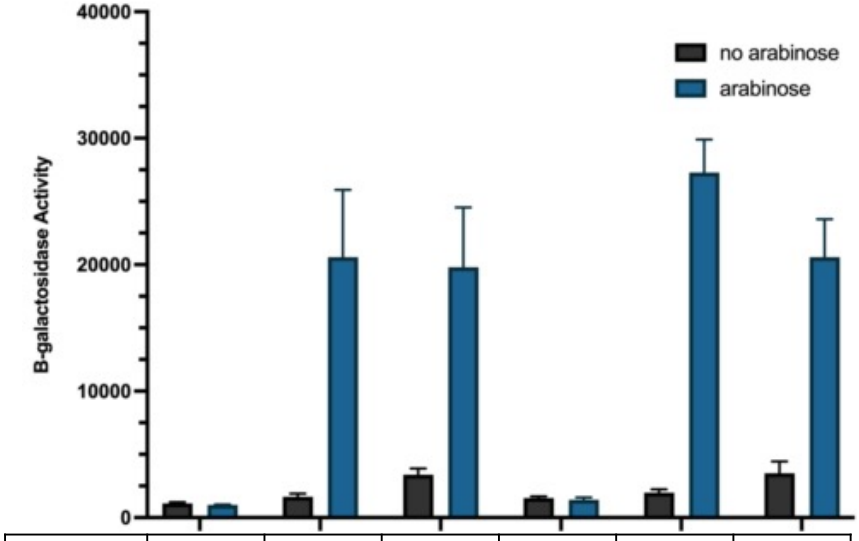
promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*caf1M down*	*caf1M down*	*caf1M down*
activator in pBAD	_	AggR 042	CfaD H149/5	_	AggR 042	CfaD H149/5

D



promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*caf1M up consensus*	*caf1M up consensus*	*caf1M up consensus*
activator in pBAD	–	AggR 042	CfaD H149/5	–	AggR 042	CfaD H149/5

E



promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*caf1M up*	*caf1M up*	*caf1M up*
activator in pBAD	–	AggR 042	CfaD H149/5	–	AggR 042	CfaD H149/5

Figure 4.9 Mutational analysis of the *caf1M* regulatory region

A. The figure shows the DNA sequence of the *caf1M* promoter sequence from EAEC strain H149/5. The DNA sites for AggR and promoter -10 elements are highlighted in blue and purple respectively. Bases where mutations were made are highlighted in red and mutant names are included alongside the corresponding sequence.
B. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW50*caf1M* or pRW50*caf1M-both*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The b-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of

bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

C. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW50*caf1M* or pRW50*caf1M-down KO*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

D. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW50*caf1M* or pRW50*caf1M-up consensus*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

E. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW50*caf1M* or pRW50*caf1M-up*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

and Rns, but not CfaD, and compared them with the sequences of promoters that are activated by AggR, Rns and CfaD. Figure 4.10 shows the weblogos for the different collations. Thus, promoters activated by AggR 042, Rns ETEC and CfaD H149/5 were compared (Figure 4.10A) separately to those regulated only by AggR 042 and Rns ETEC (Figure 4.10B). The height of each letter corresponds to the number of times it is present in the regulatory regions at certain positions. The DNA binding site for AggR consistently follows the consensus pattern for all the promoters, with 22 bp between the AggR binding site and the -10 element. The -10 hexamers are not consensus; however, this has not been shown to negatively affect expression.

Since I found that potential upstream AggR-sites are not required for CfaD-dependence, I decided to transplant part of a CfaD-independent promoter into a CfaD-dependent promoter to attempt to destroy dependence. Figure 4.11A shows an experiment with the *afaB/caf1M* hybrid promoter, where the sequence upstream from the AggR DNA binding site in the *afaB* promoter is transplanted into the sequence upstream of the AggR binding site in the *caf1M* promoter, using Q5 site-directed mutagenesis.

The recombinant plasmid, pRW50*caf1M*, and the plasmid containing the mutant sequence, pRW50*afaB-caf1M* hybrid, were transformed into BW25113 with either pBAD30 empty vector, pBAD*aggR* or pBAD*cfaD*. This allowed direct comparison of the mutant *caf1M* with the wild-type *caf1M* promoter. The β-galactosidase activity was measured to indicate promoter activity. The results in Figure 4.11B show that the β-galactosidase levels were high in cells containing the pRW50*afaB-caf1M* hybrid recombinant plasmid in the presence of AggR

(21296 units) or CfaD (16249 units), the increase was 27-fold and 21-fold, respectively, compared to the pBAD30 empty vector (772 units). The activity levels are slightly lower than in cells containing pRW50*caf1M* wild-type with CfaD expression induced (19975 units), however, the induction compared to the empty vector basal levels (965 units) was 21-fold, similar to that seen with *afaB-caf1M* hybrid. Therefore, my results show that the upstream *afaB* promoter sequences in the *afaB-caf1M* hybrid promoter do not block activation by CfaD or AggR. This argues that the determinants that confer CfaD-dependent activation lie downstream of the potential upstream site (Figure 4.11A).

The sequence compilations in Figure 4.10 identify the motif 'GTG', positioned just upstream of the AggR DNA binding site, in the CfaD-activated promoters (Crooks *et al.*, 2004; Schneider and Stephens, 1990). This motif is present in H149/5 promoters and *aafD100*, but it is not present in the promoters that CfaD failed to activate. Therefore, Q5 site-directed mutagenesis was utilised to mutate this GTG in the *caf1M* promoter, to CCG, as found at this location in the *afaB* promoter (Figure 4.12A). Hence, the pRW50*caf1M-CCG* recombinant plasmid was transformed into BW25113 with pBAD30, pBAD*aggR* or pBAD*cfaD*. Cultures were grown in the presence or absence of arabinose, the starting pRW50*caf1M* plasmid was also included as a comparison, and the β-galactosidase levels were measured to indicate promoter activity. In Figure 4.12B, the results show that cells containing the pRW50*caf1M-CCG* hybrid recombinant plasmid had high β-galactosidase levels, compared to empty vector basal levels (848 units), when CfaD (15682 units) or AggR (21628 units) expression was induced, with 18-fold

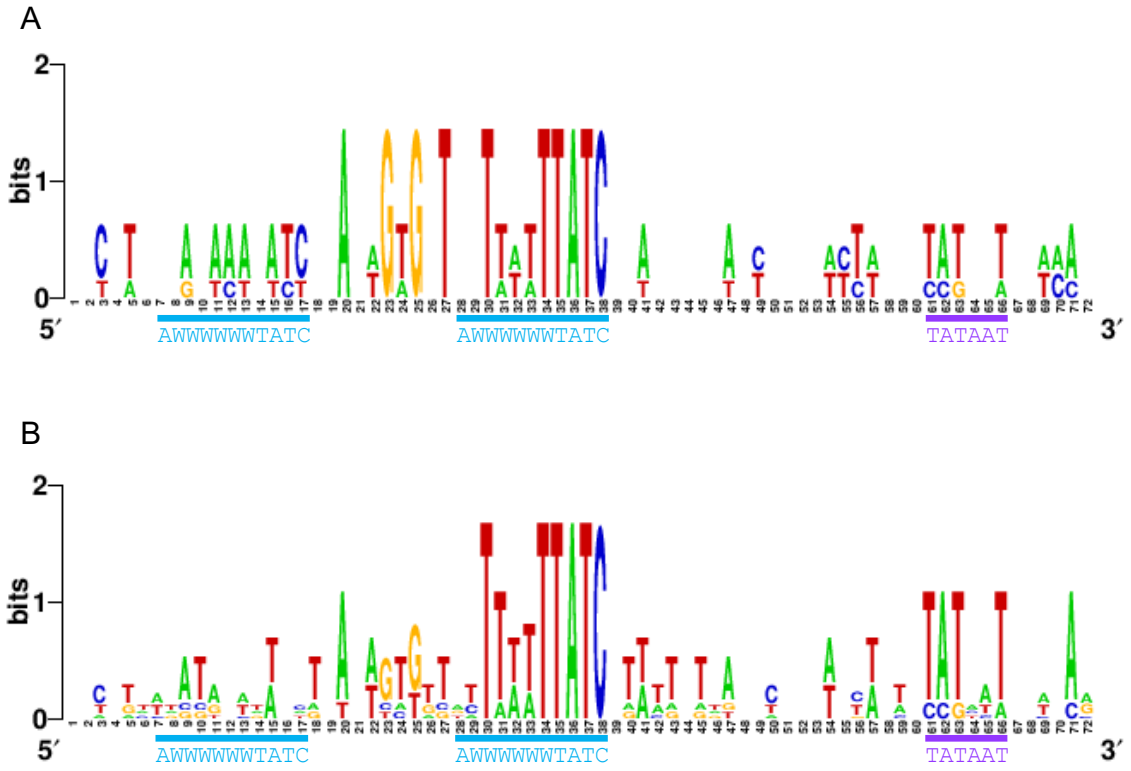


Figure 4.10 Classification of AggR activated promoters

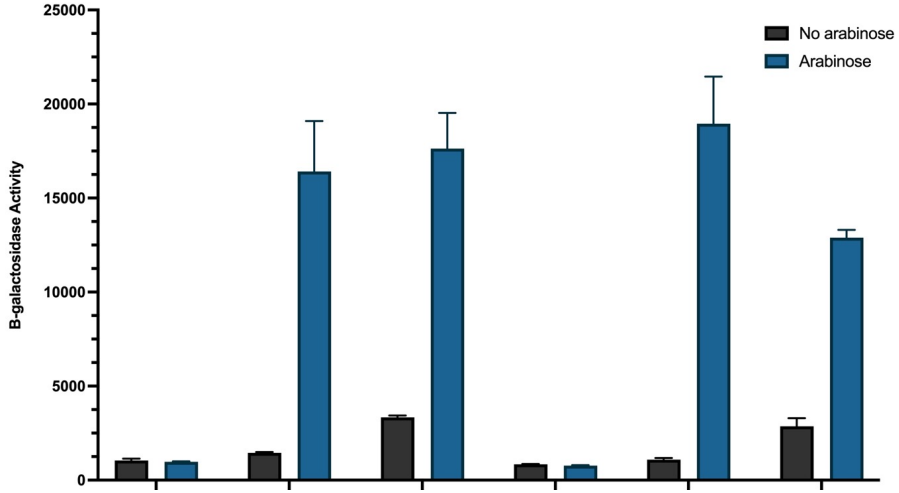
Each panel shows the weblogo for a group of promoters. The panels illustrate the promoters compared from EAEC. The weblogo was generated using http://weblogo.berkeley.edu website. The height of each letter corresponds to number of times each letter is present in the promoter regions. The position of the DNA binding site for AggR is indicated beneath the weblogo with the consensus sequence highlighted in blue, the potential upstream site is also indicated in blue. The -10 element is in purple underneath the weblogo, indicating the position of the bases. Adapted from Crooks *et al.* (2004).

A. The sections of DNA from promoters that are activated by AggR 042 (EAEC strain 042), Rns ETEC and CfaD H149/5 (EAEC strain H149/5). The promoters used to generate this weblogo were from EAEC strains H149/5 and *aafD100* from strain 042.
B. The sections of DNA from promoters that are activated by AggR 042 (EAEC strain 042) and Rns ETEC but not CfaD H149/5. The promoters used to generate this weblogo were from EAEC strains H149/5, strain 042 and strain C1010-00.

A

```
                    AWWWWWWTATC          AWWWWWWTATC                    TATAAT
afaB 042            CCTGTTTTATATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAG

caf1M H149/5        AACTAAATGGAATAATCTTAGTGTGTTGCTTTTTTATCTGACGTTAATTCCGCACTTGCTTATTATGCACAG
afaB-caf1M hybrid   ACTGTTTTATATTTTAATGAGTGTGTTGCTTTTTTATCTGACGTTAATTCCGCACTTGCTTATTATGCACAG
```

B



promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*afaB-caf1M hybrid*	*afaB-caf1M hybrid*	*afaB-caf1M hybrid*
activator in pBAD	–	AggR 042	CfaD H149/5	–	AggR 042	CfaD H149/5

Figure 4.11 Analysis of an *afaB-caf1M* hybrid promoter

A. The figure shows the DNA sequence of the *caf1M* promoter sequence from EAEC strain H149/5 and the *afaB* promoter sequence from EAEC strain 042. The DNA sites for AggR and promoter -10 elements are highlighted in blue and purple

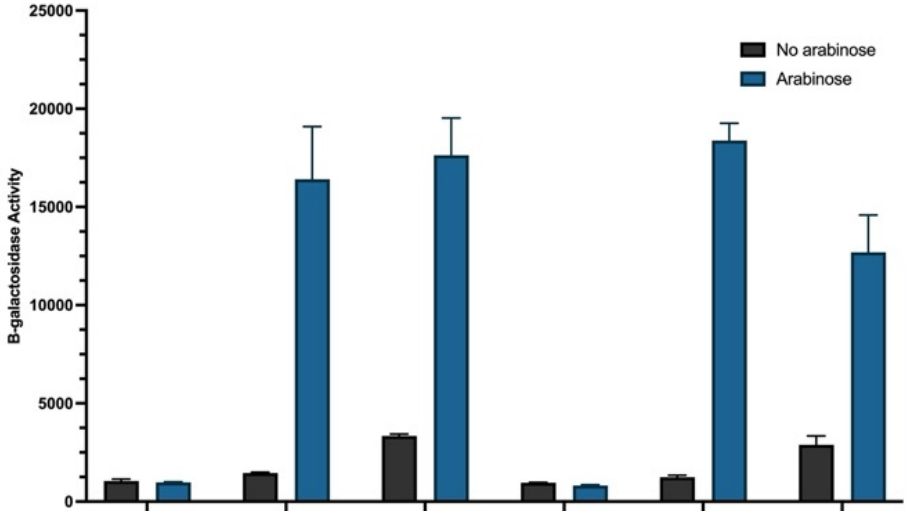
respectively, with the consensus sequences above. Bases where mutations were made are highlighted in red and mutant name, *afaB-caf1M* hybrid, is included alongside the corresponding sequence.

B. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW50*caf1M* or pRW50*afaB-caf1M hybrid*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A

```
                    AWWWWWWTATC          AWWWWWWTATC                              TATAAT
afaB 042            CCTGTTTTATATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAG

caf1M H149/5        AACTAAATGGAATAATCTTAGTGTGTTGCTTTTTTATCTGACGTTAATTCCGCACTTGCTTATTATGCACAG
caf1M-CCG           AACTAAATGGAATAATCTTAGTCCGTTGCTTTTTTATCTGACGTTAATTCCGCACTTGCTTATTATGCACAG
```

B



promoter in pRW50	*caf1M H149/5*	*caf1M H149/5*	*caf1M H149/5*	*caf1M CCG*	*caf1M CCG*	*caf1M CCG*
activator in pBAD	–	AggR 042	CfaD H149/5	–	AggR 042	CfaD H149/5

Figure 4.12 Analysis of a *caf1M-CCG* promoter mutant

A. The figure shows the DNA sequence of the *caf1M* promoter sequence from EAEC strain H149/5 and the *afaB* promoter sequence from EAEC strain 042. The DNA sites for AggR and promoter -10 elements are highlighted in blue and purple respectively, with the consensus sequences above. Bases where mutations were made are highlighted in red and mutant name, *caf1M-CCG*, is included alongside the corresponding sequence.

B. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW50*caf1M* or pRW50*caf1M-CCG*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

and 25-fold increases, respectively. However, the levels are slightly lower than in the cells containing the pRW50*caf1M* wild-type recombinant plasmid with CfaD (20-fold, 19975 units) compared to the basal levels with pBAD30 empty vector (965 units). The results indicate that the *caf1M-CCG* hybrid promoter is activated by CfaD H149/5. Hence CfaD-dependent activation is not altered when the GTG sequence upstream of the DNA-binding site for AggR is mutated.

4.3 An active promoter, an AggR-dependent promoter, and an AggR-independent promoter

In Chapter 3, I discussed AggR-dependent promoters that follow the organisation described by Muhammad Yasir (Yasir *et al.*, 2019), and the focus was also on characterising the more complex bi-directional promoter in the *yicS-nlpA* intergenic region (Section 3.3). This section will investigate some of the differences between a promoter identified in H149/5, *aap*, which had been found to be a very powerful promoter in other EAEC strains (Yasir *et al.*, 2019; Ellis *et al.*, 2019), and a promoter that was suggested as a target for AggR activation, *bssS* (Yasir *et al.*, 2019).

4.3.1 *aap*: A very strong promoter

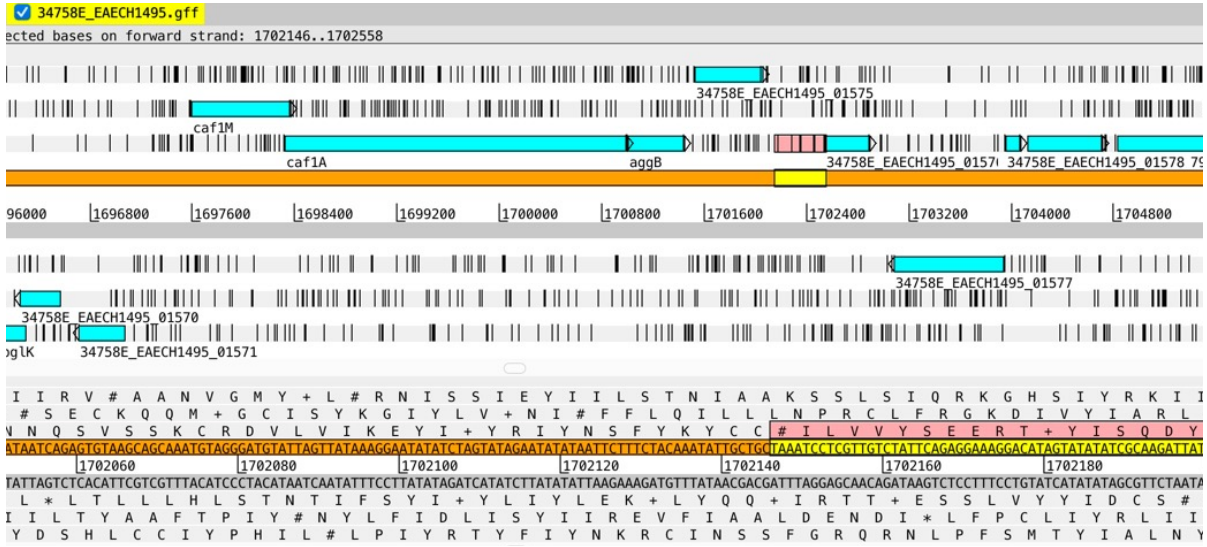
The *aap* gene, which encodes the dispersin (Aap) surface protein, has previously been investigated in EAEC 042 and its promoter was shown to be AggR-activated (Morin *et al.*, 2013; Yasir *et al.*, 2019). Following whole genome sequencing of EAEC H149/5, the *aap* promoter and gene were identified in strain H149/5 (Figure 4.13A and Figure 4.13B). The regulatory region of the H149/5 *aap* gene was found to contain several potential DNA binding sites for AggR (Figure 4.13C).

Thus, the regulatory region was amplified and cloned into pRW50, using primers that introduced EcoRI and HindIII sites upstream and downstream, respectively. The pRW50*aap* recombinant plasmid was transformed into BW25113 with either pBAD30 empty vector, pBAD*aggR*, pBAD*rns* or pBAD*cfaD*. Cultures were grown in the absence and presence of 0.2% (*w/v*) arabinose and β-galactosidase activities were measured to gauge promoter activity. The results in Figure 4.13D show high β-galactosidase levels in cells containing the pRW50*aap* plasmid, in the presence or absence of AggR (14882 and 13434 units, respectively), Rns (14962 and 14337 units, respectively) or CfaD (14288 and 13510 units, respectively). There was a slight increase in β-galactosidase levels in the presence of AggR, Rns and CfaD, however, this is difficult to distinguish due to the high background levels. This is consistent with the previous work where the strain 042 *aap* promoter was found to be a very active promoter, but a single base change downstream of the -10 element greatly reduces the high constitutive level of expression revealing a promoter that is dependent on AggR (Ellis *et al.*, 2019; Yasir, 2017). Due to lack of time, this observation was not followed up.

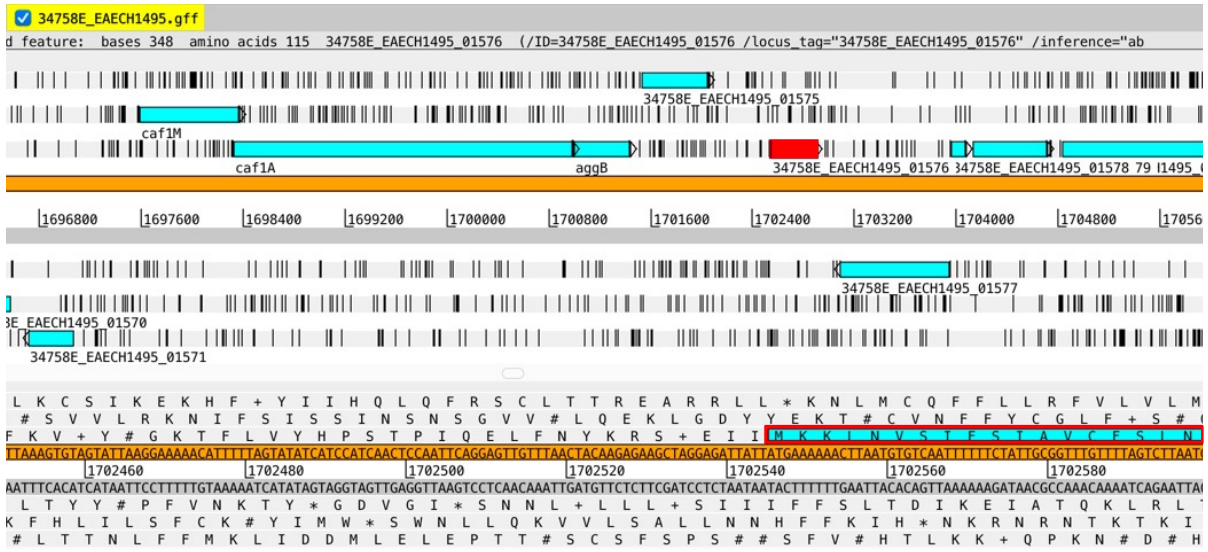
4.3.2 *bssS*: A putative DNA binding site for AggR/A simple promoter

Another potential AggR-dependent promoter in EAEC 042 had been identified adjacent to the *bssS* gene (Yasir *et al.*, 2019). This is also present in *E. coli* K-12 and a potential DNA binding site for AggR was identified. BssS, formerly known as YceP, has been implicated in regulating biofilm formation (Domka *et al.*, 2006). The alignments of the *bssS* 042 and K-12 regulatory regions indicate that the DNA sequences are almost identical, except for a few base changes, though the

A


34758E_EAECH1495.gff
ected bases on forward strand: 1702146..1702558
34758E_EAECH1495_01575
caf1M
caf1A
aggB
34758E_EAECH1495_01577
34758E_EAECH1495_01570
34758E_EAECH1495_01571

B

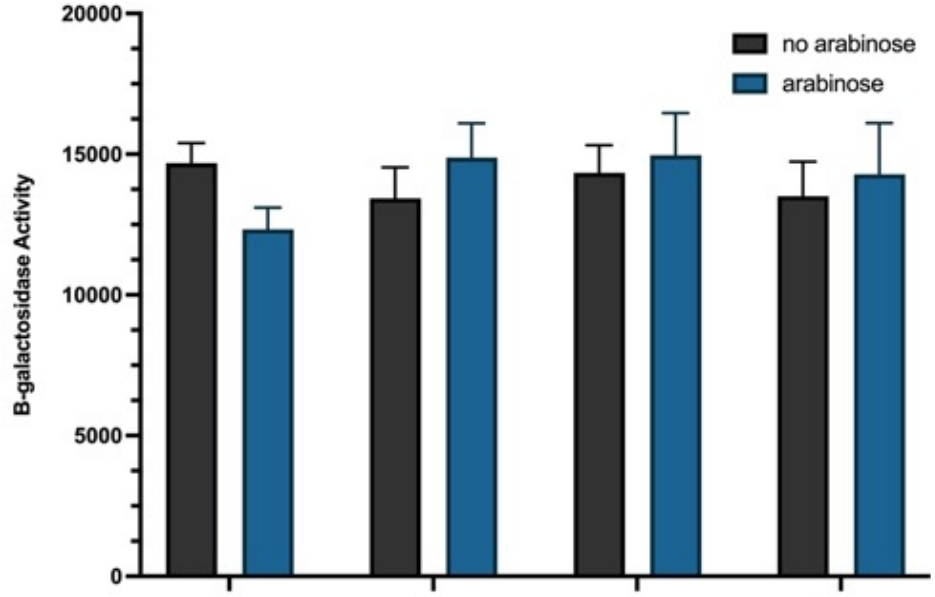

34758E_EAECH1495.gff
d feature: bases 348 amino acids 115 34758E_EAECH1495_01576 (/ID=34758E_EAECH1495_01576 /locus_tag="34758E_EAECH1495_01576" /inference="ab
34758E_EAECH1495_01575
caf1M
caf1A
aggB
34758E_EAECH1495_01577
34758E_EAECH1495_01571

C

aap(H149/5)

```
        .   .         .         .         .         .         .         .
GAATTCCTAAATCCTCGTTGTCTATTCAGAGGAAAGGACATAGTATATATCGCAAGATTATAGGATGAAAGGCCTC
      414 410       400

    .         .         .         .         .         .         .         .
GGGGTCTGTTTTATCCAACGAGGTCAGTCCTTTCCTCTCTATTTGTTCAAACTCCAGAATATGTGTTAGATCGTAT
                                              300

        .         .         .         .         .         .         .
TCTTTTCAAAAAGAGAGGAGATTTCCGTTTGGCGGAAATTTCAGAACAATATCCATTAGTTAGAATAGATAAAATT
                                                                    200

   .         .         .         .         .         .         .         .
GCATTTAATTGTCAGTTGCATCTTTTGCTGGGTAGAAAAGCAGGATCTAACGGTGAAGTTTTATTATCTATCTGCT

      .         .         .         .         .         .         .
AATACTTTTAAAGTGTAGTATTAAGGAAAAACATTTTTAGTATATCATCCATCAACTCCAATTCAGGAGTTGTTTA
                  100

.         .         .         .         .
ACTACAAGAGAAGCTAGGAGATTATTATGAAAAAACTTAAAAGCTT
                                        1
```

D



promoter in pRW50	*aap* *H149/5*	*aap* *H149/5*	*aap* *H149/5*	*aap* *H149/5*
activator in pBAD	–	AggR 042	Rns ETEC	CfaD H149/5

Figure 4.13 Analysis of the *aap* regulatory region

A. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below

these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The pink highlighted section in the forward open reading frame third row indicates the *aap* regulatory region. Adapted from Carver *et al.* (2008).

B. A screenshot of the main editor window of Artemis showing an annotated section of the EAEC H149/5 genome sequence. The solid orange and grey lines in the centre represent the forward and reverse DNA strands. The three forward and three reverse open reading frames are above and below these bars, respectively. Genes are illustrated by blue boxes and the vertical lines indicate stop codons. Below is a zoomed in view of the amino acid sequence, following the same organisation as above. The red gene box and the corresponding highlighted amino acid sequence indicate the *aap* gene. Adapted from Carver *et al.* (2008).

C. The DNA sequence covering the *aap* promoter from EAEC H149/5 was cloned into pRW50. The fragment contains an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The -10 element is highlighted in purple, the DNA binding site for AggR is highlighted in blue and the coding sequence is in bold. There are multiple potential AggR binding sites and another potential -10 element, these are highlighted in bold and underlined.

D. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*aap*, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*rns* ETEC or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

highlighted suggested DNA binding sites for AggR, and the -10 elements, are identical (Figure 4.14). The *lac* expression vector, pRW224 (Figure 4.15A), was utilised here, and the regulatory region for EAEC 042 *bssS* (Figure 4.15B) and *E. coli* K-12 *bssS* (Figure 4.15C) were cloned upstream of the *lacZ* gene, using primers that introduced an EcoRI restriction site upstream and a HindIII site downstream. The recombinant plasmids were transformed into *E. coli* BW25113 with either pBAD30 or pBAD*aggR*, and compared to pRW50*afaB*, as a positive control. Promoter activities were determined by measuring the β-galactosidase activity of lysed cells. The results in Figure 4.16 show that cells containing the recombinant plasmid pRW224*bssS* 042 had low β-galactosidase levels when the expression of AggR was induced by the presence of arabinose (4.6 units). Similarly, the β-galactosidase levels were also low in cells containing pRW224*bssS* K-12 (4.5 units). β-galactosidase levels in cells containing pRW224*bssS* 042 and pRW224*bssS* K-12 were as low as the basal levels in cells containing empty vector (4.2 and 4.8 units, respectively). The results show that, while the control *afaB* promoter is AggR-dependent, there is no evidence for promoter activity in the likely regulatory region of *bssS* from EAEC 042 nor *bssS* from *E. coli* K-12, despite a potential DNA binding site for AggR. Hence, the presence of a putative DNA binding site for AggR does not necessarily indicate AggR-activation of a promoter.

4.3.3 A consideration of the factors that confer AggR-dependence

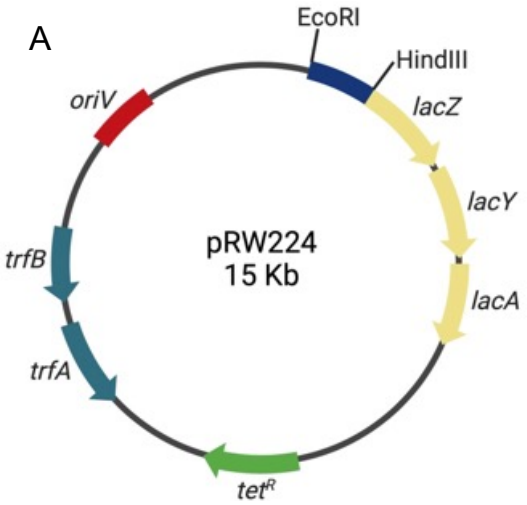
As it had been shown that the presence of an AggR DNA binding site did not confer AggR-dependence on the *bssS* promoter, I decided to compare the organisation of the *bssS* regulatory region with an AggR-activated promoter.

```
bssS042         GAATTCTTGACGCCCTGGCTGGCGAACTTTCCCGCCGTATTCAGTATGCGTTTCCTGATA
bssSK-12        GAATTCTTGACGCCCTGGCTGGCGAACTTTCCCGCCGTATTCAGTATGCGTTTCCTGATA
                ************************************************************

bssS042         ATGAAGGCCACGTATCGGTACGTTATGCCGCAGCGAATAATTTATCGGTTATTGGCGCAA
bssSK-12        ATGAAGGCCACGTATCGGTACGTTATGCCGCAGCGAATAATTTATCGGTTATTGGCGCGA
                ************************************************************ *

bssS042         CAAAAGAAGATAAACAGCGCATTAGCGAAATTCTCCAGGAAACGTGGGAAAGCGCCGATG
bssSK-12        CAAAAGAAGATAAACAGCGCATTAGTGAAATTCTCCAGGAAACGTGGGAAAGCGCCGATG
                ************************ ********************************

bssS042         ACTGGTAAGCTT
bssSK-12        ACTGGTAAGCTT
                ************
```

Figure 4.14 Alignment of the regulatory regions of *bssS* 042 and *bssS* K-12

The figure shows an alignment of the regulatory regions from *bssS* EAEC strain 042 and *bssS E. coli* strain K-12 cloned into pRW50. The fragment contains an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The DNA binding sites for AggR are highlighted in blue and the -10 hexamers are in purple.

A



B

bssS(042)

```
.         .         .         .         .         .         .
GAATTCTTGACGCCCTGGCTGGCGAACTTTCCCGCCGTATTCAGTATGCGTTTCCTGATAATGAAGGCCACGTATC
      180
.         .         .         .         .         .         .         .
GGTACGTTATGCCGCAGCGAATAATTTATCGGTTATTGGCGCAACAAAAGAAGATAAACAGCGCATTAGCGAAATT
         100
    .         .         .         .
CTCCAGGAAACGTGGGAAAGCGCCGATGACTGGTAAGCTT
                                  1
```

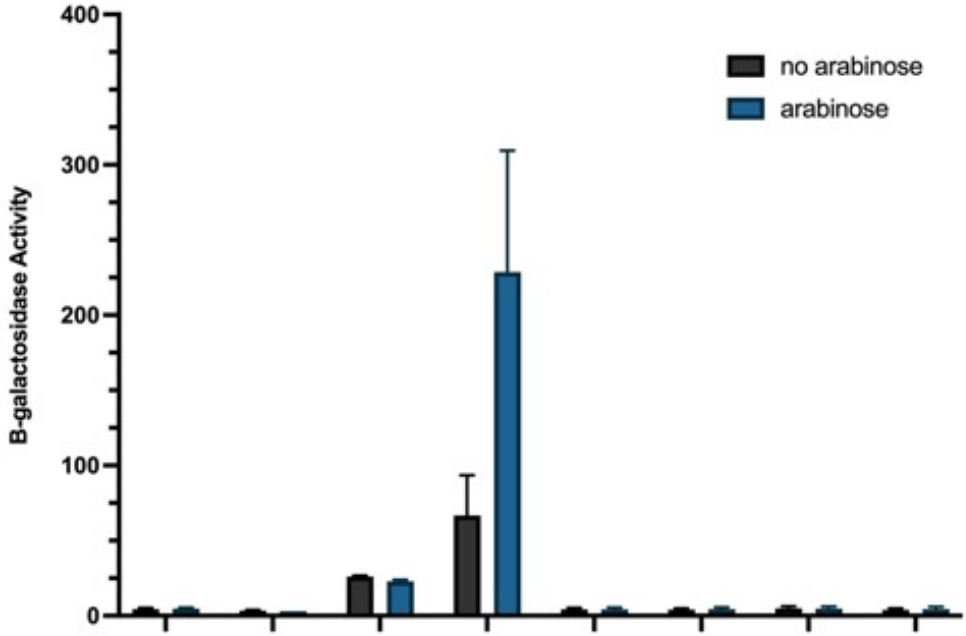
C

bssS(K-12)

```
.         .         .         .         .         .         .
GAATTCTTGACGCCCTGGCTGGCGAACTTTCCCGCCGTATTCAGTATGCGTTTCCTGATAATGAAGGCCACGTATC
      180
.         .         .         .         .         .         .         .
GGTACGTTATGCCGCAGCGAATAATTTATCGGTTATTGGCGCGACAAAAGAAGATAAACAGCGCATTAGTGAAATT
      100
    .         .         .         .
CTCCAGGAAACGTGGGAAAGCGCCGATGACTGGTAAGCTT
                                  1
```

Figure 4.15 DNA sequences of *bssS* 042 and K-12 in pRW224

A. The figure shows a plasmid map of the pRW224 *lacZ* expression vector, located upstream of the *lacZYA* operon are the EcoRI and HindIII restriction sites that were used to clone promoter fragments. The plasmid carries tetracycline resistance (*tet*^{R}). Created with BioRender.com.
B. The DNA sequence covering the *bssS* promoter from EAEC strain 042 was cloned into pRW224. The fragment contains an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The -10 element is highlighted in purple and the DNA binding site for AggR is highlighted in blue.
C. The DNA sequence covering the *bssS* promoter from *E. coli* strain K-12 was cloned into pRW224. The fragment contains an EcoRI site upstream and a HindIII site downstream, highlighted in grey. The -10 element is highlighted in purple and the DNA binding site for AggR is highlighted in blue.



activator in pBAD	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
promoter in pRW50	–	–	*afaB 042*	*afaB 042*	*bssS 042*	*bssS 042*	*bssS K-12*	*bssS K-12*

Figure 4.16 Measurement of *bssS* 042 and *bssS* K-12 promoter activities

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying either the recombinant *lac* expression plasmid pRW224 (empty vector, negative control), pRW50*afaB* (positive control), pRW224*bssS* 042 or pRW224*bssS* K-12, in addition to either pBAD30 (empty vector), pBAD*aggR* 042, or pBAD*cfaD* H149/5. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

Figure 4.17A shows the alignment of the regulatory regions from the AggR-dependent promoter, *afaB*, and the AggR-independent promoters, *bssS* from 042 and K-12. The *bssS* regulatory region is almost identical in 042 and K-12, though neither are AggR-dependent. There's a spacing of 22 bp between the AggR-binding site and -10 hexamer in *afaB* and *bssS*. The highlighted DNA-binding site for AggR resembles the consensus in each of these promoters. The -10 element, however, does not match consensus in any of the promoters and there is little consistency between promoters. The *bssS* promoters have identical -10 elements, GATAAA, the guanine at the first position in the -10 is likely the reason this promoter is non-functional. This is consistent with the result in Section 3.3.1 where I successfully conferred AggR-dependence on the previously non-functional *yicS* K-12 promoter by altering the non-consensus -10 element.

I can conclude from the results and comparison of the regulatory regions that simply having an AggR-binding site positioned at an optimum distance from a -10 element does not indicate that a promoter is functional or dependent on AggR.

4.4 Discussion

The prototypical EAEC strain 042 has been the focus of much of the published research on EAEC. However, worldwide, there is vast variation in EAEC strains, and it can easily be argued that much can be learned from studying this variation (Boisen *et al.*, 2020; Abdelwahab *et al.*, 2021). Many pathogenic strains from Brazil, isolated from children with acute diarrhoea, have been sequenced and analysed (Braga *et al.*, 2017; Alves *et al.*, 2010; Franca *et al.*, 2013; Pereira *et al.*, 2008). Of the strains that I analysed, strains H9/3 and I18/2 were found to carry the *aggR* gene, though the H149/5 strain does not. However, strain H149/5

Figure 4.17 Alignment of the base sequences at AggR-activated and non-activated promoters

The figure shows an alignment of the regulatory regions from pRW50*afaB* 042 (EAEC strain 042), *bssS* 042 (EAEC strain 042) and *bssS* K-12 (*E. coli* K-12). The DNA binding sites for AggR are highlighted in blue and the -10 hexamers are highlighted in purple.

carries a chromosomally encoded AggR-like transcription factor, labelled CfaD. Sequence alignments (Figure 4.2) show that the AggR-like transcription factors identified in I18/2 and H9/3 are almost identical to AggR from strain 042. However, CfaD H149/5 is somewhat different to AggR 042.

The AraC-family transcription factor, Rns, from ETEC, also has sequence homology to AggR (Munson and Scott, 1999). Results in Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7 and Figure 4.8 show that AggR-dependent promoters are also dependent on Rns, but, in some cases, AggR activates better than Rns.

The results in Figure 4.4, Figure 4.5 and Figure 4.6 also show that H149/5 promoters can be activated by strain 042 AggR and strain H149/5 CfaD. The binding affinities for each regulator with the promoters in H149/5 are different, the highest background levels with 'leaky' pBAD*aggR* or pBAD*cfaD* are seen with the *aatP* promoter, which therefore must be the tightest binder, followed by *cfaD* then *caf1M*. Results in Figure 4.6 show that CfaD autoactivates its own expression, which is consistent with AggR that also activates its own expression at the *aggR* promoter (Morin *et al.*, 2010; Abdelwahab *et al.*, 2021).

The *aatP* promoter has been characterised from the genomes of strain 042 as well as in two Egyptian strains. Since the strain 042 *aatP* promoter -10 hexamer (TACATT) was not consensus, the adenines at position 2 and 4 were mutated. Mutating the adenine at position 2 reduced, but did not destroy promoter activity, whereas additionally mutating position 4 stopped the activity of this promoter. This indicates that the -10 hexamer with ideal spacing of 21 bp to the AggR DNA binding site, and the other potential -10 element, with 23 bp spacing, are ambiguous (Abdelwahab *et al.*, 2021). The alignment of the H149/5 and 042 *aatP*

promoters is shown in Figure 4.18, the AggR DNA binding sites are highlighted and the -10 hexamers. The H149/5 *aatP* promoter has a consensus DNA binding site for AggR 21 bp upstream of the -10 hexamer, TATATT, which is close to consensus. As the H149/5 *aatP* promoter more closely resembles a consensus AggR-activated promoter, compared to 042, this could explain the tight binding at the promoter. However, due to time constraints, it was not possible to follow up comparison of the 042 and H149/5 *aatP* promoters.

The *aatP* and *cfaD* promoters have A/T rich, consensus DNA binding sites for AggR, however, *caf1M* has guanine and cytosine positioned at the beginning of the AggR-binding site (GCTTTTTTATC). In Chapter 3 (Section 3.4), I hypothesised that it is important for the first base to be adenine or thymine, as *yicS-nlpA* EAEC H92/3 has guanine at the first position and low expression levels. Additionally, it has been shown that expression levels were highest at the 042 *aafD* promoter when an adenine was the first base of the DNA binding site for AggR (Yasir, 2017). Therefore, I conclude that the weaker induction at the *caf1M* promoter, compared to *paatP* and *pcfaD*, is due to the non-consensus AggR DNA binding site, with the guanine and cytosine at position 1 and 2.

EAEC strain H149/5 has been characterised as an ‘atypical’ strain due to the discrete aggregative adherence pattern when colonising host intestinal mucosa. Additionally, previous analysis failed to find AAF fimbrial genes. However, following whole genome sequencing, it was possible to find the *caf1M* promoter and gene (Figure 4.5A and Figure 4.5B). The BLAST search shown in Figure 4.5C indicates that Caf1M is homologous to fimbrial genes. This indicates that we have identified a novel AAF gene in strain H149/5.

Figure 4.18 Alignment of the regulatory regions of *aatP* EAEC strain 042 and H149/5

The figure shows an alignment of the regulatory regions from pRW50*aatP* 042 (EAEC strain 042) and *aatP* H149/5 (EAEC strain H149/5). The DNA binding sites for AggR are highlighted in blue, the -10 hexamers are in purple, and the coding sequence is in bold.

The results in Figure 4.7 show that the *afaB* promoter from strain 042 and the *agg4D* promoter from strain C1010-00 fail to be activated by CfaD. The cells containing pRW50*yicS* had slightly increased β-galactosidase levels in the presence of CfaD, though this increase was marginal compared to the background levels. The β-galactosidase activity in cells containing pRW50*nlpA* were slightly reduced in the presence of CfaD. This indicates that CfaD can regulate the *yicS-nlpA* bi-directional promoter, though very weakly. This is consistent with the conclusions in Chapter 3 (Section 3.4) that it is very straightforward for this bi-directional promoter to drift in and out of the AggR-regulon. This is also consistent with the "occupying time theory", of Mohamed El-Robh, which explains how bi-directional promoters can function independently. Thus, RNAP binds at different times to the bi-directional promoters and this time at one promoter is so brief that it does not interfere with RNAP binding at the other promoter (El-Robh and Busby, 2002). The *yicS-nlpA* bi-directional promoter was identified in EAEC H149/5 but the regulatory region is identical to *E. coli* K-12. This indicates that *yicS* expression has been removed from the CfaD-regulon in H149/5, presumably because it is unnecessary for infection.

The results in Figure 4.8 show that the AggR-dependent strain 042 promoter *aafD100* can also be activated by both Rns ETEC and CfaD H149/5. This is an interesting result because the other strain 042 AggR-dependent promoters, *afaB* and *agg4D*, are not activated by CfaD. DNA sequence alignments of the regulatory regions, shown in Figure 4.8, suggested a potential upstream binding site, similar to the AggR consensus DNA binding site, that might confer CfaD-dependence on promoters. This idea, however, was challenged by mutational

analysis. Mutating the suggested upstream site to match consensus, or knocking it out, did not impact either AggR- or CfaD-mediated activation.

The weblogos in Figure 4.10 suggested further potential targets that might be required to enable CfaD-mediated activation. These targets were mutated with the intention to stop the H149/5 *caf1M* promoter being activated by CfaD. However, data in Figure 4.11 and Figure 4.12 show that creating *afaB-caf1M* hybrid promoters did not impact CfaD-activation and sequences directly upstream of the DNA-binding site for AggR do not directly confer CfaD-mediated activation on promoters. My mutational analysis indicates that more work is needed to pinpoint the factors that contribute to conferring CfaD-activation, and the downstream sequence should be considered. Additionally, some Rns-dependent promoters require a second DNA binding site for Rns further upstream from the binding site at a Class II position (Midgett *et al.*, 2021), and this could be considered when comparing CfaD-activated promoters with CfaD-independent promoters.

Previously, it was reported that members of the subgroup of AraC-family transcription factors that include AggR EAEC, Rns ETEC, CsvR ETEC, CfaR ETEC and VirF *S. flexneri*, are homologous and can be substituted at dependent promoters (Munson and Scott, 1999; Munson *et al.*, 2001). The results in Figure 4.7 confirm that Rns ETEC can be substituted for AggR at AggR-activated promoters, and activity is similar or higher than with AggR. Equally, Rns and AggR can be interchanged with strain H149/5 CfaD at H149/5 promoters. However, as CfaD H149/5 could not activate the 042 *afaB* or the C1010-00 *agg4D* promoter, and only showed weak effects at the bi-directional *yicS-nlpA*

promoters, I conclude that CfaD cannot be substituted for AggR at all AggR-regulated promoters. Note that this does not argue that CfaD H149/5 cannot belong to this subgroup, because, for example, Rns can substitute for VirF, but this is not reciprocal because VirF cannot substitute for Rns (Munson *et al.*, 2001). Figure 4.19 shows alignments of some members of the AraC family AggR subgroup, with the 'signature' stretch of homologous amino acids, that shows the greatest region of sequence identity, highlighted. However, Table 4.2 shows that overall sequence identity varies within the group. CfaD H149/5 has around 30% sequence identity with each of the other factors, showing that there are less similarities to the other subgroup members. This could argue that CfaD H149/5 is not in this subgroup, purely based on the sequence identity, however, VirF from *S. flexneri* also has around 30% sequence identity with the other transcription factors. On this basis, I believe that CfaD H149/5 belongs to the subgroup as it shows similar identity as VirF and CfaD and can be substituted by Rns and AggR, though this relationship is not consistently reciprocated.

Since factors in this subgroup are interchangeable, it has been suggested that these regulators recognise similar DNA binding sites (Munson and Scott, 1999; Munson *et al.*, 2001). Therefore, as I showed in the *afaB-caf1M* hybrid experiments that the upstream sequence was not conferring CfaD-dependence, we can assume that the downstream sequence, potentially including the confirmed AggR-binding site, is important. Thus, further research needs to focus on conferring CfaD-dependence on a CfaD-independent promoter, *afaB* or *agg4D*, by transplanting the downstream promoter sequences from an H149/5 promoter, such as *caf1M*, in order to pinpoint the key determinants.

Figure 4.19 Alignment of AraC subgroup with CfaD H149/5

The figure shows an alignment of the amino acid sequences of the AraC family subgroup: AggR 042 (EAEC strain 042), CfaD H149/5 (EAEC strain H149/5), Rns from ETEC, CsvR ETEC, CfaR ETEC, and VirF *S. flexneri*.

Table 4.2 Sequence identity

Sequences compared	Aligned score
AggR 042: CfaD H149/5	33.58%
AggR 042: Rns ETEC	66.42%
AggR 042: CfaR ETEC	60.38%
AggR 042: CsvR ETEC	61.13%
AggR 042: VirF *S. flexneri*	34.22%
CfaD H149/5: Rns ETEC	33.21%
CfaD H149/5: CfaR ETEC	33.46%
CfaD H149/5: CsvR ETEC	33.83%
CfaD H149/5: VirF *S. flexneri*	28.14%
Rns ETEC: CfaR ETEC	75.47%
Rns ETEC: CsvR ETEC	76.23%
Rns ETEC: VirF *S. flexneri*	33.08%
CfaR ETEC: CsvR ETEC	99.00%
CfaR ETEC: VirF *S. flexneri*	30.80%
CsvR ETEC: VirF *S. flexneri*	31.18%

The factors that contribute to conferring AggR-dependence appear to differ between promoters, though previously it was shown that simply transplanting an AggR-binding site into the *CCmelR* promoter could confer AggR-dependence on that promoter (Yasir *et al.*, 2019). However, the results in Figure 4.16 have shown that the presence of an AggR-binding site in the *bssS* promoter is insufficient for activation by AggR. The promoter alignments in Figure 4.17 show that the *bssS* promoters from 042 and K-12 have identical putative AggR binding sites, however, the identical -10 elements, GATAAA, do not match consensus. These promoters are not dependent on AggR, and this could simply be due to the guanine at the first position. The first base, thymine in the consensus -10 hexamer (T_{-12} A_{-11} T_{-10} A_{-9} A_{-8} T_{-7}), is important for contacting σ^{70} (Hook-Barnard and Hinton, 2007; Feklistov and Darst, 2011; Saecker *et al.*, 2021), and in the *bssS* promoters this is guanine. Therefore, I would conclude that this guanine is likely preventing this promoter from being functional.

Another interesting promoter, *aap*, from EAEC strain 042 has previously been analysed (Ellis *et al.*, 2019; Yasir *et al.*, 2019). The *aap* promoter has now also been identified in H149/5, following whole genome sequencing. This is interesting because *aap* is encoded on pAA2 in 042, but H149/5 lacks a pAA virulence plasmid, and so this promoter had not previously found in H149/5 (Pereira *et al.*, 2008). A one base mutation in the strain 042 *aap* promoter reduced promoter activity overall but unmasked an AggR-activated promoter (Yasir, 2017). Comparison of *aap* regulatory regions in Figure 4.17 indicates that this promoter could have an AggR-binding site located 22 bp upstream of the -10 element, the optimal spacing for the DNA-binding site for AggR, without conferring AggR-

dependence on a promoter, potentially due to a non-optimal -10 hexamer. This suggests that a transplant experiment, as with the *CCmelR* promoter (Yasir *et al.*, 2019), would not necessarily confer AggR-dependence on any promoter, especially as the organisation of the transplanted DNA was similar to *bssS*, the -10 element was also non-consensus (CATAAT), with a base change in the same position as *bssS*. Thus, the factors contributing to AggR-dependence are more complicated than previously expected.

The work detailed in this Chapter highlights that the AraC subgroup of transcription factors have a non-reciprocal relationship to allow an exchange of regulation. CfaD H149/5 belongs to this subgroup due to the homology and substitution relationships. Further investigation is necessary to determine the important promoter elements in H149/5 promoter that confer CfaD-dependence, the key to this will be in the downstream promoter sequences. This offers further insight into AggR-independent promoters, since an AggR-binding site does not necessarily confer dependence. This highlights the importance of a functional -10 element, particularly the first position, thymine (-12), which flips out and sits in the RNAP W-pocket (Feklistov and Darst, 2011; Saecker *et al.*, 2021); the other factors that might contribute to the lack of activation at a suggested AggR-activated promoter include: high basal activity, a neighbouring strong promoter, and non-optimal spacing between the AggR DNA-binding site and the -10 element.

Chapter 5 A Study of Defective AggR Mutants

5.1 Introduction

AggR is a member of the AraC family of transcriptional regulators. Members of this family are typically 250 to 300 residues long and follow similar structural patterns (Gallegos *et al.*, 1997). Many AraC regulators contain two domains: an N-terminal domain, for effector binding and multimerization, and a C-terminal domain, for site-specific DNA binding and interaction with RNAP. For AraC, the N-terminal domain covers residues 1 to 170 and the CTD covers residues 178 to 286 (Gallegos *et al.*, 1997). The NTD, along with the central linker regions, is not present in all AraC-members.

The hallmark of the AraC family members is a region of 99 amino acids that is found at the C-terminal domain in AraC. This domain has high sequence homology of over 20% between family members, therefore, it can be assumed that the tertiary structure is identical. This domain contains two HTH motifs. The first HTH motif (HTH1) is supposed to be the DNA binding motif, conservation of sequence in HTH1 is low and can be highly divergent between unrelated regulators. This variability is thought to be due to the role of HTH1 in recognising specific base sequences in target promoter DNA. The second HTH (HTH2) motif is more conserved, therefore, it is thought to have a common role across the regulators, likely for interacting with holo RNAP (Gallegos *et al.*, 1997; Porter and Dorman, 2002).

Thus far, my data has focused on how AggR regulates at its target promoters. Now, I would like to shift the focus to how the structure of AggR is important for its function, specifically, whether it is possible to disrupt AggR-mediated activation and block AggR-activated promoters by mutating the *aggR* gene. By co-expressing mutant AggR with wild-type AggR to block activation at target promoters, it could be possible to block AggR-activation of virulence factor gene transcription. Therefore, the structural information, previously gained from research of AraC-family transcriptional factors, was utilised to develop a strategy for generating different AggR mutants.

5.2 The signal for AggR-mediated activation

It was previously found that AggR functions as a monomer at a single DNA binding site during AggR-dependent activation at target promoters (Yasir *et al.*, 2019), but it has not been determined whether expression of AggR requires an external signal. Therefore, to investigate whether there is a signal that activates expression of AggR, I designed a system with AggR carried on the pLG339 vector, with the expression of the *aggR* gene under the control of its own promoter (Figure 3.8B).

The pLG339 empty vector, pLG339*aggR* 042*,* and arabinose-inducible pBAD*aggR* recombinant plasmids (Figure 3.5) were transformed separately into *E. coli* K-12 BW25113 Δ*lac* cells that carried either pRW50*afaB* or pRW50*agg4D*. Cultures were grown with shaking to mid-exponential phase, in the presence or absence of arabinose. The β-galactosidase levels in lysates of cells containing the pRW50*afaB* and pRW50*agg4D* plasmids were measured to determine promoter activities. The results in Figure 5.1 show that cells containing

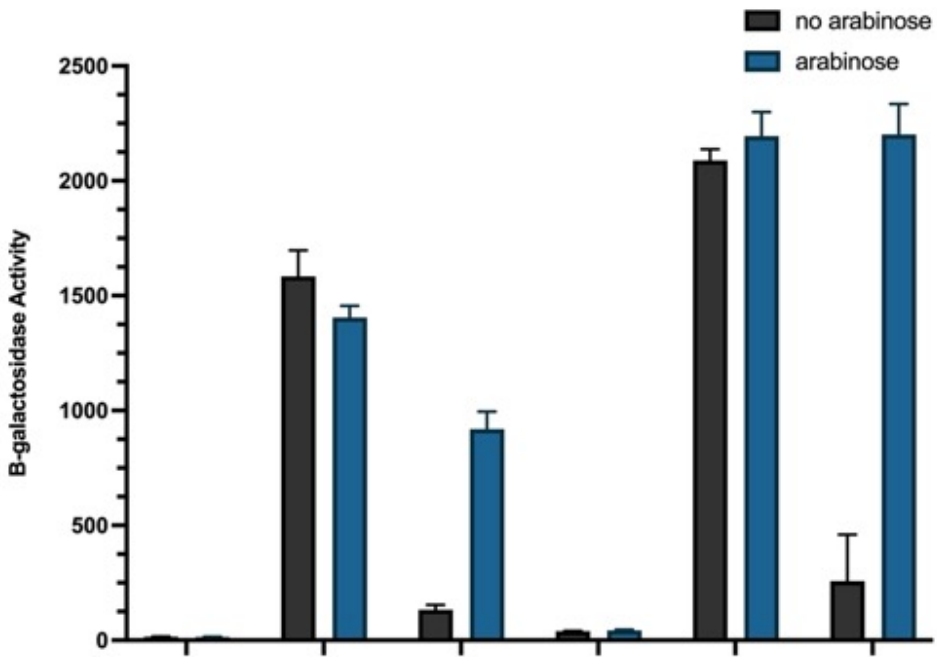
pRW50*afaB* and pLG339*aggR* 042 in the presence of arabinose had higher β-galactosidase levels compared to cells containing pBAD*aggR* (1405 units and 919 units, respectively). The β-galactosidase levels were similar for pRW50*agg4D* with either pLG339*aggR* 042 or pBAD*aggR*: 2194 and 2202 units, respectively. In fact, the cells with pRW50*afaB* or pRW50*agg4D* and pLG339*aggR* 042 also had high β-galactosidase levels in the absence of arabinose (1584 and 2088 units, respectively; one-way ANOVA, $p<0.0001$; Tukey test, $p<0.05$). This indicates that an external signal is not required to induce expression of AggR when it is expressed under the control of its own promoter.

5.3 A system for searching for *trans*-dominance

Since I had determined that AggR could be expressed without an external signal, I decided to set-up a system in which AggR was expressed both from pLG339*aggR* 042 and from arabinose-inducible pBAD*aggR*. Figure 5.2 shows a schematic of this new system with two plasmids carrying the *aggR* gene: pLG339*aggR* 042 and pBAD*aggR*. The expression of AggR, from pBAD*aggR*, is under the control of the *paraBAD* promoter, which is induced by arabinose. The recombinant plasmid, pLG339*aggR 042*, carries the *aggR* gene under the control of its own promoter. AggR is expressed from the recombinant plasmids and binds to the target promoter on the pRW50 plasmid, this activates the production of β-galactosidase.

5.3.1 Changes in the AggR protein

The new system then allowed me to determine how expression at promoters would be affected by co-expression of wild-type AggR and an AggR mutant. The



promoter in pRW50	*afaB*	*afaB*	*afaB*	*agg4D*	*agg4D*	*agg4D*
	pLG339	pLG339 AggR 042*	pBAD *aggR*	pLG339	pLG339 AggR 042*	pBAD *aggR*

Figure 5.1 Analysis of promoter activity when AggR expression is controlled by the *aggR* promoter

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB* or pRW50*agg4D*, in addition to either pLG339 (empty vector), pLG339*aggR* 042 or pBAD*aggR* 042. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 24) = 333.4$). A post-hoc Tukey's HSD test showed that the promoter activities were significantly different between cells carrying pLG339 empty vector or pLG339*aggR* 042 ($p<0.05$).
* Expression of the *aggR* gene is under the control of the aggR promoter

AggR amino acid sequence and predicted secondary structure is shown in Figure 5.3, I have highlighted the residues that were changed. Initially, I introduced mutations into the *aggR* gene through random mutagenesis, then I utilised site-directed mutagenesis to target specific codons. Mutations were present throughout the *aggR* gene: I introduced many mutations in the C-terminal domain, focusing on the two HTH motifs, the presence of which is conserved in AraC family members; and I have also introduced mutations in the non-conserved N-terminal domain and the region between the domains.

Firstly, I revisited AggR mutants I had isolated through random mutagenesis, during a previous study. Briefly, I utilised error-prone PCR to randomly introduce mutations in the *aggR* gene. Then, I selected for mutants by transforming into BW25113 cells carrying pRW50 containing an AggR-dependent promoter and I plated on MacConkey agar with appropriate antibiotic. AggR mutants defective for activation at target promoters were identified by their phenotype, a white colony, on plates. Following sequencing of the mutant *aggR* fragments, I determined the locations of mutations in the *aggR* gene. The resulting mutants are Y92C and Y92R, carrying changes in the same N-terminal domain residue; D86N and L221S, an AggR mutant with two-point mutations, one in the N-terminal and one in the C-terminal domain; and Q230R and K263STOP, an AggR mutant with a point mutation in the first α-helix of the second HTH and a 3-residue truncation at the end of the sequence.

To determine whether the randomly isolated AggR mutants are able to activate

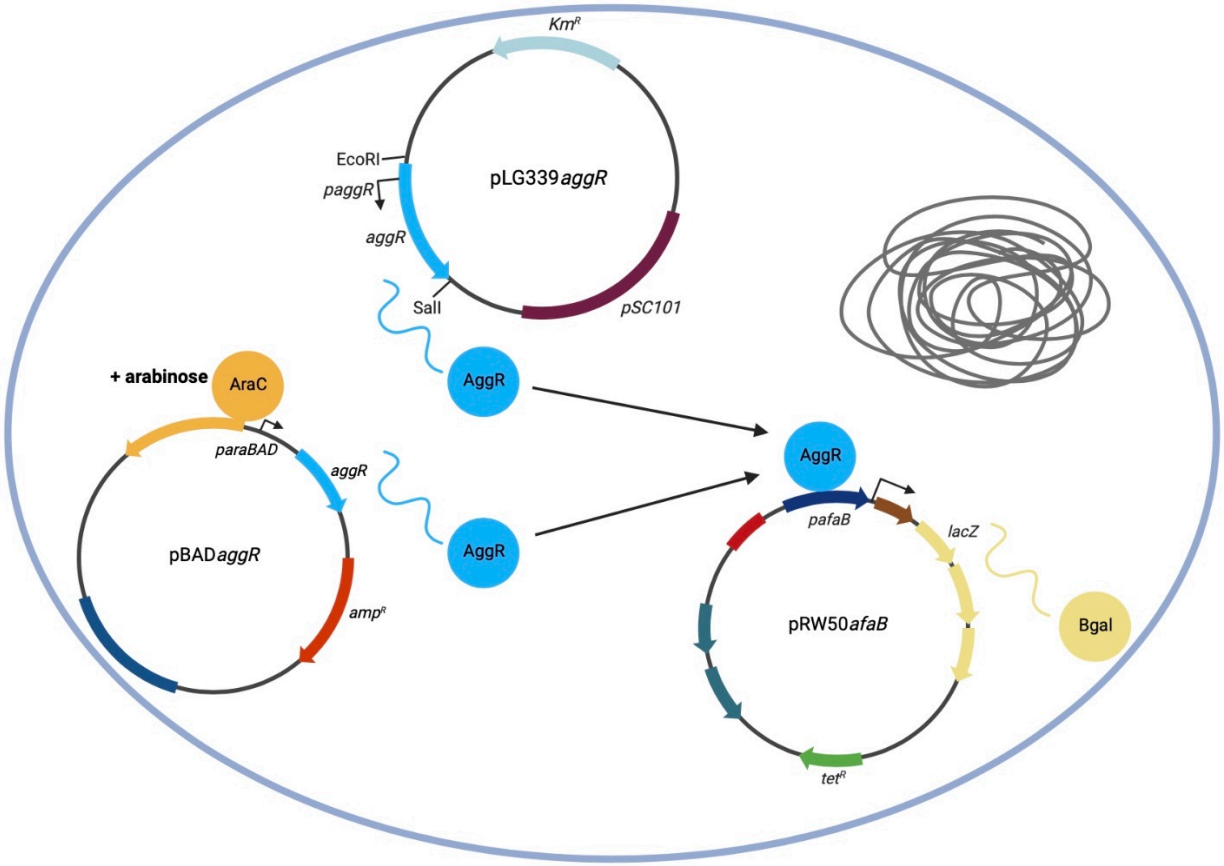


Figure 5.2 A new system to measure promoter activity

The figure illustrates the AggR-mediated activation of the target promoter on the pRW50 vector. For pLG339*aggR* 042: expression of the *aggR* gene is under the control of the *aggR* promoter. For induction of the pBAD*aggR*: arabinose binds to the *araC* gene, which activates the *paraBAD* promoter and subsequent expression of the *aggR* gene. AggR then binds to the target promoter on the pRW50 vector, activating the expression of *lacZ*, producing β-galactosidase. Created with BioRender.com.

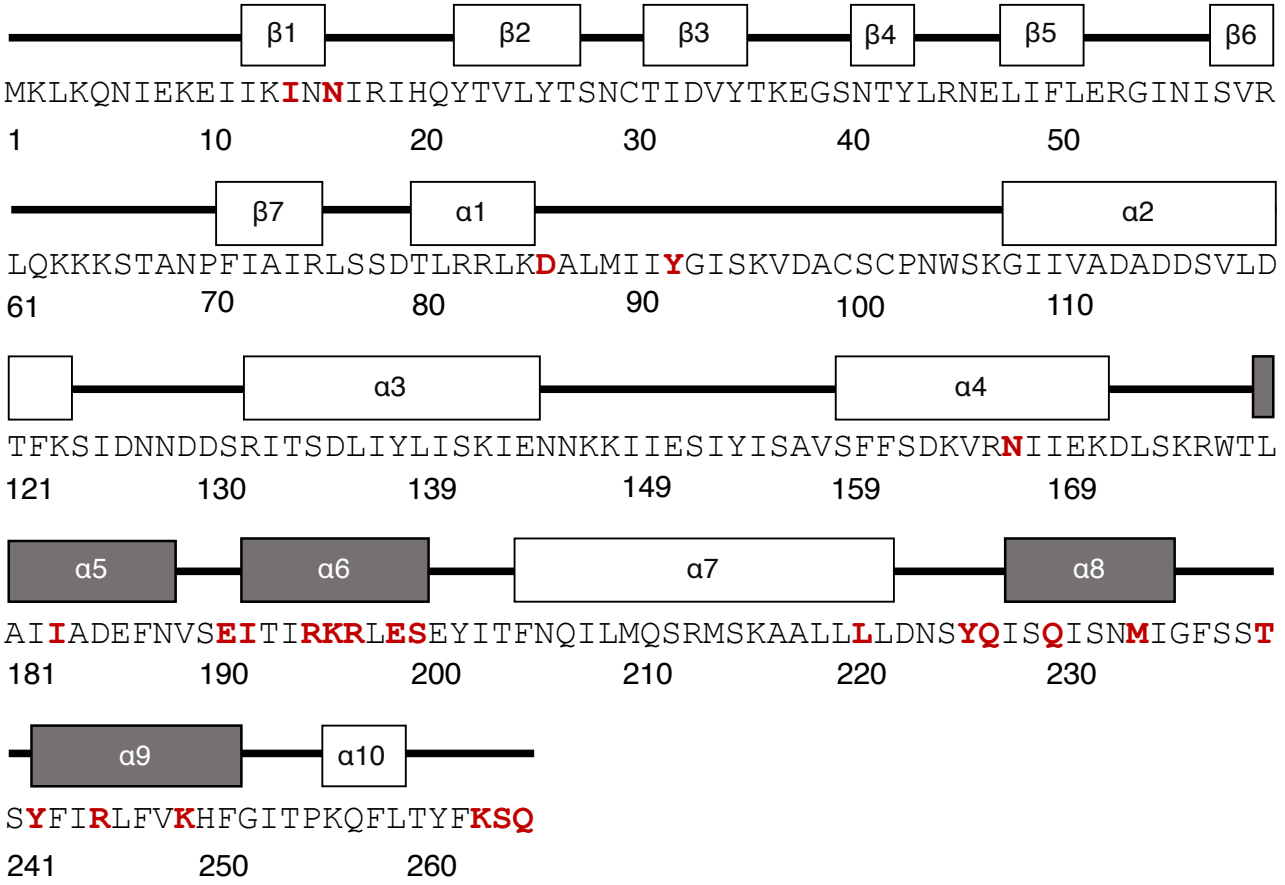


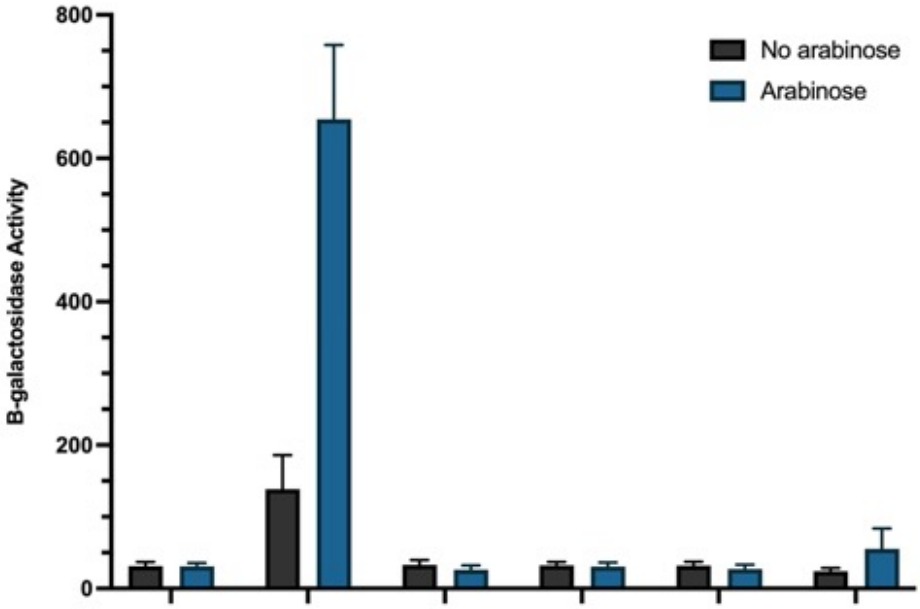
Figure 5.3 Amino acid structure of AggR

The figure shows the amino acid sequence of AggR 042 (EAEC strain 042). The predicted secondary structure located above the amino acid sequence is modelled on Rns (ETEC) with the α helices and β sheets labelled. For the tertiary structure, the two HTH motifs that are part of the AraC homologous region are highlighted in grey. The amino acids highlighted in red indicate the residues targeted for mutational analysis. Adapted from Mahon *et al.* (2010).

at an AggR-dependent promoter, I transformed pRW50*afaB* into *E. coli* K-12 BW25113 Δ*lac* cells carrying either pBAD30 empty vector, pBAD*aggR*, pBAD*aggR-Y92C*, pBAD*aggR-Y92R*, pBAD*aggR-D86N-L221S* or pBAD*aggR-Q230R-K263STOP*. Cells were grown with shaking to mid-exponential phase in the presence and absence of arabinose and the β-galactosidase levels were measured to indicate promoter activity. The results in Figure 5.4 show that β-galactosidase levels were comparable to basal levels with empty vector for each of the AggR mutants.

As I had shown that the randomly isolated AggR mutants were defective for activation at target promoters, I decided to introduce the mutants in *trans* to wild-type AggR (pLG339*aggR* 042).

The recombinant plasmid pRW50*agg4D* was transformed into *E. coli* K-12 BW25113 Δ*lac* cells carrying either pLG339 empty vector or pLG339*aggR 042*, and the arabinose-inducible pBAD30, pBAD*aggR*, pBAD*aggR-Y92C*, pBAD*aggR-Y92R*, pBAD*aggR-D86N-L221S* or pBAD*aggR-Q230R-K263STOP*. Cells were grown in the presence or absence of arabinose with shaking till mid-exponential phase. The β-galactosidase levels were measured to determine promoter activity. The β-galactosidase levels in cells containing pLG339 empty vector were as low as basal levels in the presence of AggR-Y92C and AggR-Y92R (29 and 24 units, respectively, compared to empty vector, 38 units) (Figure 5.5A) or AggR-D86N-L221S and AggR-Q230R-K263STOP (18 and 42 units, respectively, compared to empty vector, 38 units) (Figure 5.5B). The reductions



promoter in pRW50	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*
activator in pBAD	–	AggR 042	Y92C	Y92R	D86N L221S	Q230R K263 STOP

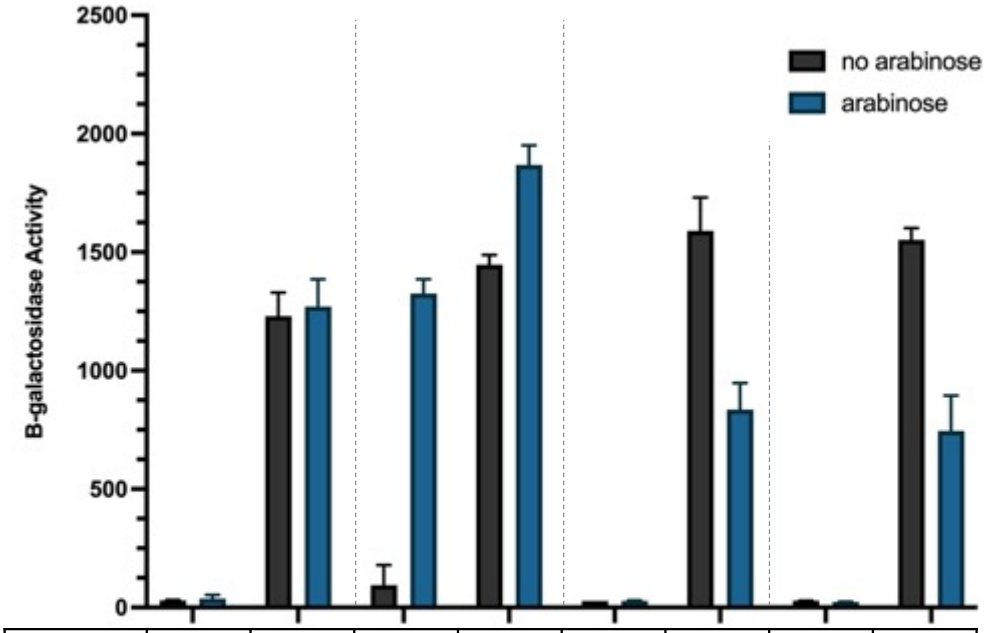
Figure 5.4 Measurement of promoter activity with AggR mutants

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* containing pRW50*afaB* 042 carrying either pBAD30 (empty vector), pBAD*aggR* 042, pBAD*aggR-Y92C*, pBAD*aggR-Y92R*, pBAD*aggRD86N-L221S* or pBAD*aggR-Q230R-K263STOP*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (w/v) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

in β-galactosidase levels due to AggR-Y92C or AggR-Y92R were similar in cells containing pLG339*aggR* 042: 1.9 fold in the presence of AggR-Y92C, from 1590 units to 835; and 2-fold in the presence of AggR-Y92R, from 1552 units to 746 (Figure 5.5A; one-way ANOVA, $p<0.0001$; Tukey test, $p<0.05$). The β-galactosidase levels in cells containing pLG339*aggR* 042 in the presence of AggR-D86N-L221S (636 units) were lower than in the presence of AggR-Q230R-K263STOP (1173 units), the reduction compared to cells in the absence of arabinose was greatest with AggR-D86N-L221S (decrease of 2-fold compared to 1.3-fold) (Figure 5.5B). Most notably, the results in Figure 5.5A and Figure 5.5B show that the β-galactosidase levels were increased when AggR expression was induced, by the addition of arabinose to the medium, in cells containing pLG339*aggR* 042.

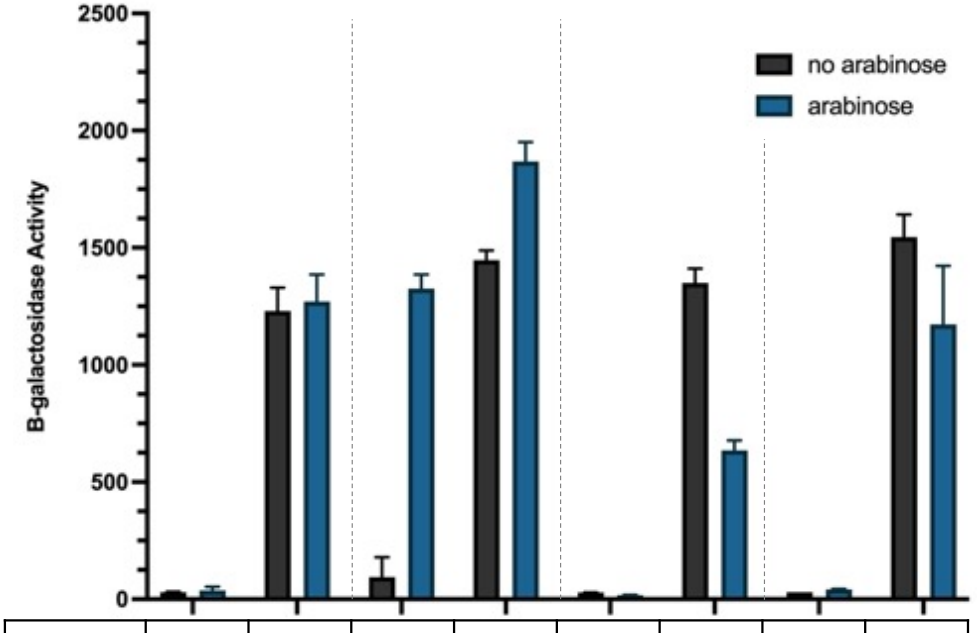
Next, I transformed pBAD30, pBAD*aggR*, pBAD*aggR-Y92C*, pBAD*aggR-Y92R*, pBAD*aggR-D86N-L221S* and pBAD*aggR-Q230R-K263STOP*, into *E. coli* K-12 BW25113 cells carrying pRW50*afaB* with either pLG339 empty vector or pLG339*aggR* 042. Again, cells were grown in the presence and absence of 0.2% (*w/v*) arabinose to mid-exponential phase, and β-galactosidase levels were measured from cell lysates to indicate promoter activity. The results in Figure 5.6A show that the cells containing pRW50*afaB* and pLG339 empty vector in the presence of pBAD*aggR-Y92C* (12 units) or pBAD*aggR-Y92R* (10 units) had very low β-galactosidase levels compared to cells with pBAD*aggR*, which showed increased β-galactosidase levels when AggR expression was induced (368 units). However, expression levels were affected when pLG339*aggR* 042 was co-expressed and β-galactosidase levels were high in the absence of arabinose.

A



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	Y92C	Y92C	Y92R	Y92R

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	D86N L221S	D86N L221S	Q230R K263 STOP	Q230R K263 STOP

Figure 5.5 Analysis of *agg4D* promoter activity with previously identified AggR mutants

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmids pRW50*agg4D*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*agg4D* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-Y92C* or pBAD*aggR-Y92R.* A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 32) = 240.3$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pLG339*aggR* 042 and carrying either pBAD*aggR-Y92C* or pBAD*aggR-Y92R* in the presence and absence of arabinose ($p<0.05$).
B. Cells containing pRW50*agg4D* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-D86N-L221S* or pBAD*aggR-Q230R-K263STOP*.

Cells containing pLG339*aggR* 042 had a 1.7-fold decrease in β-galactosidase activity when expression of AggR, from pBAD*aggR*, was induced by the addition of arabinose (874 units in the absence of arabinose and 510 units in presence). This slight reduction of β-galactosidase activity was also seen in cells containing pBAD*aggR-Y92C* or pBAD*aggR-Y92R*, a 1.4-fold (516 units compared to 363 units) and 1.8-fold decrease (492 units compared to 279 units), respectively, however, this was to a lesser extent than with cells containing wild-type pBAD*aggR*. The reduction in β-galactosidase levels measured was greatest in cells containing pLG339*aggR* 042 in the presence of AggR, followed by AggR-Y92R or AggR-Y92C respectively, however, the fold-decrease was less than seen with pRW50*agg4D* (Figure 5.5A).

The data in Figure 5.6B show that β-galactosidase levels were low in the presence and absence of AggR-D86N-L221S (10 and 13 units, respectively) or AggR-Q230R-K263STOP (29 and 17 units, respectively) in cells containing pLG339 empty vector. The reductions in β-galactosidase levels seen in Figure 5.6A, in cells containing pLG339*aggR* 042, were also seen in Figure 5.6B. The greatest reduction, 1.6-fold, was seen with pLG339*aggR* 042 in the presence of AggR-D86N-L221S, 301 units compared to 482 units in the absence of arabinose. AggR-Q230R-K263STOP had a slightly lower reduction of 1.5-fold, from 557 units in the absence of arabinose to 377 units when expression of AggR-Q230R-K263STOP was induced. The data show that the AggR mutants fail to activate the *afaB* promoter in cells containing pLG339 empty vector. In the presence of pLG339*aggR*, expression levels were decreased slightly due to expression of mutant AggR (one-way ANOVA, $p<0.0001$; Tukey test, $p>0.05$).

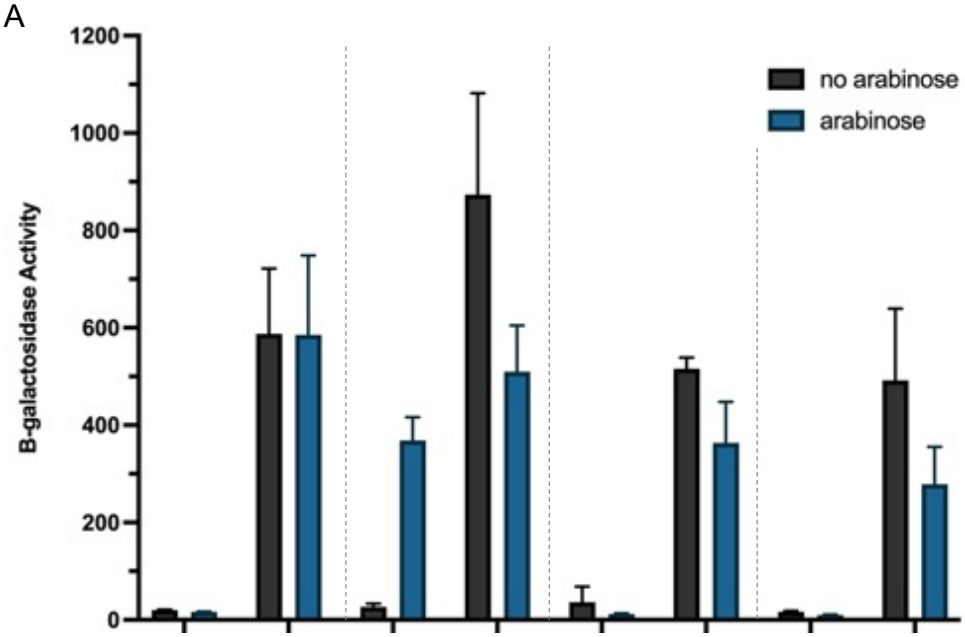
Most remarkably, cells containing pLG339*aggR* 042 and pBAD*aggR* in the presence of arabinose had increased β-galactosidase levels with pRW50*agg4D* (1.3-fold increase, 1448 units to 1868) and reduced β-galactosidase levels with pRW50*afaB* (1.7-fold decrease, 874 units to 510). This result with pRW50*afaB* is unexpected because producing extra wild-type AggR from pBAD*aggR* leads to a reduction in *afaB* promoter activity.

The results indicate that the AggR mutants were defective in activation of the *afaB* and *agg4D* promoters: AggR-Y92C and AggR-Y92R, N-terminal domain; AggR-D86N-L221S, N-terminal domain and C-terminal domain linker; and AggR-Q230R-K263STOP, first α-helix of HTH and 3 residue truncation. The data show that there was a reduction in β-galactosidase activity in cells where AggR from pLG339*aggR* 042 was expressed concurrently with the AggR-mutants from pBAD*aggR*. Though none of the mutants were strongly *trans*-dominant, the results suggest there is interference from the AggR mutants with wild-type AggR-dependent activation at the *afaB* and *agg4D* promoters.

5.3.2 Changes in AggR guided by previously isolated AggR mutants

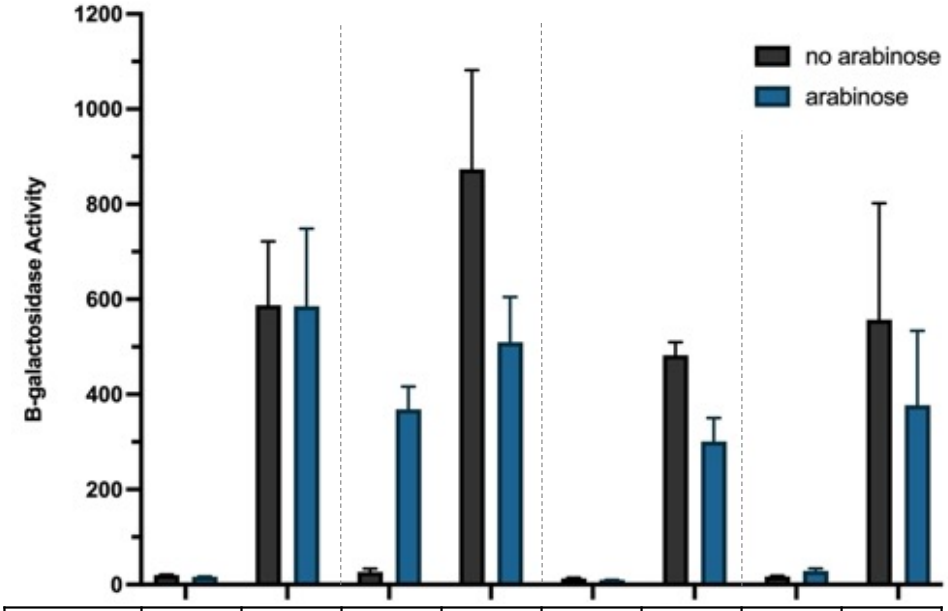
Changing amino acid residues in AggR can have implications for protein folding and resulting conformational changes can have disastrous effects on DNA binding and interactions with RNAP, and subsequent activation at promoters. In Figure 5.7, the PyMOL molecular graphic of the predicted structure of AggR is shown, the pink and purple highlighted α-helices represent the first and second HTH motifs. The location of three of the AggR changes discussed in section 5.3.1, Y92C, Y92R and Q230R, are indicated by arrows. I decided to investigate some

A



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	Y92C	Y92C	Y92R	Y92R

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	D86N L221S	D86N L221S	Q230R K263 STOP	Q230R K263 STOP

Figure 5.6 Analysis of *afaB* promoter activity with previously identified AggR mutants

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmids pRW50*afaB*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-Y92C* or pBAD*aggR-Y92R.* A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 32) = 27.85$). A post-hoc Tukey's HSD test showed that there was no significant difference between promoter activities in cells carrying pLG339*aggR* 042 and containing either pBAD*aggR-Y92C* or pBAD*aggR-Y92R* in the presence and absence of arabinose ($p>0.05$).
B. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-D86N-L221S* or pBAD*aggR-Q230R-K263STOP*.

of these residues further, by substituting alanine or serine at position 92, and reconstructing the Q230R mutant without any C-terminal deletion. I also examined the effects of the N168D substitution, located in the region between the N- and C-terminal domains, that I had previously obtained from error-prone PCR.

Mutant *aggR* genes were amplified using Megaprimer PCR, utilising a forward primer containing the mutation and the pRW50 reverse primer, introducing a HindIII restriction site, for the first PCR and the second PCR used the purified product from the first PCR and the pRW50 forward primer, introducing an EcoRI restriction site. The mutated *aggR* gene fragments were ligated into pBAD30.

The pBAD*aggR* mutant recombinant plasmids were transformed into *E. coli* BW25113 cells containing pRW50*afaB*; pBAD*aggR* and pBAD30 empty vector were also transformed into BW25113 cells as positive and negative controls, respectively. Cells were grown in the presence or absence of 0.2% (*w/v*) arabinose and promoter activity was determined by measuring the β-galactosidase activity in lysed cells. The results in Figure 5.8A show that cells containing pRW50*afaB* in the presence of AggR, AggR-N168D or AggR-Q230R had high β-galactosidase levels, in fact, the expression levels were higher in the presence of AggR-Q230R (866 units) than AggR (560 units) or AggR-N168D (586 units). The β-galactosidase levels when expression of AggR-Y92A and AggR-Y92S was induced by addition of arabinose to the medium were low, 26 units and 39 units, respectively, and comparable to the basal levels, 29 units.

Next, I transformed pBAD30, pBAD*aggR*, pBAD*aggR-N168D*, pBAD*aggR-Q230R*, pBAD*aggR-Y92A* and pBAD*aggR-Y92S* into *E. coli* BW25113 Δ*lac* cells

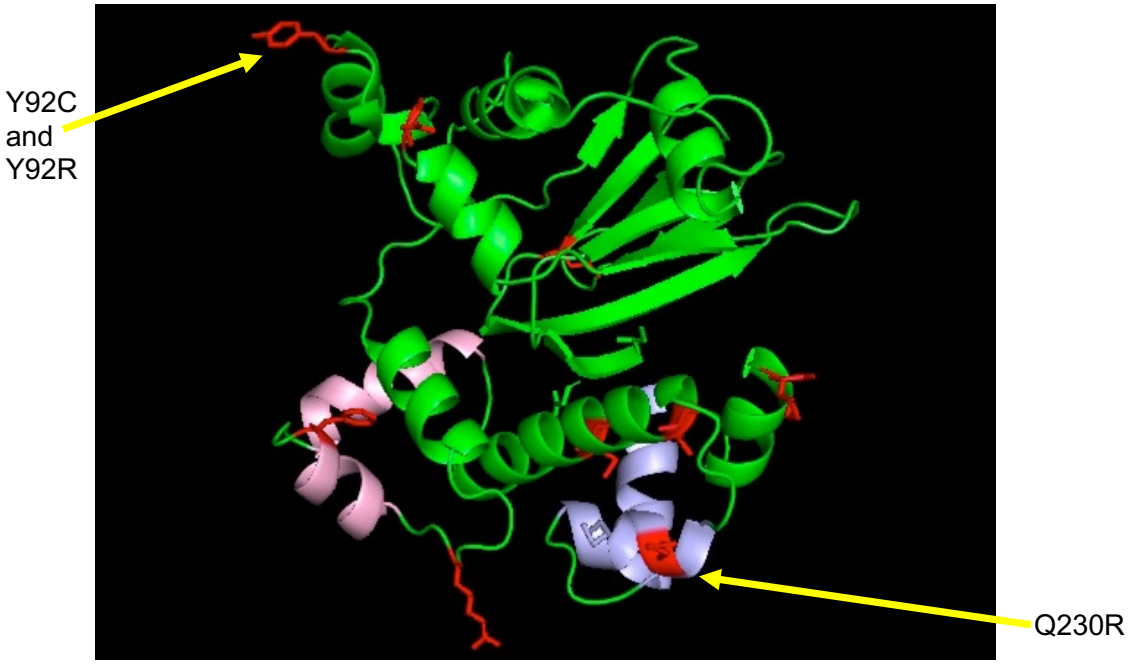


Figure 5.7 The predicted structure of AggR

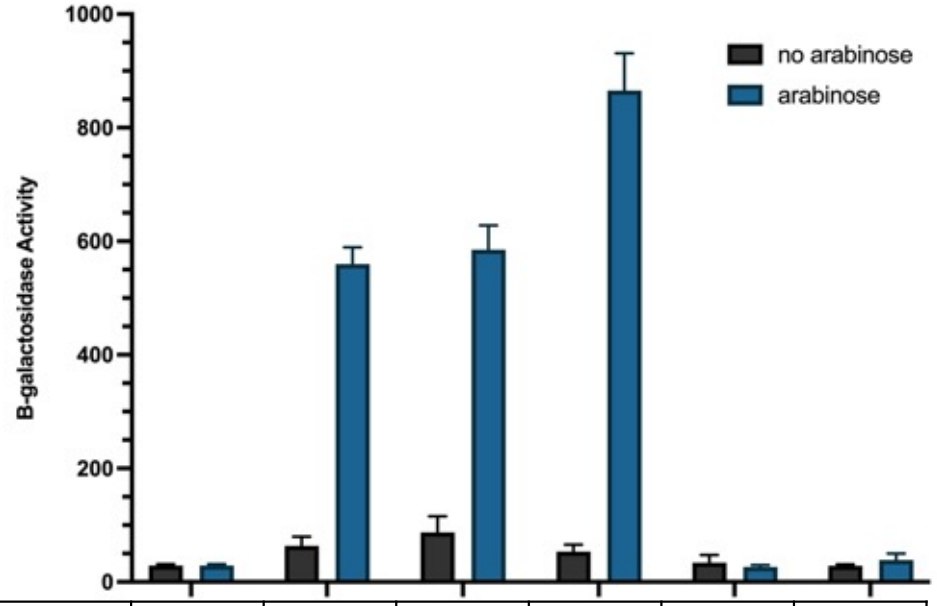
A model showing the proposed structure of the AggR protein. The homology model was generated from the crystal structure of ToxT (*Vibrio cholerae*). HTH1 is highlighted in pink and HTH2 is blue. The red amino acids indicate residues that have previously been targeted for mutagenesis, the yellow arrows show the location of the N-terminal mutation at residue 92 and in the HTH2 at position 230. The figure was constructed using PyMol (the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

containing pRW50*nlpA*. Assays were completed with the *nlpA* promoter to confirm that mutations introduced into AggR did not affect DNA binding at target promoters. Cultures were grown to mid-exponential stage with each culture grown in duplicate in the presence and absence of 0.2% (*w/v*) arabinose. The β-galactosidase levels were measured in lysed cells to indicate promoter activities. The results in Figure 5.8B show that the β-galactosidase levels were decreased when expression of each of the AggR constructs was induced by arabinose, including wild-type and the mutants, though cells with wild-type AggR showed the greatest reduction (2-fold, 78 units compared to empty vector, 153 units). The AggR mutants had the following reductions compared to empty vector: AggR-N168D, 1.7-fold and 89 units; AggR-Q230R, 1.7-fold and 90 units; AggR-Y92A, 1.3-fold and 114 units; and AggR-Y92S, 1.4-fold and 112 units. Interestingly, β-galactosidase levels also decreased in the presence of arabinose in cells containing pBAD30 empty vector. The results indicate that each of these AggR mutants can repress the *nlpA* promoter. Thus, although the AggR Y92A and Y92C mutants are defective for activating target promoters (Figure 5.8A), they are able to bind to promoter DNA.

Since cells containing AggR-Q230R showed unexpectedly high β-galactosidase levels (Figure 5.8A), compared to the low β-galactosidase levels with AggR-Q230R-K263STOP (Figure 5.5B and Figure 5.6B), I decided to introduce further changes at position 230, using Megaprimer PCR. Note that position 230 corresponds to an important residue in MelR that interacts with the RNAP σ^{70} subunit, see section 5.4.3 (Grainger *et al.*, 2004). Hence, glutamine 230 in 042 AggR was mutated to either alanine, leucine, or methionine, and the resultant

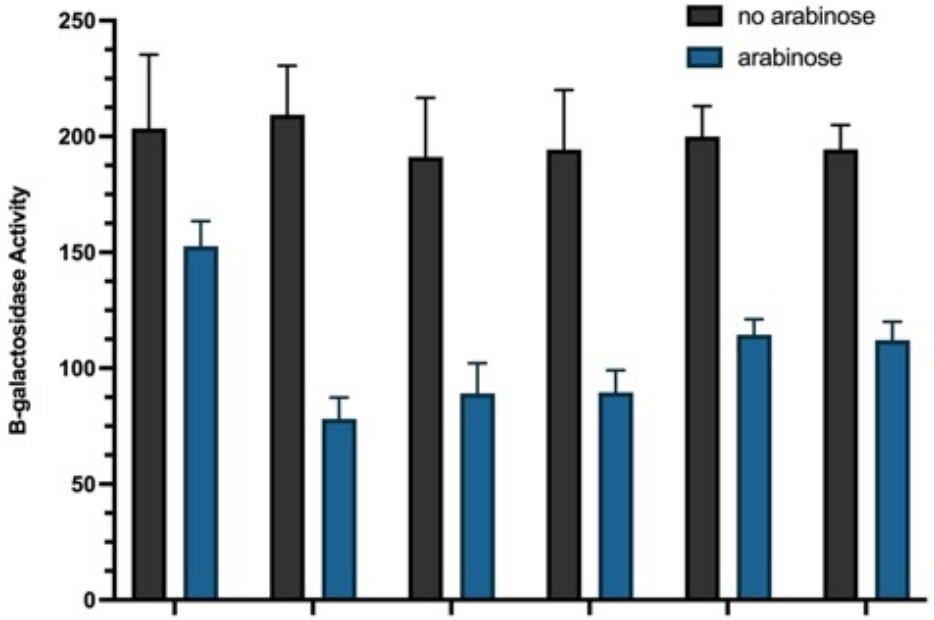
pBAD*aggR* mutant recombinant plasmids were transformed into *E. coli* BW25113 Δ*lac* cells containing pRW50*afaB*. Cells were grown in duplicate in the presence and absence of 0.2% (*w/v*) arabinose until the OD_{650} reached between 0.4 and 0.6. β-galactosidase levels were measured to indicate promoter activity. The results in Figure 5.8C show that there was an increase in β-galactosidase activity when expression of AggR-Q230A (17-fold and 368 units), AggR-Q230L (13-fold and 270 units) and AggR-Q230M (17-fold and 361 units) was induced, compared to empty vector control (21 units). The data suggest that mutating glutamine at position 230 does not stop activation at AggR-dependent promoters. Hence the results with the AggR-Q230R-K263STOP mutant, shown in Figure 5.5B and Figure 5.6B, highlight the importance of the C-terminal 3-residue truncation of AggR. For this reason, I introduced this 3-residue truncation into AggR, carrying glutamine at position 230. Megaprimer PCR was utilised to introduce the 3-residue truncation into *aggR*, using primers with an EcoRI site upstream and an XbaI site downstream. Recombinant plasmids pBAD*aggR-Q230R* and pBAD*aggR-K263STOP* were transformed into *E. coli* K-12 BW25113 Δ*lac* carrying pRW50*afaB* and either pLG339 empty vector or pLG339*aggR* 042. Cells were grown in the presence and absence of 0.2% (*w/v*) arabinose and β-galactosidase levels were measured in cell lysates to indicate promoter activity. Cells containing pLG339 empty vector had increased β-galactosidase levels when expression of pBAD*aggR-Q230R* (91-fold, 1145 units) and pBAD*aggR-K263STOP* (25-fold, 317 units) was induced, compared to pBAD30 empty vector (13 units) (Figure 5.9A). This shows that AggR-K263STOP is seriously defective for activation function. Cells containing pLG339*aggR* 042 with the expression of

A



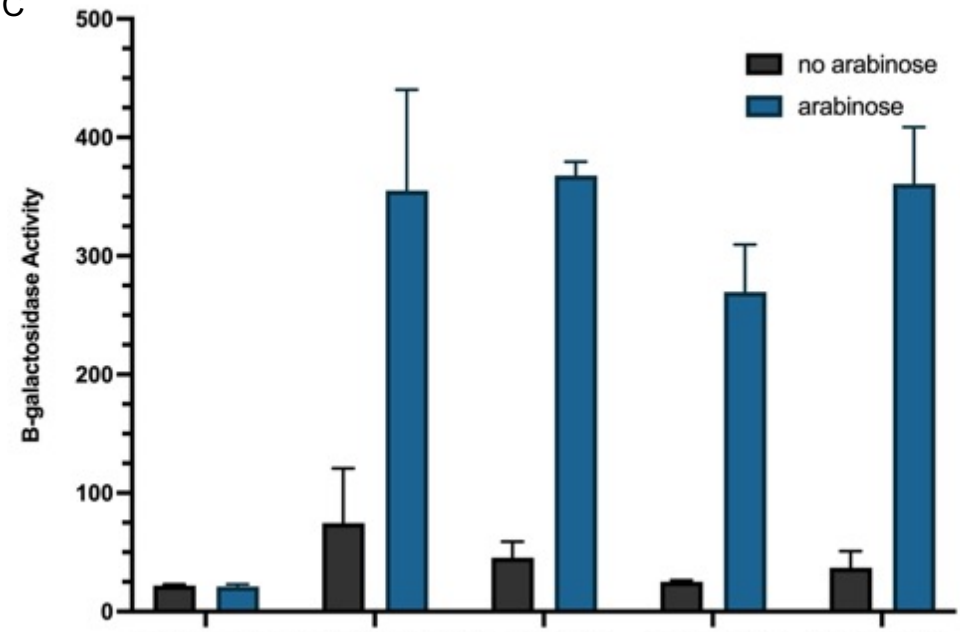
promoter in pRW50	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*
activator in pBAD	–	AggR 042	N168D	Q230R	Y92A	Y92S

B



promoter in pRW50	*nlpA*	*nlpA*	*nlpA*	*nlpA*	*nlpA*	*nlpA*
activator in pBAD	–	AggR 042	N168D	Q230R	Y92A	Y92S

C



promoter in pRW50	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*
activator in pBAD	–	AggR 042	Q230A	Q230L	Q230M

Figure 5.8 Analysis of promoter activity with AggR mutants

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50 carrying AggR-activated promoters in addition to either pBAD30 (empty vector), pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

A. Cells containing pRW50*afaB* with pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-N168D*, pBAD*aggR-Q230R*, pBAD*aggR-Y92A* or pBAD*aggR-Y92S.*
B. Cells containing pRW50*nlpA* with pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-N168D*, pBAD*aggR-Q230R*, pBAD*aggR-Y92A* or pBAD*aggR-Y92S*.
C. Cells containing pRW50*afaB* with pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-Q230A*, pBAD*aggR-Q230R* or pBAD*aggR-Q230M*.

the pBAD*aggR* induced had minimal reduction in β-galactosidase activity compared to cells in the absence of arabinose. For pBAD*aggR-Q230R*, there was a 1.2-fold decrease, with 1937 units in the absence of arabinose and 1604 units in its presence. The reduction was lower for pBAD*aggR-K263STOP*, just over 1-fold, with 1465 units in the absence of arabinose compared to 1334 units. The data indicate that separating the Q230R and K263STOP mutations negatively impacts the level of interference compared to AggR-Q230R-K263STOP (1.5-fold reduction), which shows that these mutants are not *trans*-dominant. However, the results also show that separating the mutations actually restored activation function compared to the combined mutant, this suggests that the combined mutant is defective in DNA binding, or there was a detrimental conformational change.

During constructions, another AggR-mutant was obtained by chance, AggR-QIS230-232KIP, with a three-residue substitution at positions 230-232 from QIS to KIP. pBAD*aggR-QIS230-232KIP* was transformed into *E. coli* K-12 BW25113 containing pRW50*afaB* and either pLG339 empty vector or pLG339*aggR* 042. Cells were grown in the presence and absence of 0.2% (*w/v*) arabinose and the β-galactosidase levels were measured to show promoter activity. The results in Figure 5.9B show that the β-galactosidase levels are low in the presence and absence of arabinose in cells containing pLG339 empty vector and pBAD*aggR-QIS230-232KIP*, 121 and 50 units, respectively. However, cells containing pLG339*aggR* 042 and pBAD*aggR-QIS230-232KIP* showed just over 1-fold difference in the presence and absence of arabinose, 1539 and 1654 units, respectively. The results indicate that the QIS230-232KIP AggR mutant

effectively blocked activation, however, when AggR and the QIS230-232KIP mutant were grown in *trans*, there was no reduction in activity. This suggests that the QIS230-232KIP mutant is severely defective for activation but is not *trans*-dominant.

Data show that residue Y92 is important for AggR-medicated activation, similarly, changing positions 230-232 prevents AggR-activation. While truncating the final three residues greatly reduced expression levels, activation still occurred at a lower level. Interestingly, Q230, which has been shown to be important in MelR (Grainger *et al.*, 2004), has been shown to not be vital for AggR activation at target promoters.

In summary, in this section I have isolated AggR mutants that are severely defective for activation at target promoters, however, none of the changes in AggR resulted in strong *trans*-dominance.

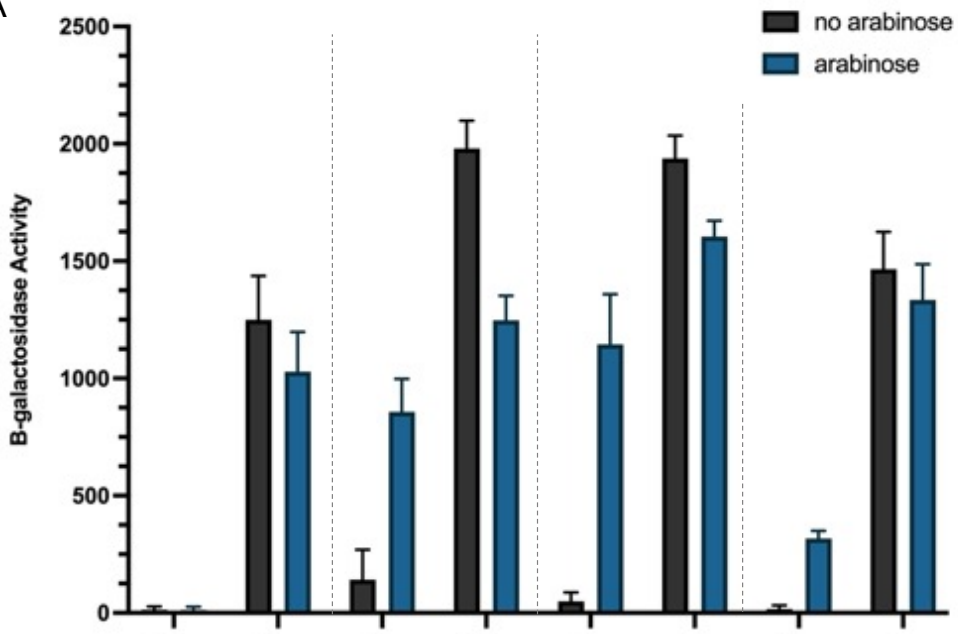
5.4 Mutations guided by other AraC family transcription factors

The changes described above that I introduced in AggR did not have a strong *trans*-dominant effect when co-expressed with wild-type AggR. Thus, I decided to investigate mutations that had been identified in other AraC members as defective for activation at target promoters.

5.4.1 Mutations according to Rns

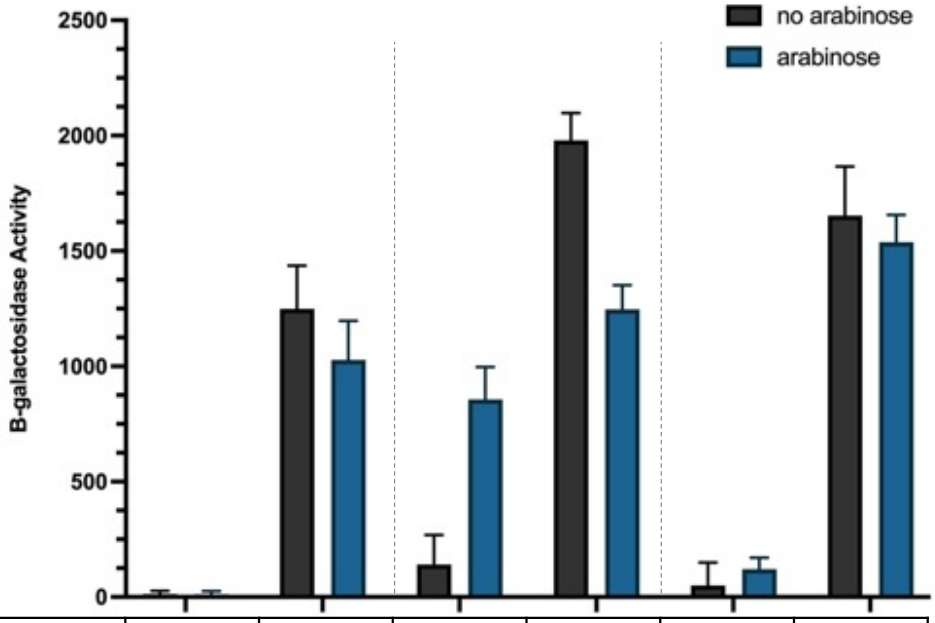
The first mutations looked at here were from Rns ETEC. As discussed in Section 4.2.1, Rns has sequence homology to AggR and can be used interchangeably at

A



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	Q230R	Q230R	K263 STOP	K263 STOP

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	QIS 230-232 KIP	QIS 230-232 KIP

Figure 5.9 Analysis of *afaB* promoter activity with co-expression of AggR mutants targeting HTH2

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmids pRW50*afaB*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-Q230R* or pBAD*aggR-K263STOP.*
B. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR* or pBAD*aggR-QIS230-232KIP*.

AggR-dependent promoters, due to recognition of similar DNA sequences. Therefore, I decided to utilise my pLG339 assay system to determine whether mutations identified in Rns would effectively block activation and act in a *trans*-dominant manner. Two Rns mutants, I14T and N16D, had previously been shown to be defective for activation (Basturea *et al.*, 2008), these are indicated in Figure 5.10, the residues at position 14 and 16 are the same in Rns and AggR.

The corresponding changes were introduced into AggR, utilising primers with an EcoRI site upstream and XbaI site downstream. The *pBADaggR-I14T* and pBAD*aggR-N16D* recombinant plasmids were transformed into *E. coli* K-12 BW25113 Δ*lac* containing pRW50*afaB* and either pLG339 empty vector or pLG339*aggR* 042. Cells were grown in the presence and absence of 0.2% (*w/v*) arabinose and β-galactosidase levels in cell lysates were measured to determine promoter activity. The results in Figure 5.11A show that the β-galactosidase levels were low in cells containing pLG339 and pBAD*aggR-I14T* or pBAD*aggR-N16D* both in the presence (15 and 25 units, respectively) and absence of arabinose (15 and 16 units, respectively), compared to empty vector (14 and 15 units, respectively).

Cells containing pLG339*aggR* 042 and pBAD*aggR-I14T* in the absence of arabinose had 1200 β-galactosidase units, this was reduced 2-fold in the presence of arabinose (551 β-galactosidase units; one-way ANOVA, $p<0.0001$; Tukey test, $p<0.05$). For pBAD*aggR-N16D*, there was just under a 2-fold difference, from 1227 β-galactosidase units in the absence of arabinose to 618 units in the presence of arabinose (Tukey test, $p<0.05$). Interestingly, cells containing pLG339*aggR* 042 and pBAD*aggR* also had over a 2-fold difference,

from 1304 β-galactosidase units to 601 units. The results indicate that AggR-I14T and AggR-N16D interfere with AggR expressed from pLG339, though these AggR mutants are not strongly *trans*-dominant.

Then, I transformed pBAD*aggR-I14T* and pBAD*aggR-N16D* into *E. coli* K-12 BW25113 cells containing pRW50*agg4D* and either pLG339 or pLG339*aggR* 042. Cells were grown with shaking until mid-exponential phase in the presence and absence of arabinose and promoter activity was determined by measuring β-galactosidase levels. The results in Figure 5.11B show that cells containing pLG339 and pBAD*aggR-I14T* had low levels of induction, 4.6-fold from 28 β-galactosidase units in the absence of arabinose to 127 units in its presence. With pBAD*aggR-N16D*, there was a 4.9-fold induction from 28 β-galactosidase units in the absence of arabinose to 137 units in its presence. β-galactosidase levels were much higher with pBAD*aggR* with 1293 units in the presence of arabinose. There was a 1.4-fold difference for cells containing pLG339*aggR* 042 and pBAD*aggR-I14T* (1063 β-galactosidase units in the presence of arabinose from 1523 units in its absence) or pBAD*aggR-N16D* (1118 units from 1594 units).

Changing residues I14 and N16 resulted in AggR mutants that were defective in activation, consistent with results obtained for Rns (Basturea *et al.*, 2008), indicating that these residues, and the N-terminal domain, are vital for AggR-dependent activation. Though neither AggR-I14T nor AggR-N16D were strongly *trans*-dominant, there was a 2-fold reduction at the *afaB* promoter and 1.4-fold reduction at the *agg4D* promoter when co-expressed with wild-type AggR. This indicates that both changes in AggR, guided by Rns, that resulted in mutants defective in activation also interfered with wild-type AggR.

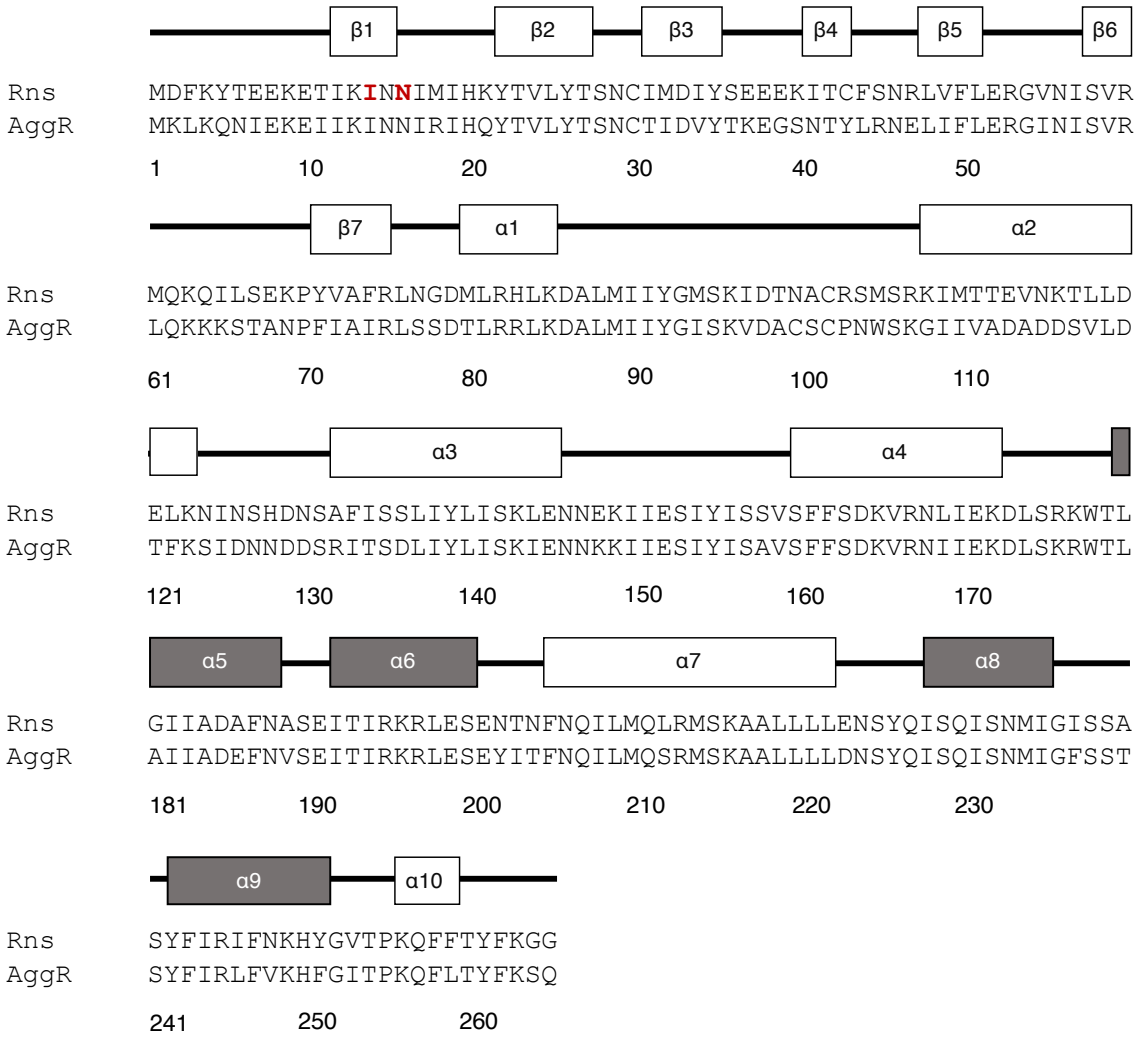
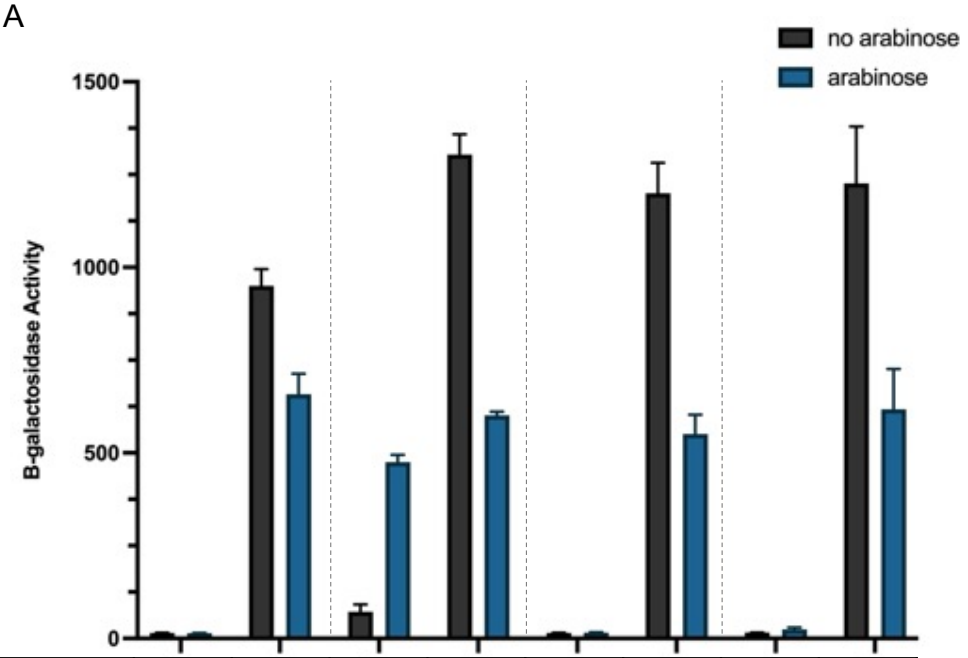


Figure 5.10 Alignment of Rns ETEC and AggR EAEC strain 042 amino acid sequences

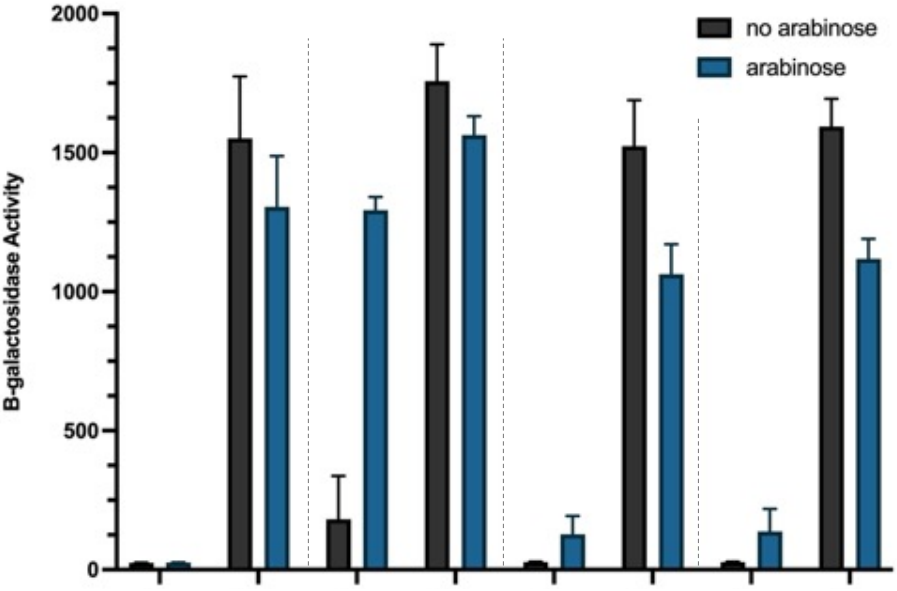
The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042) and Rns from ETEC. The predicted secondary structure located above the amino acid sequence is modelled on Rns (ETEC) with the α helices and β sheets labelled. For the tertiary structure, the two HTH motifs that are part of the AraC homologous region are highlighted in grey. The amino acids highlighted in red indicate the residues targeted for mutational analysis guided by Rns. Adapted from Mahon *et al.* (2010).

A



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	I14T	I14T	N16D	N16D

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	I14T	I14T	N16D	N16D

Figure 5.11 Analysis of promoter activity with AggR mutants guided by Rns

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmids pRW50*afaB* (A) or pRW50*agg4D* (B), in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-I14T* or pBAD*aggR-N16D.* A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 32) = 209.8$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pBAD*aggR-I14T* or pBAD*aggR-N16D* in the presence and absence of arabinose ($p<0.05$).
B. Cells containing pRW50*agg4D* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-I14T* or pBAD*aggR-N16D*. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 32) = 123.0$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pBAD*aggR-I14T* or pBAD*aggR-N16D* in the presence and absence of arabinose ($p<0.05$).

5.4.2 Mutations guided by the MelR DNA-binding domain

The changes I introduced, guided by Rns from ETEC, located in the N-terminal domain, did not result in strong *trans*-dominant AggR mutants. Therefore, I decided to target key structural regions that are conserved across the AraC family. I utilised a study by David Grainger that had mutagenized MelR by targeting residues important for DNA binding and interacting with RNAP σ70, see section 5.4.3 (Grainger *et al.*, 2004). MelR and AggR do not have high sequence identity, however, interesting MelR substitutions are located in the 99 amino acid region that is conserved between AraC family members (4.2.1).

Here, I focus on changes in MelR that affected its DNA binding function, and the corresponding changes in AggR (Figure 5.12A). My goal was to find AggR positive control mutants that prevented activation at target promoters. It has long been assumed that AggR functions as a simple Class II activator that binds to the promoter, overlapping the -35 element, and recruits RNAP to the DNA for transcription initiation. Thus, preventing binding will stop promoter activation by AggR. Mutations that resulted in small negative effects on DNA binding by MelR, highlighted in pink in Figure 5.12B, corresponded to residues E199, S200 and R245 in AggR. The greatest effects were seen at positions corresponding to residues I192, T240 and Y242 in AggR. Hence, alanine was introduced, using Megaprimer PCR, at each of these positions, using primers with an EcoRI site upstream and XbaI site downstream. The mutant recombinant plasmids, pBAD*aggR-I192A,* pBAD*aggR-E199A*, pBAD*aggR-S200A,* pBAD*aggR-T240A,* pBAD*aggR-Y242A* and pBAD*aggR-R245A,* were transformed into *E. coli* K-12 BW25113 Δ*lac* containing pRW50*afaB* (Figure 5.12C) and pRW50*agg4D* (Figure

5.12D). Cells were grown in the presence and absence of 0.2% (*w/v*) arabinose, and β-galactosidase levels in cells lysates were measured to determine promoter activity.

The data in Figure 5.12C show that there is an increase in β-galactosidase levels in the presence of arabinose, compared to empty vector (40 units), in cells containing pBAD*aggR-I192A* (6.7-fold, 271 units), pBAD*aggR-E199A* (12.1-fold, 485 units), pBAD*aggR-S200A* (11.7-fold, 467 units), pBAD*aggR-T240A* (10.5-fold, 420 units), and pBAD*aggR-R245A* (19.5-fold, 779 units). However, β-galactosidase levels in cells containing pBAD*aggR-Y242A* were low in the presence and absence of arabinose (27 and 30 units, respectively), and comparable to basal levels (40 and 29 units, respectively). The results indicate that only the pBAD*aggR-Y242A* mutant negatively affected activation at *afaB*, which is consistent with similar experiments with MelR, where the change at this position had a big effect on DNA binding. Interestingly, I192A and T240A were expected greatly to greatly affect DNA binding, as in MelR, but activation still occurred. However, β-galactosidase levels were lower with I192A compared to wild-type AggR, which had a 10.7-fold increase to 427 β-galactosidase units. The R245A mutation actually increased the fold induction, 19.4-fold compared to 10.7-fold with wild-type AggR.

The results in Figure 5.12D show a similar story at the *agg4D* promoter: β-galactosidase levels in the presence and absence of arabinose in cells containing pBAD*aggR-Y242A* (59 and 56 units, respectively) were comparable to basal levels (52 and 55 units, respectively). The remaining AggR mutants showed an increase in β-galactosidase levels in the presence of arabinose compared to

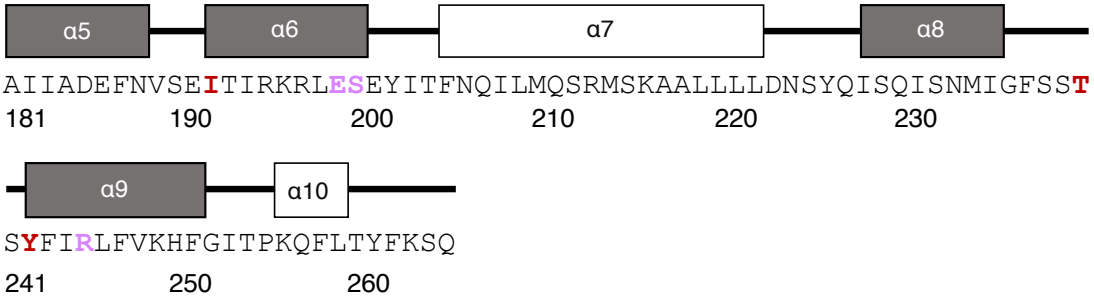
empty vector: pBAD*aggR-I192A* (23.4-fold, 1244 units), pBAD*aggR-E199A* (37.1-fold, 1929 units), pBAD*aggR-S200A* (34.1-fold, 1768 units), pBAD*aggR-T240A* (32.8-fold, 1701 units), and pBAD*aggR-R245A* (24.5-fold, 1273 units). Again, there was a slight reduction with AggR-I192A compared to AggR wild-type, which had a 30.4-fold increase with 1580 units. Interestingly, the T240A mutation in MelR had a big effect on DNA binding, but here, there was an increase compared to wild-type AggR. This was consistent with AggR-E199A and AggR-S200A, though these mutations in MelR had a moderately negative impact on DNA binding. Overall, these results show that the mutations in AggR did not behave as expected according to MelR.

Following these experiments with potential DNA binding mutants of AggR, I decided to utilise the pLG339 co-expression assay system. Thus, I transformed the AggR mutants into *E. coli* K-12 BW25113 cells containing pRW50*afaB* and either pLG339 empty vector or pLG339*aggR* 042. Cells were grown in LB in the presence and absence of arabinose, and β-galactosidase activities were measured.

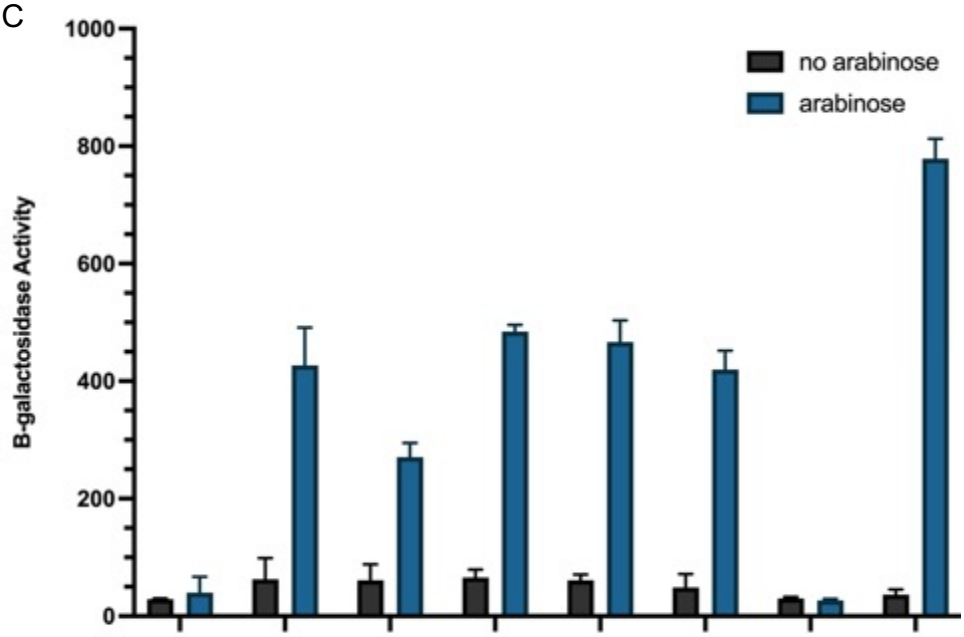
The results in Figure 5.13A show that the DNA binding mutants, that had a small effect in MelR, expressed in *trans* with wild-type AggR reduced β-galactosidase levels. The reductions between cells, in the presence and absence of arabinose, containing pBAD*aggR-E199A* (1456 and 2222 units, respectively), pBAD*aggR-S200A* (1358 and 2259 units, respectively) and pBAD*aggR-R245A* (1544 and 2179 units, respectively) were 1.5-fold, 1.7-fold, and 1.4-fold, respectively. However, cells containing pLG339*aggR* 042 and pBAD*aggR* gave a 1.4-fold reduction, from 2193 to 1582 units.

A

B

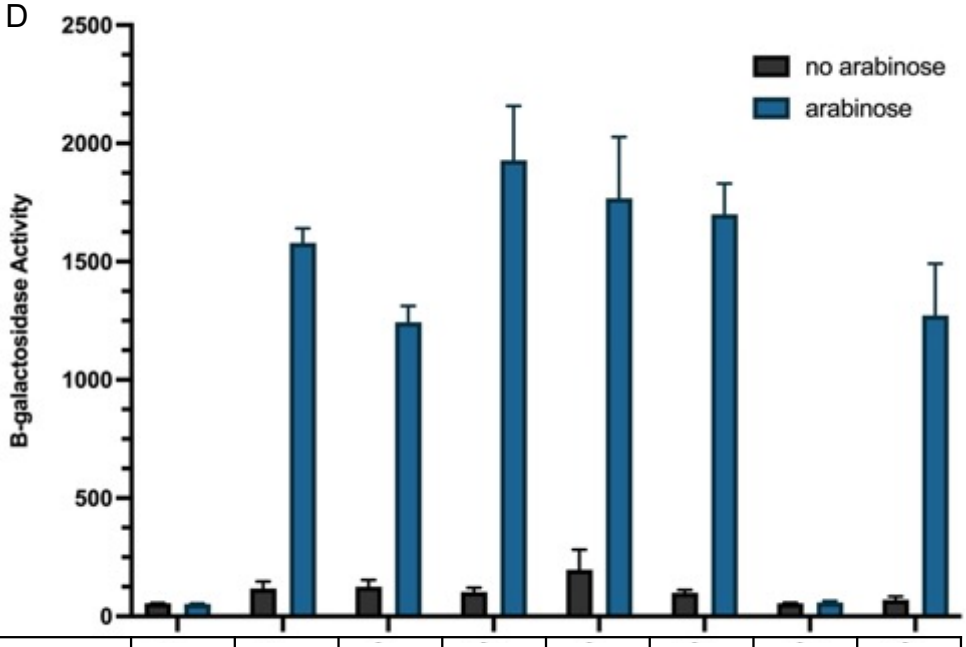


C



promoter in pRW50	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*	*afaB*
activator in pBAD	–	AggR 042	I192A	E199A	S200A	T240A	Y242A	R245A

D



promoter in pRW50	*agg4D*	*agg4D*	*agg4D*	*agg4D*	*agg4D*	*agg4D*	*agg4D*	*agg4D*
activator in pBAD	–	AggR 042	I192A	E199A	S200A	T240A	Y242A	R245A

Figure 5.12 Mutational analysis and promoter activity with AggR mutants, guided by MelR DNA binding mutants

A. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042) and MelR (*E. coli*) for the second α-helix in HTH1 and first α-helix in HTH2. The arrows indicate residues that were targeted in MelR for mutational analysis. The amino acids are highlighted according to properties; light blue, hydrophobic; pink, glycine; dark grey, basic; and dark blue, polar and aromatic. Adapted from Grainger *et al.* (2004).
B. The figure shows the amino acid sequence of AggR 042 (EAEC strain 042), the predicted secondary structure located above the amino acid sequence is modelled on Rns (ETEC) with the α helices and β sheets labelled. The two HTH motifs that are part of the AraC homologous region are highlighted in grey. The amino acids highlighted in pink and red indicate the residues targeted for mutational analysis guided by MelR; residues highlighted in red had a big effect on DNA binding in MelR, and a small effect if highlighted in pink.
C. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB* in addition to either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-I192A,* pBAD*aggR-E199A,* pBAD*aggR-S200A,* pBAD*aggR-T240A,* pBAD*aggR-Y242A* or pBAD*aggR-R245A*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*)

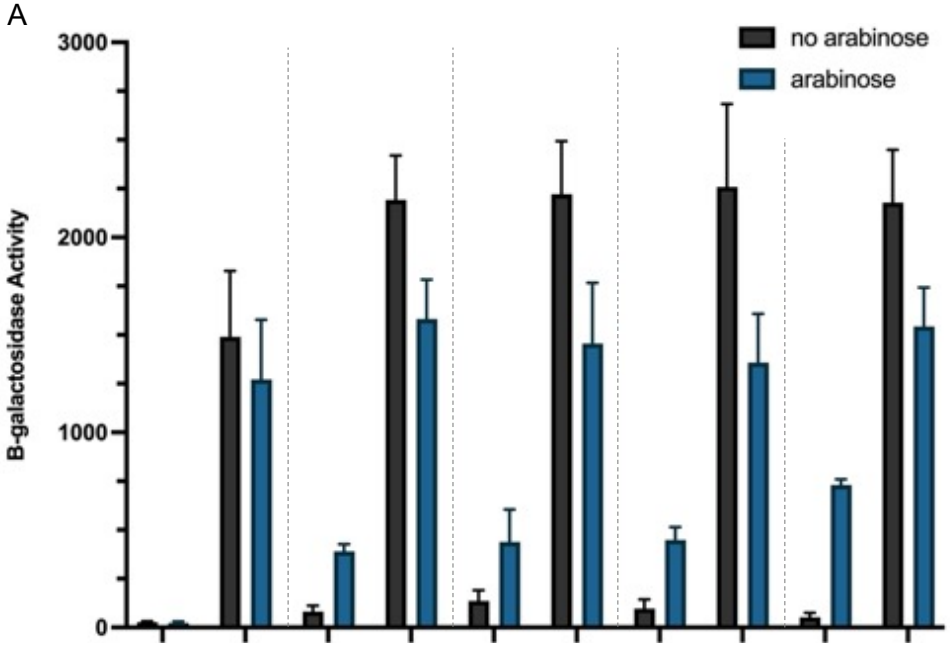
arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

D. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*agg4D* in addition to either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-I192A,* pBAD*aggR-E199A,* pBAD*aggR-S200A,* pBAD*aggR-T240A,* pBAD*aggR-Y242A* or pBAD*aggR-R245A*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

Next, the DNA binding mutants with bigger effects on MelR-DNA binding, were co-expressed with pBAD*aggR* and showed reduced β-galactosidase activity (Figure 5.13B). Cells containing pBAD*aggR-I192A* (1249 units) in the presence of arabinose had a 1.8-fold reduction compared to cells grown in the absence of arabinose (2249 units), this was the greatest reduction seen. Cells with expression of AggR-T240A induced had a 1.5-fold reduction, with 1423 and 2109 units. Cells containing pLG339 and pBAD*aggR-Y242A* had very low β-galactosidase levels comparable to basal levels. However, there was only a 1.5-fold decrease in β-galactosidase activity when AggR-Y242A expression was induced, 1196 units, compared to cells grown in the absence of arabinose, 1791 units.

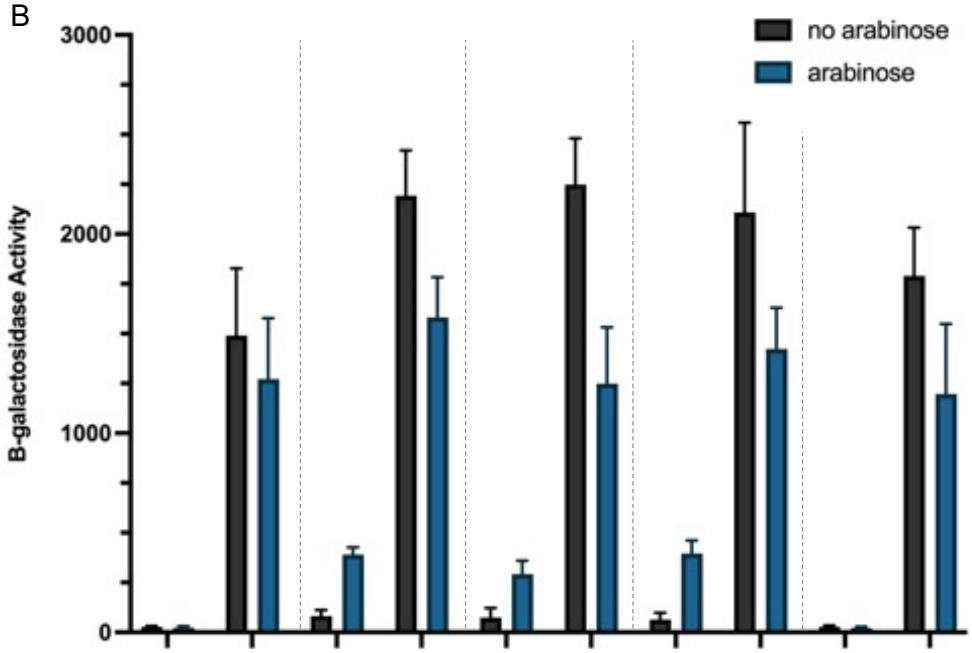
The data show that the mutations introduced in AggR, guided by mutations in the MelR DNA binding domain, behave differently than in MelR. Despite the fact that the mutations E199A, S200A and R245A had weak effects in MelR, the results in this study indicate that these mutations do not prevent activation at target promoters or, presumably, DNA binding. Similarly, the mutations that had big effects in MelR, I192A and T240A do not have a negative impact on AggR binding affinity to DNA. Only the Y242A change in AggR actually destroyed activation at the *afaB* promoter. When the AggR mutants were expressed in *trans* with wild-type AggR, the fold-decrease in β-galactosidase activity was less than 2-fold for each mutant AggR. This indicates that, while there is some interference, none of these mutants are strongly *trans*-dominant.

A



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	E199A	E199A	S200A	S200A	R245A	R245A

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	I192A	I192A	T240A	T240A	Y242A	Y242A

Figure 5.13 Analysis of promoter activity with co-expression of AggR mutants, guided by MelR DNA binding mutations

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmids pRW50*afaB* in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-E199A*, pBAD*aggR-S200A* or pBAD*aggR-R245A.*
B. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-I192A*, pBAD*aggR-T240A* or pBAD*aggR-Y242A*.

5.4.3 Mutations guided by the MelR σ70 interaction

Since AggR functions as a Class II activator, binding to promoter DNA overlapping the -35 element and, likely, interacting with the RNAP σ70 domain, I considered targeting residues, guided by MelR, involved in this interaction. As I was searching for a *trans*-dominant mutant, I wanted to introduce changes that either improve or destroy the interaction, thereby either sequestering RNAP or stopping its recruitment at target promoters.

The positions important for MelR σ70 interaction correspond to residues Y226, Q227, Q230 and M234 in AggR (Figure 5.14A and Figure 5.14B). The residue located at Q230 has already been subjected to mutagenesis (Section 5.3.2) and shown to not be *trans*-dominant. Each of the other positions, Y226, Q227 and M234, was then changed to alanine utilising Megaprimer PCR. Primers were used with an EcoRI restriction site upstream and an XbaI site downstream. The recombinant plasmids, pBAD*aggR-Y226A*, pBAD*aggR-Q227A* and pBAD*aggR-M234A*, were transformed into *E. coli* K-12 BW25113 Δ*lac* cells containing pRW50*afaB* and either pLG339 empty vector or pLG339*aggR* 042. Cells were grown in LB, in the presence and absence of arabinose, to mid-exponential phase. Promoter activity was determined by measuring β-galactosidase levels in cell lysates.

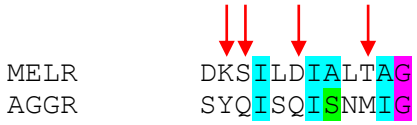
The results in Figure 5.14C show a 31.7-fold increase in β-galactosidase levels in cells containing pLG339 with pBAD*aggR-Y226A* (592 units) in the presence of arabinose, compared to pBAD30 empty vector (19 units). The β-galactosidase levels are comparable to wild-type AggR, with a 32.3-fold induction, 605 units. In contrast, there was a 1.3-fold reduction in cells containing pLG339*aggR* 042 with

expression of AggR-Y226A induced by arabinose (1555 units), compared to cells in the absence of arabinose (2077 units). This is compared to a 1.4-fold reduction in β-galactosidase activity in cells containing pLG339*aggR* 042 and pBAD*aggR*, with 2139 and 1484 units. The results indicate that mutating tyrosine at position 226 to alanine does not stop interaction with RNAP σ70, additionally, AggR-Y226A is not *trans*-dominant.

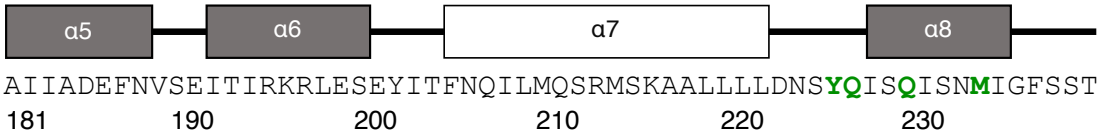
Next, the AggR-Q227A and AggR-M234A mutants were assayed. Figure 5.14D shows that β-galactosidase levels were similar in cells containing pLG339 empty vector with pBAD*aggR* or pBAD*aggR-M234A*, however, the levels were lower with pBAD*aggR-Q227A*. For pBAD*aggR*, the fold-induction was 32.3 with 605 units, compared to 19 units with pBAD30 empty vector. In contrast, pBAD*aggR-Q227A* and pBAD*aggR-M234A* had 9.1-fold (171 units) and 30.6-fold induction (573 units), respectively. This indicates that, when expression of AggR-Q227A was induced, there was a 20-fold lower increase in β-galactosidase activity compared to wild-type AggR. Therefore, mutating the glutamine at position 227 to alanine impacts AggR activation at the *afaB* promoter.

When AggR, AggR-Q227A and AggR-M234A were co-expressed with wild-type AggR, there was a 1.4, 1.6 and 1.3-fold reduction, respectively. Cells containing AggR-Q227A had 1256 units in the presence and 1955 units in the absence of arabinose, AggR-M234A had 1474 and 1928 units. The results indicate that AggR-Q227A expressed in *trans* with AggR showed the greatest reduction in β-galactosidase activity, however, this mutant is not strongly *trans*-dominant. However, position 227 has been shown to be important for AggR activation as there was a dramatic reduction in β-galactosidase levels.

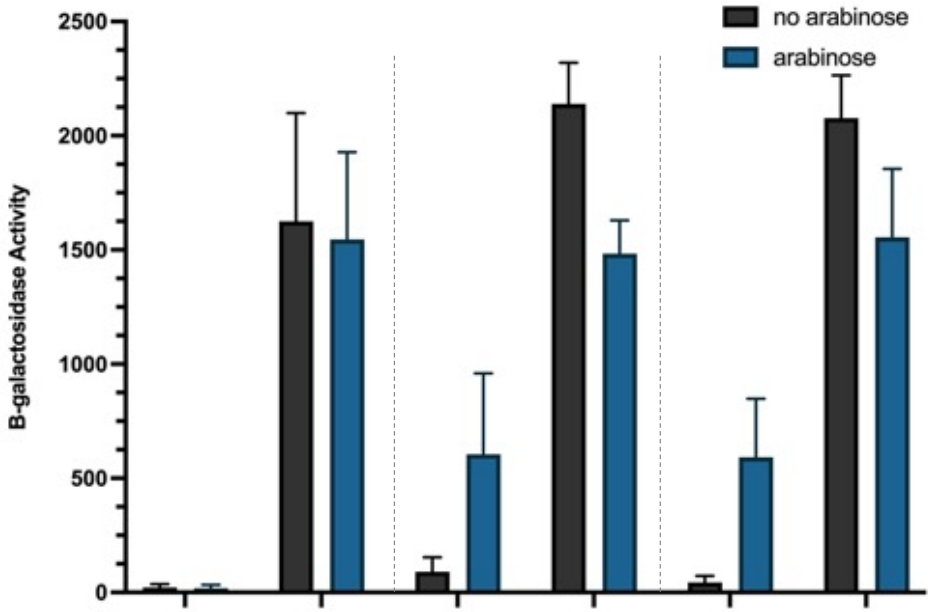
A



B

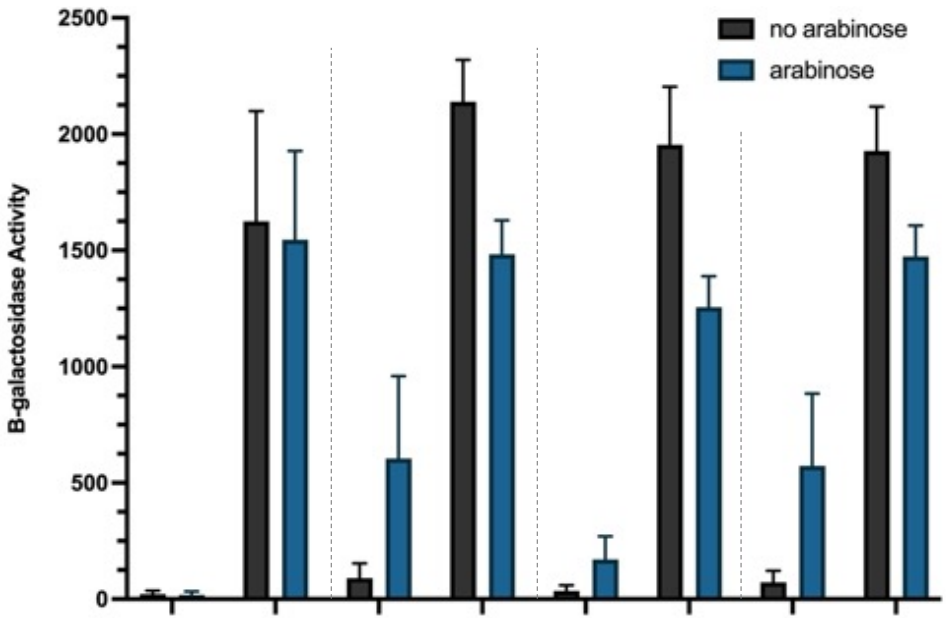


C



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	Y226A	Y226A

D



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	Q227A	Q227A	M234A	M234A

Figure 5.14 Mutational analysis and promoter activity with co-expression of AggR mutants, guided by MelR σ70 interaction domain mutants

A. The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042) and MelR (*E. coli*) for the second α-helix in HTH1 and first α-helix in HTH2. The arrows indicate residues that were targeted in MelR for mutational analysis. The amino acids are highlighted according to properties; light blue, hydrophobic; pink, glycine; and green, polar and aliphatic. Adapted from Grainger *et al.* (2004).
B. The figure shows the amino acid sequence of AggR 042 (EAEC strain 042), the predicted secondary structure located above the amino acid sequence is modelled on Rns (ETEC) with the α helices and β sheets labelled. The two HTH motifs that are part of the AraC homologous region are highlighted in grey. The amino acids highlighted in green indicate the residues targeted for mutational analysis guided by MelR that had an effect on σ70 interaction.
C. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR* or pBAD*aggR-Y226A*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per

milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

D. The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR,* pBAD*aggR-Q227A* or pBAD*aggR-M234A*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

My data indicates that “one size does not fit all”, and what works for MelR doesn’t necessarily work for AggR: only AggR-Q227A greatly affected expression levels, indicating that this residue is important for AggR-activation function, potentially for contacting RNAP. While some interference was seen when mutants were co-expressed with wild-type AggR, this did not stop AggR-activation at *afaB*.

5.4.4 Mutations of charged residues

In the next part of my work, I studied the effects of changing several charged residues in the two HTH motifs of AggR that are conserved in AraC family of transcription factors and might make contacts with specific DNA sequences (Bhende and Egan, 1999; Mahon *et al.*, 2010; Schuller *et al.*, 2012). Figure 5.15 highlights these residues.

The glutamic acid at position 191 was targeted as it is at the start of the predicted second ‘recognition’ α-helix of HTH1, and this negatively charged residue has been identified as important for contacting DNA in other AraC family transcription factors (Bhende and Egan, 1999). An alanine codon was introduced by using Megaprimer PCR with primers that contained an EcoRI restriction site upstream and HindIII site downstream. The recombinant plasmid, pBAD*aggR-E191A*, was transformed into *E. coli* K-12 BW25113 Δ*lac* cells containing pRW50*afaB* and either pLG339 or pLG339*aggR* 042. Cells were grown in the presence and absence of 0.2% (*w/v*) arabinose and β-galactosidase levels were measured from cell lysates to determine promoter activity. The results in Figure 5.16A show that β-galactosidase levels were low in cells containing pLG339 and pBAD*aggR-E191A*, in the presence and absence of arabinose (101 and 44 β-galactosidase

units, respectively), comparable to empty vector control (19 and 22 units, respectively). Cells containing pLG339*aggR* 042 and pBAD*aggR-E191A* in the presence of arabinose had 1151 units, compared to 1793 units in the absence of arabinose, which was a 1.6-fold reduction. In contrast, pLG339*aggR* 042 and pBAD*aggR* co-expressed had a 1.4-fold decrease, from 2139 units to1484. This indicates that expressing AggR-E191A in *trans* with wild-type AggR, from pLG339*aggR* 042, had a greater effect than wild-type AggR. However, AggR-E191A is not strongly *trans*-dominant in this system.

Next, positions 195-197 were targeted due to the charged amino acids arginine, lysine, arginine (RKR) at these positions, which are conserved in other AraC family transcriptional regulators, including Rns (ETEC) (Schuller *et al.*, 2012; Mahon *et al.*, 2010); additionally, these positions have been identified as contacting DNA in other AraC members (Bhende and Egan, 1999). Each of these residues was separately mutated to alanine, the three positions were also mutated together to alanine to determine how removing this stretch of electrically charged amino acids would affect AggR-activation. Mutations were introduced into AggR using megaprimer PCR with primers containing an EcoRI restriction site upstream and HindIII site downstream.

The recombinant plasmids, pBAD*aggR-R195A*, pBAD*aggR-K196A* and pBAD*aggR-R197A*, were transformed into *E. coli* K-12 BW25113 cells containing pRW50*afaB* with pLG339 empty vector or pLG339*aggR* 042. Cells were grown in LB to mid-exponential phase, samples were grown in the absence and presence of arabinose. Promoter activities were determined by measuring β-galactosidase levels in cell lysates.

The data in Figure 5.16B show that each of the pBAD*aggR* mutants prevented AggR-dependent activation and β-galactosidase levels are like basal levels with pBAD30 empty vector. The reduction in β-galactosidase levels were similar in cells containing pLG339*aggR* 042 and pBAD*aggR*, pBAD*aggR-R195A*, pBAD*aggR-K196A* and pBAD*R197A*: 1.4, 1.4, 1.5 and 1.4-fold reductions, respectively. β-galactosidase levels measured in the presence and absence of arabinose were: 1468 and 2021 units, respectively, for pBAD*aggR*; 1129 and 1624 units, respectively, for pBAD*aggR-R195A*; 1333 and 1940, respectively, for pBAD*aggR-K196A*; and 998 and 1425 units, respectively, for pBAD*aggR-R197A*. The results indicate that individually altering positions 195-197 in AggR prevents AggR-activation at *afaB*, however, the reduction in β-galactosidase levels when AggR mutants were co-expressed with wild-type AggR was minimal, comparable to overexpression of wild-type AggR from the different plasmids. Therefore, the 195, 196 and 197 position mutants are not *trans*-dominant.

Next, I constructed a mutant where all three of the charged residues were changed to alanine. Again, I used Megaprimer PCR and primers with EcoRI and HindIII sites, upstream and downstream, respectively. The pBAD*aggR-RKR195-197AAA* recombinant plasmid was transformed into *E. coli* K-12 BW25113 cells containing pRW50*afaB* and either pLG339 or pLG339*aggR* 042. Cells were grown in the presence and absence of arabinose. Figure 5.16C shows that mutating all three residues stops activation at the *afaB* promoter. However, in cells containing pLG339*aggR* 042 and pBAD*aggR-RKR195-197AAA*, arabinose triggered a reduction in β-galactosidase expression by only 1.3-fold. This indicates that mutating positions 195-197 does not result in a *trans*-dominant

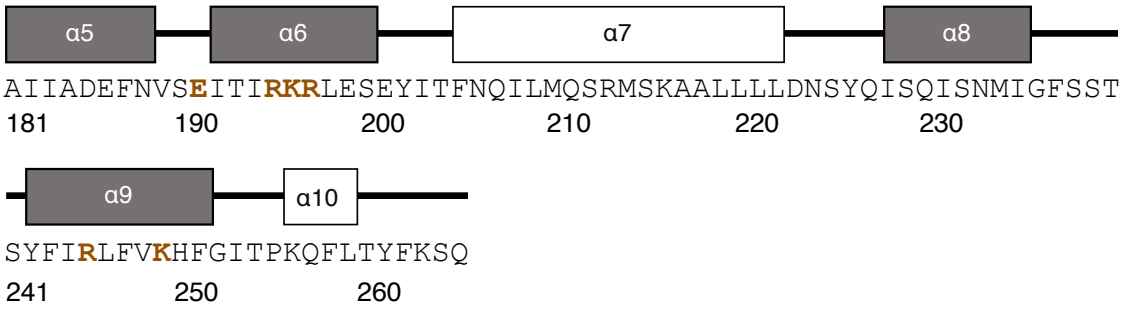
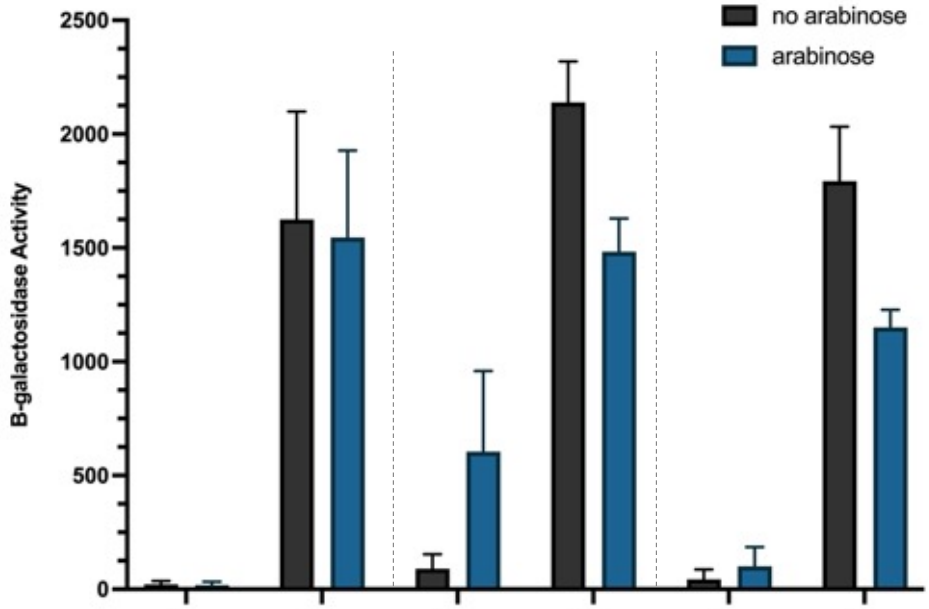


Figure 5.15 Mutational analysis of AggR, targeting charged residues

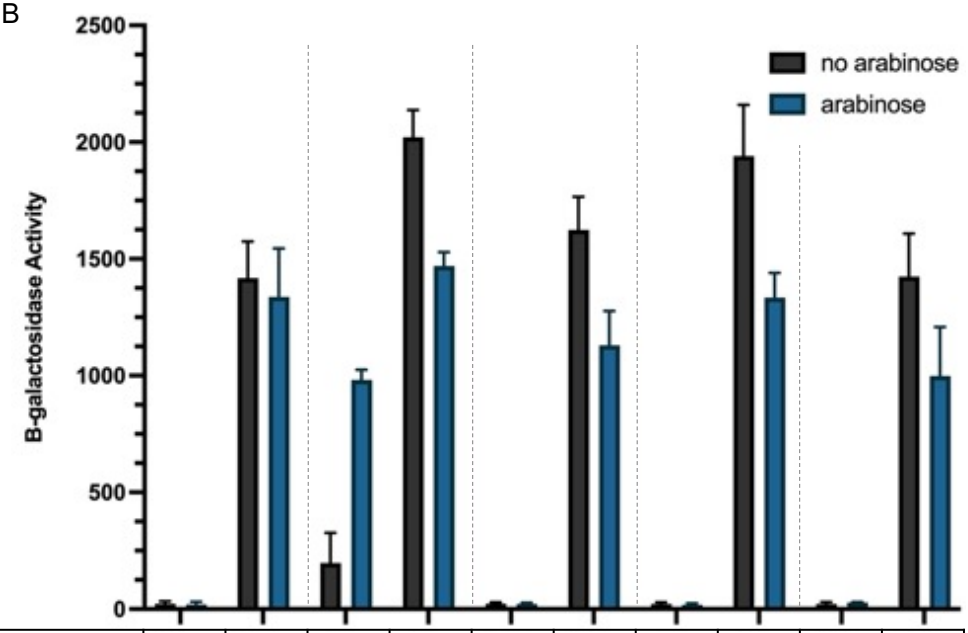
The figure shows the amino acid sequence of AggR 042 (EAEC strain 042), the predicted secondary structure located above the amino acid sequence is modelled on Rns (ETEC) with the α helices and β sheets labelled. The two HTH motifs that are part of the AraC homologous region are highlighted in grey. The amino acids highlighted in brown indicate the residues targeted for mutational analysis due to these positions being conserved charged residues. Adapted from Mahon *et al.* (2010).

A



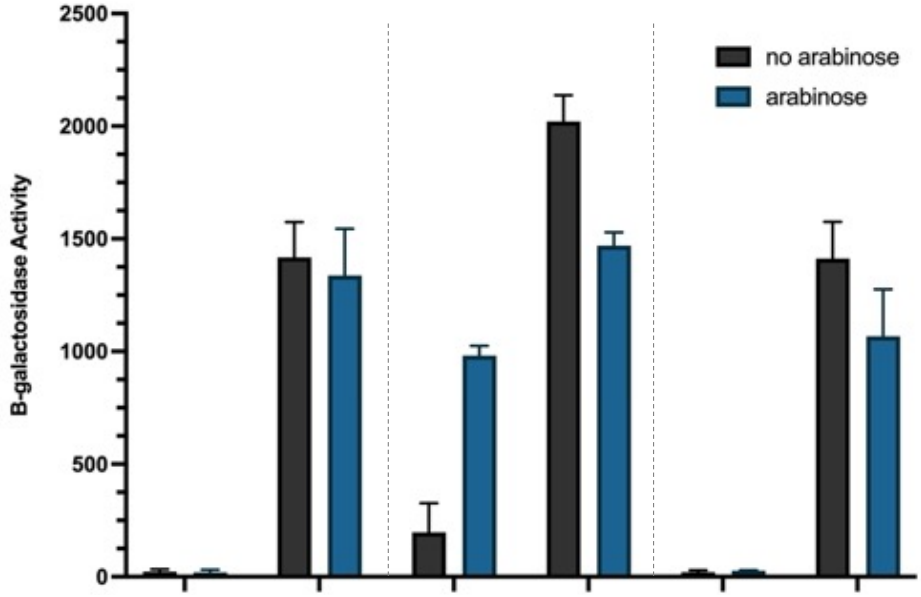
activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	E191A	E191A

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	R195A	R195A	K196A	K196A	R197A	R197A

C



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	RKR 195-197 AAA	RKR 195-197 AAA

Figure 5.16 Analysis of *afaB* promoter activity with co-expression of AggR mutants targeting HTH1

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmids pRW50*afaB*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR* or the pBAD*aggR* mutants. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR* or pBAD*aggR-E191A.* A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 60) = 76.6$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pLG339*aggR* 042 and pBAD*aggR-E191A* in the presence and absence of arabinose ($p<0.05$).
B. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR*, pBAD*aggR-R195A*, pBAD*aggR-K196A* or pBAD*aggR-R197A*.
C. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR* or pBAD*aggR-RKR195-197AAA.*

AggR mutant and the reduction in β-galactosidase levels is actually less than that seen when the residues at 195, 196 and 197 were individually changed.

Next, the second α-helix in the second HTH motif was targeted. Again, charged amino acids, arginine and lysine, were targeted, at positions 245 and 249, respectively. Mutations were introduced into *aggR* using megaprimer PCR with primers that containing an EcoRI restriction site upstream and HindIII site downstream. Cells were grown to mid-exponential phase in the presence and absence of arabinose and promoter activity was determined by measuring β-galactosidase levels in cell lysates. Firstly, with cells containing pLG339 empty vector, the results in Figure 5.17 show that cells containing pBAD*aggR-R245K* (1128 β-galactosidase units) had the highest induction levels, 64-fold, in the presence of arabinose. This was followed by pBAD*aggR*, 46-fold induction to 814 β-galactosidase units, then pBAD*aggR-K249A* with 34-fold induction to 607 units. The results indicate that neither AggR-R245K nor AggR-K249A stopped AggR-activation at *afaB*, in fact, AggR-R245K increased activation. Next, with the cells containing pLG339*aggR* 042, the data show that there was only a 1.1-fold reduction in β-galactosidase levels when expression of AggR-R245K and AggR-K249A was induced, from 1516 to 1382 units, and from 1404 to 1277 units, respectively. When wild-type AggR was co-expressed, there was a 1.4-fold reduction. The results show that removing these charged residues in the second α-helix of the second HTH motif does not negatively impact on activation at target promoters or result in a *trans*-dominant AggR mutant.

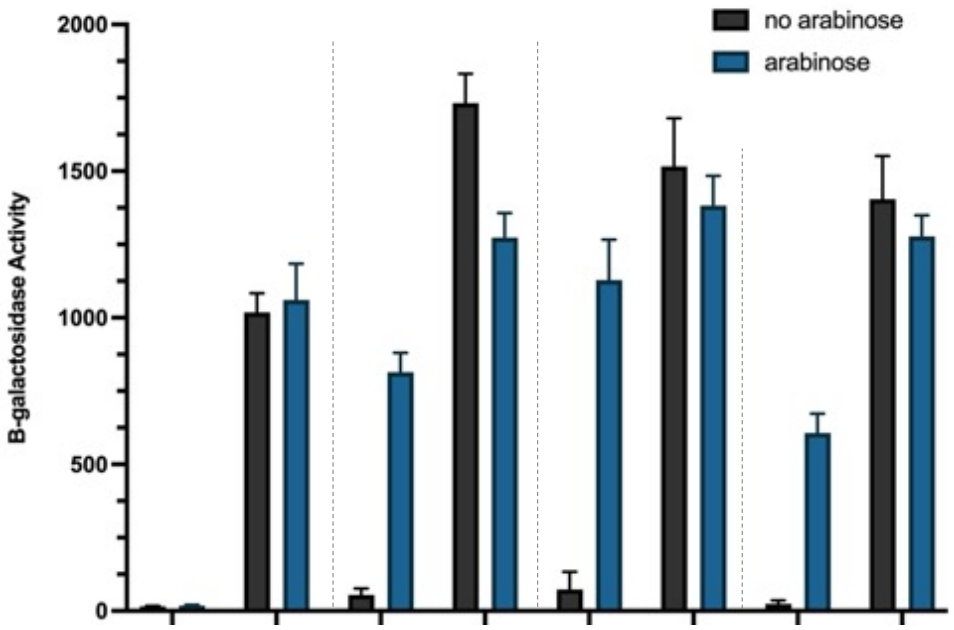
In summary, the results show that, while none of the AggR mutants are strongly *trans*-dominant, there were dramatic effects on AggR-mediated activation:

changing the charged residues in the second α-helix of HTH1 completely prevented AggR-activation at the *afaB* promoter, however, changing charged residues in the second α-helix of HTH2 did not prevent activation. This indicates that the charged residues in HTH1 are vital for AggR activation function, presumably for DNA binding.

5.4.5 Mutations guided by VirF

After considering changes in AggR in its N-terminal domain, guided by Rns, and in its C-terminal domain, guided by MelR, I considered a previous report of *trans*-dominant mutants in *Shigella flexneri* VirF (Porter and Dorman, 2002). As discussed in Section 4.4, AggR and VirF share 34% sequence identity and belong to the same AraC subgroup as Rns, CsvR and CfaR from ETEC (Munson and Scott, 1999; Munson *et al.*, 2001). The study by Porter and Dorman identified two changes in VirF that resulted in *trans*-acting VirF mutants, when co-expressed with wild-type. The first, VirF-I180N, located in the first α-helix of HTH1, altered a conserved hydrophobic residue, which affected the formation of the DNA-protein complex. The second, Y224STOP, located just preceding the first α-helix of HTH2, introduced a premature stop codon, resulting in deletion of HTH2, which is involved in contacting RNAP (Porter and Dorman, 2002).

The sequence alignment in Figure 5.18 shows Rns ETEC, AggR EAEC 042 and VirF *S. flexneri*, with the predicted secondary structure, indicating the position of the two HTH motifs. I have highlighted the locations where changes in AggR were previously introduced in AggR. In VirF, a number of changes that affect its activity



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	R245K	R245K	K249A	K249A

Figure 5.17 Analysis of *afaB* promoter activity with co-expression of AggR mutants targeting HTH2

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB* in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and pBAD*aggR,* pBAD*aggR-R245K* or pBAD*aggR-K249A*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.
* Expression of the *aggR* gene is under the control of the *aggR* promoter

are highlighted in blue, and the locations of the I180N and Y224STOP changes are highlighted in red. Correspondingly in AggR, I engineered an isoleucine to asparagine change at position 183, and also introduced a premature stop codon at position 227. Note that I had previously introduced a change at this position (see Section 5.4.3: glutamine to alanine), when making changes guided by MelR. To introduce these changes in AggR, I utilised Q5 site-directed mutagenesis. The pBAD*aggR-I183N* and pBAD*aggR-Q227STOP* recombinant plasmids were transformed into *E. coli* K-12 BW25113 Δ*lac* cells containing pRW50*afaB* and either pLG339 or pLG339*aggR* 042. Cells were grown in LB, and in the presence and absence of arabinose, to mid-exponential phase and β-galactosidase activity was measured as an indicator of promoter activity.

Figure 5.19A shows the results with AggR-I183N. With cells containing pLG339 empty vector, the data show that the AggR-I183N mutant gives low β-galactosidase levels in the presence and absence of arabinose comparable to empty vector controls. Cells containing pBAD*aggR* in the presence of arabinose (277 units) had a 7-fold induction over cells in the absence of arabinose (39 units). In cells containing pLG339*aggR* 042 with pBAD*aggR*, the β-galactosidase levels decreased 1.2-fold, from 1436 units to 1164 units when expression of AggR was induced. There was also a 1.2-fold decrease with AggR-I183N: from 1146 units in the absence of arabinose to 943 units when expression of AggR-I183N was induced (one-way ANOVA, $p>0.0001$; Tukey test, $p>0.05$). The results indicate that the change at position 183 blocks AggR-activation at the *afaB* promoter, however, AggR-I183N is not *trans*-dominant to wild-type AggR.

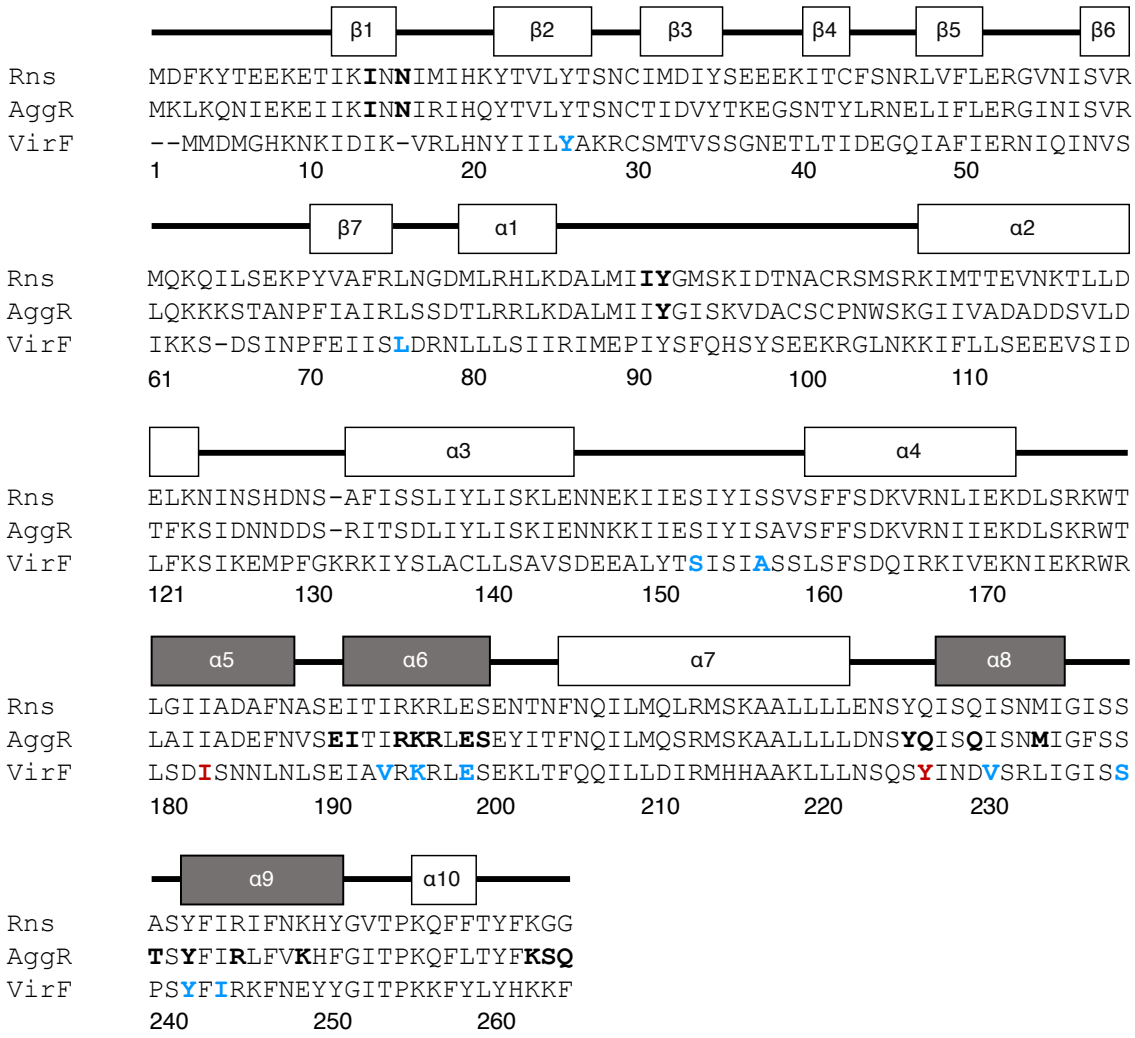


Figure 5.18 Alignment of AggR, Rns and VirF amino acid sequences

The figure shows an alignment of the amino acid sequences of AggR 042 (EAEC strain 042), VirF (*S. flexneri*) and Rns (ETEC). The predicted secondary structure located above the amino acid sequence is modelled on Rns (ETEC) with the α helices and β sheets labelled. For the tertiary structure, the two HTH motifs that are part of the AraC homologous region are highlighted in grey. The amino acids highlighted in bold in the AggR and Rns sequences indicate the residues previously targeted for mutational analysis. For VirF, residues highlighted in blue were defective VirF mutants and red indicates the residues that resulted in a *trans*-dominant phenotype. Adapted from Mahon *et al.* (2010).

Figure 5.19B shows the effects of AggR truncation with AggR-Q227STOP. With cells containing pLG339 empty vector, as with AggR-I183N, this AggR mutant stopped AggR-activation at the *afaB* promoter. Cells containing pLG339*aggR* 042 and pBAD*aggR* showed similar levels of β-galactosidase expression in the presence and absence of arabinose, 1404 and 1460 units, respectively. There was a marginal reduction, 1.1-fold, when expression of AggR-Q227STOP was induced in cells carrying pLG339*aggR* 042 (one-way ANOVA, $p<0.0001$; Tukey test, $p>0.05$). The results indicate that AggR-Q227STOP is not *trans* dominant to wild-type AggR.

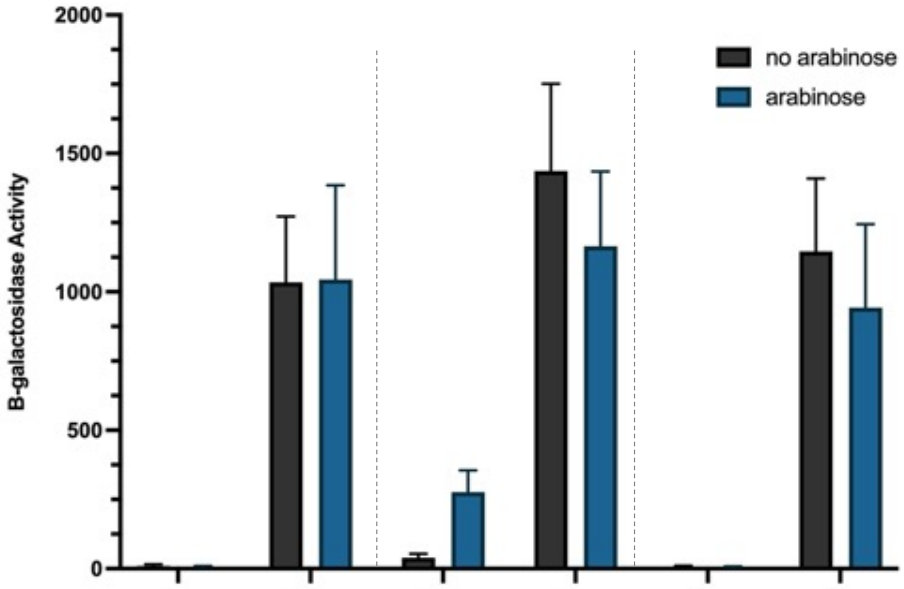
The results in this section indicate that what works for VirF does not work for AggR; when transferred to AggR, the changes that were *trans* dominant in VirF did not produce *trans* dominant AggR mutants, i.e., the AggR mutants did not block wild-type AggR activation of *afaB*. However, the data have shown that I183, in the first α-helix of HTH1, is vital for AggR-dependent activation at the *afaB* promoter, presumably for contacting promoter DNA. Additionally, the presence of HTH2 is also vital for AggR-activation, presumably for contacting RNAP.

5.5 Flipping the script – AggR mutants in pLG339

The *aggR* mutations described so far vary in effectiveness at stopping AggR-activation and cause different levels of interference when expressed in *trans* to wild-type AggR. One consistent feature is that the β-galactosidase levels have been higher when AggR was expressed from pLG339*aggR* 042 compared to pBAD*aggR*. For this reason, I investigated whether this difference in expression levels might be masking *trans* dominant effects. Thus, I used Q5 site-directed

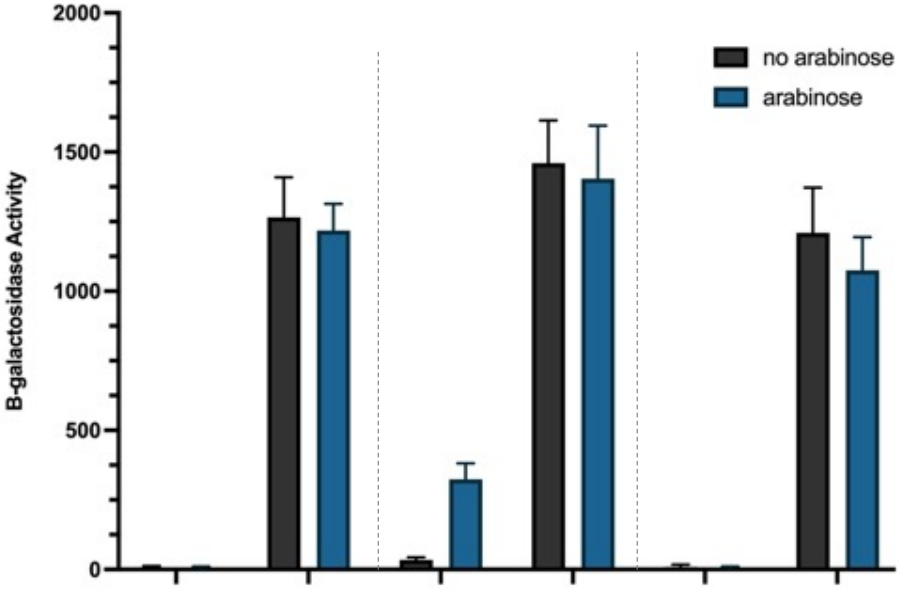
mutagenesis to introduce selected mutations in pLG339*aggR* 042, under the control of the *aggR* promoter, picking *aggR* mutants that were the most effective

A



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	I183N	I183N

B



activator in pLG339*	–	AggR 042	–	AggR 042	–	AggR 042
activator in pBAD	–	–	AggR 042	AggR 042	Q227 STOP	Q227 STOP

Figure 5.19 Analysis of *afaB* promoter activity with co-expression of AggR mutants guided by VirF

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB*, in addition to either pLG339 (empty vector) or pLG339*aggR* 042 and carrying either pBAD*aggR*, pBAD*aggR-I183N* (A) or pBAD*aggR-Q227STOP* (B). Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the *aggR* promoter

A. Cells containing pRW50*afaB* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR* or pBAD*aggR-I183N.* A one-way ANOVA was calculated using the promoter activities, showing the analysis was not significant ($p>0.0001$, $F(11, 60) = 0.2348$). A post-hoc Tukey's HSD test showed that there was no significant difference between promoter activities in cells containing pLG339*aggR* 042 and carrying pBAD*aggR-I183N* in the presence and absence of arabinose ($p>0.05$).
B. Cells containing pRW50*agg4D* with pLG339 or pLG339*aggR* 042 and carrying either pBAD30 (empty vector), pBAD*aggR* or pBAD*aggR-Q227STOP*. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 60) = 222.4$). A post-hoc Tukey's HSD test showed that there was no significant difference between promoter activities in cells containing pLG339*aggR* 042 and carrying pBAD*aggR-Q227STOP* in the presence and absence of arabinose ($p>0.05$).

in blocking activation at promoters: AggR-Y92C and AggR-Y92R (Section 5.3.1), AggR-I14T and AggR-N16D (Section 5.4.1), and AggR-E191A (Section 5.4.4).

First, I transformed pLG339*aggR-Y92C* and pLG339*aggR-Y92R* into *E. coli* K-12 Δ*lac* BW25113 cells containing pRW50*afaB* and either pBAD30 or pBAD*aggR*. Cells were grown to mid-exponential phase in the presence and absence of arabinose, and β-galactosidase levels were measured to indicate promoter activity. The data in Figure 5.20 show that cells containing pBAD30 empty vector and pLG339*aggR-Y92C* or pLG339*aggR-Y92R* had β-galactosidase levels as low as empty vector controls. In contrast, there was a 13-fold increase in β-galactosidase levels in cells with pBAD*aggR* in the presence of arabinose, from 91 units to 1221 units. When wild-type AggR was expressed from both pBAD*aggR* and pLG339*aggR* 042, 1429 β-galactosidase units were measured, which is a 1.5-fold reduction from the 2141 units measured in the absence of arabinose. Cells containing pBAD*aggR* and pLG339*aggR-Y92C* showed a 7.8-fold increase in β-galactosidase levels, from 86 units in the absence of arabinose to 667 units in its presence (one-way ANOVA, $p<0.0001$; Tukey test, $p<0.05$). Similarly, with AggR-Y92R, there was an 8.2-fold increase from 79 to 659 units (one-way ANOVA, $p<0.0001$; Tukey test, $p<0.05$). This induction is less, in the presence of AggR-Y92R or AggR-Y92C, than with wild-type AggR, indicating that, though neither AggR mutant is strongly *trans* dominant, there is significant interference.

Next, the pLG339*aggR-I14T* and pLG339*aggR-N16D* recombinant plasmids were transformed into *E. coli* K-12 BW25113 cells containing pRW50*afaB* and either pBAD30 or pBAD*aggR*. Cells were grown in LB to mid-exponential phase,

in the presence or absence of arabinose, and the promoter activity was determined by measuring β-galactosidase levels. The results, illustrated in Figure 5.21, show that cells containing pBAD30 empty vector and either pLG339*aggR-I14T* or pLG339*aggR-N16D* had low β-galactosidase levels, comparable to empty vector. The induction level was 8.7-fold for pBAD*aggR* with pLG339 empty vector. For cells containing pBAD*aggR* and pLG339*aggR* 042, there was a 1.7-fold reduction from cells in the absence of arabinose to cells with AggR expression induced from pBAD*aggR* (2573 and 1504 units, respectively). In cells containing both pBAD*aggR* and the pLG339*aggR* mutants, there was a 10-fold induction for pBAD*aggR-I14T* (from 60 to 605 units; Tukey test, $p<0.05$) and a 9.7-fold induction for pBAD*aggR-N16D* (from 67 units to 651; Tukey test, $p<0.05$). For pLG339*aggR-I14T*, this is a 2.4-fold reduction in β-galactosidase levels compared to the levels of pLG339*aggR* 042 with empty vector. Similarly, AggR-N16D also interfered with wild-type AggR, giving a 2.2-fold reduction.

Finally, I transformed pLG339*aggR-E191A* into *E. coli* K-12 BW25113 cells containing pRW50*afaB* and either pBAD30 empty vector or pBAD*aggR*. Cells were grown in the presence and absence of arabinose and β-galactosidase levels were measured to determine promoter activity. The results in Figure 5.22 show that cells containing pBAD30 empty vector and pLG339*aggR-E191A* had low β-galactosidase levels in the presence and absence of arabinose, both 80 units, compared to empty vector controls, 36 and 44 units, respectively. There was a 3.7-fold increase in β-galactosidase levels, from 297 to 1100 units in the absence and presence of arabinose, respectively, in cells containing pLG339 and pBAD*aggR*. In cells containing pLG339*aggR-E191A* with expression of

pBAD*aggR* induced by the addition of arabinose to the media, there was a 3.2-fold increase (834 units) over the cells grown in the absence of arabinose (258 units). This was a 1.7-fold reduction in β-galactosidase activity compared to the cells containing pLG339*aggR* 042 and pBAD30 empty vector (Tukey test, $p<0.05$). Therefore, this AggR mutant is not *trans* dominant.

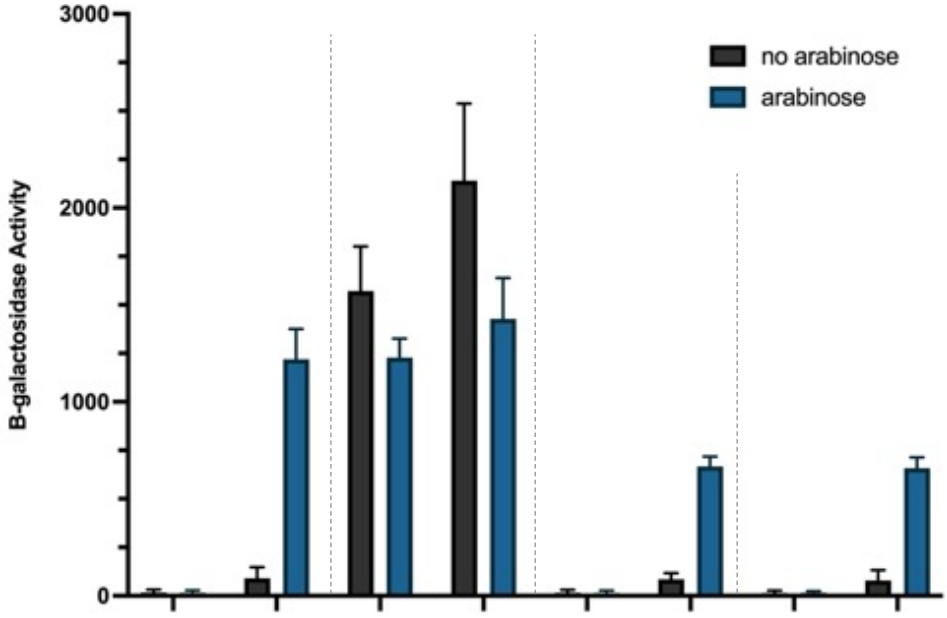
The results in this section indicate that there is some interference with wild-type AggR, when the mutant AggR is expressed from the pLG339*aggR* 042 vector, under the control of the *aggR* promoter.

5.6 Discussion

The AggR regulon contains many virulence factors involved in the pathogenicity of EAEC. My starting intention was to use AggR mutants to block AggR-activation at promoters, thus preventing expression of virulence factors and opening up a possible avenue for the treatment of EAEC infections without using antibiotics; for example, by delivering non-functional AggR in *trans* with wild-type AggR to block activation at target promoters. Through extensive mutational analysis of the *aggR* gene, and subsequent measurements of AggR-dependent activation at promoters, I have exhausted efforts to find a *trans*-dominant mutant.

Data in Figure 5.1 shows expression and activity of AggR independent of any external signal, and activation levels at the tested target promoters were the same or higher than with arabinose-induced pBAD*aggR*. This is an interesting result because it indicates that AggR expression must be turned on constantly, at least to low levels. This is consistent with the idea that AggR activity must reach a threshold for target promoters to be switched on, see Chapter 6.

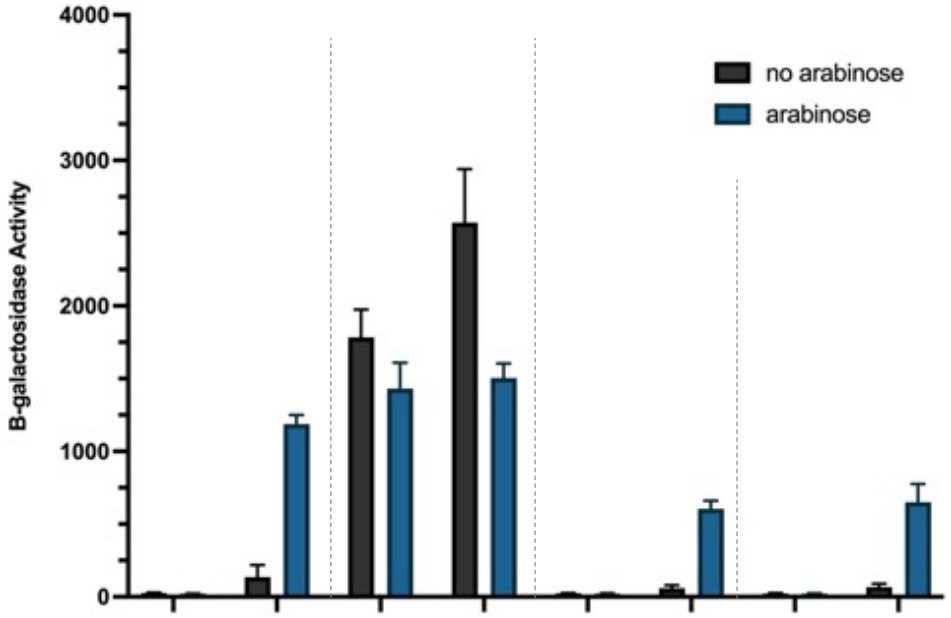
I subjected AggR to extensive mutagenesis, both random and site-directed,



activator in pBAD	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pLG339*	–	–	AggR 042	AggR 042	Y92C	Y92C	Y92R	Y92R

Figure 5.20 Analysis of *afaB* promoter activity with co-expression of pLG339*aggR* 042 mutants targeting the N-terminal domain

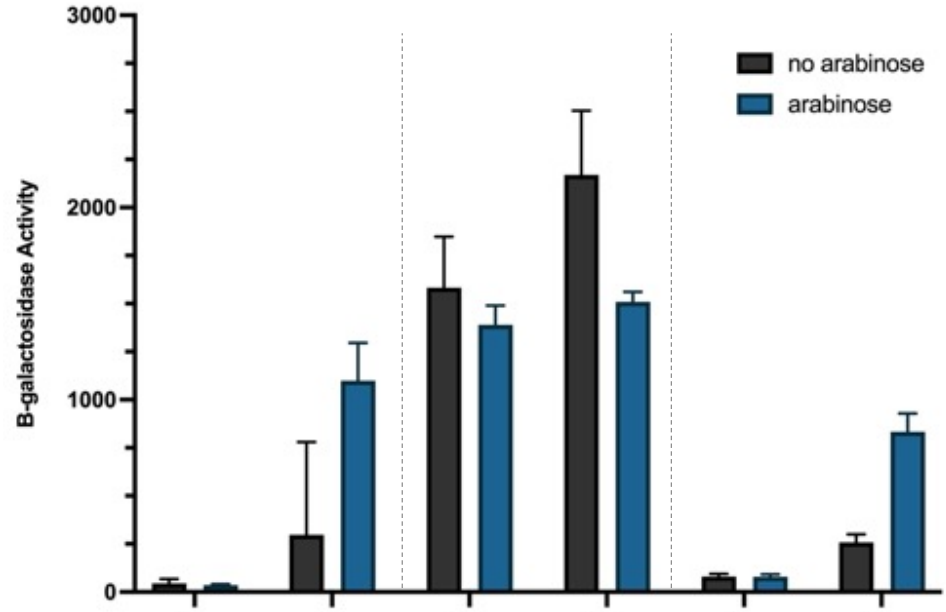
The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB*, in addition to either pBAD30 (empty vector) or pBAD*aggR* and carrying either pLG339, pLG339*aggR*, pLG339*aggR-Y92C* or pLG339*aggR-Y92R*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 80) = 160.8$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pBAD*aggR* 042 and carrying either pLG339 empty vector, pLG339*aggR-Y92C*, or pLG339*aggR-Y92R* in the presence and absence of arabinose ($p<0.05$).
* Expression of the *aggR* gene is under the control of the *aggR* promoter



activator in pBAD	–	AggR 042	–	AggR 042	–	AggR 042	–	AggR 042
activator in pLG339*	–	–	AggR 042	AggR 042	I14T	I14T	N16D	N16D

Figure 5.21 Analysis of *afaB* promoter activity with co-expression of pLG339*aggR* 042 mutants guided by Rns

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB*, in addition to either pBAD30 (empty vector) or pBAD*aggR* and carrying either pLG339, pLG339*aggR*, pLG339*aggR-I14T* or pLG339*aggR-N16D*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(15, 80) = 259.8$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pBAD*aggR* 042 and carrying either pLG339 empty vector, pLG339*aggR-I14T*, or pLG339*aggR-N16D* in the presence and absence of arabinose ($p<0.05$).
* Expression of the *aggR* gene is under the control of the *aggR* promoter



activator in pBAD	–	AggR 042	–	AggR 042	–	AggR 042
activator in pLG339*	–	–	AggR 042	AggR 042	E191A	E191A

Figure 5.22 Analysis of *afaB* promoter activity with co-expression of pLG339*aggR* 042 mutants targeting HTH1

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB*, in addition to either pBAD30 (empty vector) or pBAD*aggR* and carrying either pLG339, pLG339*aggR* or pLG339*aggR-E191A*. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 60) = 84.34$). A post-hoc Tukey's HSD test showed that there was a significant difference between promoter activities in cells containing pBAD*aggR* 042 and carrying either pLG339 empty vector or pLG339*aggR-E191A* ($p<0.05$).
* Expression of the *aggR* gene is under the control of the *aggR* promoter

throughout the *aggR* gene (Figure 5.3). In Table 5.1, AggR mutants that I have constructed and assayed are organised by the location of the change in AggR. The changes in the 33 mutants that I have investigated are located throughout AggR, both in the N- and C-terminal domains, and I found changes that maintained, improved, and stopped activation at the tested target promoters. When mutant defective AggR was co-expressed with wild-type AggR, interference with AggR-dependent activation was consistently below 2-fold. Research with other AraC family members, and the predicted structure of AggR, has guided my search for defective AggR mutants, and targeting key residues can directly interfere with DNA binding or contacts with RNAP.

For analysis of the AggR N-terminal domain, I utilised mutants that stopped AggR-mediated activation at target promoters, *afaB* (Y92C, and Y92R, Figure 5.5 and Figure 5.8) and *agg4D* (Y92C and Y92R, Figure 5.6). AggR-N168D, which is located at the end of the N-terminal domain, does not stop activation (Figure 5.8). Using my new 2-plasmid genetic system to introduce AggR mutants in *trans* to wild-type AggR, I showed that the changes at residue 92 reduced activity less than 2-fold when co-expressed with wild-type AggR, indicating that there was interference but not strong *trans*-dominance.

When I assayed these mutants with the *nlpA* promoter, there was a proportional reduction in activity compared to wild-type AggR. The use of the *nlpA* promoter provides a test for whether mutant AggRs are positive control mutants, thus still able to bind to DNA, even though defective for activation. I conclude that AggR-Y92A and AggR-Y92S are positive control mutants.

As discussed in Section 4.2.1, Rns from ETEC and AggR from EAEC strain 042

Table 5.1 AggR mutants and their characteristics

Location	Mutation	Activation at *afaB*?	Reduction when co-expressed	Source and predicted function
N-terminal	I14T	no	2-fold	Rns ETEC, 5.4.1
	N16D	no	2-fold	Rns ETEC, 5.4.1
	Y92A, Y92C, Y92R, Y92S	no	Y92C: 1.4-fold Y92R: 1.8-fold	5.3.1, 5.3.2
	N168D	yes	N/A	5.3.2
HTH1	I183N	no	1.2-fold	VirF *S. flexneri,* 5.4.5
	E191A	no	1.6-fold	Conserved, 5.4.4
	I192A	yes	1.8-fold	MelR *E. coli,* 5.4.2 DNA binding
	R195A	no	1.4-fold	Conserved, 5.4.4
	K196A	no	1.5-fold	Conserved, 5.4.4
	R197A	no	1.4-fold	Conserved, 5.4.4
	RKR195-197AAA	no	1.3-fold	Conserved, 5.4.4
	E199A	yes	1.5-fold	MelR *E. coli,* 5.4.2 DNA binding

	S200A	yes	1.7-fold	MelR *E. coli,* 5.4.2 DNA binding
HTH2	Y226A	yes	1.3-fold	MelR *E. coli,* 5.4.3 σ70 interaction
	Q227A	yes	1.6-fold	MelR *E. coli,* 5.4.3 σ70 interaction
	Q227STOP	no	1.1-fold	VirF *S. flexneri,* 5.4.5
	Q230A, Q230L, Q230M, Q230R	yes	Q230R: 1.2-fold	MelR *E. coli,* 5.4.3 σ70 interaction
	QIS230-232KIP	no	~1-fold	5.3.2
	M234A	yes	1.3-fold	MelR *E. coli,* 5.4.3 σ70 interaction
	T240A	yes	1.5-fold	MelR *E. coli,* 5.4.2 DNA binding
	Y242A	no	1.5-fold	MelR *E. coli,* 5.4.2 DNA binding
	R245A	yes	1.4-fold	MelR *E. coli,* 5.4.2 DNA binding
	R245K	yes	1.1-fold	Conserved, 5.4.4
	K249A	yes	1.1-fold	Conserved, 5.4.4

C-terminal	K263STOP	yes	~1 -fold	5.3.2
N-terminal and linker	D86N-L221S	no	1.6-fold	5.3.1
C-terminal and HTH2	Q230R-K263STOP	no	1.5-fold	5.3.1

have high sequence identity, and, in Figure 5.10, I show the alignment of both transcription factors with the predicted secondary structure. Two Rns variants, with changes in the N-terminal domain, were identified as positive control mutants (Basturea *et al.*, 2008), and I introduced the corresponding changes into AggR: AggR-I14T and AggR-N16D prevented activation at the *afaB* and *agg4D* promoters; both mutants also interfered with wild-type AggR, around 2-fold.

The mutational analysis of the AggR N-terminus indicates that mutating this region does not result in a strong *trans*-dominant mutant, this could be due to the predicted function of this domain: effector binding or multimerization. As I have shown that AggR does not require an external signal for expression of *aggR* and a single DNA-binding site for AggR is required for activation, I would conclude that the changes in this region might interfere with the proper folding of the AggR protein or with holding the C-terminal domain in the correct conformation. The changes at residues I14, N16 and Y92 resulted in AggR mutants defective for AggR-dependent activation at target promoters, this indicates that the N-terminal domain is vital for AggR function, and, in particular, these three residues.

Interestingly, fold-difference when AggR-I14T and AggR-N16D were expressed in *trans* with wild-type AggR was greater at the *afaB* promoter than the *agg4D* promoter (Figure 5.11). However, there was 1.4-fold difference at the *agg4D* promoter, indicating that there was some interference. In Section 3.4, I discussed AggR binding affinities at different promoters and concluded that the affinity was greater at the *agg4D* promoter than at the *afaB* promoter. This might explain why the interference at each promoter is different, perhaps the greater the affinity, the lesser the *trans*-acting effect, i.e., the lack of interference from co-expression is

more pronounced at *agg4*D than *afaB*. This could explain why it does not seem possible to isolate a *trans*-dominant AggR mutant and this lack of *trans*-dominance is just more pronounced at *agg4D* than *afaB*.

Members of the AraC family of transcriptional regulators have been studied to different extents to determine the importance of the HTH motifs in the C-terminal domain and the effect of altering key residues: HTH1 for recognition of specific promoter DNA and HTH2 for contacting RNAP (Gallegos *et al.*, 1997). MelR is an AraC-family transcription regulator controlled by melibiose. A mutational study of MelR identified six mutations in HTH1 and HTH2 that impacted DNA binding, three had a big effect and three had a small effect (Grainger *et al.*, 2004). As a DNA binding mutant, one would expect to find a mutant that binds more strongly to the DNA, thus blocking wild-type AggR from binding. In Figure 5.12, the amino acid sequences for part of the HTH motifs from MelR and AggR are aligned showing the corresponding residues in AggR. AggR-Y242A, located in HTH2, was the only mutant that prevented AggR-activation at *afaB* and *agg4D*. AggR-I192A, in HTH1, decreased expression levels slightly compared to wild-type AggR. Both changes in AggR corresponded to MelR changes that had a big effect on DNA binding. Interestingly, I192 also corresponds to a residue in Rns important for DNA binding (Mahon *et al.*, 2010). When co-expressed with wild-type AggR, none of these DNA binding mutants were *trans*-dominant and the greatest interference was seen with AggR-I192A. Though these changes are located in both of the HTH motifs for AggR and MelR, the relative importance for contacting DNA differs for each amino acid.

The MelR study also identified residues that interact with RNAP. Interestingly, I had previously obtained a random mutant, AggR-Q230R-K263STOP, that prevented AggR-activation and interfered with wild-type AggR when expressed in *trans* (Figure 5.5). The glutamine at position 230 in HTH2 had been identified in MelR as interacting with the RNAP σ^{70} domain. Thus, I changed residue 230 to different amino acids but I did not see any negative effects on AggR function (Figure 5.8 and Figure 5.9). One of these changes, glutamine at position 230, had been previously examined (see Section 5.3.2). Taken together, my data show that changes to this amino acid have no effect on activation by AggR. This position is not as important in AggR compared to MelR. Therefore, I conclude that AggR operates differently to MelR and, while this amino acid is in HTH2, which, in MelR, interacts with RNAP σ70, its importance is negligible in AggR.

Subsequently (Figure 5.14), I analysed mutations in *aggR* that correspond to the other locations in MelR that interact with the RNAP σ^{70} domain. The four changes, including Q230, were located just before and within the first α-helix of HTH2 (Grainger *et al.*, 2004). I focused on the other three changes, corresponding to Y226, Q227 and M234 in AggR: none of these stopped AggR-activation and they were not *trans*-dominant. This was not entirely unexpected, considering the result with AggR-Q230R, when I had concluded that MelR and AggR are too different in the way that they function for them to behave similarly when mutated. I also measured the effects of changing conserved charged residues that are adjacent or neighbouring the residues suggested from MelR. In Figure 5.16, I showed that mutating these residues that are located in the second α-helix in HTH1 had a big effect on AggR-activity. However, none of these AggR mutants were strongly

trans-dominant. Similarly, targeting two charged residues in the second α-helix of HTH2 had but small effects on activation by wild-type AggR. This indicates that, despite targeting regions with amino acids with similar properties, and stopping AggR-activation at the *afaB* promoter, there were no strong *trans*-dominant effects. From this, I conclude that simply changing residues that are conserved or important for interacting with DNA or RNAP is not sufficient to make a *trans*-dominant mutant.

In a final effort to find a *trans*-dominant AggR mutant, I looked at a study that found two *trans*-dominant mutants in *Shigella flexneri* VirF (Porter and Dorman, 2002). I chose this because VirF belongs to the same AraC subgroup as AggR, though they have lower sequence identity than AggR and Rns (Munson and Scott, 1999; Barbieri *et al.*, 2001). One of the mutants identified by Porter and Dorman carried a single change in the first α-helix of HTH1 whilst the other lacked HTH2 due to a premature stop codon. The corresponding changes in AggR, however, did not result in *trans*-dominance despite both changes abolishing AggR-mediated activation (Figure 5.19). Neither mutant could stop wild-type AggR from activating at the *afaB* promoter.

While expression of AggR did not require any external signal, I consistently saw that expression levels were higher with pLG339*aggR* 042 than when expression of AggR from pBAD*aggR* was induced by arabinose. Therefore, I wondered if the level of expression of wild-type AggR from pLG339*aggR* 042 was masking *trans*-dominant effects following arabinose-induced expression of mutant AggRs. Data in Figure 5.20, Figure 5.21 and Figure 5.22 show that AggR-I14T, AggR-N16D, AggR-Y92C, AggR-Y92R and AggR-E191A clearly interfere to some extent with

AggR-dependent activation at the *afaB* promoter. Table 5.2 shows the levels of reduction in activity when wild-type AggR was co-expressed with AggR mutants, both when the mutants were under the control of the *aggR* promoter and induced by arabinose. The results indicate that, for most of the AggR mutants, the reduction in activity was more pronounced when the mutant AggR was introduced in pLG339*aggR* 042. The explanation for this phenomenon is complicated, though it is clear that effects are similar regardless of any masking by overexpression.

In the study of VirF *trans*-acting mutants, it was considered that expression reductions of 50% or more indicated *trans*-dominance, this was identified as a strong effect (Porter and Dorman, 2002). However, I conclude that a *trans*-dominant effect is not strong if it is only 50%. There is little work into classifying the exact conditions of *trans*-dominance, but I would expect stronger reductions in expression in order to class an AggR mutant as strongly *trans*-dominant. It has been hypothesised that the VirF mutants targeting HTH1 and HTH2 are *trans*-dominant because wild-type VirF binds to DNA as a monomer and then forms a dimer with the VirF mutant (Di Martino *et al.*, 2016), as is the case with MelR (Bourgerie *et al.*, 1997). For AggR, the largest reductions were observed with AggR-I14T and AggR-N16D; these were the greatest levels of *trans*-action (Table 5.1).

It had previously been claimed that many of the changes in Rns result in similar disruption to activation function as seen in AggR (Mahon *et al.*, 2010). Therefore, as these transcription factors are so similar, and I have shown that AggR mutants

Table 5.2 AggR mutants in pLG339 and pBAD30

Location	**Mutation**	**Reduction with co-expression of*:**		**Section**
		pLG339*aggR* 042	pBAD*aggR*	
N-terminal	I14T	2-fold (Figure 5.10)	2.4-fold (Figure 5.21)	5.4.1, 5.5
	N16D	2-fold (Figure 5.10)	2.2-fold (Figure 5.21)	5.4.1, 5.5
	Y92C	1.4-fold (Figure 5.4)	1.9-fold (Figure 5.20)	5.3.1, 5.5
	Y92R	1.8-fold (Figure 5.4)	1.8-fold (Figure 5.20)	5.3.1, 5.5
HTH1	E191A	1.6-fold (Figure 5.16)	1.7-fold (Figure 5.22)	5.4.4, 5.5

* Reduction of expression levels at *afaB* promoter with co-expression of AggR and AggR mutants.

are not *trans*-acting, I would conclude that both function differently to normal Class II activation, where the activator binds to the target promoter overlapping the -35 element and then recruits the RNAP; instead, I propose that AggR, and by extension Rns, functions by pre-recruitment, though more work would be required to confirm this. I suggest that it may not be possible to make a *trans*-dominant AggR mutant because AggR does not function as a normal Class II activator, and functions by pre-recruitment. In pre-recruitment, AggR first contacts RNAP and the AggR-RNAP complex then scans the DNA for AggR-DNA binding sites (Taliaferro *et al.*, 2012; Griffith *et al.*, 2002; Griffith and Wolf, 2004; Martin *et al.*, 2002). The I14T and N16D mutations in both Rns and AggR had strong negative effects on expression levels, but neither mutation in either transcription factor was *trans*-dominant. In Section 3.4, I discussed that AggR could be functioning by first binding to RNAP then scanning DNA to find a dependent promoter, which is also the mechanism for Rob, SoxS and MarA (Taliaferro *et al.*, 2012; Griffith *et al.*, 2002; Griffith and Wolf, 2004; Martin *et al.*, 2002).

In summary, in this Chapter, I have described many changes in AggR that stopped AggR-activation at target promoters: in the N-terminal domain, I14T, N16D and Y92A (C, S or R); in the C-terminal domain in HTH1, I183N, E191A, R195A, K196A and R197A; and in HTH2, Q227A (and a premature stop codon) and Y242A. The results have indicated the importance of both the N-terminal and C-terminal domains for AggR function. However, I have only seen minimal negative *trans*-dominance when AggR mutants were co-expressed with wild-type AggR. Additionally, I have shown that simply overexpressing wild-type AggR,

from both pBAD*aggR* and pLG339*aggR* 042, also reduced expression levels. This indicates that an excess of AggR, wild-type or mutant, is interfering with AggR-activation rather than there being an actual *trans*-dominant effect. Thus, I conclude that AggR functions by pre-recruitment and it is not possible to make a true strong *trans*-dominant AggR mutant.

Chapter 6 The Regulation of AggR

6.1 Introduction

It has been suggested that the regulatory network for AggR is formed of an interlinked positive and negative feedback loop: a similar network is also responsible for controlling MarA (Garcia-Bernardo and Dunlop, 2013). Expression of the *aggR* gene is thought to be driven by a bi-stable switch, dependent on positive and negative feedback, which has two stable states: 'on' and 'off' (Raj and van Oudenaarden, 2008; Lebar *et al.*, 2014; Abdelwahab *et al.*, 2021). Due to autoregulation, basal expression of AggR can activate further expression and, if level of protein reaches a critical threshold, this 'flips the switch' to the 'on' state. Presumably, this metaphorical, and not physical, 'switch flipping' could occur due to fluctuations in gene expression (Raj and van Oudenaarden, 2008) and stochastic variations of repressor levels (Lebar *et al.*, 2014).

6.2 Aar – the AggR regulator

The Nataro lab have suggested that the activity of AggR is not dependent on a ligand and is in fact modulated by a small protein called Aar (Mickey and Nataro, 2020). Aar, which stands for AggR-activated regulator, is a negative regulator of AggR and was the first AraC Negative Regulator (ANR) to be identified (Santiago *et al.*, 2014). The large ANR family of bacterial gene regulators act as anti-activators that down-regulate AraC partners in clinically important Gram-negative bacteria (Santiago *et al.*, 2014; Santiago *et al.*, 2016). ANRs do not have DNA binding ability and are predicted to bind directly to their cognate master

regulators, to the central linker region, thus blocking dimerization function and preventing DNA binding (Santiago *et al.*, 2016; Hodson *et al.*, 2017).

ANR members are organised into 5 phylogenetic clusters and share common features such as a low molecular mass (<10 kDa) and certain highly conserved residues (Santiago *et al.*, 2016). Their secondary structure is predicted to consist of 3 highly conserved α-helix motifs that span the length of the protein, and these motifs are conserved within each phylogenetic cluster (Santiago *et al.*, 2016).

The *aar* gene is located near the *aggR* gene on the pAA2 virulence plasmid and is one of the most up-regulated genes in the AggR regulon (Santiago *et al.*, 2014). Aar is conserved and prevalent across EAEC strains. The Aar protein is 63 amino acids, 7.23 kDa, formed of three α helical domains for oligomerisation. Aar does not appear to have DNA binding ability; thus, Aar is thought to bind directly to AggR to inhibit binding to target promoter DNA (Hodson *et al.*, 2017; Santiago *et al.*, 2014). Aar also regulates the expression of other proteins, including the global regulatory protein, H-NS, which silences transcription of some AraC regulators (Mickey and Nataro, 2020; Santiago *et al.*, 2017; Morin *et al.*, 2010). Aar has been shown to bind directly to AggR and to H-NS (Mickey and Nataro, 2020).

Aar homologs include anti-regulators encoded by ETEC, *Vibrio cholerae, Citrobacter rodentium* and *Salmonella enterica*, in fact, these homologs have been shown to rescue defective *aar* mutants in EAEC 042 (Santiago *et al.*, 2017). The Rns/CfaD negative regulator from ETEC, known as Cnr or Rnr-1, actually down-regulates AggR when overexpressed (Santiago *et al.*, 2014; Santiago *et al.*, 2017; Hodson *et al.*, 2017). As with the relationship between Cnr (Rnr-1) and auto-activating CfaD in ETEC, Aar prevents excess production of AggR, indirectly

suppressing virulence gene expression. In ETEC, Cnr (Rnr-1) is required during the late stages of host infection to facilitate the transition from the host intestine to the external environment (Hodson *et al.*, 2017).

The control of EAEC virulence relies on an interplay between AggR and Aar: the feed-forward up-regulation of AggR expression is conserved, but the downregulation by Aar is not conserved. A study by Radwa Abdelwahab, investigating EAEC strains isolated in Egypt, E36 and E42, showed that both strains have characteristic AAF fimbrial genes. Strain E36 encodes the *aar* gene, but strain E42 does not. Comparison of Aar from E36 and 042 showed that inducing overexpression of E36 Aar decreased biofilm production in EAEC strain 042, whereas 042 Aar had little effect. Sequence alignments showed that the ribosome binding sites (RBS) are different (Abdelwahab *et al.*, 2021). Recall that the RBS contains a domain called the Shine-Dalgarno sequence which helps to recruit the ribosome to the mRNA translation initiation site (Chen *et al.*, 1994). 042 has a suboptimal Shine-Dalgarno sequence, therefore, the RBS was altered to an improved Shine-Dalgarno sequence in 042 *aar* to generate the *aar*$^{+}$ gene: Aar^{+} completely interrupted EAEC 042 biofilm formation. Thus, it was concluded that the poor RBS resulted in lower levels of Aar expression in EAEC strain 042 (Abdelwahab *et al.*, 2021).

6.2.1 The proposed mechanism

Since Aar+ effectively disrupted biofilm formation, I wanted to determine whether Aar or Aar^{+} would similarly disrupt AggR activity when AggR was under the control of its own promoter (as in pLG339*aggR* 042). Hence DNA fragments encoding Aar and Aar^{+} were each cloned into the pBAD30 vector using primers

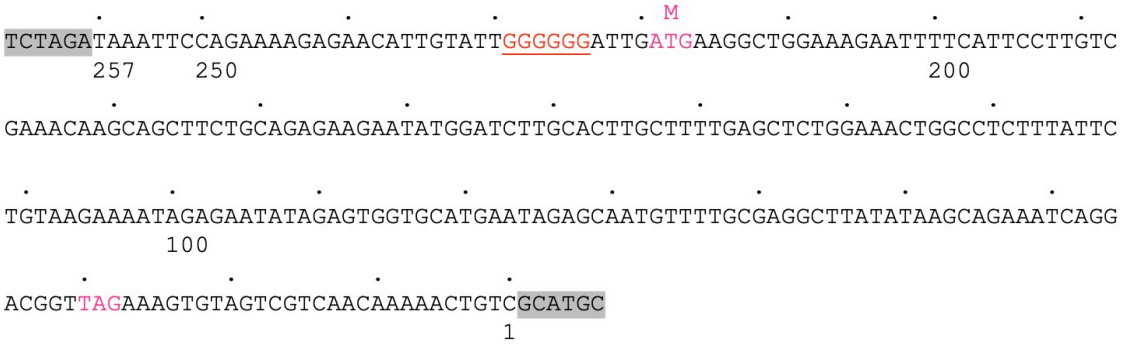
that introduced an upstream XbaI restriction site and a downstream SphI site, so that the *aar* and *aar*$^{+}$ genes were under the control of the arabinose-inducible *araBAD* promoter (Figure 6.1). Then, to measure effects on AggR, I transformed pBAD*aar* and pBAD*aar*$^{+}$ into *E. coli* K-12 BW23115 Δ*lac* cells containing pRW50*afaB* with either pLG339 empty vector or pLG339*aggR* 042. Cells were grown with shaking in LB to mid-exponential phase, and β-galactosidase levels were measured in cell lysates to determine promoter activity. The data in Figure 6.2A show that there is a very small decrease in β-galactosidase levels when expression of Aar was induced, from 448 to 384 units (one-way ANOVA, $p<0.0001$; Tukey test, $p>0.05$). There is a greater decrease with induction of Aar^{+} (which has the improved Shine-Dalgarno sequence), from 469 to 319 units, approximately a 1.5-fold reduction (Tukey test, $p>0.05$). Nonetheless, neither Aar nor Aar^{+} greatly reduced β-galactosidase levels which indicates that neither is effective at stopping AggR-dependent activation at the *afaB* promoter.

I then transformed pBAD*aar* and pBAD*aar*$^{+}$ into *E. coli* K-12 BW23115 Δ*lac* cells containing pRW50*agg4D* with either pLG339 empty vector or pLG339*aggR* 042. β-galactosidase levels were measured in cell lysates to determine promoter activity. The data in Figure 6.2B show that there is a small decrease in β-galactosidase levels when Aar expression was induced (611 to 452 units; one-way ANOVA, $p<0.0001$; Tukey test, $p>0.05$), but a greater decrease with induction Aar^{+}, over 2-fold (884 to 421 units; Tukey test, $p<0.05$).

To conclude, altering the Shine-Dalgarno sequence to create *aar*$^{+}$ improves the suppression of AggR activity. My data, together with previous research by Radwa Abdelwahab, suggest that Aar is not an effective repressor of AggR.

A

pBAD*aar*



B

pBAD*aar*$^{+}$

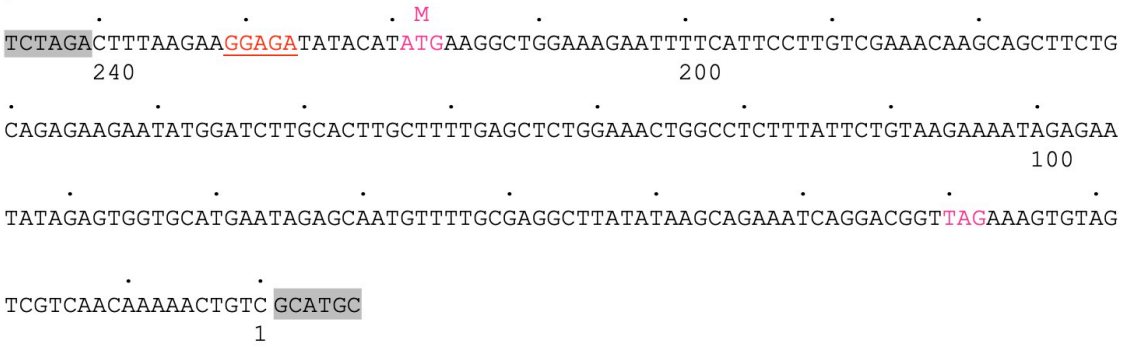


Figure 6.1 DNA sequences of *aar* and *aar*$^{+}$ genes cloned into pBAD30

The DNA sequence of the *aar* (A) and *aar*$^{+}$ (B) genes cloned into pBAD30 utilising XbaI and SphI restriction sites, highlighted in grey. The translation start and stop codons are highlighted in pink. The Shine-Dalgarno sequences are highlighted in red.

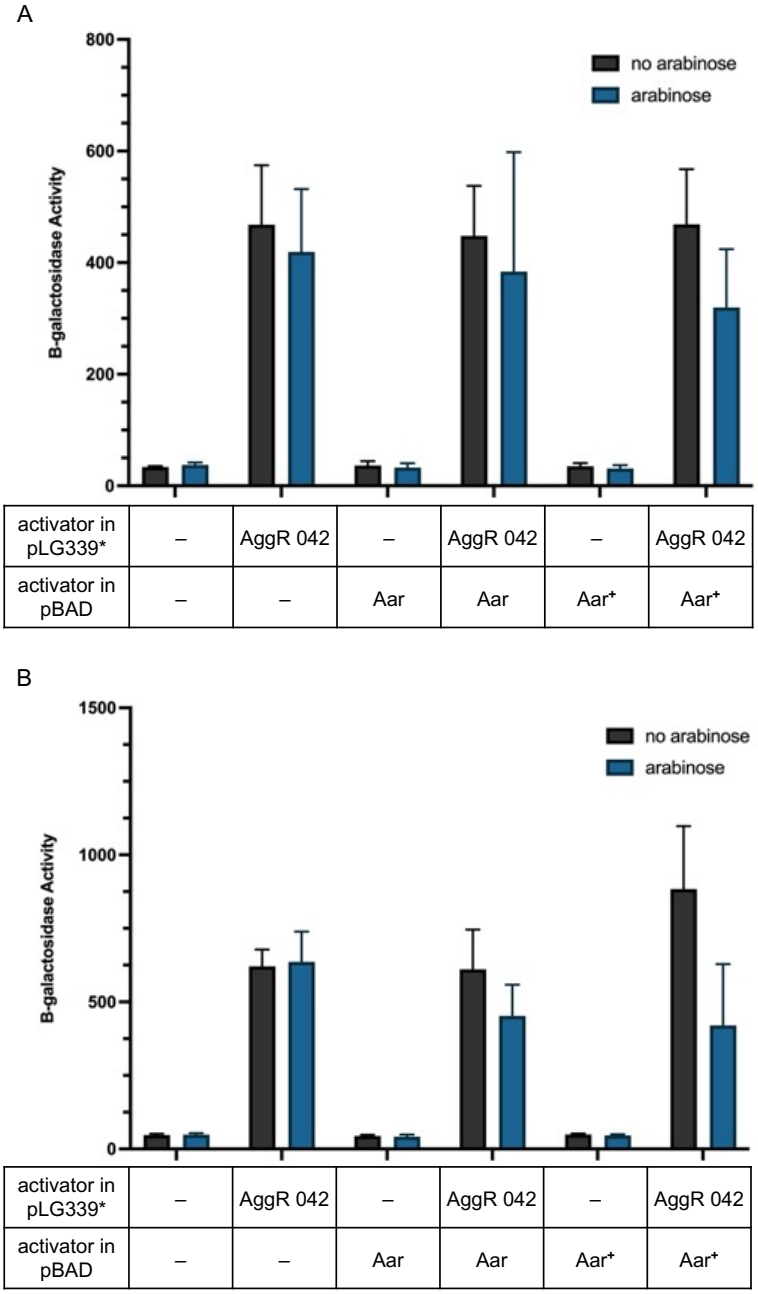


Figure 6.2 Analysis of *afaB* and *agg4D* promoter activity with AggR and either Aar or Aar$^+$

The figure illustrates measured β-galactosidase activities in *E. coli* K-12 BW25113 Δ*lac* carrying the recombinant *lac* expression plasmid pRW50*afaB* (A) or pRW50*agg4D* (B), and either pLG339 (empty vector) or pLG339*aggR* 042 with pBAD30, pBAD*aar* or pBAD*aar*$^+$. Cells were grown in LB in the absence (black bars) or presence (blue bars) of 0.2% (*w/v*) arabinose. The β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass. The results are the calculated means of three independent determinations, and the standard deviations are shown for each data point.

* Expression of the *aggR* gene is under the control of the aggR promoter
    A. Cells containing pRW50*afaB* with either pLG339 (empty vector) or pLG339*aggR* 042, in addition to pBAD30, pBAD*aar* or pBAD*aar*$^{+}$. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 60) = 30.25$). A post-hoc Tukey's HSD test showed that there was no significant difference between cells carrying pLG339*aggR* 042 and containing pBAD*aar* or pBAD*aar*$^{+}$ in the presence and absence of arabinose ($p>0.05$).
    B. Cells containing pRW50*agg4D* with either pLG339 (empty vector) or pLG339*aggR* 042, in addition to pBAD30, pBAD*aar* or pBAD*aar*$^{+}$. A one-way ANOVA was calculated using the promoter activities, showing the analysis was significant ($p<0.0001$, $F(11, 24) = 26.36$). A post-hoc Tukey's HSD test showed that there was no significant difference between cells carrying pLG339*aggR* 042 and containing pBAD*aar* in the presence and absence of arabinose ($p>0.05$), however, the difference was significant for cells containing pBAD*aar*$^{+}$ ($p<0.05$).

6.3 Single cell analysis

Thus far, my research has focused on 'ensemble' measurements of AggR activity in bacterial populations by measuring promoter activity or biofilm formation. I have shown previously (5.2) that expression of AggR does not require any external signal, and so, I wanted to investigate whether expression was stochastic; whether there are threshold levels that are reached at different times to 'switch' gene expression 'on' and 'off' (Ambriz-Avina *et al.*, 2014). However, cell-to-cell variation, and thus the stochastic nature of gene expression, cannot be determined using 'ensemble' experiments, such as β-galactosidase assays. Therefore, I needed to analyse gene expression at the single-cell level (Longo and Hasty, 2006).

6.3.1 Flow cytometry

Fluorescence flow cytometry is a high-throughput method to study single-cell gene expression in bacteria (Galbusera *et al.*, 2020). A flow cytometer functions by passing single cells through a channel where a light beam impacts bacterial cells. Light is absorbed, and the resulting scatters, at right angles or in the forward direction (forward scatter), and cell fluorescence are then measured (Ambriz-Avina *et al.*, 2014; Galbusera *et al.*, 2020). There are three major parts of a flow cytometer: fluidics, to help ensure cells pass through one at a time; optics, the light source to impact on a single cell and scatter light, which is then collected and filtered through a band pass filter to a photomultiplier tube (PMT); and electronics, the signal collected by the PMT is used by photo detectors to produce current pulses which are then amplified, converted to voltage and then converted to numbers by analogue signal processing electronics (Ambriz-Avina *et al.*,

2014). Each measured 'event' typically corresponds to measurement of a single cell and every second hundreds of events are generated, analysis of the signals is then required to generate plots and data for interpretation (Galbusera *et al.*, 2020; Ambriz-Avina *et al.*, 2014).

6.3.2 The potential stochasticity of *aggR* expression

To measure gene expression at the single cell level, flow cytometry was used to detect GFP fluorescence, utilising an AggR-activated target promoter fused to the *gfp* gene (Bryant *et al.*, 2014). Thus, the *aafD100* regulatory region (Figure 3.3C) was cloned into the GFP expression plasmid, pRW400 (Figure 6.3A), upstream of the *gfp* gene, using primers containing an upstream EcoRI restriction site and a downstream HindIII site (Figure 6.3B).

For flow cytometry, fluorescent dyes are commonly used to distinguish cells from background debris. Here, the fluorescent cell marker SYTO84 was used, which emits orange fluorescence when intercalated with DNA. To accurately measure fluorescence inside cells, and separate from background noise, a gating strategy was used (Figure 2.1). Briefly, SYTO84 fluorescence (channel YL1-H) was plotted against forward scatter (FSC-H) on a dot plot and the $SYTO84^{+}$ population was then gated, highlighted by a polygon, and labelled in orange. The $SYTO84^{+}$ population was then plotted on a dot plot with: $SYTO84^{+}$ fluorescence (YL1-H) against GFP fluorescence (BL1-H). A quadrat gate was added to separate the populations based on $SYTO84^{+}$ or $SYTO84^{-}$ and GFP^{+} or GFP^{-}.

6.3.3 Regulation by AggR

The pLG339*aggR* 042 and pLG339 (Figure 3.8) recombinant plasmids were transformed into *E. coli* K-12 BW25113 Δ*lac* cells containing pRW400*aafD100*.

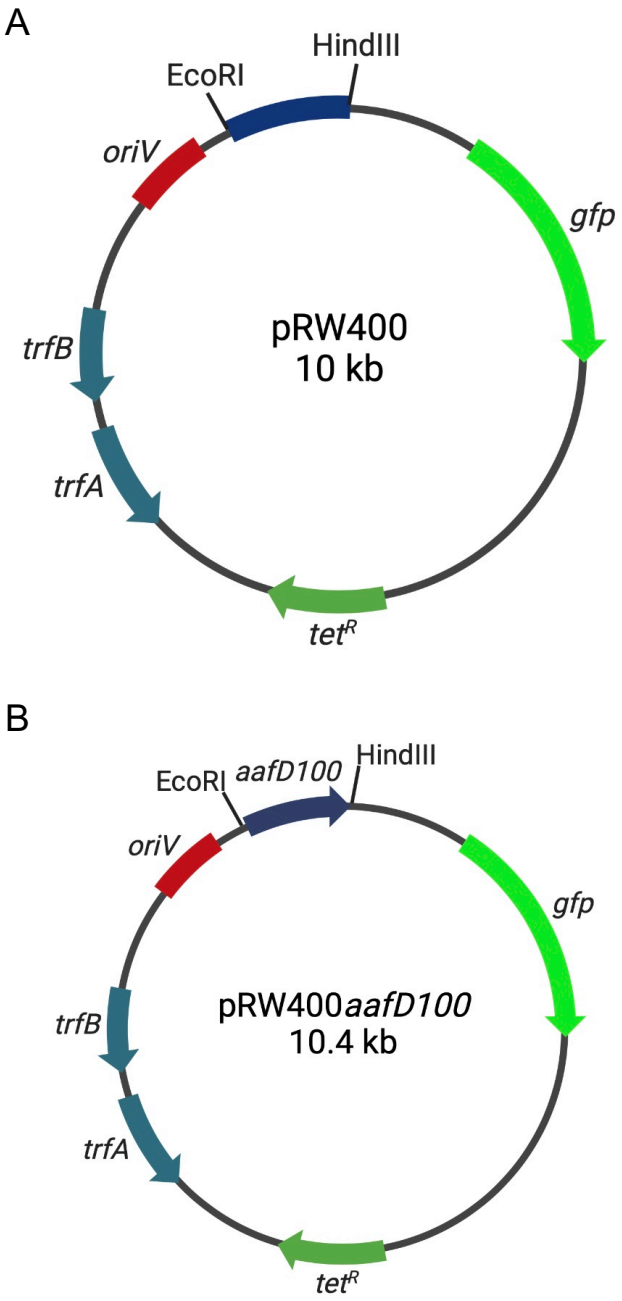


Figure 6.3 The pRW400 and pRW400*aafD100* low copy number vectors

A. The figure shows a plasmid map of the pRW400 *gfp* expression vector, with EcoRI and HindIII restriction sites upstream of *gfp* that were used to clone promoter fragments. The plasmid carries a tetracycline resistance determinant (*tet*^{R}). Created with BioRender.com.
B. The figure shows a plasmid map of the pRW400 *gfp* expression vector with the *aafD100* promoter cloned upstream of *gfp* using the EcoRI and HindIII restriction sites. The plasmid carries a tetracycline resistance determinant (*tet*^{R}). Created with BioRender.com.

Cells were grown in LB with shaking until mid-exponential phase and cells were then washed with HBS; each sample had an unstained and SYTO84-stained duplicate.

The results, illustrated in Figure 6.4A and Figure 6.4B, show an example of one sample of unstained cells carrying pLG339*aggR* 042 to indicate the gating strategy for the experiment to separate the background from the events (this is clear in the stained samples). Figure 6.4B and Figure 6.4D show the combined data from the biological replicates. The gate is outlined by a polygon and labelled SYTO84^{+} in orange, more than 10,000 events were collected in this gate in each of the biological replicates. The panel shows the progression of separating according to stain and GFP fluorescence, firstly, the forward-scatter FSC-H against SYTO-84 (YL1-H) dot plot indicates the separation of events from background debris (Figure 6.4A and Figure 6.4C). Then, in Figure 6.4B and Figure 6.4D, I took the SYTO84^{+} gate and plotted GFP fluorescence (BL1-H) against SYTO84 (YL1-H) with a quadrat gate to indicate the percentage of events identified with the combinations of unstained (SYTO-84^{-}), stained (SYTO-84^{+}), no GFP signal (GFP^{-}) and GFP signal (GFP^{+}). The plots indicate that the gating strategy has successfully identified the background debris: unstained and no GFP fluorescence detected.

The data in Figure 6.5A, Figure 6.5B and Figure 6.5C show an example of a SYTO-84^{+} pLG339*aggR* 042 sample, the panels are organised as in Figure 6.4. Briefly, the dot-plot in Figure 6.5A and Figure 6.5D indicate the SYTO-84^{+} cells in the polygon gate separated from background debris, this gate was then used to create the middle quadrat dot-plot, with the quadrats: top left (Q1), SYTO-84^{+}

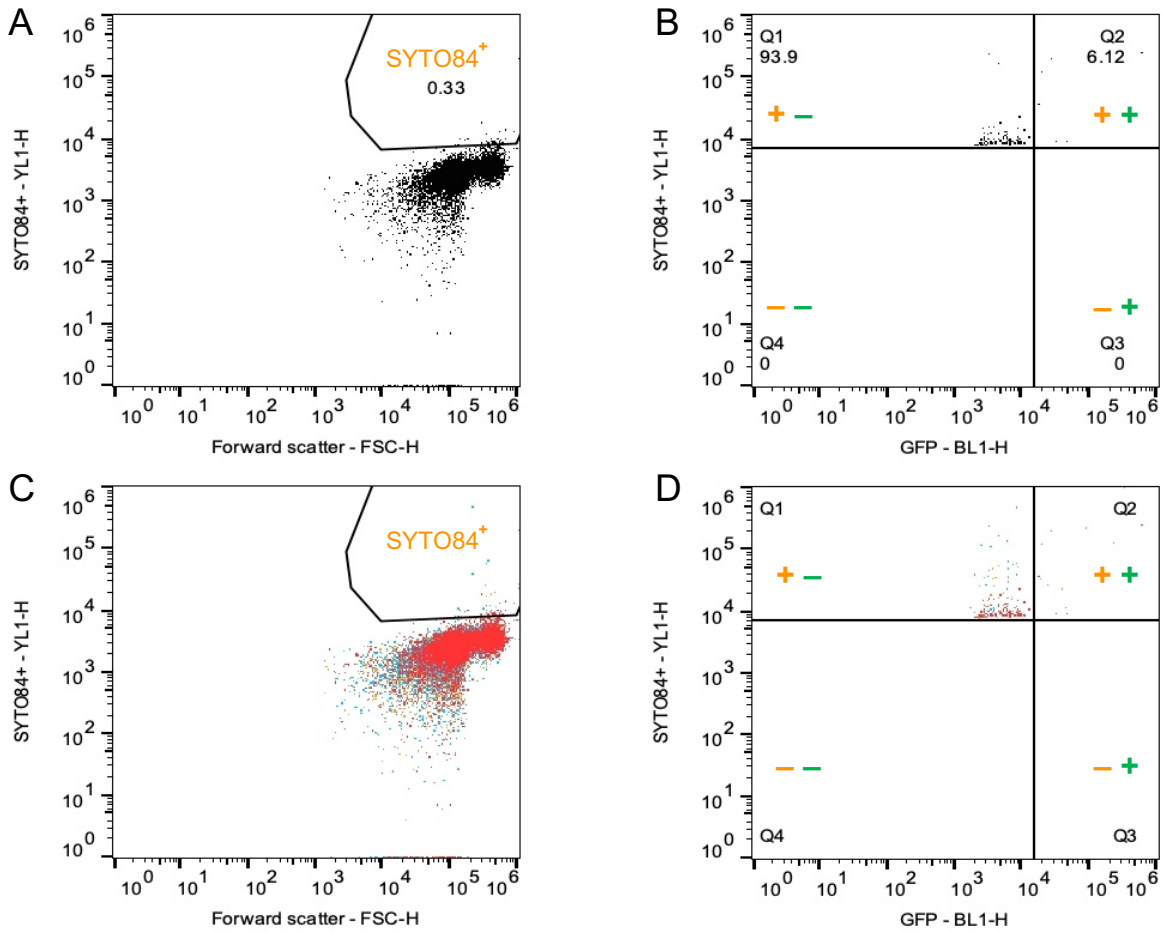


Figure 6.4 Flow cytometry plots of unstained control samples

The flow cytometry plots of the negative control unstained BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042. In (A) and (C), cells were separated from background using channels YL1-H against FSC-H, highlighted by a polygon, and labelled SYTO84⁺ in orange. In (B) and (D), cells in the SYTO84⁺ gate were then analysed using channels YL1-H against BL1-H and a quadrat gate separated the SYTO84 (labelled in orange) and GFP signals (labelled in green): Q1 for SYTO-84⁺ GFP⁻ (+ –); Q2 for SYTO-84⁺ GFP⁺ (+ +); Q3 for SYTO-84⁻ GFP⁺ (– +); and Q4 for SYTO-84⁻ GFP⁻ (– –). Data was analysed and plots were made using FlowJo.

A. An example of one sample of BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042. The percentage of events in the polygon gate of SYTO84⁺ is indicated.
B. An example of one sample of BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042. The percentage of events in each of the quadrat gates is indicated.
C. Three samples of BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042 combined, each sample highlighted in different colours.
D. Three samples of BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042 combined, each sample highlighted in different colours.

GFP$^-$; top right (Q2), SYTO-84$^+$ GFP$^+$; bottom right (Q3), SYTO-84$^-$ GFP$^+$; and bottom left (Q4), SYTO-84$^-$ GFP$^-$ (Figure 6.5B and Figure 6.5E). Finally, the fluorescence of the SYTO84$^+$ GFP$^-$ and SYTO84$^+$ GFP$^+$ populations were plotted on a logarithmic histogram to indicate GFP fluorescence intensity with distinct peaks (Figure 6.5C and Figure 6.5F). The presence of distinct peaks, with a shift from the GFP$^-$ to the GFP$^+$ peak, indicates that the main population does not have GFP fluorescence but there is a subpopulation that does, thus showing the heterogeneity within the population.

Data was summarised in graphs as shown in Figure 6.6 to indicate the number of events in each sample that resulted in a GFP fluorescent signal. The data show that there was no fluorescent signal detected in the SYTO-84$^-$ controls, with empty vector or pLG339*aggR* 042, which indicates that the background debris has been removed. The SYTO-84$^+$ with pLG339 control has events in the GFP$^-$ and no fluorescent signal was detected. For the SYTO-84$^+$ samples containing pLG339*aggR* 042, there are also events for GFP$^-$, however, there are also a small number of GFP$^+$ events: a mean of 1088 events from 13363 total events, 12.3%. This indicates that in a small proportion of the cells containing AggR, *aafD100* promoter activity is activated by AggR.

6.4 Discussion

The AggR-activated repressor, Aar, is thought to bind directly to AggR to prevent DNA binding (Santiago *et al.*, 2016). Previously, it was shown that improving the Shine-Dalgarno sequence in Aar from EAEC strain 042 improved the ability of Aar to repress AggR (Abdelwahab *et al.*, 2021). Thus, improved expression of

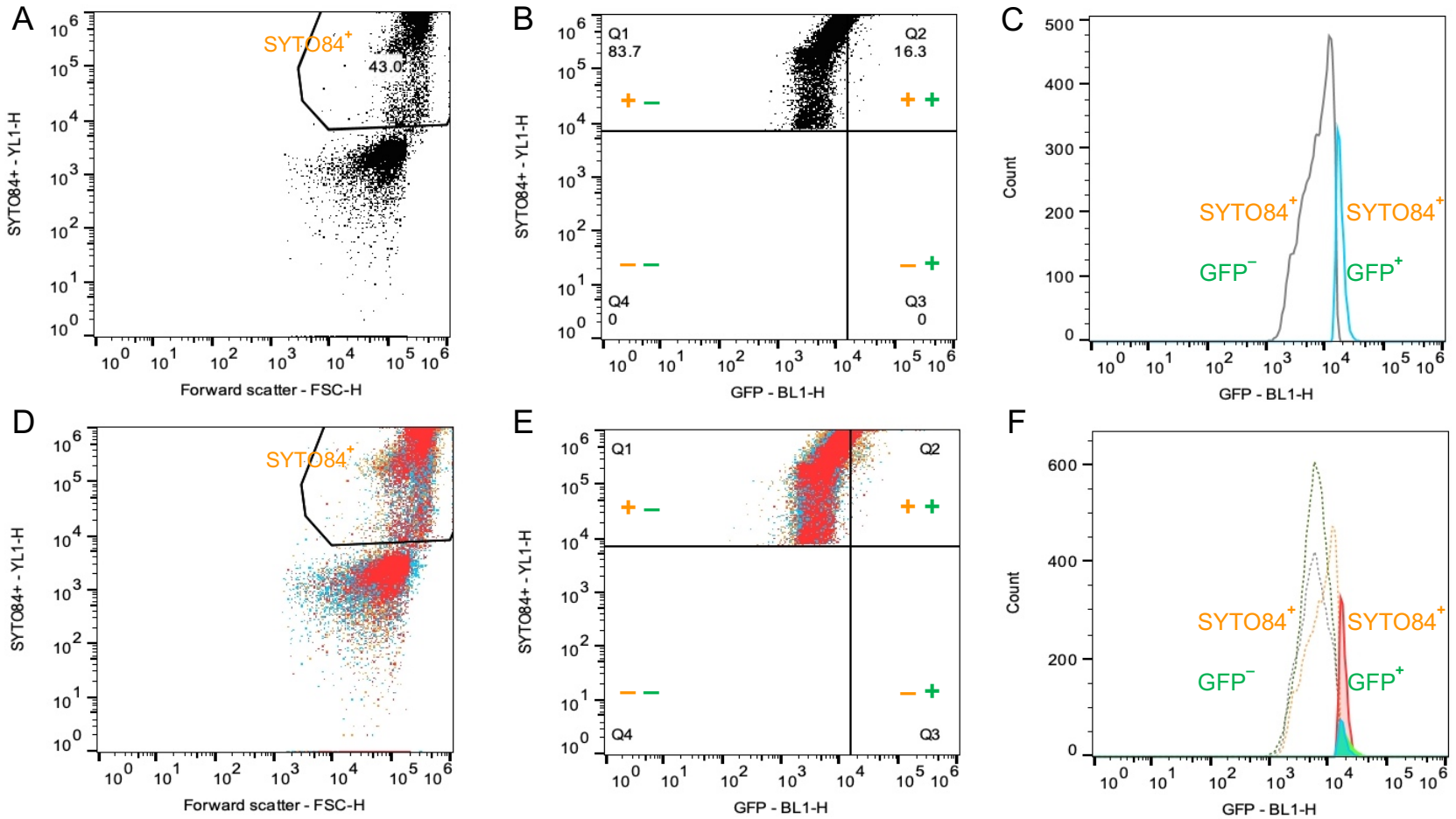
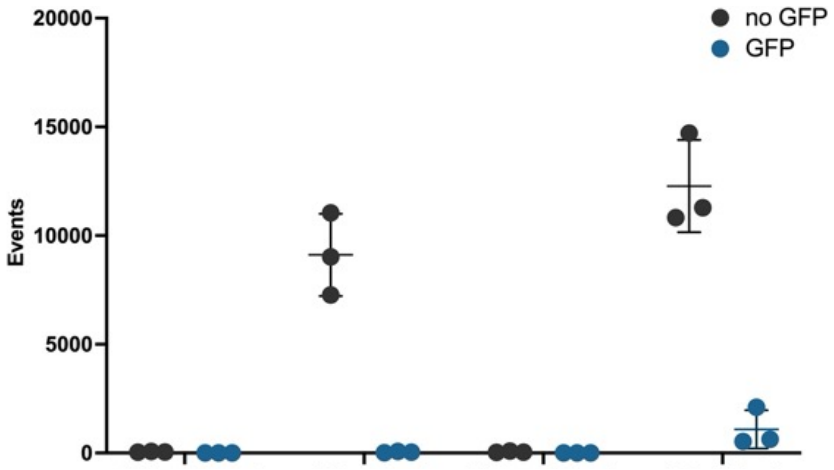


Figure 6.5 Flow cytometry plots of stained AggR samples

The flow cytometry plots of the stained BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042. Data was analysed and plots were made using FlowJo.

A. An example of one sample of BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042. Cells were separated from background using channels YL1-H against FSC-H. A SYTO84^{+} population of over 10,000 events was gated, highlighted by a polygon, and labelled SYTO84^{+} in orange. The percentage of events in the polygon gate of SYTO84^{+} is indicated.
B. Cells in the SYTO84^{+} gate (A) were analysed using channels YL1-H against BL1-H and a quadrat gate separated the SYTO84 (labelled in orange) and GFP signals (labelled in green): Q1 for SYTO-84^{+} GFP^{-} (+ –); Q2 for SYTO-84^{+} GFP^{+} (+ +); Q3 for SYTO-84^{-} GFP^{+} (– +); and Q4 for SYTO-84^{-} GFP^{-} (– –). The percentage of events in each of the quadrat gates is indicated.
C. The SYTO-84^{+} GFP^{-} and SYTO-84^{+} GFP^{+} quadrat gates (B) were plotted on a logarithmic histogram against BL1-H. The grey peak indicates the cells from the SYTO-84^{+} GFP^{-} quadrat gate and the blue peak indicates the cells from the SYTO-84^{+} GFP^{+} quadrat gate.
D. Three samples of BW25113 cells containing pRW400*aafD100* and pLG339*aggR* 042 combined, each sample highlighted in different colours. Cells were separated from background using channels YL1-H against FSC-H. A SYTO84^{+} population of over 10,000 events was gated, highlighted by a polygon, and labelled SYTO84^{+} in orange.
E. Cells in the SYTO84^{+} gate (A) were analysed using channels YL1-H against BL1-H and a quadrat gate separated the SYTO84 (labelled in orange) and GFP signals (labelled in green): Q1 for SYTO-84^{+} GFP^{-} (+ –); Q2 for SYTO-84^{+} GFP^{+} (+ +); Q3 for SYTO-84^{-} GFP^{+} (– +); and Q4 for SYTO-84^{-} GFP^{-} (– –).
F. The SYTO-84^{+} GFP^{-} and SYTO-84^{+} GFP^{+} quadrat gates (B) were plotted on a logarithmic histogram against BL1-H. The dotted line peaks indicate the cells from the SYTO-84^{+} GFP^{-} quadrat gate and the solid line peaks indicate the cells from the SYTO-84^{+} GFP^{+} quadrat gate.



Syto84	–	+	–	+
AggR 042 in pLG339*	–	–	+	+

Figure 6.6 The heterogeneity of GFP fluorescence

The single circles represent the biological replicates from the SYTO84^{+} population of cells, the black bar indicates the mean of the replicates +/- standard error of the mean (SEM). There are 3 biological replicates for each sample. In cells containing pLG339 empty vector, the mean number of SYTO-84^{+} GFP^{-} events was 9112. In cells containing pLG339*aggR* 042, the mean number of SYTO-84^{+} GFP^{-} events was 12275 and the mean number of SYTO-84^{+} GFP^{+} events was 1088.
* Expression of the *aggR* gene is under the control of the aggR promoter

Aar gave more repression activity. However, both Aar and Aar^{+} showed weak repression when AggR was expressed under the control of its own promoter, pLG339*aagR* 042 (Figure 6.2).

The EAEC strains from Egypt support the idea that there is a feedforward mechanism of activation shared between strains, whereby the control of virulence by AggR is triggered either by an external signal or random fluctuations in protein levels (Abdelwahab *et al.*, 2021). The probability of switching into the virulent 'on' state relies on basal levels of AggR, due to AggR-independent expression of the *aggR* gene, and the level of Aar. The level of Aar, in turn, depends on AggR-dependent activation of the *aar* promoter and translation of Aar from *aar* gene messenger (Abdelwahab *et al.*, 2021). However, the *aar* gene is not present in all EAEC strains, nor is the level of repression equal between Aar from different strains, it is presumably an addition to the AggR feedforward loop, and it would explain why Aar is not a strong repressor because it toggles, and does not destroy, AggR levels.

To further understand the mechanism of AggR regulation, and the potential bi-stable switch, it was important to determine whether *aggR* gene expression was due to stochasticity or if a ligand is responsible for triggering AggR activity. To show this, it was necessary to determine whether there was a difference in gene expression between single cells. Unfortunately, there are limitations to flow cytometry: bacteria are very small which can make it difficult to distinguish these small cells from cellular debris and machine background noise (Ambriz-Avina *et al.*, 2014). To mitigate this, I utilised staining to aid in the separation of background noise and cells. However, it would also be useful to combine the

fluorescent flow cytometry data with subsequent fluorescent activated cell sorting (FACS) or microscopy analysis to further confirm my results (Davey and Kell, 1996). Due to time constraints, and the limitations due to the COVID-19 lockdown restrictions of lab work and training, it was not possible to follow up on this.

In Figure 6.5, there are two distinct populations, one that does not activate at the target promoter and one that does. Therefore, I would conclude that AggR functions stochastically. However, the proportion of individual cells that are 'on' in Figure 6.6 is lower than the proportion of 'off' cells, this suggests that a smaller number of cells reach the minimal threshold required to trigger the 'flipping' of the bi-stable switch to 'on', thus activating at the target promoter.

The data suggest that there is no ligand or trigger but that the control of AggR activity depends on a bi-stable switch flipped in a stochastic way. The frequency of 'switch flipping' is dependent on two principal factors: AggR independent expression of AggR, which varies from strain to strain, and the activity of Aar. Thus, a certain threshold of AggR, due to uninduced expression of AggR, is needed to trigger the *aggR* promoter and produce more AggR – the feedforward loop, i.e. AggR levels increase to reach the threshold level which 'flips the switch' to the 'on' state, though this might happen rarely (as seen in Figure 6.6). Alongside production of AggR, the expression of *aar* is also induced, which then dampens the activity of AggR by acting as an anti-activator to modulate AggR – the feedback loop with dampening (Abdelwahab *et al.*, 2021).

The critical threshold that must be reached explains why certain strains are virulent while others are not. In fact, a recent study by the Nataro lab of over 100 EAEC strain genomes searched for a crucial gene that conferred virulence,

however, they concluded that virulent strains must be defined by their phenotype rather than their genotype, which is unique to EAEC within diarrheagenic *E. coli* strains (Boisen *et al.*, 2020). The prevalence of the *aar* gene in the strains analysed was below 50%, though it is common in non-diarrhoeal EAEC isolates (Modgil *et al.*, 2021; Boisen *et al.*, 2020). Notably, the *aggR* gene was present in 78.9% of the >100 EAEC strains investigated, however, at least one gene in each of the >100 EAEC strains was under the control of AggR. This is consistent with the results detailed in Chapter 4, CfaD from EAEC strain H149/5 was identified as an AggR-like transcription factor and there were several AggR-activated promoters in strain H149/5, despite the fact that this strain is considered atypical because it lacks the *aggR* gene and has a discreet pattern of aggregative adherence. Following the comprehensive study of the EAEC pan-genome (Boisen *et al.*, 2020), the theory is that the virulent phenotype is due to this bi-stable switch which is thrown stochastically. Thus, I would conclude that different strains flip into virulence modes at different frequencies and, interestingly, induction is a random event rather than guided by a ligand. This is remarkable because transcription factors generally respond to a specific stimulus and instead, there is this stochastic induction of virulence in a pathogen that does not require a ligand.

In this Chapter, I have highlighted that Aar (EAEC strain 042) and Aar^{+}, with increased expression, slightly reduce AggR activity, Aar^{+} more than Aar, however, this reduction is marginal. This is due to the role of Aar in modulating AggR activity in the feedforward and feedback loop by which AggR activates expression of *aggR* and *aar*, respectively. Furthermore, work in this Chapter

detailed the stochastic nature of *aggR* gene expression and, after AggR production reached a critical threshold, the subsequent activation at target promoters. This offered insight into the remarkable behaviour of this transcription factor: the AggR regulon contains many virulence factors, yet AggR expression is induced randomly and without an external signal or stimulus.

Chapter 7 Final Discussion

EAEC has been studied extensively to improve our knowledge its epidemiology and physiology. EAEC genomes have been examined to locate their virulence determinants, and the regulon of the AggR transcription regulator has been investigated to identify 'simple' AggR-dependent promoters, their essential promoter elements, and key activation determinants. In this study, I have analysed some more complex promoters, and identified the mechanism by which AggR activates, i.e., AggR interacts with RNAP via pre-recruitment (Figure 7.1).

7.1 EAEC strain differences EAEC strain differences

A study by the Nataro lab, which analysed over 100 EAEC strains, indicated that virulent strains are defined by their phenotype rather than their genotype (Boisen *et al.*, 2020). For EAEC strain H149/5, as highlighted in Chapter 4, its pattern of adherence to epithelial cells is not the typical "stacked-brick" form that is characteristic of EAEC and it does not carry the *aggR* gene, however, this strain is virulent and was isolated from patients suffering from acute diarrhoea. Two further strains, H9/3 and I192/3, that display a typical adherence pattern, were isolated and sequenced alongside strain H149/5. In 4.2.1, I showed that both H9/3 and I192/3 encode a transcription factor that is almost identical to AggR, and while H149/5 carries an AggR-like factor, CfaD, which has low sequence identity to AggR. The fact that all three strains are pathogenic supports the conclusion that pathogenic strains are defined by their phenotype rather than their genotype.

Interestingly, the H149/5 strain does not carry a pAA virulence plasmid, the *aggR* gene or known AAF genes. However, after whole-genome sequencing, I identified a gene homologous to fimbrial genes, indicating the discovery of a novel AAF gene in EAEC strain H149/5. The novel genes identified in this study indicate that whole-genome sequencing is important to understand the pathogenesis and heterogeneity of different EAEC strains.

EAEC strains that do not carry the *aggR* gene still encode genes that are regulated by AggR (Boisen *et al.*, 2020). Hence, in Section 4.2.2, I showed that there are several AggR-activated genes present in EAEC strain H149/5.

Rns and AggR belong to the AraC-subgroup (Nataro *et al.*, 1994; Yasir *et al.*, 2019). As outlined in Section 4.2.1, they have high sequence similarity and are thought to recognise similar promoter sequences. In Section 4.2, I showed that Rns can substitute for AggR at AggR-activated promoters, which corresponds with the reported characteristics for this subgroup. Additionally, the AggR-activated promoters were also activated by Rns from ETEC and CfaD from strain H149/5. In Section 4.4, I highlighted that the sequence similarity between AggR 042 and CfaD H149/5 is low, comparable to VirF and AggR, however, as AggR can be substituted for CfaD at CfaD-activated promoters and CfaD could be substituted for AggR at the *aafD100* promoter from EAEC strain 042, I determined that CfaD is a member of this AraC subgroup. My study underscores that whole-genome sequencing can enhance understanding of pathogenicity, particularly when a strain does not follow the typical patterns of adherence.

7.2 AggR activation at target promoters

Several AggR-dependent promoters were previously characterised by Muhammad Yasir, and he showed that the optimal spacing between the -10 hexamer and the single DNA binding site for AggR (consensus 5'-AWWWWWWTATC-3' located at a Class II position) is 22 bp (Yasir *et al.*, 2019). In this study, I investigated more complex promoters, including a bi-directional promoter, to further our understanding of AggR-dependent promoters. It had previously been reported that, at bi-directional promoters, one promoter blocks expression of the other promoter, as RNAP cannot bind in both directions for transcription (Warman *et al.*, 2021). However, my results show that the *yicS-nlpA* bi-directional promoters are not obstructed by each other. This result is consistent with a previous study of CRP-dependent bi-directional promoters, which showed that the proportion of time occupied by RNAP heading in either direction was so small, so there was no interference (El-Robh and Busby, 2002).

It has long been assumed that the presence of a regulatory region adjacent to a gene indicates that it must be expressed, however, as the evolution of bacterial genomes occurs rapidly, it stands to reason that not all genes are expressed (Minchin and Busby, 2009), as is the case with the *bssS* gene in EAEC 042 and *E. coli* K-12, despite carrying a DNA site for AggR. In fact, this also appears to be the case for the *yicS-nlpA* bi-directional promoter. Previously, it was identified that the *yicS* gene is important for pathogenicity in APEC (Verma *et al.*, 2018), however, its role in EAEC was not known. I showed in Section 3.3.3 that removing the *yicS* gene from the EAEC strain 042 chromosome does not prevent biofilm formation. As it has been suggested that bi-directional transcription is a property

of newly acquired DNA (Warman *et al.*, 2021), this might explain why the *yicS-nlpA* promoter has since been removed from some strains.

My work shows that different factors affect dependence of a promoter on AggR. These include neighbouring strong promoters, the sequence of the AggR-DNA binding site and the -10 element sequence, as well as its AggR-independent basal activity.

In my studies, I highlighted the importance of the first base in the DNA binding site for AggR, which seems to have an effect on the expression levels and the AggR binding affinity. In Section 3.3.1, the importance of the first base compared to the second is clear at the *yicS* promoter, as the expression from *yicS* strain 042 is stronger than at *yicS* strain H92/3. This indicates that simple changes in the promoter can allow it to dip in and out of the AggR regulon. This is not exclusive to the DNA site for AggR, the -10 element can also be altered to destroy AggR-dependence, as with the *E. coli* K-12 *yicS* promoter.

Thus, this study has enhanced the understanding of more complex AggR-dependent promoters and the key factors that confer dependence. In this study, I have highlighted that there are different factors that confer dependence or lack thereof.

7.3 Pre-recruitment and the structure of AggR

The structure of AggR can offer insight into key residues for contacting RNAP and target promoters, but structures for AraC family members are few and far between. However, some information can be obtained by comparing AggR to related AraC-family members.

There are over 40 virulence genes in the AggR regulon, and disrupting the expression of these genes could have positive outcomes for patients suffering from acute diarrhoea due to EAEC colonisation of the intestine and subsequent infection. I had hoped to do this by exploiting changes in key residues in AggR. Through extensive mutational analysis, partly guided by other AraC-family transcriptional regulators, located residues both in the NTD and the CTD, that are essential for AggR function.

For example, in the NTD, changes at positions I14 and N16, guided by Rns, and Y92 destroyed activation function. These residues are located in the region that is supposed to be important for effector binding or multimerization. However, AggR has been shown to function as a monomer (Yasir *et al.*, 2019) and there is no evidence that AggR binds an effector, consistent with Rns (Basturea *et al.*, 2008), nonetheless, the results in the study are important as I have shown that this region is still required for AggR-mediated activation.

All AraC-family transcription regulators have a 100 amino acid stretch of sequence homology, consisting of two HTH motifs, and this domain is thought to be responsible for activation function at target promoters (Gallegos *et al.*, 1997; Porter and Dorman, 2002). For AggR, I found that both HTH1 and HTH2 are independently required for AggR-mediated activation.

However, my work highlights the importance of both domains of the AggR protein for its activation function. This argues for interactions between the domains.

There are two main mechanisms proposed for transcription-factor dependent activation at bacterial target promoters: recruitment and pre-recruitment. For recruitment, the transcription activator binds to the promoter DNA before

recruiting the RNAP to the promoter. For pre-recruitment, the transcriptional activator interacts with the RNAP in solution then directs the RNAP to the promoter binding sites (Figure 7.1) (Cortes-Avalos *et al.*, 2021). Rob has been shown to function via pre-recruitment, hence, Rob forms a large complex with RNAP in the absence of promoter DNA. Interestingly, Rob also binds to DNA in the absence of RNAP, so it is thought to be able to function by both recruitment and pre-recruitment (Shi *et al.*, 2022). Some AraC members have been shown to favour this pre-recruitment method: SoxS interacts with the RNAP α-subunit which inhibits the interactions with UP elements, while directing the RNAP to soxbox sites in functional promoters (Cortes-Avalos *et al.*, 2021; Zafar *et al.*, 2011). MarA also functions via pre-recruitment (Taliaferro *et al.*, 2012). In Section 3.4, I introduced the idea that AggR functions via pre-recruitment as I had showed that AggR weakly repressed the *nlpA* promoter; essentially, AggR binds as a complex with RNAP to the AggR-DNA binding site at the *yicS-nlpA* promoter, therefore, repression is weak.

Interestingly, mutations in AggR, guided by MarA, do not function as expected or as in MarA (Grainger *et al.*, 2004). This indicates that though they both function by this unusual DNA-scanning method of recruiting RNAP to the promoter for transcription, the proteins are too different to directly indicate the significance of each residue.

SoxS and MarA are composed only of the activation/DNA-binding domain and, though Rob contains both domains, the structural elements and subsequent roles of each domain are swapped (Shi *et al.*, 2022; Lee *et al.*, 2012). It has previously been determined that AggR functions as a monomer (Yasir *et al.*, 2019), this is

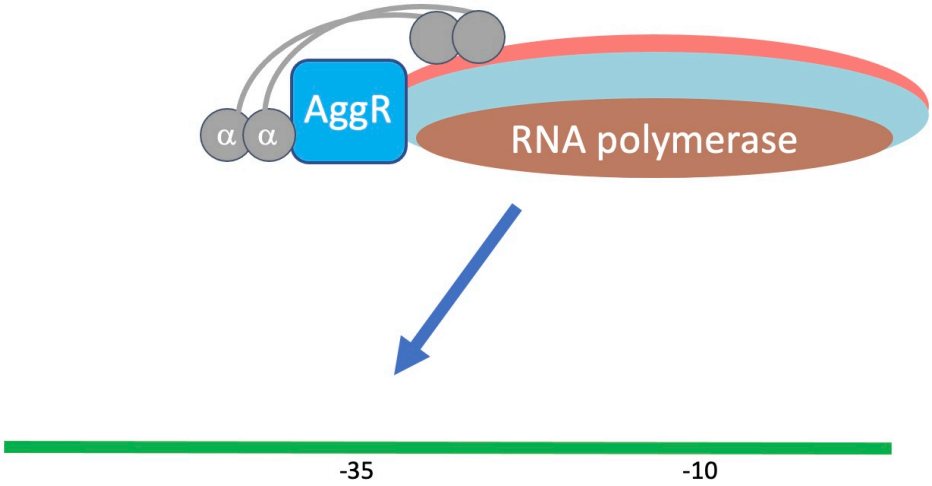


Figure 7.1 Mechanism of transcription activation by pre-recruitment

The figure shows the binding of AggR to the RNAP σ region 4, with the AggR-NTD required to maintain the AggR-CTD into the active conformation for interaction with RNAP and promoter DNA. Then, the AggR-RNAP complex scans the DNA for an AggR-dependent promoter. AggR then binds at a position overlapping the -35 element, like a Class II activator, and the RNAP makes sequence-specific contacts to the promoter elements. Adapted from Browning and Busby (2004).

also the case for AraC transcriptional activators Rob, SoxS and MarA (Taliaferro *et al.*, 2012). Thus, I suggest that AggR, MarA, SoxS and Rob belong to a AraC subgroup defined by the mechanism of activation, i.e., RNAP recruitment.

As with SoxS distinguishing soxboxes in promoters from those that lie outside of promoters, the mechanism of pre-recruitment could allow AggR to recruit RNAP and then bring RNAP exclusively to functional promoters. As I have shown that the presence of an AggR binding site doesn't necessarily indicate that the promoter is AggR-dependent, this mechanism prevents AggR from binding at AggR-binding sites that are not present in functional promoters.

Analysis of the AggR protein has shown that substitutions in the NTD and CTD, in both HTH1 and HTH2, have a negative effect on AggR activity at promoters. However, none of the AggR mutants are strongly *trans*-dominant when co-expressed with wild-type AggR, which suggests that AggR functions via pre-recruitment, a unique mechanism utilised by a selection of AraC transcriptional regulators, placing AggR in this unique group.

7.4 AggR regulation

A major aim of my study was to understand how AggR activity is regulated. In Section 5.2, I showed that the expression of AggR is not triggered by an external signal, this is an important discovery for a regulator of virulence in a pathogenic strain of *E. coli*. Since there is no direct evidence that they activity of AggR is controlled by any ligand or by post-transcriptional modification, I wanted to determine if expression of AggR was subject to stochastic bursts and hence AggR-dependent virulence would depend on a bistable switch. Thus, in Section 6.3, I showed that expression of AggR differs between individual cells, despite

the lack of an external signal. Interestingly, as I had showed that there was a lower proportion of cells that expressed GFP, this supported the idea that the production of AggR must reach a critical threshold to flip the bistable switch to 'on' for expression of virulence genes. This contradicts the long-standing Jacob and Monod model that transcription factors are deterministic (Jacob and Monod, 1961). Additionally, the weak repressor of AggR activity, Aar, acts as a dampener in the feedforward loop to modulate the AggR-virulence switch, as outlined in Chapter 6 (Figure 7.2) (Abdelwahab *et al.*, 2021). This modulation of activity has also recently been identified in *Shigella* flexneri, with VirF, whereby VirF-dependent expression of the *virB* gene, which is required during secretion and invasion of host cells, might be dampened to improve *Shigella* persistence within host cells (McKenna *et al.*, 2022).

In fact, the pre-recruitment mechanism of transcription activation is consistent with the threshold theory; appropriation of RNAP by AggR could help EAEC efficiently switch on virulence genes. An example of a transcription factor appropriating RNAP for virulence is the activator, AsiA, which binds to and remodels σ^{70} domain 4, which allows MotA to interact, thus allowing the development of T4 infection (Lee *et al.*, 2012). In fact, the different binding affinities between AggR and dependent promoters could indicate how, despite the lack of external signal to trigger expression of AggR, it is possible for AggR to prioritise genes important for virulence. Thus, I would suggest that the regulation of the AggR regulon is related to the strength of the promoter rather than a signal.

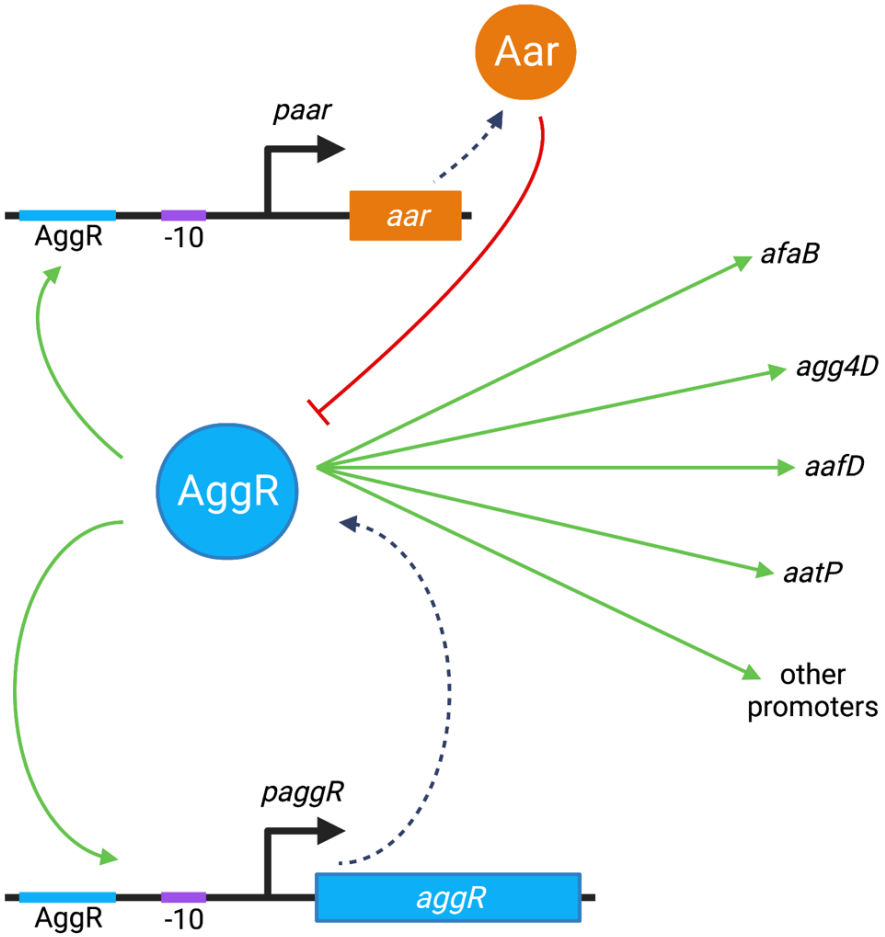


Figure 7.2 AggR regulation: a feedforward loop with dampening

The figure shows the AggR-dependent virulence 'switch': AggR-independent expression occurs at low levels which activates the production of more AggR – feedforward loop. AggR activates the expression of the *aar* gene and the activity of AggR is modulated by the Aar – dampening. When the level of AggR production reaches a critical threshold, the genes in the AggR-regulon are activated. Adapted from Abdelwahab *et al.* (2021). Created with BioRender.com.

These results are important for understanding the mechanism of AggR regulation but have wider implications for combating EAEC virulence. Though it is not possible to introduce a dominant AggR mutant in *trans* to disrupt wild-type AggR, thereby disrupting the AggR regulon, further understanding of the AggR-activated virulence switch will show potential targets for disruption of virulence gene expressions.

The results in this study enhance the understanding of the AggR-dependent virulence switch. This is particularly important as the prevalence of antimicrobial resistant strains increases (Abdelwahab *et al.*, 2021) and will aid in the development of alternatives treatments for patients, particularly as research shifts to focus on developing alternatives to antibiotics (Ghosh *et al.*, 2019; Kumar *et al.*, 2021).

7.5 Concluding remarks

AggR is the master regulator of virulence in EAEC, and its regulon includes genes encoded on the chromosome and on a virulence plasmid. My study has improved understanding of the AggR regulation mechanism, including the important features in the AggR protein, and the key features that confer dependence at promoters, with the following novel findings:

- The key features of an AggR dependent promoter, key bases in the DNA binding site for AggR and the -10 element, and how these key bases are targeted during evolution to remove genes from the AggR regulon.
- The *aap* gene and *caf1M* gene in H149/5, indicating the presence of dispersin and fimbrial genes, respectively.

- An AggR-like transcription factor, CfaD, from EAEC H149/5 that has sequence similarity to Rns and AggR and should belong to the AraC-subgroup. AggR and Rns can substitute for this transcription factor at promoters.
- Key residues for AggR-dependent activation at target promoters.
    - The NTD: I14, N16, Y92.
    - The CTD:
        - HTH1: I183, E191, R195, K196, R197.
        - HTH2: Q227, Y242.
- AggR mutants that are weakly *trans*-dominant and slightly suppress wild-type AggR function when co-expressed.
- AggR functions via pre-recruitment by first interacting with the RNAP σ subunit and then scans DNA, as an AggR-RNAP complex, for a target promoter.
- There is no external signal that induces expression of AggR.
- Expression of the *aggR* gene is stochastic.
- A small proportion of cells reach a critical threshold of AggR production to 'flip' a bistable switch 'on'.

References

Abdelwahab, R., Yasir, M., Godfrey, R. E., Christie, G. S., Element, S. J., Saville, F., Hassan, E. A., Ahmed, E. H., Abu-Faddan, N. H., Daef, E. A., Busby, S. J. W. and Browning, D. F. (2021) Antimicrobial resistance and gene regulation in Enteroaggregative Escherichia coli from Egyptian children with diarrhoea: Similarities and differences, *Virulence*, 12(1), pp. 57-74.

Alsharif, G., Ahmad, S., Islam, M. S., Shah, R., Busby, S. J. and Krachler, A. M. (2015) Host attachment and fluid shear are integrated into a mechanical signal regulating virulence in Escherichia coli O157:H7, *Proc Natl Acad Sci U S A*, 112(17), pp. 5503-8.

Alves, J. R., Pereira, A. C. M., Souza, M. C., Costa, S. B., Pinto, A. S., Mattos-Guaraldi, A. L., Hirata, R., Rosa, A. C. P. and Asad, L. M. B. O. (2010) Iron-limited condition modulates biofilm formation and interaction with human epithelial cells of enteroaggregative Escherichia coli (EAEC), *Journal of Applied Microbiology*, 108(1), pp. 246-255.

Ambriz-Avina, V., Contreras-Garduno, J. A. and Pedraza-Reyes, M. (2014) Applications of flow cytometry to characterize bacterial physiological responses, *Biomed Res Int*, 2014, p. 461941.

Ammar, E. M., Wang, X. and Rao, C. V. (2018) Regulation of metabolism in Escherichia coli during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose, *Sci Rep*, 8(1), p. 609.

Andrade, F. B., Abreu, A. G., Nunes, K. O., Gomes, T. A. T., Piazza, R. M. F. and Elias, W. P. (2017) Distribution of serine protease autotransporters of Enterobacteriaceae in typical and atypical enteroaggregative Escherichia coli, *Infection Genetics and Evolution*, 50, pp. 83-86.

Asadi Karam, M. R., Rezaei, A. A., Siadat, S. D., Habibi, M. and Bouzari, S. (2017) Evaluation of Prevalence, Homology and Immunogenicity of Dispersin among Enteroaggregative Escherichia coli Isolates from Iran, *Iranian biomedical journal*, 21(1), pp. 40-47.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol Syst Biol*, 2, p. 2006 0008.

Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. and Darst, S. A. (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex, *Elife*, 4.

Bamidele, O., Jiang, Z.-D. and Dupont, H. (2019) Occurrence of putative virulence-related genes, aatA, aggR and aaiC, of Enteroaggregative Escherichia coli (EAEC) among adults with travelers' diarrhea acquired in Guatemala and Mexico, *Microbial Pathogenesis*, 128, pp. 97-99.

Barbieri, J. T., Munson George, P., Holcomb Lisa, G. and Scott June, R. (2001) Novel Group of Virulence Activators within the AraC Family That Are Not Restricted to Upstream Binding Sites, *Infection and Immunity*, 69(1), pp. 186-193.

Basturea, G. N., Bodero, M. D., Moreno, M. E. and Munson, G. P. (2008) Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding, *J Bacteriol*, 190(7), pp. 2279-85.

Berger, P., Knodler, M., Forstner, K. U., Berger, M., Bertling, C., Sharma, C. M., Vogel, J., Karch, H., Dobrindt, U. and Mellmann, A. (2016) The primary transcriptome of the Escherichia coli O104:H4 pAA plasmid and novel insights into its virulence gene expression and regulation, *Scientific Reports*, 6.

Berriman, M. and Rutherford, K. (2003) Viewing and annotating sequence data with Artemis, *Brief Bioinform*, 4(2), pp. 124-32.

Bhargava, S., Johnson, B. B., Hwang, J., Harris, T. A., George, A. S., Muir, A., Dorff, J. and Okeke, I. N. (2009) Heat-Resistant Agglutinin 1 Is an Accessory Enteroaggregative Escherichia coli Colonization Factor, *Journal of Bacteriology*, 191(15), pp. 4934-4942.

Bhende, P. M. and Egan, S. M. (1999) Amino acid-DNA contacts by RhaS: an AraC family transcription activator, *J Bacteriol*, 181(17), pp. 5185-92.

Bodero, M. D., Pilonieta, M. C. and Munson, G. P. (2007) Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic Escherichia coli, *J Bacteriol*, 189(5), pp. 1627-32.

Boisen, N., Krogfelt, K. A. and Nataro, J. P. (2013) 'Chapter 8 - Enteroaggregative Escherichia coli', in Donnenberg, M. S. (ed.) *Escherichia coli (Second Edition)*. Boston: Academic Press, pp. 247-273.

Boisen, N., Melton-Celsa, A. R., Scheutz, F., O'Brien, A. D. and Nataro, J. P. (2015) Shiga toxin 2a and Enteroaggregative Escherichia coli--a deadly combination, *Gut Microbes*, 6(4), pp. 272-8.

Boisen, N., Osterlund, M. T., Joensen, K. G., Santiago, A. E., Mandomando, I., Cravioto, A., Chattaway, M. A., Gonyar, L. A., Overballe-Petersen, S., Stine, O. C., Rasko, D. A., Scheutz, F. and Nataro, J. P. (2020) Redefining enteroaggregative Escherichia coli (EAEC): Genomic characterization of epidemiological EAEC strains, *Plos Neglected Tropical Diseases*, 14(9).

Boisen, N., Struve, C., Scheutz, F., Krogfelt, K. A. and Nataro, J. P. (2008) New adhesin of enteroaggregative Escherichia coli related to the Afa/Dr/AAF family, *Infect Immun*, 76(7), pp. 3281-92.

Boll, E. J., Struve, C., Boisen, N., Olesen, B., Stahlhut, S. G. and Krogfelt, K. A. (2013) Role of enteroaggregative Escherichia coli virulence factors in uropathogenesis, *Infect Immun*, 81(4), pp. 1164-71.

Bourgerie, S. J., Michan, C. M., Thomas, M. S., Busby, S. J. and Hyde, E. I. (1997) DNA binding and DNA bending by the MelR transcription activator protein from Escherichia coli, *Nucleic Acids Res*, 25(9), pp. 1685-93.

Braga, R. L. L., Pereira, A. C. M., Santos, P. A. D., Freitas-Almeida, A. C. and Rosa, A. C. P. (2017) Ex Vivo Model of Rabbit Intestinal Epithelium Applied to the Study of Colonization by Enteroaggregative Escherichia Coli, *Arq Gastroenterol*, 54(2), pp. 130-134.

Brown, N. L., Stoyanov, J. V., Kidd, S. P. and Hobman, J. L. (2003) The MerR family of transcriptional regulators, *FEMS Microbiol Rev*, 27(2-3), pp. 145-63.

Browning, D. F. and Busby, S. J. (2004) The regulation of bacterial transcription initiation, *Nat Rev Microbiol*, 2(1), pp. 57-65.

Browning, D. F. and Busby, S. J. (2016) Local and global regulation of transcription initiation in bacteria, *Nat Rev Microbiol*, 14(10), pp. 638-50.

Bryant, J. A., Sellars, L. E., Busby, S. J. and Lee, D. J. (2014) Chromosome position effects on gene expression in Escherichia coli K-12, *Nucleic Acids Res*, 42(18), pp. 11383-92.

Brzuszkiewicz, E., Thurmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F. D., Boelter, J., Petersen, H., Gottschalk, G. and Daniel, R. (2011) Genome sequence analyses of two isolates from the recent Escherichia coli outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic Escherichia coli (EAHEC), *Arch Microbiol*, 193(12), pp. 883-91.

Buchholz, U., Bernard, H., Werber, D., Bohmer, M. M., Remschmidt, C., Wilking, H., Delere, Y., an der Heiden, M., Adlhoch, C., Dreesman, J., Ehlers, J., Ethelberg, S., Faber, M., Frank, C., Fricke, G., Greiner, M., Hohle, M., Ivarsson, S., Jark, U., Kirchner, M., Koch, J., Krause, G., Luber, P., Rosner, B., Stark, K. and Kuhne, M. (2011) German outbreak of Escherichia coli O104:H4 associated with sprouts, *N Engl J Med*, 365(19), pp. 1763-70.

Busby, S., Kotlarz, D. and Buc, H. (1983) Deletion mutagenesis of the Escherichia coli galactose operon promoter region, *J Mol Biol*, 167(2), pp. 259-74.

Carver, T., Berriman, M., Tivey, A., Patel, C., Bohme, U., Barrell, B. G., Parkhill, J. and Rajandream, M. A. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database, *Bioinformatics*, 24(23), pp. 2672-6.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J. and McQuillan, J. A. (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data, *Bioinformatics*, 28(4), pp. 464-9.

Casadaban, M. J. and Cohen, S. N. (1980) Analysis of gene control signals by DNA fusion and cloning in Escherichia coli, *J Mol Biol*, 138(2), pp. 179-207.

Chaudhuri, R. R., Sebaihia, M., Hobman, J. L., Webber, M. A., Leyton, D. L., Goldberg, M. D., Cunningham, A. F., Scott-Tucker, A., Ferguson, P. R., Thomas, C. M., Frankel, G., Tang, C. M., Dudley, E. G., Roberts, I. S., Rasko, D. A., Pallen, M. J., Parkhill, J., Nataro, J. P., Thomson, N. R. and Henderson, I. R. (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative Escherichia coli strain 042, *PLoS One*, 5(1), p. e8801.

Chen, H., Bjerknes, M., Kumar, R. and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs, *Nucleic Acids Res*, 22(23), pp. 4953-7.

Chen, J., Boyaci, H. and Campbell, E. A. (2021) Diverse and unified mechanisms of transcription initiation in bacteria, *Nat Rev Microbiol*, 19(2), pp. 95-109.

Choy, H. E. and Adhya, S. (1992) Control of gal transcription through DNA looping: inhibition of the initial transcribing complex, *Proc Natl Acad Sci U S A*, 89(23), pp. 11264-8.

Clements, A., Young, J. C., Constantinou, N. and Frankel, G. (2012) Infection strategies of enteric pathogenic Escherichia coli, *Gut Microbes*, 3(2), pp. 71-87.

Cortes-Avalos, D., Martinez-Perez, N., Ortiz-Moncada, M. A., Juarez-Gonzalez, A., Banos-Vargas, A. A., Estrada-de Los Santos, P., Perez-Rueda, E. and Ibarra, J. A. (2021) An update of the unceasingly growing and diverse AraC/XylS family of transcriptional activators, *FEMS Microbiol Rev*, 45(5).

Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004) WebLogo: a sequence logo generator, *Genome Res*, 14(6), pp. 1188-90.

Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M. and Finlay, B. B. (2013) Recent Advances in Understanding Enteric Pathogenic Escherichia coli, *Clinical Microbiology Reviews*, 26(4), pp. 822-880.

Czeczulin, J. R., Balepur, S., Hicks, S., Phillips, A., Hall, R., Kothary, M. H., NavarroGarcia, F. and Nataro, J. P. (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative Escherichia coli, *Infection and Immunity*, 65(10), pp. 4135-4145.

Davey, H. M. and Kell, D. B. (1996) Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses, *Microbiol Rev*, 60(4), pp. 641-96.

Davies, J. (1996) Origins and evolution of antibiotic resistance, *Microbiologia*, 12(1), pp. 9-16.

Davis, M. C., Kesthely, C. A., Franklin, E. A. and MacLellan, S. R. (2017) The essential activities of the bacterial sigma factor, *Can J Microbiol*, 63(2), pp. 89-99.

Dhiman, A. and Schleif, R. (2000) Recognition of overlapping nucleotides by AraC and the sigma subunit of RNA polymerase, *J Bacteriol*, 182(18), pp. 5076-81.

Di Martino, M. L., Falconi, M., Micheli, G., Colonna, B. and Prosseda, G. (2016) The Multifaceted Activity of the VirF Regulatory Protein in the Shigella Lifestyle, *Front Mol Biosci*, 3, p. 61.

Dias, R. C. B., Tanabe, R. H. S., Vieira, M. A., Cergole-Novella, M. C., Dos Santos, L. F., Gomes, T. A. T., Elias, W. P. and Hernandes, R. T. (2020) Analysis of the Virulence Profile and Phenotypic Features of Typical and Atypical Enteroaggregative Escherichia coli (EAEC) Isolated From Diarrheal Patients in Brazil, *Front Cell Infect Microbiol*, 10, p. 144.

Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M. and Sentenac, A. (1995) A universally conserved region of the largest subunit participates in the active site of RNA polymerase III, *EMBO J*, 14(15), pp. 3766-76.

Domka, J., Lee, J. and Wood, T. K. (2006) YliH (BssR) and YceP (BssS) regulate Escherichia coli K-12 biofilm formation by influencing cell signaling, *Appl Environ Microbiol*, 72(4), pp. 2449-59.

Dudley, E. G., Thomson, N. R., Parkhill, J., Morin, N. P. and Nataro, J. P. (2006) Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative Escherichia coli, *Mol Microbiol*, 61(5), pp. 1267-82.

Ebright, R. H., Werner, F. and Zhang, X. (2019) RNA Polymerase Reaches 60: Transcription Initiation, Elongation, Termination, and Regulation in Prokaryotes, *J Mol Biol*, 431(20), pp. 3945-3946.

Egan, S. M. (2002) Growing repertoire of AraC/XylS activators, *J Bacteriol*, 184(20), pp. 5529-32.

El-Hajj, Z. W. and Newman, E. B. (2015) An Escherichia coli Mutant That Makes Exceptionally Long Cells, *Journal of Bacteriology*, 197(8), pp. 1507-1514.

El-Robh, M. S. and Busby, S. J. (2002) The Escherichia coli cAMP receptor protein bound at a single target can activate transcription initiation at divergent promoters: a systematic study that exploits new promoter probe plasmids, *Biochem J*, 368(Pt 3), pp. 835-43.

Elias, W. P., Czeczulin, J. R., Henderson, I. R., Trabulsi, L. R. and Nataro, J. P. (1999) Organization of biogenesis genes for aggregative adherence fimbria II

defines a virulence gene cluster in enteroaggregative Escherichia coli, *Journal of Bacteriology*, 181(6), pp. 1779-1785.

Ellis, S. J., Crossman, L. C., McGrath, C. J., Chattaway, M. A., Holken, J. M., Brett, B., Bundy, L., Kay, G. L., Wain, J. and Schuller, S. (2020) Identification and characterisation of enteroaggregative Escherichia coli subtypes associated with human disease, *Sci Rep*, 10(1), p. 7475.

Ellis, S. J., Yasir, M., Browning, D. F., Busby, S. J. W. and Schuller, S. (2019) Oxygen and contact with human intestinal epithelium independently stimulate virulence gene expression in enteroaggregative Escherichia coli, *Cell Microbiol*, 21(6), p. e13012.

Estrada-Garcia, T., Perez-Martinez, I., Bernal-Reynaga, R. and Zaidi, M. B. (2014) Enteroaggregative coli: A Pathogen Bridging the North and South, *Curr Trop Med Rep*, 1(2), pp. 88-96.

Feklistov, A. and Darst, S. A. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit, *Cell*, 147(6), pp. 1257-69.

Forquet, R., Pineau, M., Nasser, W., Reverchon, S. and Meyer, S. (2021) Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters, *mSystems*, 6(4), p. e0097821.

Franca, F. L., Wells, T. J., Browning, D. F., Nogueira, R. T., Sarges, F. S., Pereira, A. C., Cunningham, A. F., Lucheze, K., Rosa, A. C., Henderson, I. R. and de Luna, M. (2013) Genotypic and phenotypic characterisation of enteroaggregative Escherichia coli from children in Rio de Janeiro, Brazil, *PLoS One*, 8(7), p. e69971.

Fujiyama, R., Nishi, J., Imuta, N., Tokuda, K., Manago, K. and Kawano, Y. (2008) The shf Gene of a Shigella flexneri Homologue on the Virulent Plasmid pAA2 of Enteroaggregative Escherichia coli 042 Is Required for Firm Biofilm Formation, *Current Microbiology*, 56(5), pp. 474-480.

Galbusera, L., Bellement-Theroue, G., Urchueguia, A., Julou, T. and van Nimwegen, E. (2020) Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria, *Plos One*, 15(10).

Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J. L. (1997) Arac/XylS family of transcriptional regulators, *Microbiol Mol Biol Rev*, 61(4), pp. 393-410.

Garcia-Bernardo, J. and Dunlop, M. J. (2013) Tunable stochastic pulsing in the Escherichia coli multiple antibiotic resistance network from interlinked positive and negative feedback loops, *PLoS Comput Biol*, 9(9), p. e1003229.

Ghosh, C., Sarkar, P., Issa, R. and Haldar, J. (2019) Alternatives to Conventional Antibiotics in the Era of Antimicrobial Resistance, *Trends Microbiol*, 27(4), pp. 323-338.

Grainger, D. C., Webster, C. L., Belyaeva, T. A., Hyde, E. I. and Busby, S. J. (2004) 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(5), pp. 1297-309.

Griffith, K. L., Shah, I. M., Myers, T. E., O'Neill, M. C. and Wolf, R. E. (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, *Biochemical and Biophysical Research Communications*, 291(4), pp. 979-986.

Griffith, K. L. and Wolf, R. E. (2004) Genetic Evidence for Pre-recruitment as the Mechanism of Transcription Activation by SoxS of Escherichia coli: The Dominance of DNA Binding Mutations of SoxS, *Journal of Molecular Biology*, 344(1), pp. 1-10.

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter, *J Bacteriol*, 177(14), pp. 4121-30.

Hacker, J. and Blum-Oehler, G. (2007) In appreciation of theodor escherich, *Nature Reviews Microbiology*, 5(12), pp. 902-902.

Hamilton, E. P. and Lee, N. (1988) Three binding sites for AraC protein are required for autoregulation of araC in Escherichia coli, *Proc Natl Acad Sci U S A*, 85(6), pp. 1749-53.

Harrington, S. M., Dudley, E. G. and Nataro, J. P. (2006) Pathogenesis of enteroaggregative Escherichia coli infection, *FEMS Microbiol Lett*, 254(1), pp. 12-8.

Havt, A., Lima, I. F., Medeiros, P. H., Clementino, M. A., Santos, A. K., Amaral, M. S., Veras, H. N., Prata, M. M., Lima, N. L., Di Moura, A., Leite, A. M., Soares, A. M., Filho, J. Q., Houpt, E. R., Nataro, J. P., Guerrant, R. L. and Lima, A. A. (2017) Prevalence and virulence gene profiling of enteroaggregative Escherichia coli in malnourished and nourished Brazilian children, *Diagn Microbiol Infect Dis*, 89(2), pp. 98-105.

Hebbelstrup Jensen, B., Olsen, K. E., Struve, C., Krogfelt, K. A. and Petersen, A. M. (2014) Epidemiology and clinical manifestations of enteroaggregative Escherichia coli, *Clin Microbiol Rev*, 27(3), pp. 614-30.

Hodson, C., Yang, J., Hocking, D. M., Azzopardi, K., Chen, Q. Y., Holien, J. K., Parker, M. W., Tauschek, M. and Robins-Browne, R. M. (2017) Control of Virulence Gene Expression by the Master Regulator, CfaD, in the Prototypical Enterotoxigenic Escherichia coli Strain, H10407, *Frontiers in Microbiology*, 8.

Hook-Barnard, I. G. and Hinton, D. M. (2007) Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters, *Gene Regul Syst Bio*, 1, pp. 275-93.

Huttener, M., Prieto, A., Espelt, J., Bernabeu, M. and Juarez, A. (2018) Stringent Response and AggR-Dependent Virulence Regulation in the Enteroaggregative Escherichia coli Strain 042, *Front Microbiol*, 9, p. 717.

Hüttener, M., Prieto, A., Espelt, J., Bernabeu, M. and Juárez, A. (2018) Stringent Response and AggR-Dependent Virulence Regulation in the Enteroaggregative Escherichia coli Strain 042, *Frontiers in Microbiology*, 9.

Ibarra, J. A., Perez-Rueda, E., Segovia, L. and Puente, J. L. (2008) The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors, *Genetica*, 133(1), pp. 65-76.

Ishihama, A. (2012) Prokaryotic genome regulation: a revolutionary paradigm, *Proc Jpn Acad Ser B Phys Biol Sci*, 88(9), pp. 485-508.

Islam, M. S., Pallen, M. J. and Busby, S. J. (2011) A cryptic promoter in the LEE1 regulatory region of enterohaemorrhagic Escherichia coli: promoter specificity in AT-rich gene regulatory regions, *Biochem J*, 436(3), pp. 681-6.

Ito, K., Matsushita, S., Yamazaki, M., Moriya, K., Kurazono, T., Hiruta, N., Narimatsu, H., Ueno, N., Isobe, J., Yatsuyanagi, J., Kumagai, N., Hashimoto, M. and Ratchtrachenchai, O. A. (2014) Association between aggregative adherence

fimbriae types including putative new variants and virulence-related genes and clump formation among aggR-positive Escherichia coli strains isolated in Thailand and Japan, *Microbiology and Immunology*, 58(8), pp. 467-473.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins, *J Mol Biol*, 3, pp. 318-56.

Jenkins, C., van Ijperen, C., Dudley, E. G., Chart, H., Willshaw, G. A., Cheasty, T., Smith, H. R. and Nataro, J. P. (2005) Use of a microarray to assess the distribution of plasmid and chromosomal virulence genes in strains of enteroaggregative Escherichia coli, *Fems Microbiology Letters*, 253(1), pp. 119-124.

Jensen, B. and Hebbelstrup Rye Rasmussen, S. (2017) Genetic virulence profile of enteroaggregative Escherichia coli isolated from children with either acute og persistent diarrhea, *Frontiers in Cellular and Infection Microbiology*, 7.

Jonsson, R., Struve, C., Boll, E. J., Boisen, N., Joensen, K. G., Sorensen, C. A., Jensen, B. H., Scheutz, F., Jenssen, H. and Krogfelt, K. A. (2017) A Novel pAA Virulence Plasmid Encoding Toxins and Two Distinct Variants of the Fimbriae of Enteroaggregative Escherichia coli, *Frontiers in Microbiology*, 8.

Kahramanoglou, C., Webster, C. L., El-Robh, M. S., Belyaeva, T. A. and Busby, S. J. W. (2006) Mutational analysis of the Escherichia coli melR gene suggests a two-state concerted model to explain transcriptional activation and repression in the melibiose operon, *Journal of Bacteriology*, 188(9), pp. 3199-3207.

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. and Ebright, R. H. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism, *Science*, 314(5802), pp. 1144-7.

Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004) Pathogenic Escherichia coli, *Nat Rev Microbiol*, 2(2), pp. 123-40.

Kumar, A., Mallik, D., Pal, S., Mallick, S., Sarkar, S., Chanda, A. and Ghosh, A. S. (2015) Escherichia coli O8-antigen enhances biofilm formation under agitated conditions, *FEMS Microbiol Lett*, 362(15), p. fnv112.

Kumar, M., Sarma, D. K., Shubham, S., Kumawat, M., Verma, V., Nina, P. B., Jp, D., Kumar, S., Singh, B. and Tiwari, R. R. (2021) Futuristic Non-antibiotic Therapies to Combat Antibiotic Resistance: A Review, *Front Microbiol*, 12, p. 609459.

Lang, C., Fruth, A., Holland, G., Laue, M., Muhlen, S., Dersch, P. and Flieger, A. (2018) Novel type of pilus associated with a Shiga-toxigenic E. coli hybrid pathovar conveys aggregative adherence and bacterial virulence, *Emerg Microbes Infect*, 7(1), p. 203.

Lebar, T., Bezeljak, U., Golob, A., Jerala, M., Kadunc, L., Pirs, B., Strazar, M., Vucko, D., Zupancic, U., Bencina, M., Forstneric, V., Gaber, R., Lonzaric, J., Majerle, A., Oblak, A., Smole, A. and Jerala, R. (2014) A bistable genetic switch based on designable DNA-binding domains, *Nat Commun*, 5, p. 5007.

Lee, D. J., Bingle, L. E., Heurlier, K., Pallen, M. J., Penn, C. W., Busby, S. J. and Hobman, J. L. (2009) Gene doctoring: a method for recombineering in laboratory and pathogenic Escherichia coli strains, *BMC Microbiol*, 9, p. 252.

Lee, D. J., Minchin, S. D. and Busby, S. J. (2012) Activating transcription in bacteria, *Annu Rev Microbiol*, 66, pp. 125-52.

Li, J. and Zhang, Y. (2014) Relationship between promoter sequence and its strength in gene expression, *The European Physical Journal E*, 37(9), p. 86.

Lim, J. Y., Yoon, J. and Hovde, C. J. (2010) A brief overview of Escherichia coli O157:H7 and its plasmid O157, *J Microbiol Biotechnol*, 20(1), pp. 5-14.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N. R. (1992) Broad host range plasmids carrying the Escherichia coli lactose and galactose operons, *FEMS Microbiol Lett*, 74(2-3), pp. 271-6.

Longo, D. and Hasty, J. (2006) Dynamics of single-cell gene expression, *Molecular Systems Biology*, 2.

Mahon, V., Smyth, C. J. and Smith, S. G. J. (2010) Mutagenesis of the Rns regulator of enterotoxigenic Escherichia coli reveals roles for a linker sequence and two helix-turn-helix motifs, *Microbiology (Reading)*, 156(Pt 9), pp. 2796-2806.

Martin, R. G., Gillette, W. K., Martin, N. I. and Rosner, J. L. (2002) Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in Escherichia coli, *Molecular Microbiology*, 43(2), pp. 355-370.

Martinez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria, *Curr Opin Microbiol*, 6(5), pp. 482-9.

Mazumder, A., Ebright, R. H. and Kapanidis, A. N. (2021) Transcription initiation at a consensus bacterial promoter proceeds via a 'bind-unwind-load-and-lock' mechanism, *Elife*, 10.

McHale, T. M., Garciarena, C. D., Fagan, R. P., Smith, S. G. J., Martin-Loches, I., Curley, G. F., Fitzpatrick, F. and Kerrigan, S. W. (2018) Inhibition of Vascular Endothelial Cell Leak Following Escherichia coli Attachment in an Experimental Model of Sepsis, *Crit Care Med*, 46(8), pp. e805-e810.

McKenna, J. A., Karney, M. M. A., Chan, D. K., Weatherspoon-Griffin, N., Becerra Larios, B., Pilonieta, M. C., Munson, G. P. and Wing, H. J. (2022) The AraC/XylS Protein MxiE and Its Coregulator IpgC Control a Negative Feedback Loop in the Transcriptional Cascade That Regulates Type III Secretion in Shigella flexneri, *J Bacteriol*, 204(7), p. e0013722.

Mejia-Almonte, C., Busby, S. J. W., Wade, J. T., van Helden, J., Arkin, A. P., Stormo, G. D., Eilbeck, K., Palsson, B. O., Galagan, J. E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria, *Nature Reviews Genetics*, 21(11), pp. 699-714.

Mellies, J. L. and Barron, A. M. (2006) Virulence Gene Regulation in Escherichia coli, *EcoSal Plus*, 2(1).

Menard, L. P. and Dubreuil, J. D. (2002) Enteroaggregative Escherichia coli heat-stable enterotoxin 1 (EAST1): a new toxin with an old twist, *Crit Rev Microbiol*, 28(1), pp. 43-60.

Miajlovic, H. and Smith, S. G. (2014) Bacterial self-defence: how Escherichia coli evades serum killing, *FEMS Microbiol Lett*, 354(1), pp. 1-9.

Mickey, A. S. and Nataro, J. P. (2020) Dual Function of Aar, a Member of the New AraC Negative Regulator Family, in Escherichia coli Gene Expression, *Infect Immun*, 88(6).

Midgett, C. R., Talbot, K. M., Day, J. L., Munson, G. P. and Kull, F. J. (2021) Structure of the master regulator Rns reveals an inhibitor of enterotoxigenic Escherichia coli virulence regulons, *Scientific Reports*, 11(1), p. 15663.

Minchin, S. D. and Busby, S. J. (2009) Analysis of mechanisms of activation and repression at bacterial promoters, *Methods*, 47(1), pp. 6-12.

Modgil, V., Chaudhary, P., Bharti, B., Mahindroo, J., Yousuf, M., Koundal, M., Mohan, B. and Taneja, N. (2021) Prevalence, Virulence Gene Profiling, and Characterization of Enteroaggregative Escherichia coli from Children with Acute Diarrhea, Asymptomatic Nourished, and Malnourished Children Younger Than 5 Years of Age in India, *J Pediatr*, 234, pp. 106-114 e5.

Mohamed, J. A., Huang, D. B., Jiang, Z. D., DuPont, H. L., Nataro, J. P., Belkind-Gerson, J. and Okhuysen, P. C. (2007) Association of putative enteroaggregative Escherichia coli virulence genes and biofilm production in isolates from travelers to developing countries, *Journal of Clinical Microbiology*, 45(1), pp. 121-126.

Morin, N., Santiago, A. E., Ernst, R. K., Guillot, S. J. and Nataro, J. P. (2013) Characterization of the AggR regulon in enteroaggregative Escherichia coli, *Infect Immun*, 81(1), pp. 122-32.

Morin, N., Tirling, C., Ivison, S. M., Kaur, A. P., Nataro, J. P. and Steiner, T. S. (2010) Autoactivation of the AggR regulator of enteroaggregative Escherichia coli in vitro and in vivo, *FEMS Immunol Med Microbiol*, 58(3), pp. 344-55.

Moxon, E. R. (2009) Bacterial variation, virulence and vaccines, *Microbiology (Reading)*, 155(Pt 4), pp. 997-1003.

Munson, G. P., Holcomb, L. G. and Scott, J. R. (2001) Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites, *Infect Immun*, 69(1), pp. 186-93.

Munson, G. P. and Scott, J. R. (1999) Binding site recognition by Rns, a virulence regulator in the AraC family, *J Bacteriol*, 181(7), pp. 2110-7.

Naryshkina, T., Mustaev, A., Darst, S. A. and Severinov, K. (2001) The beta ' subunit of Escherichia coli RNA polymerase is not required for interaction with initiating nucleotide but is necessary for interaction with rifampicin, *J Biol Chem*, 276(16), pp. 13308-13.

Nataro, J. P., Baldini, M. M., Kaper, J. B., Black, R. E., Bravo, N. and Levine, M. M. (1985) Detection of an adherence factor of enteropathogenic Escherichia coli with a DNA probe, *J Infect Dis*, 152(3), pp. 560-5.

Nataro, J. P., Deng, Y., Cookson, S., Cravioto, A., Savarino, S. J., Guers, L. D., Levine, M. M. and Tacket, C. O. (1995) Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers, *J Infect Dis*, 171(2), pp. 465-8.

Nataro, J. P., Yikang, D., Giron, J. A., Savarino, S. J., Kothary, M. H. and Hall, R. (1993) Aggregative adherence fimbria I expression in enteroaggregative Escherichia coli requires two unlinked plasmid regions, *Infection and immunity*, 61(3), pp. 1126-1131.

Nataro, J. P., Yikang, D., Yingkang, D. and Walker, K. (1994) AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative Escherichia coli, *Journal of bacteriology*, 176(15), pp. 4691-4699.

Navarro-Garcia, F. and Elias, W. P. (2011) Autotransporters and virulence of enteroaggregative E. coli, *Gut Microbes*, 2(1), pp. 13-24.

Navarro-Garcia, F., Ruiz-Perez, F., Cataldi, A. and Larzabal, M. (2019) Type VI Secretion System in Pathogenic Escherichia coli: Structure, Role in Virulence, and Acquisition, *Front Microbiol*, 10, p. 1965.

Nickels, B. E., Garrity, S. J., Mekler, V., Minakhin, L., Severinov, K., Ebright, R. H. and Hochschild, A. (2005) The interaction between sigma70 and the beta-flap of Escherichia coli RNA polymerase inhibits extension of nascent RNA during early elongation, *Proc Natl Acad Sci U S A*, 102(12), pp. 4488-93.

Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., Rose, D. J., Blattner, F. R. and Nataro, J. P. (2003) The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system, *J Biol Chem*, 278(46), pp. 45680-9.

Okeke, I. N., Lamikanra, A., Czeczulin, J., Dubovsky, F., Kaper, J. B. and Nataro, J. P. (2000) Heterogeneous virulence of enteroaggregative Escherichia coli strains isolated from children in Southwest Nigeria, *J Infect Dis*, 181(1), pp. 252-60.

Olesen, B., Neimann, J., Bottiger, B., Ethelberg, S., Schiellerup, P., Jensen, C., Helms, M., Scheutz, F., Olsen, K. E., Krogfelt, K., Petersen, E., Molbak, K. and Gerner-Smidt, P. (2005) Etiology of diarrhea in young children in Denmark: a case-control study, *J Clin Microbiol*, 43(8), pp. 3636-41.

Opalka, N., Mooney, R. A., Richter, C., Severinov, K., Landick, R. and Darst, S. A. (2000) Direct localization of a beta-subunit domain on the three-dimensional structure of Escherichia coli RNA polymerase, *Proc Natl Acad Sci U S A*, 97(2), pp. 617-22.

Palmer, M. E., Lipsitch, M., Moxon, E. R. and Bayliss, C. D. (2013) Broad conditions favor the evolution of phase-variable loci, *mBio*, 4(1), pp. e00430-12.

Pereira, A. C. M., Britto, J. D., de Carvalho, J. J., de Luna, M. D. and Rosa, A. C. P. (2008) Enteroaggregative Escherichia coli (EAEC) strains enter and survive within cultured intestinal epithelial cells, *Microbial Pathogenesis*, 45(5-6), pp. 310-314.

Petushkov, I., Pupov, D., Bass, I. and Kulbachinskiy, A. (2015) Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription, *Nucleic Acids Res*, 43(12), pp. 5798-809.

Pis Diez, C. M., Juncos, M. J., Villarruel Dujovne, M. and Capdevila, D. A. (2022) Bacterial Transcriptional Regulators: A Road Map for Functional, Structural, and Biophysical Characterization, *Int J Mol Sci*, 23(4).

Porter, M. E. and Dorman, C. J. (2002) In vivo DNA-binding and oligomerization properties of the Shigella flexneri AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis, *J Bacteriol*, 184(2), pp. 531-9.

Prager, R., Lang, C., Aurass, P., Fruth, A., Tietze, E. and Flieger, A. (2014) Two novel EHEC/EAEC hybrid strains isolated from human infections, *PLoS One*, 9(4), p. e95379.

Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences, *Cell*, 135(2), pp. 216-26.

Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., Joukov, V., Lotov, V., Pannu, R., Rudnev, D., Shkeda, A., Weitz, E. M. and Schneider, V. A. (2021) Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV), *Genome Res*, 31(1), pp. 159-169.

Ray-Soni, A., Bellecourt, M. J. and Landick, R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End, *Annu Rev Biochem*, 85, pp. 319-47.

Rhodius, V. A. and Mutalik, V. K. (2010) Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, sigmaE, *Proc Natl Acad Sci U S A*, 107(7), pp. 2854-9.

Rhodius, V. A., West, D. M., Webster, C. L., Busby, S. J. and Savery, N. J. (1997) Transcription activation at class II CRP-dependent promoters: the role of different activating regions, *Nucleic Acids Res*, 25(2), pp. 326-32.

Roberts, J. W. (2019) Mechanisms of Bacterial Transcription Termination, *J Mol Biol*, 431(20), pp. 4030-4039.

Roberts, P. A., Huebinger, R. M., Keen, E., Krachler, A. M. and Jabbari, S. (2018) Predictive modelling of a novel anti-adhesion therapy to combat bacterial colonisation of burn wounds, *PLoS Comput Biol*, 14(5), p. e1006071.

Rojo, F. (1999) Repression of transcription initiation in bacteria, *J Bacteriol*, 181(10), pp. 2987-91.

Rosa, A. C., Mariano, A. T., Pereira, A. M., Tibana, A., Gomes, T. A. and Andrade, J. R. (1998) Enteropathogenicity markers in Escherichia coli isolated from infants with acute diarrhoea and healthy controls in Rio de Janeiro, Brazil, *J Med Microbiol*, 47(9), pp. 781-90.

Rosano, G. L. and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges, *Front Microbiol*, 5, p. 172.

Rossiter, A. E., Browning, D. F., Leyton, D. L., Johnson, M. D., Godfrey, R. E., Wardius, C. A., Desvaux, M., Cunningham, A. F., Ruiz-Perez, F., Nataro, J. P., Busby, S. J. W. and Henderson, I. R. (2011) Transcription of the plasmid-encoded toxin gene from Enteroaggregative Escherichia coli is regulated by a novel co-activation mechanism involving CRP and Fis, *Molecular Microbiology*, 81(1), pp. 179-191.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. and Barrell, B. (2000) Artemis: sequence visualization and annotation, *Bioinformatics*, 16(10), pp. 944-5.

Saecker, R. M., Chen, J., Chiu, C. E., Malone, B., Sotiris, J., Ebrahim, M., Yen, L. Y., Eng, E. T. and Darst, S. A. (2021) Structural origins of Escherichia coli RNA polymerase open promoter complex stability, *Proc Natl Acad Sci U S A*, 118(40).

Santiago, A. E., Ruiz-Perez, F., Jo, N. Y., Vijayakumar, V., Gong, M. Q. and Nataro, J. P. (2014) A Large Family of Antivirulence Regulators Modulates the Effects of Transcriptional Activators in Gram-negative Pathogenic Bacteria, *Plos Pathogens*, 10(5).

Santiago, A. E., Yan, M. B., Hazen, T. H., Sauder, B., Meza-Segura, M., Rasko, D. A., Kendall, M. M., Ruiz-Perez, F. and Nataro, J. P. (2017) The AraC Negative Regulator family modulates the activity of histone-like proteins in pathogenic bacteria, *PLoS Pathog*, 13(8), p. e1006545.

Santiago, A. E., Yan, M. B., Tran, M., Wright, N., Luzader, D. H., Kendall, M. M., Ruiz-Perez, F. and Nataro, J. P. (2016) A large family of anti-activators accompanying XylS/AraC family regulatory proteins, *Mol Microbiol*, 101(2), pp. 314-32.

Savarino, S. J., Fox, P., Deng, Y. K. and Nataro, J. P. (1994) Identification and Characterization of a Gene-Cluster Mediating Enteroaggregative Escherichia-Coli Aggregative Adherence Fimbria-I Biogenesis, *Journal of Bacteriology*, 176(16), pp. 4949-4957.

Schleif, R. (2010) AraC protein, regulation of the L-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action, *FEMS microbiology reviews*, 34, pp. 779-96.

Schneider, T. D. and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences, *Nucleic Acids Res*, 18(20), pp. 6097-100.

Schuller, A., Slater, A. W., Norambuena, T., Cifuentes, J. J., Almonacid, L. I. and Melo, F. (2012) Computer-Based Annotation of Putative AraC/XylS-Family Transcription Factors of Known Structure but Unknown Function, *Journal of Biomedicine and Biotechnology*.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation, *Bioinformatics*, 30(14), pp. 2068-9.

Sheikh, J., Czeczulin, J. R., Harrington, S., Hicks, S., Henderson, I. R., Le Bouguenec, C., Gounon, P., Phillips, A. and Nataro, J. P. (2002) A novel dispersin protein in enteroaggregative Escherichia coli, *J Clin Invest*, 110(9), pp. 1329-37.

Shi, J., Wang, F., Li, F., Wang, L., Xiong, Y., Wen, A., Jin, Y., Jin, S., Gao, F., Feng, Z., Li, J., Zhang, Y., Shang, Z., Wang, S., Feng, Y. and Lin, W. (2022)

Structural basis of transcription activation by Rob, a pleiotropic AraC/XylS family regulator, *Nucleic Acids Res*, 50(10), pp. 5974-5987.

Shi, J., Wen, A., Zhao, M., Jin, S., You, L., Shi, Y., Dong, S., Hua, X., Zhang, Y. and Feng, Y. (2020) Structural basis of Mfd-dependent transcription termination, *Nucleic Acids Res*, 48(20), pp. 11762-11772.

Shimada, T., Ogasawara, H., Kobayashi, I., Kobayashi, N. and Ishihama, A. (2021) Single-Target Regulators Constitute the Minority Group of Transcription Factors in Escherichia coli K-12, *Front Microbiol*, 12, p. 697803.

Shin, M., Kang, S., Hyun, S. J., Fujita, N., Ishihama, A., Valentin-Hansen, P. and Choy, H. E. (2001) Repression of deoP2 in Escherichia coli by CytR: conversion of a transcription activator into a repressor, *EMBO J*, 20(19), pp. 5392-9.

States, D. J. and Gish, W. (1994) Combined use of sequence similarity and codon bias for coding region identification, *J Comput Biol*, 1(1), pp. 39-50.

Stoker, N. G., Fairweathe, N. F. and Spratt, B. G. (1982) Versatile low-copy-number plasmid vectors for cloning in Escherichia coli, *Gene*, 18(3), pp. 335-341.

Sutherland, C. and Murakami, K. S. (2018) An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase, *EcoSal Plus*, 8(1).

Taliaferro, L. P., Keen, E. F., Sanchez-Alberola, N. and Wolf, R. E. (2012) Transcription Activation by Escherichia coli Rob at Class II Promoters: Protein–Protein Interactions between Rob's N-Terminal Domain and the σ70 Subunit of RNA Polymerase, *Journal of Molecular Biology*, 419(3), pp. 139-157.

Telford, W. G., Hawley, T., Subach, F., Verkhusha, V. and Hawley, R. G. (2012) Flow cytometry of fluorescent proteins, *Methods*, 57(3), pp. 318-30.

Thomas, M. S. and Wigneshweraraj, S. (2014) Regulation of virulence gene expression, *Virulence*, 5(8), pp. 832-4.

Tierrafria, V. H., Rioualen, C., Salgado, H., Lara, P., Gama-Castro, S., Lally, P., Gomez-Romero, L., Pena-Loredo, P., Lopez-Almazo, A. G., Alarcon-Carranza, G., Betancourt-Figueroa, F., Alquicira-Hernandez, S., Polanco-Morelos, J. E., Garcia-Sotelo, J., Gaytan-Nunez, E., Mendez-Cruz, C. F., Muniz, L. J., Bonavides-Martinez, C., Moreno-Hagelsieb, G., Galagan, J. E., Wade, J. T. and

Collado-Vides, J. (2022) RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in Escherichia coli K-12, *Microb Genom*, 8(5).

Verma, R., Rojas, T. C. G., Maluta, R. P., Leite, J. L., Nakazato, G. and de Silveira, W. D. (2018) Role of hypothetical protein YicS in the pathogenicity of Avian Pathogenic Escherichia coli in vivo and in vitro, *Microbiol Res*, 214, pp. 28-36.

Wade, J. T. and Struhl, K. (2008) The transition from transcriptional initiation to elongation, *Curr Opin Genet Dev*, 18(2), pp. 130-6.

Warman, E. A., Forrest, D., Guest, T., Haycocks, J. J. R. J., Wade, J. T. and Grainger, D. C. (2021) Widespread divergent transcription from bacterial and archaeal promoters is a consequence of DNA-sequence symmetry, *Nature Microbiology*, 6(6), pp. 746-756.

Washburn, R. S. and Gottesman, M. E. (2015) Regulation of transcription elongation and termination, *Biomolecules*, 5(2), pp. 1063-78.

Winkelman, J. T., Vvedenskaya, I. O., Zhang, Y., Zhang, Y., Bird, J. G., Taylor, D. M., Gourse, R. L., Ebright, R. H. and Nickels, B. E. (2016) Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection, *Science*, 351(6277), pp. 1090-3.

Yang, X. and Lewis, P. J. (2010) The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation, *Transcription*, 1(2), pp. 66-9.

Yasir, M. (2017) *Regulation of Virulence Determination in Enteroaggregative Escherichia coli*. PhD Thesis. University of Birmingham. https://etheses.bham.ac.uk/id/eprint/7251/1/Yasir17PhD.pdf (Accessed: 10 June 2022).

Yasir, M., Icke, C., Abdelwahab, R., Haycocks, J. R., Godfrey, R. E., Sazinas, P., Pallen, M. J., Henderson, I. R., Busby, S. J. W. and Browning, D. F. (2019) Organization and architecture of AggR-dependent promoters from enteroaggregative Escherichia coli, *Mol Microbiol*, 111(2), pp. 534-551.

Zafar, M. A., Sanchez-Alberola, N. and Wolf, R. E. (2011) Genetic Evidence for a Novel Interaction between Transcriptional Activator SoxS and Region 4 of the

sigma(70) Subunit of RNA Polymerase at Class II SoxS-Dependent Promoters in Escherichia coli, *Journal of Molecular Biology*, 407(3), pp. 333-353.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E. and Ebright, R. H. (2012) Structural basis of transcription initiation, *Science*, 338(6110), pp. 1076-80.