REGULATION OF VIRULENCE DETERMINANTS IN ENTEROAGGREGATIVE *ESCHERICHIA COLI*

BY

MUHAMMAD YASIR

A THESIS

SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INSTITUTE OF MICROBIOLOGY AND INFECTION
SCHOOL OF BIOSCIENCES
UNIVERSITY OF BIRMINGHAM
MARCH 2017

Abstract

Escherichia coli is part of the normal flora of the human large intestine but some strains can cause intestinal and extra-intestinal infections. There are many pathogenic strains of *E. coli* and Enteroaggregative *E. coli* (EAEC) is one of them. EAEC is a common cause of diarrhoea in developing and industrialised countries. EAEC has many virulence factors but the most important virulence determinant for diarrhoeal illness is the formation of fimbriae, which help in attachment to human intestinal cells. A typical pathogenic EAEC strain also contains the transcriptional regulator, AggR (the aggregation regulator). AggR regulates the transcription of many genes in EAEC but how AggR achieves this is not still very clear. The objective of this project was to identify the promoters which control the expression of the fimbrial genes and other AggR-dependent virulence genes in EAEC strain 042 and EAEC strain 17-2, and to determine the promoter dependence on AggR, with an aim to locating AggR-binding sites at these promoters.

There are five types of adherence aggregative fimbriae (AAF) systems (*i.e.* AAF/I –AAF/V), classified on the basis of fimbrial subunit sequence, and all are regulated by AggR. In the present study, I have investigated the regulation of the promoters which control the expression of AAF/II, the anti-aggregating protein (*aap*) gene in EAEC 042, present on aggregation adherence plasmid (pAA2), and type 6 secretion system (T6SS) present on chromosome to determine the location of AggR-binding sites. I have also investigated the promoter that controls the expression of the genes encoding AAF/I from EAEC strain 17-2, which are located on the pAA plasmid. These promoters were identified and analysed on the basis of β-galactosidase activity measured from lysates of cells which carried these promoter fragments cloned in *lacZ* expression vector.

In this study, AggR-dependent fimbrial promoters from EAEC 042 were located and transcription start sites were identified. The promoters are present upstream of the fimbrial operons in Region 1 and Region 2 of plasmid pAA2. An AggR-dependent operon promoter was identified upstream of *aggD* (AAF/I) on pAA in EAEC 17-2. The promoters for *aap* and the T6SS have also been located upstream of the *aap* and the *aaiA* gene, respectively and are AggR-regulated. The AggR-binding sites and promoter elements were investigated by deletion analysis and confirmed by introducing site directed point mutations into the promoter fragments. The results were interpreted on the basis of β-galactosidase activity measured from lysates of cells carrying these promoter fragments cloned in *lacZ* expression vector.

The AggR-binding site from the *aafD,* an AggR-dependent promoter was also transplanted into a non-AggR-dependent promoter and it conferred AggR-dependent activity onto that promoter. This semi-synthetic promoter construction also enhances our understanding of AggR-dependent promoters, indicating that the distance between the AggR-binding motif and the -10 hexamer element is important and that 22 bp is the optimal distance. Moreover, AggR interaction with RNA polymerase was investigated by making point mutations in AggR and this indicated that the N-terminus is important for AggR-dependent activation.

Dedicated to

my father, Muhammad Rafi Cheema (Late)

and

my mother, Shamim Akhtar

Acknowledgement

During my PhD, I have got a chance to work with a number of colleagues, whom I want to thank for their time, support and valuable company. First of all, I want to say thanks to Prof. Steve Busby, for accepting me as a PhD student in his group and, being my mentor during all these years. He has been a great source of encouragement, support and valuable suggestions. I have deep gratitude and respect for Dr. Douglas Browning who guide me on every step in the lab. It would have been next to impossible to understand the project and complete this thesis without his inspiring support and guidance. I also want to say special thanks to Prof. Ian Henderson for being my internal assessor and for valuable discussions on the project. I want to say thanks to Dr. Jack Brayant, Dr James Haycocks, Dr Dave Lee and Dr. Tim Wells for their technical advices during laboratory work. A special thanks to Rita Godfrey, for her technical assistance, taking care of lab consumable and instruments, making the working environment safe and friendly.

I want to say thanks to Dr. Laura Sellers, Dr Amanda Rossiter, Dr. Shahid Islam, Ghadah, Bandar, Karen, Zobair, Chiara, and other past and present members of Busby's group, who made the lab environment very friendly and welcoming. I also want to extend my thanks to Prof. Cole and his lab members Clair, Ian, Basema and Jang. I must say thanks to Annette, who was being very helpful throughout my PhD in fixing meetings with Steve.

I want to say special thanks to my parents, siblings, nephews, nieces, family and friends for their support and prayers throughout my career. A special thanks to Faisal Nadeem, Nassar Cheema, Prof. A. Hannan (late) and Malik Aamir for being supportive and encouraging to start my PhD. I also have special thanks for Kiku, Sandy, Bhargav, Absar, Irfan and Sajjad for being there to support without any conditions. I also want to thank Sahara, Tania, Anmol, Kamal, Sabari, Kanika and Archontissa for making home ambience when we all are away from home during all these years. It will be hard to forget delicious dinners, just kidding jokes, arrhythmic songs, school stories, board games and then philosophical comments to support and encourage each other.

Last but not least, My PhD would have not been possible without the sposporship of Darwin Trust of Edinburgh. I want to thank Sir Kenneth Murray (Late) and other members of the Trust for the sponsorhip and support throughout my PhD.

Table of contents

List of Figures

List of Tables

List of abbreviations

A: Adenosine

AggR: Aggregation regulator

Amp^R: Amplicillin-resistance

APS: Ammonium persulfate

ATP: Adenosine triphosphate

bp: Base pair

C: Cytosine

cAMP: Cyclic adenosine monophosphate

CIP: Calf intestinal alkaline phosphate

Cm^R: Chloramphenicol-resistance

CRP: Cyclic AMP receptor protein

CTD: Carboxyl terminal domain

DNA: Deoxyribonucleic acid

DNase: Deoxyribonuclease

dNTP: Deoxynucleoside triphosphate (N= A, C, G or T)

EDTA: Ethylenediaminetetraacetic acid

G: Guanine

GTP: Guanosine-5'-triphosphate

H-NS: Histone like-nucleiod structuring protein

HTH: Helix turn helix

kDa: Kilo Dalton

kb: Kilo base pair

LB: Lysogeny broth

mRNA: messenger RNA

NTD: amino terminal domain

OD: Optical density

ONPG: Ortho-nitrophenyl-b-D-galactopyranoside

PCR: Polymerase chain reaction

(p)ppGpp: Guanosine pentaphosphate

RNA: Ribonucleic acid

RNase: Ribonuclease

RNAP: RNA polymerase

rRN Ribosomal RNA

SDS: Sodium dodecyl sulphate

SD: Shine-Dalgarno

T: Thymidine

TBE: Tris/Borate/EDTA buffer

Tet^R: Tetracycline resistance

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

tRNA: Transfer RNA

Tris: 2-Amino-2-(hydroxymethyl)-1,3-propanediol

v/v: Volume per volume

w/v: Weight per volume

Amino acids standard single letter abbreviations were used.

Chapter 1

Introduction

1.1 Preamble

Unicellular and multicellular organisms have coexisted side by side on the earth for millions of years. Unicellular organisms can live inside multicellular organisms in symbiotic relationships, however, in certain conditions this interaction can become disastrous with the unicellular organism causing detrimental effects to the multicellular organism. Some unicellular organisms such as bacteria and parasites can cause disease in plants, animals and humans and these are called pathogens. Even though, the multicellular organisms evolved hundreds of millions of years ago along with these unicellular organisms, they still are susceptible to these comparatively simpler organisms in certain conditions.

Escherichia coli is one of the bacterial species that colonise the intestinal tract of many animals. It can produce disease in certain individuals and the molecules or structures produced by *E. coli* that help in pathogenesis are called virulence factors (Johnson, 1991). The formation of these virulence factors is a multistage process and the biogenesis of certain factors, at different stages of bacterial growth, is essential for pathogenesis. Thus, the expression of virulence factors at different stages of pathogenesis is tightly controlled. These virulence factors can be controlled at any point of the expression pathway, but control at the level of transcript initiation is the most economic.

This thesis focuses on the transcription regulation mechanisms in one of the pathogenic *E. coli*, *i.e.* Enteroaggregative *E. coli* (EAEC) strains. In the first section of this chapter (1.1-1.6), I describe the characteristics of *E. coli* and EAEC, including the virulence determinants, together with master virulence regulator AggR. In the second part (1.7 and 1.8), I examine mechanisms of bacterial transcription regulation, with most of this information obtained from the studies in *E. coli* K-12. In the last part of the chapter, I detail the AraC family of

transcription regulators and one of its member, AggR, which plays an important role in EAEC virulence.

1.2 *Escherichia coli*

Bacteria are prokaryotic organisms that are found in all environments. In some environments, they co-exist with other organisms. Most bacteria are present as harmless commensals that in some situations, are beneficial to the host organisms but some bacteria cause disease. Bacteria exist in different shapes, from cocci to bacilli and comma shapes, and their sizes can vary from 1μm-20μm. One of the most studied bacterium is *E. coli* that is named after its discoverer, Theodor Escherich, a German physician (Hacker and Blum-Oehler, 2007). *E. coli* is a Gram negative rod and a member of the Enterobacteriaceae. It forms a part of the normal flora of many warm-blooded animals, including human beings. The size of a typical *E. coli* is about 0.5μm in width and about 2μm in length (El-Hajj and Newman, 2015). The bacterium is surrounded by a cell envelope made up of an inner membrane, periplasmic space and an outer membrane (Silhavy *et al.*, 2010). The membranes contain a variety of embedded proteins that help in various functions including transport of molecules across the membranes, secretion of certain proteins and antibiotic resistance (Tenover, 2006; Tseng *et al.*, 2009; Lombard, 2014). Some strains of *E. coli* are surrounded by a polysaccharide capsule that protects the bacterium from phagocytosis (Emody *et al.*, 2003). It has peritrichous flagella that help in its movement and many strains of *E. coli* are highly motile (Tecon and Or, 2016). Some strains also have fimbriae that help in the attachment to different surfaces during colonization of hosts, while other types of fimbriae are also involved in conjugation (Klemm, 1985). The bacterium has a circular chromosome and may contain plasmids that encode for different genes (Carattoli *et al.*, 2005).

Some *E. coli* strains have emerged as pathogens while most are non-pathogenic (Clements *et al.*, 2012). The *E. coli* strain that is mostly used in the laboratory as a model is the non-pathogenic, *E. coli* K-12. The *E. coli* strain K-12 is considered a model organism for the study of different biochemical and molecular processes in bacteria (Kuhnert *et al.*, 1995). Specific mechanisms and processes can be studied in *E. coli* K-12, *e.g.* promoter regulation and protein production, and the mechanisms studied in this model organism can often be applied to other pathogenic bacteria. In the present study, most of the experiments were carried out in *E. coli* K-12 and it is assumed that similar mechanism of gene expression take place in the pathogenic *E. coli.*

1.3 Pathogenic *Escherichia coli*

Most *E. coli* are commensals but some pathogenic strains can cause disease, particularly in children and immunocompromised adults. The most common infections by pathogenic *E. coli* include diarrhoea in children and urinary tract infections. The major pathotypes of *E. coli* that cause diarrhoea are EAEC, Enteroinvasive *E. coli* (EIEC), Enteropathogenic *E. coli* (EPEC), Enterotoxigenic *E. coli* (ETEC), Enterohaemorrhagic *E. coli* (EHEC) and Diffuse adhering *E. coli* (DAEC) (Kaper *et al.*, 2004; Iguchi *et al.*, 2009). These pathotypes have been identified on the basis of specific pathogenic features. EAEC is one of these common pathotypes, associated with diarrhoea in adults and children (Huang *et al.*, 2006a). The exact molecular basis for distinction between the pathogenic and the non-pathogenic EAEC is yet to be unveiled but it has been proven that some strains of EAEC are typical human pathogens (Nataro and Kaper, 1998; Kaper *et al.*, 2004). EAEC as defined by Nataro and Kaper (1998) consists of "*E. coli* strains that do not secrete enterotoxins LT (heat labile toxin) or ST (heat stable toxin), and that adhere to HEp-2 cells in an aggregative adherence (AA) pattern". This

definition seems general and it includes the classical pathogenic strains of EAEC as well as the non-pathogenic strains.

1.4 Pathogenicity of EAEC

EAEC is associated with diarrhoea in children and was reported for the first time by Nataro *et al.* (1987) in Santiago, Chile. In the following years, studies by Bhan *et al.* (1989) and Cravioto *et al.* (1991) also reported EAEC as the causative agent of diarrhoea but during the same era, some investigators found EAEC in healthy individuals as well, without diarrhoea (Echeverria *et al.*, 1992; Gomes *et al.*, 1989). Finally, this controversy about pathogenicity of EAEC was resolved by a study of Nataro *et al.* (1995) , that tested four EAEC strains, 042, JM221, 17-2 and 34b for the causative agent of diarrhoea in healthy volunteers, only EAEC 042 could produce diarrhoea, while the other three strains failed to produce symptoms in volunteers. Another volunteer study by Mathewson *et al.* (1986) has shown that JM221 can cause diarrhoea in some adult volunteers. So, these studies have established EAEC as a pathogenic bacterium that can cause intestinal disease in healthy adults.

EAEC is a cause of diarrhoea in adults and children in both developing and industrialised countries. It is also an important cause of traveller's diarrhoea, and it can cause acute and persistent diarrhoea in the immunocompromised adult patients (Adachi *et al.*, 2001; Huang *et al.*, 2006b). EAEC is reported as a cause of food-borne outbreaks in Europe and other developed countries (Grad *et al.*, 2012; Harada *et al.*, 2007). EAEC can also produce extra-intestinal infections and an outbreak of urinary tract infection is also reported in Denmark (Olesen *et al.*, 2012; Kaur *et al.*, 2010). EAEC can also cause sepsis in some cases (Herzog *et al.*, 2014).

EAEC pathogenicity is now established as one of the most common causes of diarrhoea in the

United States and a large surveillance study in the UK has proven EAEC to be an important pathogen (Nataro *et al.*, 2006; Wilson *et al.*, 2001). In addition, a retrospective analysis of intestinal infectious disease data from the UK has argued for a detailed investigation of EAEC genetic and biochemical properties (Chattaway *et al.*, 2013). Moreover, a recent outbreak of diarrhoea with haemolytic uremic syndrome in Germany that led to significant morbidity and mortality was associated with EAEC-like virulence properties. The EAEC strain O104:H4, responsible for this outbreak, is believed to have acquired a prophage that encodes for shiga toxin 2, causing haemolytic uremic syndrome (Frank *et al.*, 2011). This outbreak led to many investigations, which aimed to correlate the combinations of virulence determinants in other *E. coli* strains. Recently, two new strains have been found to contain a combination of virulence genes from EAEC and EHEC strains (Prager *et al.*, 2014).

The status of EAEC as a pathogen highlights the importance of understanding pathogenesis and the contributing factors. The pathogenesis of EAEC strain 042 (EAEC 042) has been studied in detail and revealed that the disease process involves attachment of the bacterium to human intestinal cells, hyper-secretion of mucus, the production of toxins, and inflammation which result in damage to the mucosa and intestine (Figure 1.1) (Huang *et al.*, 2004). Understanding this pathogenesis was enhanced by molecular and genetic studies, which showed that the AggR protein (aggregation regulator) is involved in the regulation of transcription of many genes important in causing diarrhoea (Sarantuya *et al.*, 2004; Nataro, 2005). Important genes that are regulated by AggR include AAF (aggregative adherence fimbriae) genes, *aap, aatPABCD,* and the genes of a type VI secretion system (T6SS).

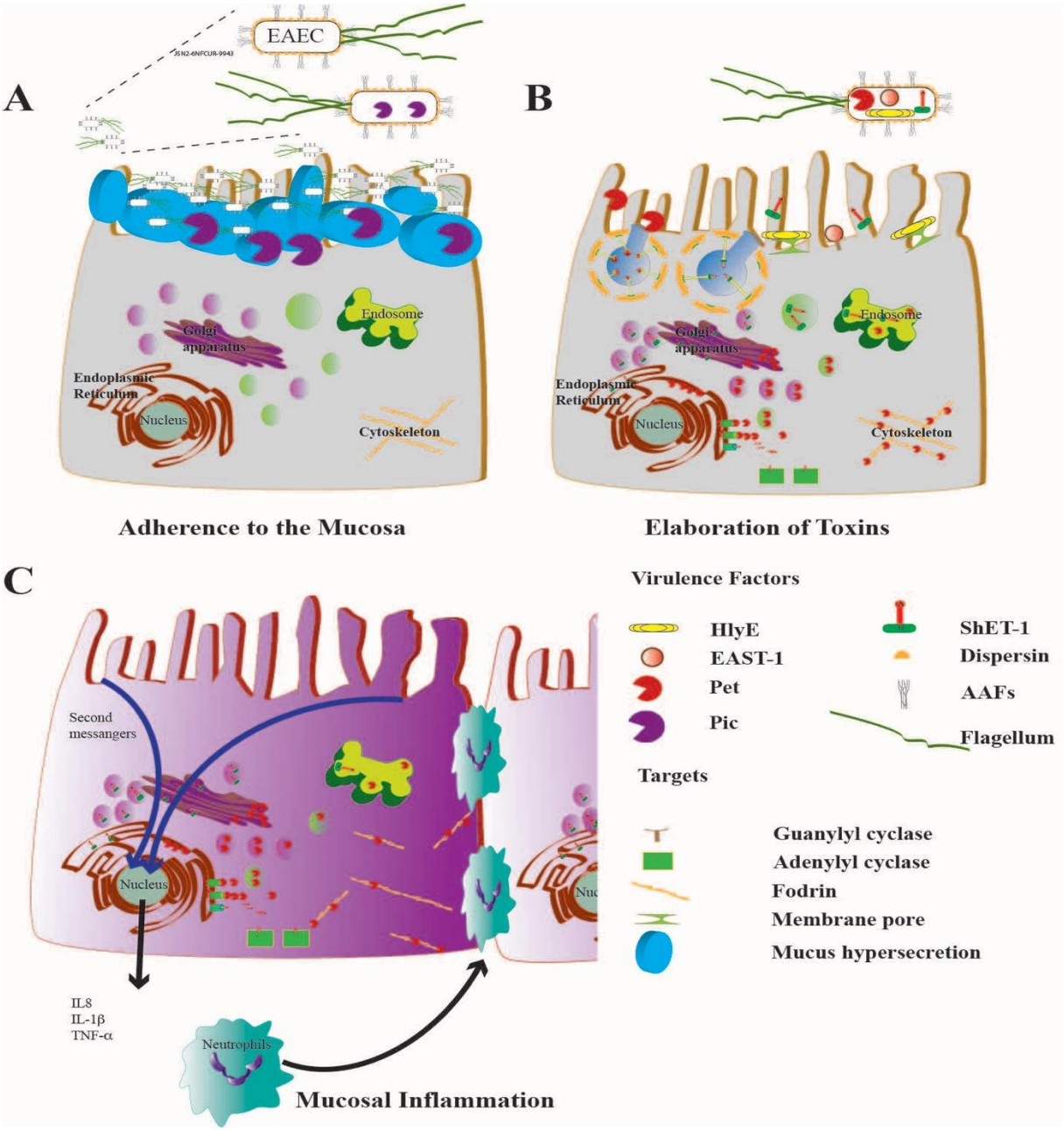


Figure 1.1 Pathogenesis of EAEC

A. The panel shows EAEC attached to human intestinal cells using attachement adherence fimbriae (AAFs). Dispersin prevents the fimbriae from collapsing on to the bacterial cell surface. The protein involved in colonization (Pic) is secreted and hypersecretion of mucus occurs. **B**. The panel illustrates the secretion of toxins such as haemolytic pore forming protein (HlyE), plasmid-encoded toxin (Pet), enteroaggregative *E. coli* heat-stable enterotoxin (EAST-1) and *Shigella* enterotoxin 1 (ShET-1) that are taken up by intestinal cells. **C.** The panel shows the last stage of pathogenesis where the toxins disrupt the cytoskeleton of the intestinal cells and the cells produce inflammatory markers like IL-8, IL-1β and TNF-α. The neutrophils invade the tissue and this leads to the symptoms of the disease. This figure is adapted from Navarro-Garcia and Elias (2011).

1.5 Identification of EAEC

EAEC is a common pathogen and its diagnosis can be achieved using a number of methods. However, the most important is the nature of its specific adherence to HEp-2 (Human Epithelial type 2) cells, *i.e.* aggregation in the form of "a stacked brick pattern". This adherence pattern distinguishes EAEC from other strains that attach to HEp-2 cells, *e.g.* EPEC and DAEC, however, the adherence of EAEC to HEp-2 or Caco cells does not distinguish pathogenic strains from strains isolated from asymptomatic individuals (Franca *et al.*, 2013).

Aggregation adherence is a complex test and it is not feasible to perform in all diagnostic laboratories, so molecular characteristics have been used to find a specific and easy way to perform diagnostic tests on EAEC (Santiago *et al.*, 2014; Aslani *et al.*, 2011). Some parameters show some specificity for EAEC, *e.g.* the presence of *aat, aap, aggR* and *aggA/aafA* within a strain (Franca *et al.*, 2013). Since AggR regulates many virulence genes, there was a possibility that *aggR* would be present in only pathogenic strains. However, it has been found that *aggR*, *aap*, *aat* and other genetic markers are present in both pathogenic and commensal strains, so identification of any of these markers is not the proof of EAEC pathogenicity (Santiago *et al.*, 2014).

In a recent study, a new protein Aar (AggR activated regulator) has been identified in EAEC, and has been shown to repress AggR activity (Santiago *et al.*, 2014). Aar is the first member of the family of AraC negative regulators (ANR), found by Santiago *et al,* (2014). ANR family members down-regulate the AraC family regulator activity by interacting with the AraC family proteins (Santiago *et al.*, 2016). The identification of ANR protein factors could possibly be used in EAEC diagnostics in the future, but, todate no molecular diagnostic tests

have been found to equal the specificity of the aggregation adherence test.

1.6 Virulence determinants of EAEC

EAEC has many virulence factors that include Aggregation adherence fimbriae (AAF), dispersin, toxins, and the transcription regulator AggR (Zamboni *et al.*, 2004). Using microarray analysis Morin *et al.* (2013), found that many virulence determinant genes are regulated by AggR. AggR is a master transcription regulator of EAEC and the *aggR* open reading frame is present on a large virulence plasmid, named as pAA2 (Jenkins *et al.*, 2005). AggR is a small protein of 30 kDa in size and is a member of the AraC family of transcription regulators (Elias *et al.*, 1999). AggR regulates several genes involved in pathogenesis present on plasmid pAA2 (*i.e.* AAF, *aap* and *aat*) (Jenkins *et al.*, 2005). A pathogenicity island of 117 kb is present on the chromosome of EAEC 042 containing the genes for a T6SS, which are also regulated by AggR (Dudley *et al.*, 2006). Thus, AggR regulates plasmid and chromosomally-encoded genes, which are important for EAEC pathogenesis. In this chapter, some of these virulence determinants such as AAF, dispersin, T6SS and plasmid encoded toxin, are discussed.

1.6.1 Attachment Adherence Fimbriae (AAF)

EAEC pathogenesis initiates with adherence to the human intestinal cells with the help of fimbriae and adhesins. For most EAEC strains, a virulence plasmid is present that encodes the genes for attachment adherence fimbriae, and the specific stacked-brick pattern of EAEC is due to AAF (Zamboni *et al.*, 2004). These AAF have some similarity in biogenesis to another fimbriae formation system, the Dr family, present in both uropathogenic *E. coli* (UPEC) and DAEC (Elias *et al.*, 1999). Fimbriae are formed by major fimbrial subunits that are exported to the surface of the bacterium and a cap subunit (adhesin or pilin) is present at the tip of each

fimbrium (Servin, 2005).

Fimbriae formation is a multi-step process, involving secretion of fimbrial and cap subunits into the periplasmic space. These subunits are transported to the periplasm by the Sec system. The Sec system is a molecular machine which transports many proteins across the inner membrane, and thus is not specific for fimbriae subunit transport (Driessen *et al.*, 1998). The fimbrial subunits are transported from the periplasm to the outer membrane by a chaperon-usher system, and assembled on the outer membrane. The chaperone protein, which is present in the periplasm, ensures that the structural subunits (fimbrial and cap subunits) are transported in a folded form. The usher protein is present in the outer membrane and this exports and assembles the structural subunits into the fimbrium (Figure 1.2) (Piatek *et al.*, 2005).

The genes for AAF in EAEC strains are organised in two regions on virulence plasmids; referred to as Region 1 and Region 2 (Figure 1.3 and Figure 1.4). In each system there is one gene for each chaperone, usher, fimbrial subunit, pilin subunit and the master transcription regulator of EAEC AggR (Czeczulin *et al.*, 1997). On the basis of sequence differences of the major subunits of the fimbriae, five AAF systems have been identified in EAEC (Prager *et al.*, 2014). The major subunits of fimbriae in AAF/I, AAF/II, AAF/III, AAF/IV and AAF/V systems are encoded by *aggA, aafA, agg3A, hdaA* and *aaf5A,* respectively (Ito *et al.*, 2014). However, there are some EAEC strains that do not have any recognised AAF system (Prager *et al.*, 2014). The AAF systems AAF/I present in EAEC strain 17-2 and AAF/II found in EAEC strain 042 have been studied extensively, and these will be discussed in more detail in this chapter.

AAF/I is found in EAEC strain 17-2 and the genes are encoded on the pAA plasmid. The

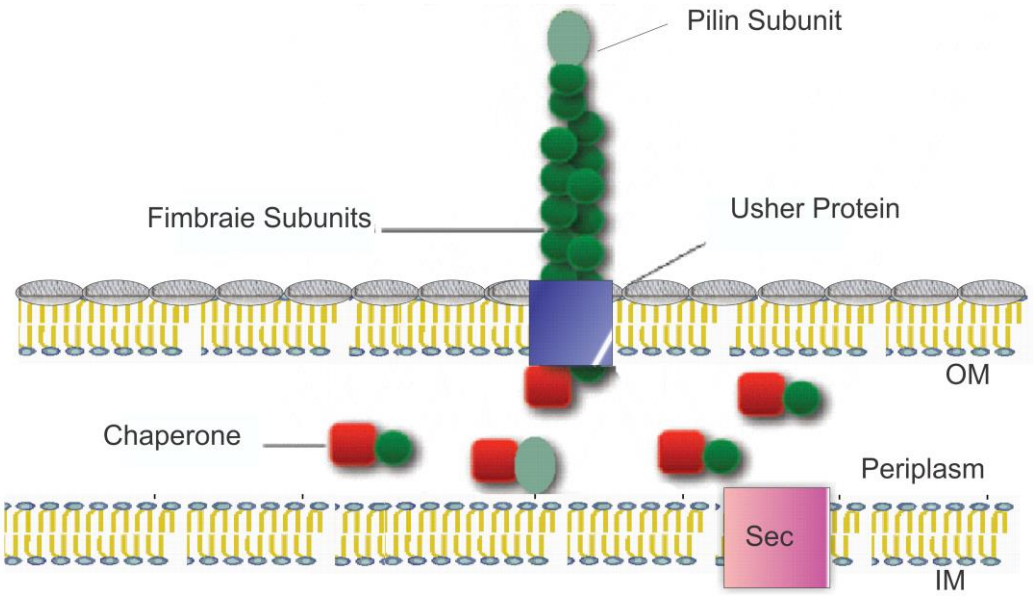


Figure 1.2 The chaperone usher system

The green circles represent fimbrial subunits and the red squares, the chaperone proteins. The chaperone proteins carry fimbrial subunits to the usher protein (blue). Usher protein exports the fimbrial subunits to make the fimbrium. The adhesion subunit (light green), helps in the attachment of EAEC to surfaces. OM and IM stand for outer and inner membrane, respectively. The pink box in the inner membrane represents the Sec system which transports subunits into the periplasm. This figure is adapted from Servin (2005).

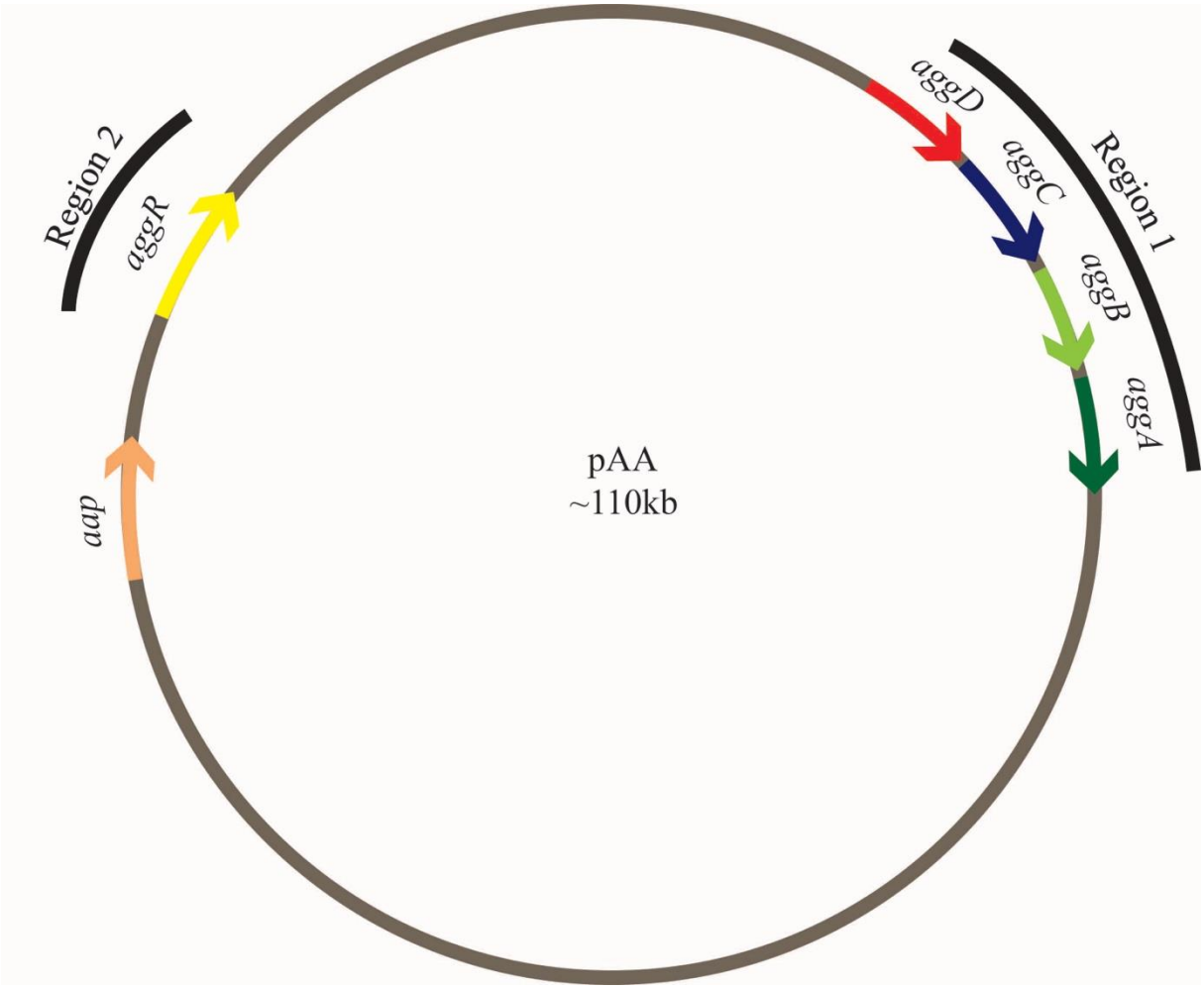


Figure 1.3 A schematic representation of the pAA plasmid

The figure shows a schematic diagram of the pAA plasmid from EAEC strain 17-2 and the two regions that are important for AAF/I (attachment adherence fimbriae). Region 1 encodes *aggD* (a chaperone protein), *aggC* (the usher protein), *aggB* (the fimbrial adhesin) and *aggA* (the fimbrial subunit) and Region 2 encodes *aggR* (a transcription activator). Another AggR regulated gene, *aap* (dispersin), is located uspstream of *aggR* (Nataro *et al.*, 1993). This diagram is not to scale.

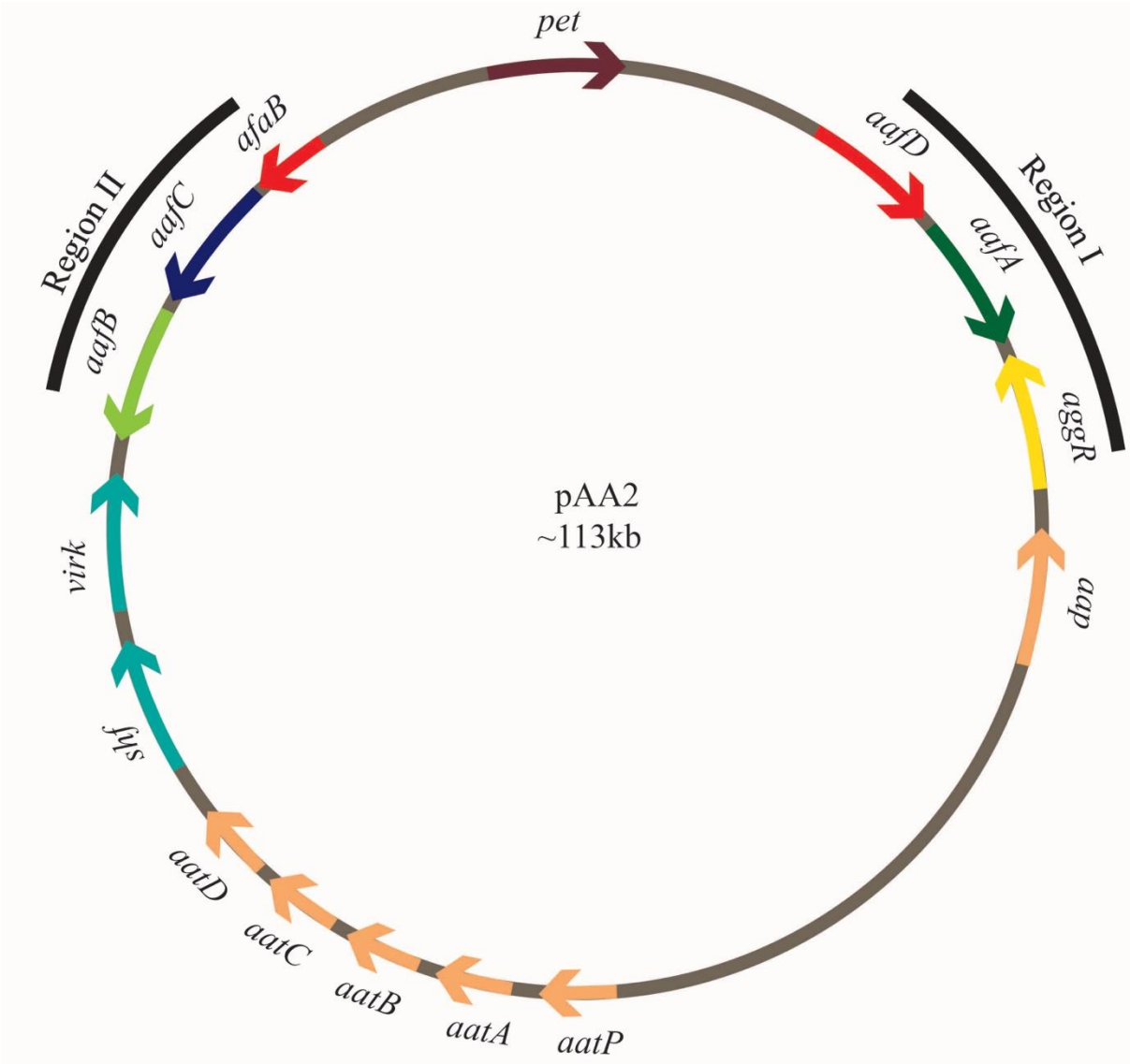


Figure 1.4 A schematic representation of the pAA2 plasmid

The figure shows a schematic diagram of the pAA2 plasmid from EAEC strain 042 and the two regions that encode AAF/II (attachment adherence fimbriae). Region 1 encodes *aafD* (a chaperone protein), *aafA* (the fimbrial subunit) and *aggR* (a transcription activator) and Region 2 encodes *afaB* (a chaperone pseudogene), *aafC* (the usher protein) and *aafB* (the fimbrial adhesin). Plasmid encoded toxin (*pet*) present between the two fimbriae encoding regions is also shown. A number of other genes involved in EAEC pathogenesis that are present on pAA2, are also shown in the diagram which includes *aap*, *aatPABCD*, *shf* and *virK* (Fujiyama *et al.*, 2008). This diagram is not to scale.

genes for AAF are present in two regions. The structural genes are present in Region 1 and include *aggDCBA,* while Region 2 codes for transcription regulator, AggR (Figure 1.3). The *aggD* gene encodes for the chaperone, *aggC* for the usher protein, *aggB* for the adhesin subunit and *aggA* for the major pilin subunit (Savarino *et al.*, 1994). The expression of this chaperone-usher system genes is regulated by AggR, encoded in Region 2. AAF/II is found in EAEC strain 042, and the genes are carried by the pAA2 plasmid. The genes for AAF/II are also present in two regions. Region 1 contains *aafD, aafA* and *aggR* while Region 2 has *afaB, aafC* and *aafB* (Figure 1.4). The genes present in Region 1 encodes for the chaperone (*aafD*), the pilin subunit (*aafA*) and a transcription regulator (*aggR*). Region 2 encodes for the usher protein (*aafC*) and the fimbrial adhesin (*aafB*). Region 2 also carries *afaB*, which is a pseudogene that does not generate functional protein. The *afaB* pseudogene has a sequence identity of 57% to the chaperone encoding gene (*aafD*) of AAF/II and also has similarity in sequence to *afaB* of Dr. family of adhesins (Elias *et al.*, 1999). Thus, although AAF/I and AAF/II are encoded by same number of genes and proteins, but there are differences in the organisation of these genes on respective plasmids.

1.6.2 Dispersin and the Aat secretion system

On plasmid pAA2, the *aap* gene is located upstream of *aggR*, and it encodes an anti-aggregation protein that associates with the EAEC cell surface (Figure 1.5) (Sheikh *et al.*, 2002). This protein prevents the bacterial fimbriae from collapsing onto the cell surface, helping cell dispersal and hence the protein has been named dispersin. It has been shown, that *aap* mutants are more prone to auto-agglutination than wild type EAEC 042, and mutants are less able to penetrate through mucin layer of the gut. Thus, the characteristics conferred by dispersin are important in the pathogenesis of EAEC 042 (Sheikh *et al.*, 2002). The precise mechanism by which dispersin is translocated is still not clear but an entero<u>a</u>ggregative <u>A</u>BC

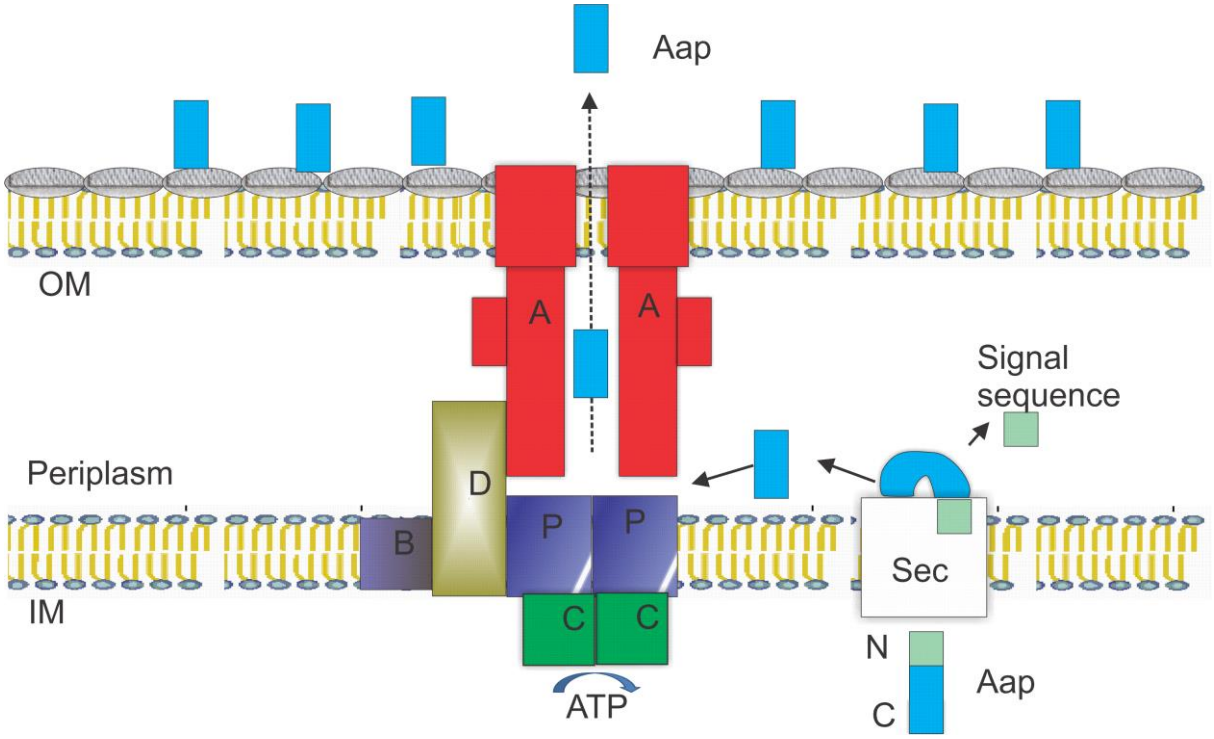


Figure 1.5 The model for dispersin secretion

The panel shows the working model of dispersin secretion. Dispersin (Aap) is transported independently of the Aat secretion system into the periplasm by the Sec system. AatA is located within the outer membrane and part of it protrudes into the periplasm, acting as the point of secretion for Aap across the outer membrane. The AatPBCD complex is embedded in the inner membrane, and it is proposed that AatD contacts AatA in the outermembrane. This figure has been adapted from Nishi *et al.* (2003).

transporter (Aat) system present in EAEC 042, has been shown to translocate dispersin across the cell envelope (Figure 1.5) (Nishi *et al.*, 2003). This system is encoded by five genes, *aatPABCD,* present on pAA2 (Figure 1.4) and deletion of any of the *aat* locus disrupts transport of dispersin outside the cell (Sheikh *et al.*, 2002). The *aat* operon and *aap* gene are located on pAA2 in different transcription units but both are regulated by the master virulence regulator, AggR (Sheikh *et al.*, 2002).

Figure 1.5 illustrates that translocation of dispersin from the cytoplasm to the periplasm is independent of the Aat system and is transported by the Sec system. However, the secretion of dispersin from the periplasm to outside the bacterium requires the Aat system. The roles of AatPABCD have been predicted on the basis of their structural resemblance to the proteins of secretion systems, and the predicted functions are supported by deletion and mutation analysis experiments (Nishi *et al.*, 2003). Thus, AatA is embedded in the outer membrane and has homology with the TolC protein, an outer membrane protein in T1SS (type 1 secretion system) found in *E. coli* (Figure 1.5) (Imuta *et al.*, 2008). AatP shows similarity to ABC transporter permeases, while AatC has structural homology with ATP-binding proteins. AatB does not show any significant amino acid homology to any known protein and AatD is similar to an acid inducible protein in *Borrelia burgdorferi* (Nishi *et al.*, 2003).

Hence, a working model for dispersin secretion is that dispersin first translocates independently of this Aat system into the periplasmic space. The subsequently secretion of dispersin from the periplasmic space to the outer environment is through the AatA channel in the outer membrane that requires energy provided by ATPases (AatP and AatC) present in inner memberane. Dispersin is “conducted” to AatA, in the outer membrane, by either AatD or by both AatD and AatB, acting as adaptor proteins.

1.6.3 The Aai Type VI secretion system of Enteroaggregative *E. coli*

Bacterial proteins and toxins are secreted from the cell, either into the environment or into other prokaryotic and eukaryotic cells by various secretion systems. The secretion of effector proteins by the T6SSs is a complex process, as the secreted molecules are injected directly into other cells (Tseng *et al.*, 2009). This requires a number of proteins organised in a complex manner across the cell envelope and effector proteins are assembled inside the cytoplasm and secreted through the entire cell envelope into the target cell (Cascales, 2008). T6SSs are encoded by large gene clusters and even though the number of the genes varies among organisms, some of the components are conserved in all the organisms that form a functional T6SS (Cascales, 2008; Murdoch *et al.*, 2011). These conserved components are called "the core components" and genes encoding these components are often named *tss* (type six secretion) (Figure 1.6A). The other components that are not present in all T6SS, are called "accessory components" and the genes encoding these are named as *tag* (type six associated genes) (Shalom *et al.*, 2007).

In both EAEC strains 17-2 and 042, the T6SS has three important components, a) the membrane complex, b) the baseplate and c) a phage-tail like complex. The membrane complex is a transmembrane assembly that holds the baseplate. The membrane complex of T6SS consist of three membrane associated components; TssL, TssM and TssJ (Aschtgen *et al.*, 2010; Durand *et al.*, 2015). A recent study by Zoued *et al.* (2016) has shown that another protein TssA is also required for the baseplate to attach to the membrane complex. The baseplate is formed of the TssEFGK proteins and a valine–glycine repeat protein G, VgrG. Note that VgrG is also serves as the spike used to puncture the host membrane (Figure 1.6B). The baseplate is a platform where the effector molecules assemble in the form of a long tube that resembles a phage tail like structure. The phage tail is formed from hexamers of

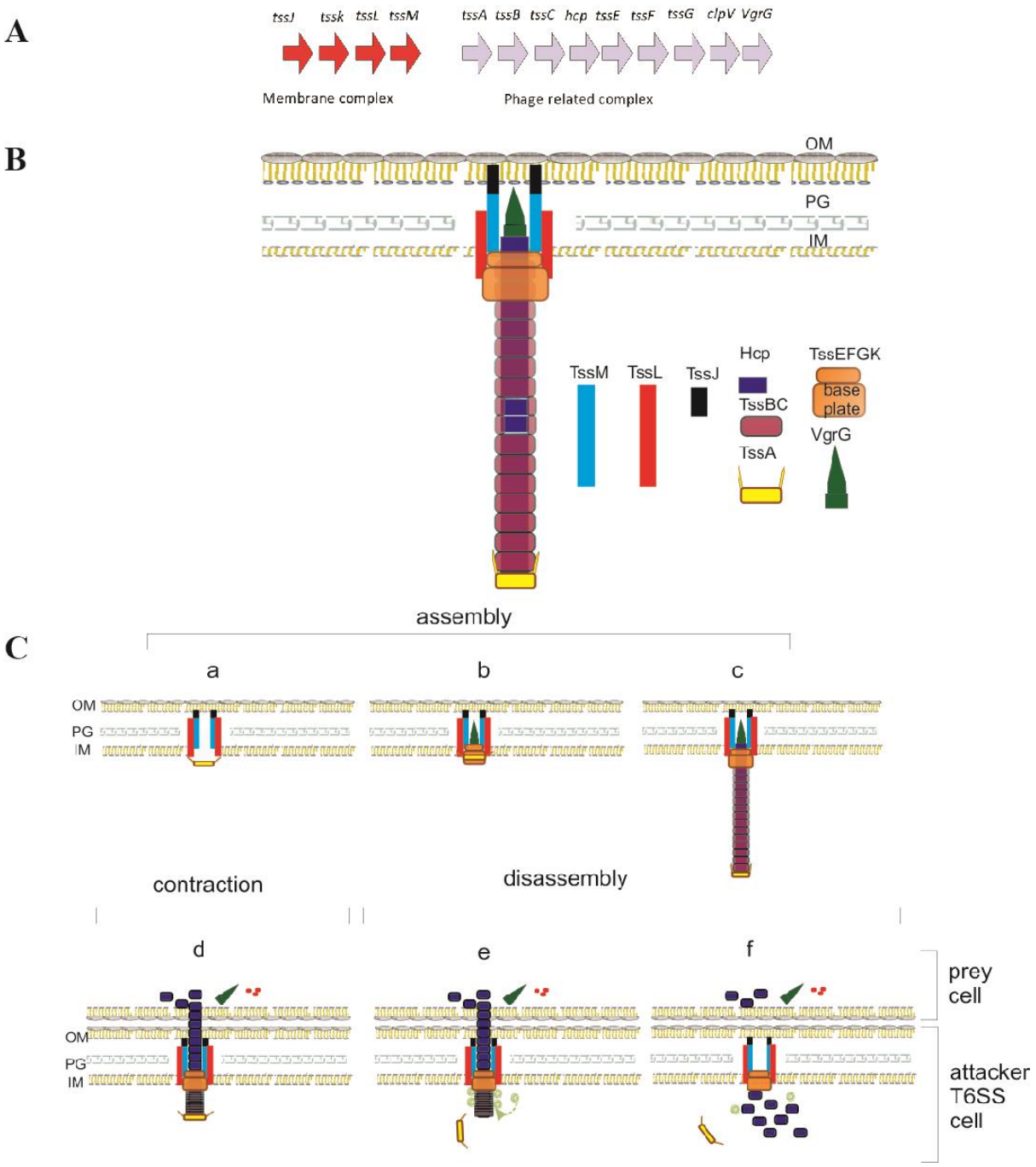


Figure 1.6 A schematic overview of T6SS protein secretion

A. The genes encoding a T6SS: The arrows represent genes encoding the core T6SS. There are two clusters of genes encoding the membrane complex and the phage related complex of T6SS.
B. **The organization of the T6SS:** This represents how the T6SS assembles and functions. TssM, TssL, and TssJ represent the membrane complex. VrgG and the baseplate is a platform on which assembly of the tail like structure is formed composed of Hcp and TssBC.
C. **A working model of T6SS protein secretion:** The figure represents the three stages of T6SS protein secretion *i.e.* assembly of all the proteins necessary for function, contraction of the tail to deliver the effector protein into the target cell and disassembly of the proteins after secretion. This figure has been adapted from Zoued *et al.* (2014) and Zoued *et al.* (2016).

haemolysin co-regulated protein (Hcp) stacked over one another to form a tube like structure on the baseplate. The interaction of Hcp hexamers with each other is not particularly strong but it is stabilised by a contractile sheath formed of TssB and TssC (Lossi *et al.*, 2013).

The T6SS works in three steps; in the first step, components (such as Hcp) assemble into the phage-tail like structure on the baseplate, and the TssBC sheath covers the tail. In the second step, the TssBC sheath contacts, making VgrG and Hcp passes through the membrane complex channel. VgrG punctures the membrane of the target cells, and the toxins (*e.g.* Hcp) are delivered into the target cell. The final step is disassembly of the system after the delivery of product (Figure 1.6C) (Silverman *et al.*, 2011).

T6SSs help bacteria to secrete effector proteins directly into other cells, which can aid in competing with other bacteria in different ecological environments. These systems can also be involved in pathogenesis by delivering effector proteins directly into eukaryotic cells. They have originally reported in animal pathogens but have now also been found in some human pathogens as well *i.e. Vibrio cholera*, *Edwardsiella tarda, Pseudomonas auroginosa* and EAEC (Pukatzki *et al.*, 2009).

Although, a T6SS can help pathogens compete with other bacteria and can influence pathogenesis, its exact role in EAEC biology is unclear (Sana *et al.*, 2012). There are two clusters of genes *i.e. sci-1* and *sci-2* which are located on the bacterial chromosome that encode for two T6SSs in EAEC. The T6SS, encoded by *sci-1* is involved in biofilm formation but the role of the T6SS encoded by *sci-2* is yet to be determined (Aschtgen *et al.*, 2008). The expression of *sci-2* is controlled by AggR and this regulation may indicate that the *sci-2* encoded T6SS has a role in pathogenesis, as AggR-regulates many other virulence factors (Dudley *et al.*, 2006).

The *sci-2* gene cluster is located in a pathogenicity island (PAI) in EAEC 042, and as genes of *sci-2* are regulated by AggR, it has been termed the AggR Activated Island (*aai*). The proteins encoded by these genes have homology with proteins that form the core system of T6SS present in other bacteria. AaiC is secreted by the *sci-2*, and AaiC is homologue of the Hcp-1/Hcp-2 proteins, secreted by the *sci-1* T6SS of EAEC 042 (Dudley *et al.*, 2006). The exact roles of many genes in the secretion of AaiC are unknown, but it has been observed that genes *aaiA* to *aaiP* are sufficient for secretion of AaiC (Dudley *et al.*, 2006). Furthermore, AaiC cannot be secreted in the absence of the *sci-2* encoded T6SS components, even when the *sci-1* is intact, indicating that the *sci-1* encoded T6SS does not recognise the AaiC as a substrate.

1.6.4 Plasmid-encoded toxin

As the name indicates, the Plasmid-encoded toxin (Pet) gene is present on the pAA2 plasmid where it is located between the two AggR-regulated fimbrial operons (Figure 1.4). Pet is a protein toxin produced by pathogenic prototype EAEC strain 042 and is important for virulence of EAEC 042 (Betancourt-Sanchez and Navarro-Garcia, 2009). It is a member of the autotransporter family of proteins and acts as a serine protease. Pet has identity with other autotransporter toxins, for example EspP from EPEC (49%), EspC from EHEC (45%) and SepA from *Shigella flexneri* (31%) (Eslava *et al.*, 1998). Moreover, Pet is not restricted to EAEC but it has also been found in atypical EPEC strains (Ruiz *et al.*, 2014).

The Pet protein consists of a signal domain, a passenger domain and β-barrel domain (Domingo Meza-Aguilar *et al.*, 2014). The Pet toxin, like other autotransporters is transloacted to the periplasm by the Sec system. As Pet is member of autoransporters, it is capable of secretion from the periplasm without the requirement of a dedicated transport system. The β-barrel domain makes a transmembrane channel and the passenger domain

moves through the channel, displaying Pet outside the bacteria. The Pet toxin then undergoes an autocleavage reaction, secreting the Pet toxin from EAEC into the intestinal environment or medium (Domingo Meza-Aguilar *et al.*, 2014).

The mode of action of Pet is the alteration of the cytoskeleton of human intestinal cells. This results in the loss of actin stress fibers followed by cell rounding and detachment of cells (Henderson *et al.*, 2004; Betancourt-Sanchez and Navarro-Garcia, 2009). The production of Pet is one of the important virulence determinants of EAEC pathogenesis, and the studies so far have shown that *pet* expression is regulated by FIS and CRP, two important global transcription regulators of the *E. coli* (Rossiter *et al.*, 2011; Rossiter *et al.*, 2015).

1.7 Bacterial transcription

Transcription is the process in which the base sequence of one strand of DNA is used as a template to produce an RNA copy. The transcription process results in the production of various types of RNA molecules, *i.e.* small RNA molecules, non-coding RNAs, ribosomal RNA (rRNA), transfer RNA (tRNA) and messenger RNA (mRNA). mRNA is used as a template to make protein, by a process termed translation. A single bacterium has thousands of genes and not all of them are transcribed at the same time. Most bacteria are very selective in transcribing their genes, only switching on genes that are required for survival in a particular environment. Transcription initiation is an important step that controls expression of different genes in bacteria. The genes are regulated in response to certain growth and environmental condition and interaction with other cells. Transcription regulation requires the transcription machinery and this includes the DNA-dependent RNA polymerase (RNAP), DNA promoter sequences and numerous transcription factors (Murakami and Darst, 2003).

1.7.1 RNA polymerase

The bacterial multisubunit DNA-dependent RNAP enzyme plays the central role in transcription. The enzyme exists in two forms, core-RNAP and holo-RNAP enzyme. In *E. coli* the core RNAP enzyme has five subunits that include the beta (β), beta prime (β'), the omega (ω) and two alpha (α) subunits (Figure 1.7). RNAP which has a sigma (σ) subunit associated, is termed the holoenzyme (Murakami *et al.*, 2002). Structural studies of RNAP have shown that the shape of the RNAP enzyme resembles a crab claw. One part of the claw is formed by the β-subunit, while the β'-subunit contributes to the other part. The two α-subunits *i.e.* αI and αII are attached to the β and β'-subunits, respectively at the "hinge" of the claw. The small ω-subunit is attached to the β'-subunit at the bottom of the claw. The small ω subunit helps in the assembly of a functional core enzyme (α2ββ'ω), acting like a chaperone (Mathew and Chatterji, 2006).

In RNAP, the β and β' subunits are held together by two alpha subunits. Each alpha subunit has two domains, amino terminal domain (αNTD) that binds to the β and β' subunits and carboxyl terminal domain (αCTD), that binds to the promoter sequences and can interact with transcription factors (Browning and Busby, 2004). The two domains of α subunits are connected by a flexible linker that provides the flexibility to the αCTD to bind at feasible position during transcription (Figure 1.7) (Blatter *et al.*, 1994; Ebright and Busby, 1995).

The transcription processes comprise of many stages: promoter recognition, open complex formation, transcription initiation, promoter escape, elongation and termination. The process starts when a holo-RNAP (core RNAP containing the σ subunit) binds a promoter sequence, and, after the initiation of transcription, the σ subunit detaches itself from core-RNAP. The last two stages of transcription namely, elongation and termination are carried out by core-

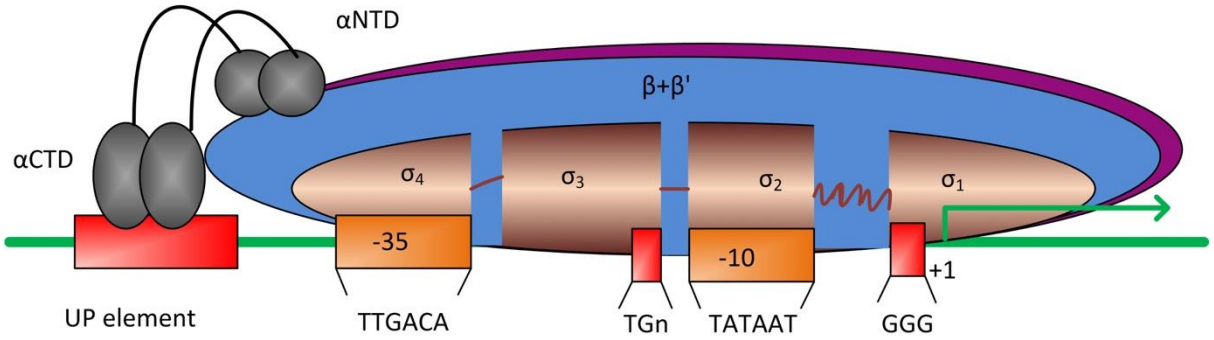


Figure 1.7 A schematic diagram of RNAP holoenzyme bound to a promoter

The figure illustrates the binding of RNAP to a promoter. The αCTDs of RNAP attached to the UP-element of the promoter and regions σ_1, σ_2, σ_3 and σ_4 of RNAP are in contact with the discriminator, the -10 hexamer (consensus TATAAT), the extended -10 (TGn) and the -35 hexamer sequences (consensus TTGACA), respectively. Adapted from Browning and Busby (2004).

RNAP. Only the holo-RNAP has the ability to initiate transcription at specific loci as the core-RNAP cannot recognise specific promoter sequences (Murakami *et al.*, 2002). The σ subunit enables the RNAP to recognise certain DNA sequences, and this subunit is often called the specificity sigma unit (Campbell *et al.*, 2002). The σ subunit is a multidomain protein consisting of independently folded domains joined by linkers (Feklistov *et al.*, 2014).

Most bacteria have multiple sigma factors that help them to recognise different types of promoters. In *E. coli* there are 7 different σ subunits and each regulates a different set of promoters. The most abundant is σ^{70} and it regulates most of the genes in *E. coli*, and so is called the housekeeping σ factor. The σ^{70} associates with the core-RNAP more often as compared to other σ factors and it confers transcription of most of the important genes of the organism (Maeda *et al.*, 2000). The other *E. coli* sigma factors include σ^{54} (nitrogen stress), σ^{38} (stationary phase), σ^{32} (heat shock), σ^{28} (flagellar), σ^{24} (periplasmic proteins and heat stress), σ^{19} (iron transport) and these regulate promoters of specific genes (Cook and Ussery, 2013). It has been observed that each σ factors binds to a specific promoter sequence, so each sigma factor regulates different set of genes (Cook and Ussery, 2013).

Structural and sequence analysis has revealed that *E. coli* σ^{70} consists of four conserved domains namely σ_1, σ_2, σ_3 and σ_4. The σ domains are attached to each other by flexible linkers and each domain recognise specific promoter element (Figure 1.7) (Murakami and Darst, 2003; Feklistov *et al.*, 2014). The σ subunit has three basic functions *i.e.* to recognise a promoter sequence, to align the RNAP to the transcription start site and to help in unwinding of the DNA (Feklistov and Darst, 2009; Bae *et al.*, 2015).

1.7.2 Promoters

RNAP binds to specific DNA sequences called promoters prior to the initiation of

transcription. It may initially bind directly to a promoter sequence or scan along the DNA until it finds one. The attachment of RNAP to the promoter results in the closed promoter complex as the DNA strands are still in the double helical form. In the next step, the double stranded DNA helix around the transcription start site is unwound and this structure is termed the open complex (Rivetti *et al.*, 1999).

A typical σ^{70} dependent promoter has two principal conserved DNA elements, the -10 hexamer element (consensus 5'-TATAAT-3') and the -35 hexamer element (consensus 5'-TTGACA-3'), located approximately 10 and 35 bp upstream of the transcription start site (+1), respectively. It is important to note that -10 or -35 hexamer elements of naturally occurring promoters rarely match this consensus sequence (Busby and Ebright, 1994). These two promoter elements are separated by 17±1 bp and the sequence of this spacer region can be important for promoter efficiency (Mulligan *et al.*, 1985; Singh *et al.*, 2011). Beside these two principal promoter elements, some promoters may have an extended -10 element or a sequence upstream of the -35 hexamer element, called the UP-element. The extended -10 element is a TG motif located one base pair upstream of the -10 hexamer and the UP-element is ~20 basepair A/T rich sequence upstream of the -35 hexamer element (Busby and Ebright, 1999). In addition to these, there can be a discriminator element (5`-GGG-3`) present in the -4 to -6 region of the promoter which can play an important role in RNAP open complex stability (Lee *et al.*, 2012; Feklistov, 2013; Gummesson *et al.*, 2013).

The different promoter elements are recognised primarily by different σ domains and each σ subunit domain is necessary for the recognition of certain promoter element (Busby and Ebright, 1994). The -10 hexamer element is recognised by domain σ_2 and the -35 hexamer sequence by domain σ_4. The extended -10 promoter element is recognised by domain σ_3 and the UP-element is recognised by the CTDs of the RNAP α-subunits (Busby and Ebright,

1994). Thus, the domains 4, 3, 2 and 1 of the σ subunit recognise the -35 hexamer element, the extended -10 hexamer element, the -10 hexamer element and the discriminator element, respectively (Lee *et al.*, 2012).

Feklistov (2013) has highlighted the fact that some of the promoter elements *e.g.* the -10 hexamer element is recognised by RNAP when the DNA is single stranded while other are recognised, when the DNA is double stranded *e.g.* extended -10 hexamer element and -35 hexamer element in *Thermus aquaticus*. The β' and σ^{A}_{3} domain of sigma subunit recognise the extended -10 hexamer element when DNA is double stranded and in promoters where the extended -10 hexamer element is not present; these domains interact with -17 to -13 nucleotides. The -10 hexamer element is recognised by σ^{A}_{2} when the DNA is single stranded (Bae *et al.*, 2015).

Beside the core promoter elements, the UP-element interaction with the αCTDs of RNAP is important and its effect on promoter activity can vary from 2-fold to more than 90-fold at naturally occurring promoters. The UP-element sequence consists of A or T tracts and the UP-element has two subsites. Each subsite interacts with one αCTD subunit. The αCTD of RNAP interacts with the UP-element by binding the minor grove (Busby and Ebright, 1999). However, there are some reports of promoters where the interaction of αCTD and the UP-element has no effect or even a negative effect on promoter expression (Ross *et al.*, 1998). This could be due to inappropriate placement of the UP-element with respect to the core promoter or the presence of an A tract on the UP-element subsite closer to -35 element (the proximal subsite) can enhance the promoter affinity for RNAP so that RNAP escape becomes difficult (Tagami and Aiba, 1999; Ellinger *et al.*, 1994). These findings suggest that UP-element interaction with αCTD can be productive as well as non-productive, depending upon the core promoter sequence and UP-element position (Gourse *et al.*, 2000).

It is worth mentioning that no naturally occurring promoter has all promoter elements which match the consensus sequence (Browning and Busby, 2004). A study by Miroslavova and Busby (2006) showed that, a semi-synthetic promoter with an exact match to consensus sequences on all the core promoter elements exhibited weak promoter activity. It was suggested that RNAP binds too tightly to the promoter such that RNAP escape becomes difficult resulting in reduced promoter activity. The strength of a promoter can be described as how efficiently the elongation complex clears the promoter and is available to the next RNAP (Brunner and Bujard, 1987). To understand the process of transcription: initiation, elongation and termination are discussed below.

1.7.3 Transcript initiation and open promoter complex formation

After promoter recognition and formation of the closed complex, the DNA double helix bends upstream of the -35 hexamer element and wraps around the back of RNAP, resulting in an intermediate promoter complex (Saecker *et al.*, 2011). The next step in transcript initiation is the transition of the intermediate complex into the open complex, that includes unwinding of the DNA strands and the template strand moving into the active site of RNAP (Dehaseth and Helmann, 1995). The stable intermediate promoter complex leads to unwinding of the DNA double helix by W433 of the σ^{70} subunit of RNAP acting like a wedge. The -10 hexamer element is recognised when the DNA is single stranded, with the A_{-11} of the -10 hexamer element on the non-template strand flipping into pocket in the σ subunit (Feklistov and Darst, 2011). This starts DNA melting and then T_{-7} (from the -10 hexamer element of the non-template strand) also flips into another pocket on the σ subunit and the template strand moves into the active site of RNAP (Feklistov, 2013). The active site is occupied by region $\sigma_{1.1}$ before open complex formation, thus, open complex formation is accompanied by movement of region $\sigma_{1.1}$ from this position to clear the channel for DNA repositioning (Murakami and

Darst, 2003). The different steps in this process, described above, are illustrated in the Figure 1.8.

The open promoter complex leads to transcript synthesis at the promoter but the RNA exit channel in RNAP is blocked by a loop, in $\sigma_{3.2,}$ which is in the way of the growing nascent RNA chain. When the RNA transcript reaches to length of 6 nucleotide, it displaces the loop from its position and RNA elongation can occur (Figure 1.8D) (Basu *et al.*, 2014). This displacement of the $\sigma_{3.2}$ loop is important for the dislodging domain σ_4 from the β flap and finally separation of domain σ_4 from the -35 hexamer element. This dissociation leads to complete separation of all the σ subunit from RNAP and the core-RNAP can continue transcription from this point (Figure 1.8E and F) (Mukhopadhyay *et al.*, 2001; Korzheva *et al.*, 2000).

1.7.4 Transcript elongation and termination

When the length of the nascent transcript reaches 8-9 nucleotides, RNAP leaves the promoter elements and starts moving along the DNA template, this is called transcription elongation. The release of RNAP from the promoter elements results in promoter escape (Robb *et al.*, 2013). The transition of the transcript initiation complex into the elongation complex can be unsuccessful sometimes and the RNAP often releases small RNA fragments as a result of abortive transcription (Kapanidis *et al.*, 2006). In this instance RNAP pulls downstream DNA into the enzyme transcribing it without releasing the upstream DNA, it results in the accumulation of single stranded DNA in the RNAP active site, producing huge stress and the process is called DNA scrunching (Zuo and Steitz, 2015). This stress can be released in two ways, first by the release of abortive transcripts with the small RNA fragments being released and RNAP stays in contact with the promoter elements. The second possibility is successful transcription initiation, this results in the escape of RNAP from promoter and the RNA-DNA

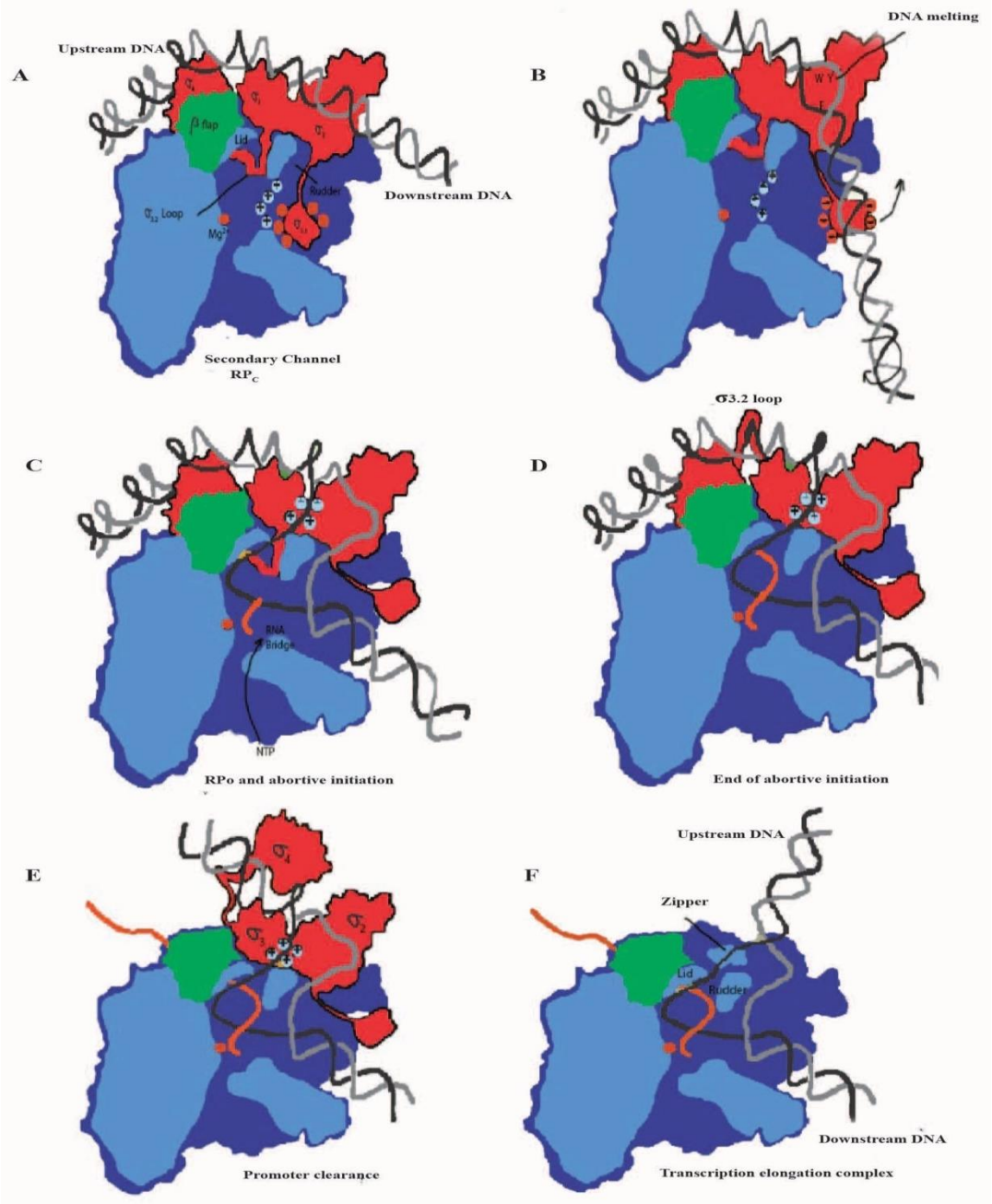


Figure 1.8 Initiation of transcription at a bacterial promoter

This figure shows the interaction of RNAP with promoter DNA during the different stages of transcription initiation.
A. The binding of RNAP to DNA and the formation of the close promoter complex
B. Bending of the DNA strand and movement of $\sigma_{1.1}$ from its position to adjust template strand into active site
C. Melting and unwinding of the DNA to form the open promoter complex.
D. Displacement of domain $\sigma_{3.2}$ by the growing RNA chain
E. The weakening of σ factor binding to the core enzyme
F The transcription elongation, movement of RNAP along the DNA template
This figure has been adapted from Murakami and Darst (2003).

hybrid stays intact, with RNAP moving into the transcript elongation phase (Kapanidis *et al.*, 2006). Transcription elongation continues until transcription termination occurs either due to presence of an intrinsic termination sequence or due to Rho-dependent termination. Intrinsic termination occurs due to the presence of an inverted repeat sequence that forms a hairpin loop in the RNA which is followed by a run of U residues. This hairpin loop with poly U tail, weakens the RNA-DNA hybrid and allows RNAP to dissociate from the template (Nedialkov *et al.*, 2013). The alternate transcription termination mechanism, involves the protein Rho, which is a hexameric protein that functions as an RNA-DNA helicase. Rho has two RNA-binding sites, *i.e.* primary and secondary sites that enable the protein to attach to the RNA and terminate the transcription (Skordalakes and Berger, 2003). The primary binding site is present towards the N-terminal of the Rho and secondary binding site is present towards the C-terminal. The Rho protein binds to the RNA at Rho utilization sites (Rut) using the primary binding site (Banerjee *et al.*, 2006). The attachment of Rho to RNA requires an optimum length of "naked" non-translated RNA (80-90 nucleotide bases). After attachment of Rho to the nascent RNA, Rho moves along the RNA towards the 3' end of RNA. This movement of Rho along the RNA requires energy and it is provided by ATP hydrolysis. When RNAP pauses during transcription, the Rho protein approaches the 3' end, detaches RNA from the template DNA and RNAP is released (Brennan *et al.*, 1987; Hitchens *et al.*, 2006; Skordalakes and Berger, 2003).

Some proteins help in reducing pausing of RNAP and regulate the elongation complex; these are called Nus factors (N-utilization substances) and these also help in rho-dependent termination process. There are four Nus factors important in *E. coli, i.e.* NusA, NusB, NusE and NusG, and NusG interacts with Rho protein (Burmann and Rosch, 2011). NusG is a dimeric protein that attaches to the RNAP by the N-terminal domain and helps in releasing σ

from the β' flap of holo-RNAP. NusG stays connected with the RNAP during transcription that is followed by translation into protein, when ribosome reaches the stop codon during translation, the rho has length of naked-mRNA to load onto the mRNA for termination. NusG protein's C-terminal domain binds to the rho and helps in rho dependent termination (Li *et al.*, 2016; Burmann and Rosch, 2011).

1.8 Transcription regulation

The concept of gene expression being activated and repressed by transcription factors was proposed by Jacob and Monod (Jacob and Monod, 1961). This emerged from the idea that a protein binding at the regulatory sequence of a gene could either increase or decrease the transcription of that gene. Proteins that bind to regulatory sequences are called transcription factors. Transcription factor encoding sequences can be present very near or at a distance in the genome from the genes that they regulate, and these transcription factor serve as "regulatory switches" for genes (Balleza *et al.*, 2009).

Gene expression can be affected by the environmental conditions and interaction of regulatory sequences to certain proteins. Transcription regulation can be very simple, *e.g.* by just a single factor, while in other cases, it may involve more than one transcription factor and each factor often relies on a different environmental signal (Browning and Busby, 2004). Most transcription factors act as activators or repressors, but some regulation factors can work as both (Perez-Rueda and Collado-Vides, 2000). Some transcription factors bind to a specific sequence at a single promoter and so regulate a single transcription unit, while others can affect the transcription of more than one promoter and so may act as global transcription regulators. For example, the cyclic adenosine monophosphate (cAMP) receptor protein (CRP)

regulates more than 200 different promoters in *E. coli* (Martinez-Antonio and Collado-Vides, 2003).

The strength of transcription regulation depends upon the affinity between the transcription factor and the regulatory sequence. The same transcription factor may produce different effects on rate of transcription at different promoters, depending upon the transcription regulatory sequences. Transcription factors are divided into two broad categories, activators and repressors, on the base of their effect on promoter activity. How each of the transcription factor can interact with a promoter is discussed below.

1.8.1 Activation of transcription by activator proteins

Studies with the CRP protein, showed that it could activate using two mechanisms: Class I and Class II. In Class I activation, CRP binds upstream of the -35 hexamer element, at specific points on the same face of the DNA helix (*e.g.* near positions of -62, -72, -83 or -93) and αCTD makes contact with a face of the activator called the activation region (Figure 1.9A) (Carattoli *et al.*, 2005; Browning and Busby, 2004). The *lac* promoter, which is activated by CRP, is an example of a Class I activated promoter, where the transcription factor binds upstream of the -35 hexamer element (Igarashi *et al.,* 1991). However, if the activator protein binding site overlaps the -35 hexamer element, it is called a Class II activated promoter. In Class II activation, the transcription factor interacts with the σ subunit of RNAP, helps in both the recruitment of RNAP and activation of transcription initiation (Figure 1.9B) (Browning and Busby, 2016). Class II activators can also bind to the αCTD and αNTD subunits in addition to the σ subunit (Feng *et al.*, 2016). The activation of the *coo* promoter in ETEC by Rns and the Pm promoter in *Pseudomonas putida* (*P. putida*) by XylS are examples of Class II activation (Murphree *et al.*, 1997; Marques *et al.*, 1999).

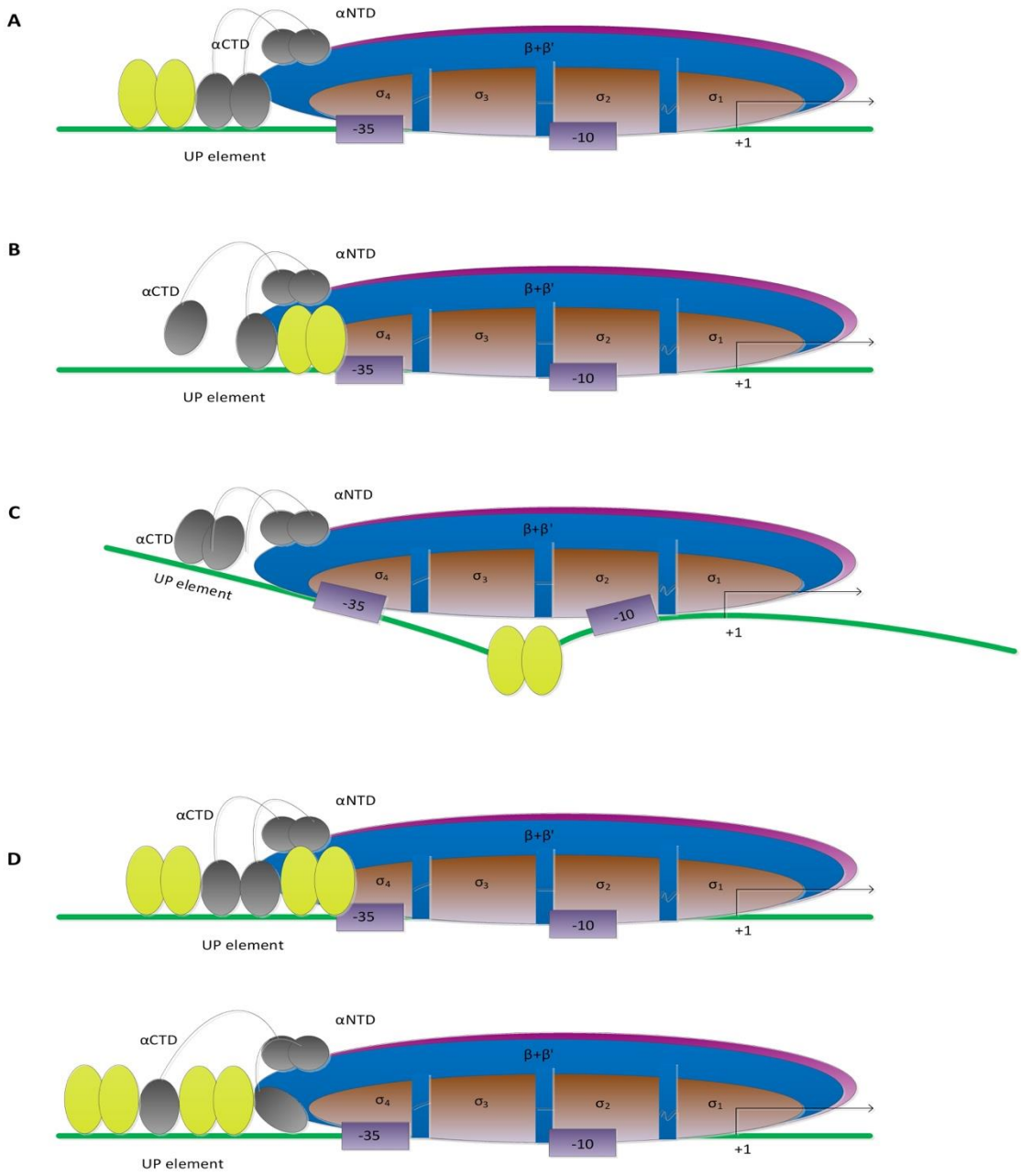


Figure 1.9 Mechanisms of transcription activation

A. Class I activation: The figure shows the binding of an activator (yellow) upstream of the -35 hexamer element, where it interacts with αCTD to promote transcription.
B. **Class II activation:** The figure shows the binding of an activator (yellow) overlapping the -35 hexamer element in such a manner that it interacts with αCTD and σ_4 to promote transcription.
C. **Activation by conformational change:** The figure shows the binding of the transcription activator between the -10 hexamer element and -35 hexamer element, which brings about a conformational change in the DNA to promote transcription.
D. Complex activation: The panel shows the binding of more than one activator (yellow) with the DNA strand either at Class I and Class II binding sites or two Class I activator binding sites.
This figure is adapted from Browning and Busby (2004).

At some promoters, the activators can bind to the DNA and bring about conformational changes that allow RNAP to bind the promoter DNA (Figure 1.9C). The majority of transcription activators that activate by this mechanism, are the MerR family of transcription regulators and they induce a conformational change in the DNA to realign the -10 and -35 hexamer elements (Brown *et al.*, 2003). An example of this class of promoters is the *copA* promoter, (a gene encoding for a copper export protein), activated by CueR (Stoyanov *et al.*, 2001).

If only one activator binds to the promoter, then it is relatively simple to explain its mode of action but at some promoters, two transcription activators bind to the promoter sequence independently to co-regulate transcription. This could be due to different combinations of activators, for example, two Class I activators on the same promoter or a Class I and a Class II activator (Figure 1.9D) (Lee *et al.*, 2012). An example of this type of activation is at the acetate synthetase promoter 2 (*acs*2), that is activated by two CRP molecules bind at positions -69.5 and -122.5 upstream of the transcription start site (Beatty *et al.*, 2003).

1.8.2 Repression by transcription factors

Repressors are proteins that bind to promoters and decrease transcription initiation by destabilising or blocking RNAP binding to the promoter. There are three simple mechanisms of RNA transcription repression. One of the mechanisms is when the repressor protein binds near the core promoter elements, and thus results in the failure of RNAP to bind to the promoter and start transcription. This kind of repressor directly minimises transcription by blocking RNAP binding (Figure 1.10A). For example, the *lac* promoter is repressed by binding of the LacI repressor, at the core promoter sequence (Muller-Hill, 1998).

The second method of transcription repression is when several repressor molecules bind at a

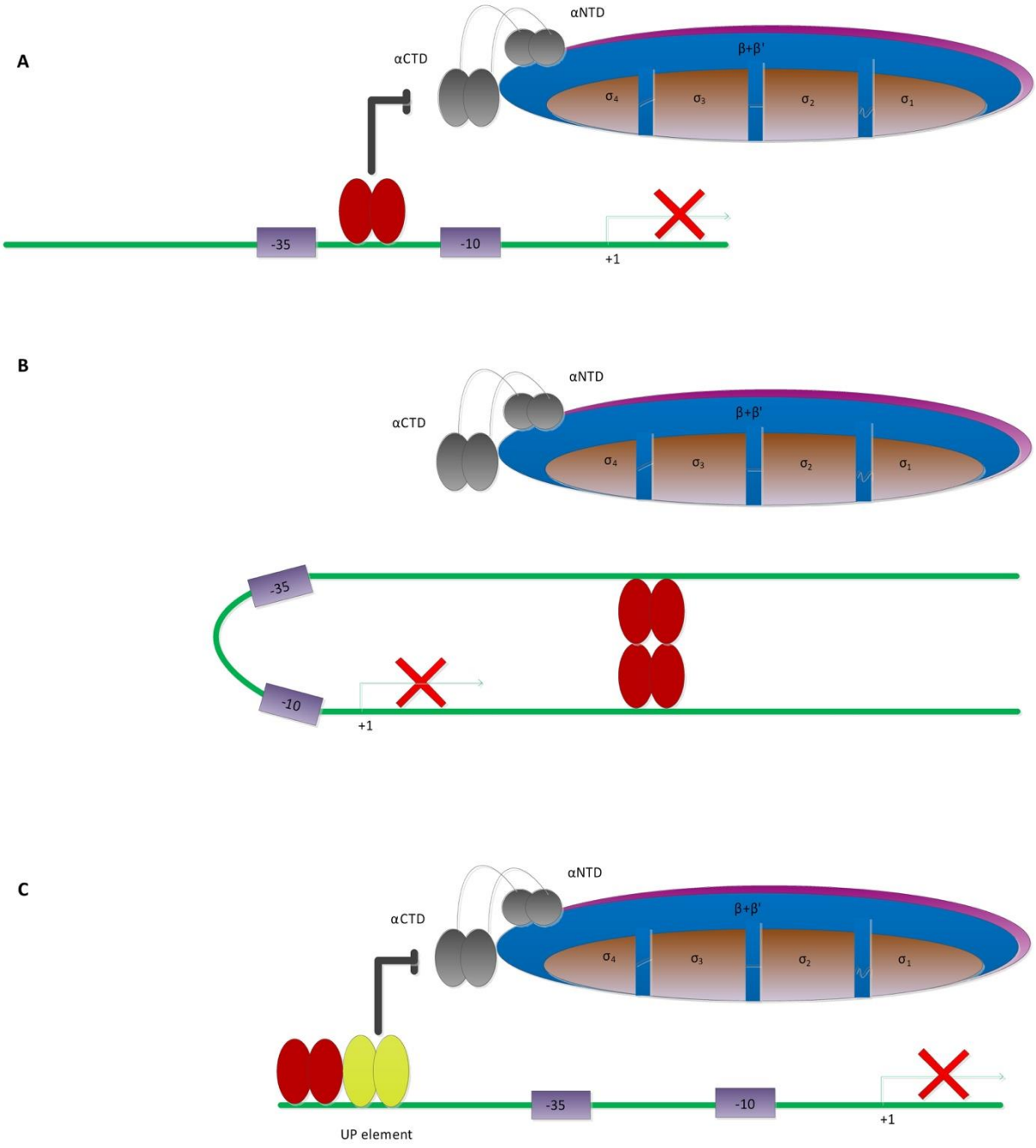


Figure 1.10 Mechanisms of transcription repression

A. The repressor (red) binds between the promoter core elements, preventing RNAP from binding to the promoter.
B. The repressors bind to the DNA making a loop, and preventing RNAP from binding the promoter.
C. The repressor (red) binds to the DNA near the activator (yellow) and modifies the activator in such a way that it cannot recruit RNAP effectively.
This figure is adapted from Browning and Busby (2004).

promoter and result in DNA looping. This can result in promoter core elements keeping trapped in the loop and RNAP cannot bind to the promoter (Figure 1.10B). For example, GalR binding induces DNA looping and repression of the *gal* promoter (Swint-Kruse and Matthews, 2009). The third type of repression mechanism involves inhibition of an activator protein. In this instance, the repressor binds to the target promoter in such a way that it prevents the activator from performing its function (Figure 1.10C). For example, the interaction of CytR with CRP to inactivate the *cytR* promoter is an example of this type of repression (Valentin-Hansen *et al.*, 1996).

1.8.3 Promoter regulation by modifications

Transcription can also be regulated by the modification of the DNA sequences. One of the most common processes is, transcription regulation by modification of the specific nucleotide bases by the Dam methylase enzyme. The Dam methylase adds the methyl group to adenine nucleotide base in GATC motif and this can effect transcription at certain promoters. The process was identified by Blyn *et al.* (1990) who showed that methylation can switch ON or OFF a promoter. For example, there are two GATC sites present in the regulatory region of pyelonephritis-associated gene pilus (*pap*) operon in UPEC. When the distal GATC site, in relation to the *pap* genes is methylated, the promoter turned OFF and the gene is not expressed. When the proximal GATC is methylated, this turns the promoter to ON in *pap* expression and pili formation.

An additional epigenetic mechanism that regulates the gene expression is switching promoter orientation (Browning and Busby, 2016). In the ON phase, the promoter is located in such a position that if RNAP binds to the promoter, and transcription will lead into the gene. In the OFF phase, the promoter is inverted with the help of inversion enzymes and other regulatory protein so it cannot express the gene. An example of such a gene regulation is the *fim* operon

that encodes for type I fimbriae in *E. coli*. The 314 bp promoter of *fimA* is inverted by FimB and FimE with the help of IHF and HNS (Blumer *et al.*, 2005).

Another mechanism that can affect transcription is the variation of spacing between promoter elements by the addition or deletion of a single nucleotide or dinucleotide repeat. An example of this type of promoter regulation is the *hifA* and *hifB* promoters that controls the expression of the fimbrial components in *Haemophilus influenzae*. The promoters have different numbers of AT repeats between the -10 hexamer element and -35 hexamer element. Thus, in bacterial population different bacteria have different spacing between these elements. This leads to the different levels of gene expression (Bruant *et al.*, 2002).

1.8.4 Transcription regulation by second messengers and factors interacting with RNAP

Bacteria can sense changes in external environment and can react to the environmental signals by a change in metabolism that alters the expression of certain genes. They often produce signalling molecules or second messangers, and important examples of these are cyclic adenosine monophosphate (cAMP), cyclic diguanylate monophosphate (c-di-GMP) and guanosine pentaphosphate (p)ppGpp (Kariisa *et al.*, 2015).

cAMP is one of the important second messengers and its concentration changes in response to glucose or lactose concentration in the medium. When the interacellular glucose concentration is low, the adenylyl cyclase present in the cytoplasm senses the dearth of the carbon source and produces cAMP from AMP. The cAMP binds to CRP and activates the transcription of genes required for alternate source of energy. The role of cAMP in *E. coli* is limited to interact with CRP. The cAMP-CRP complex is important for regulation of genes involved in catabolism (Gorke and Stulke, 2008).

Another important second messenger in *E. coli* is ppGpp that affects transcription of many

genes. When the bacterium is experiencing starvation, there are less free amino acids available to synthesise new proteins. This results in the presence of more uncharged tRNAs within the cell and this activates the ribosomal associated protein RelA to produce ppGpp (Hauryliuk *et al.*, 2015). When the cells experience starvation for other carbon sources, the cytoplasmic enzyme SpoT is activated and it also produces guanosine tetraphosphate (ppGpp). Increased levels of ppGpp shifts the pattern of transcription with more expression of stress response genes and less or no expression of genes involved in rapid growth and the cell division, *e.g.* rRNA operons. This is possible as ppGpp can stabilise the transcription bubble of certain genes and destabilise the transcription bubble of other genes. One result of increased levels of ppGpp is an increase in the expression of genes for amino acids production in bacteria (Paul *et al.*, 2005). Note that in *E. coli* the protein factor, DksA, binds to RNAP transiently and increases the RNAP-ppGpp complex stability facilitating the response to ppGpp (Paul *et al.*, 2005; Perederina *et al.*, 2004).

Transcription elongation is not a continuous process, rather there are stops and starts as the RNAP moves along the DNA. Some pauses are brief and they are resolved quickly with the resumption of elongation. However, some of the pauses are longer and can end up in RNAP backtracking. Backtracking is a process in which RNAP trails back on RNA-DNA hybrid from 1 nucleotide to few nucleotides (Nudler, 2012). When backtracking occurs, transcription is arrested and the transcription factors GreA and GreB are able to resolve this complex. These factors cleave the nascent RNA from the 3' end and relieve the transcription blockage (Roberts *et al.*, 2008). GreA and GreB are similar in structure to DksA and all of these proteins regulate RNAP by interacting through the secondary channel (Browning and Busby, 2016).

1.9 Overview of AraC transcription activation

Transcription regulation in EAEC is complex and global regulators and local transcription regulators are important for growth, multiplication, metabolism and survival. AraC is a transcription regulator of *E. coli* and it regulates the genes for arabinose metabolism. A number of transcription factors have been classified as AraC/XylS family members on the basis of their sequence homology to AraC of *E. coli* and XylS of *P. putida*. The AraC/XylS family of transcription regulators has more than 800 members. Family members have a conserved DNA binding domain in the C-terminal that generally consists of two helix turn helix (HTH) motifs which interact with promoter DNA. Members of this family can also exist either as monomers or dimers, depending upon the particular protein (Egan, 2002). Family members are involved in the regulation of a diverse range of genes (Munson and Scott, 2000). Some members regulate carbon metabolism and responses to environmental stress, while others may regulate virulence factor expression in animal and plant pathogens.

Many AraC family members activate expression of a range of virulence determinants. Fimbriae are important virulence determinants and AraC family members that increase the expression of fimbrial genes include AggR from EAEC and CsvR and Rns from ETEC (Gallegos *et al.*, 1997). As AggR regulates the expression of many virulence determinants in EAEC, a brief review of AraC family of transcription regulators is necessary (Nataro *et al.*, 1994).

In *E. coli*, the *araBAD* operon encodes for arabinose metabolising enzymes whilst *araC* encodes for the AraC, a transcription regulator (Figure 1.11A). AraC exists as a dimeric protein and there are three binding sites for AraC upstream of the *araBAD* promoter *i.e.* *araO2*, *araO1* and *araI* (Figure 1.11A). The *araI* site is split into two subsites *araI1* and

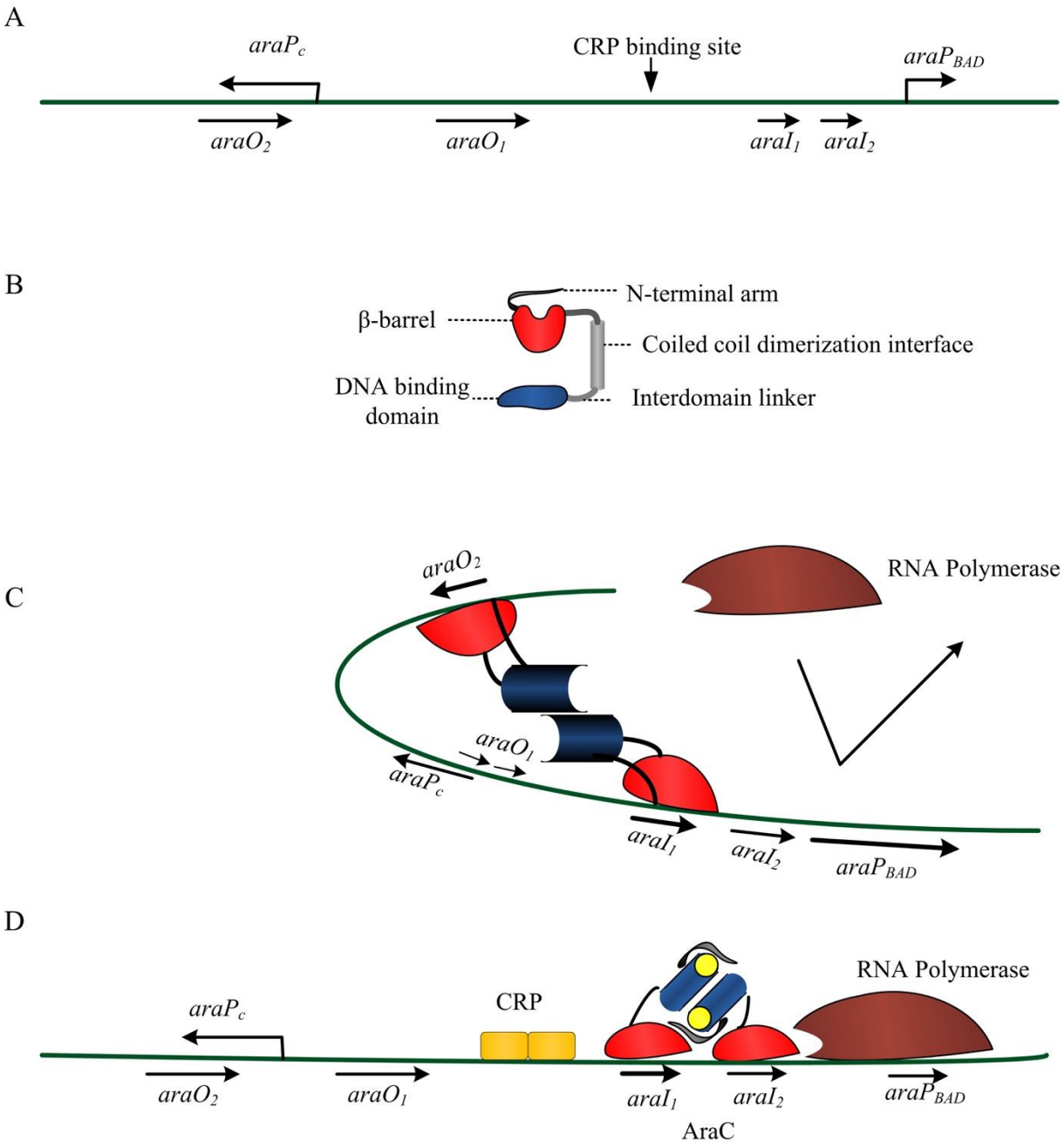


Figure 1.11 Regulation of the *araBAD* promoter by AraC

A. A map of the *ara* control region. There are four AraC binding sites (*araO*1, *araO*2, *aral*1, and *aral*2). The bent arrows show the transcriptional units regulated by AraC.
B. Domain organization of a single subunit of AraC protein.
C. Negative control of *araBAD* expression. In the absence of arabinose, AraC binds to *araO*2 and *aral*1, blocking access of RNAP to the promoter.
D. Positive control of *araBAD* expression: arabinose binds to AraC, changing its shape so it binds as a dimer to *aral*1 and *aral*2. This allows RNAP to bind to the promoter and active transcription can occur. This figure is adapted from Seedorff and Schleif (2011).

araI2, each of these regions can be bound by one monomer of AraC (Hamilton and Lee, 1988). Whilst cells are growing in the medium without arabinose, they do not require the enzymes for arabinose metabolism and so AraC represses transcription of *araBAD* by forming a loop, binding to *araO2* and *araI1* (Figure 1.11B and C). In this arrangement the binding of RNAP to the *araBAD* promoter is blocked by DNA loop formation. When cells are grown in the presence of arabinose, AraC undergoes a conformational change, as arabinose binds to it, and this results in the binding of AraC to *araI1* and *araI2,* instead of binding to $araO_2$ (Figure 1.11D). The switching of AraC from one conformation to another has been also termed the light switch model (Schleif, 2010). This alternate AraC binding arrangement results in the up regulation of *araBAD* promoter as the repression loop is released and RNAP is able to bind to the promoter. As, AraC binding site also overlaps the -35 hexamer element, it is suggested AraC interacts with RNAP in the transcription initiation (Dhiman and Schleif, 2000). AraC can also bind to the $araO_1$ binding site but this binding site autoregulates its own transcription from the *araC* promoter (Gallegos *et al.*, 1997).

1.10 An overview of the AggR regulon in EAEC

AggR is a member of the AraC family of transcriptional regulators and the *aggR* gene is present on the large virulence plasmids present in EAEC strains (Nataro *et al.*, 1994; Boll *et al.*, 2012). AggR has high level of amino acid similarity with CfaD (68%) and Rns (66%) from ETEC (Munson, 2013). To date, there is little direct information regarding the mechanism of regulation by AggR but related regulators, such as Rns, have been studied in some detail. A study by Mahon *et al.* (2012) has shown that Rns can exist as a dimer, and on the basis of *in vivo* experiment with a bacterial two hybrid system, it has been suggested that AggR may also dimerise when it activates transcription (Santiago *et al.*, 2016). But, on the other hand, studies on the Rns have shown that it is binds to DNA as a monomer and it exists

as monomer in *in vitro*, so these members of AraC family are monomeric or dimeric, is still under consideration (Basturea *et al.*, 2008; Munson and Scott, 1999; Munson, 2013).

The AraC family of regulators are defined on the basis of their conserved DNA binding domains, present in the C-terminus, (*i.e.* two HTH regions). Each HTH consists of a 21 amino acid motif and AraC/XylS transcription factors interact with the DNA through these HTH regions. Figure 1.12 shows the sequence of the HTHs of AggR compared to the other closely related members (*i.e.* Rns, CsvR and VirF). The HTH regions sequences of these closely related transcription factors show high amino acid identity to AggR. Figure 1.12 also shows the sequence alignment of AraC, MelR and AggR HTH domains (Gallegos *et al.* (1997).

Little is known about AggR-dependent regulation but the AggR promoter has been identified and investigated (Morin *et al.*, 2010). Two AggR-binding sites have been identified at the *aggR* promoter and both are similar to the Rns consensus sequence (*i.e.*, AnnnnnnTATC), moreover, it is also shown that AggR also regulates its expression from these sites. It has also been demonstrated that the *E. coli* global regulators FIS (Factor Inversion Stimulation) and H-NS (Histone like-nucleiod structuring protein), play a role in regulating *aggR* expression (Morin *et al.*, 2010). Santiago *et al.* (2016) found that AraC family members are negatively regulated by AraC negative regulators (ANR) which are small proteins that interact with AraC family members. The first member of the ANR family was found in EAEC 042 and it was named as Aar (AggR activated regulator) (Santiago *et al.*, 2014). Expression of the Aar negative regulator is activated by AggR and once expressed Aar binds to AggR and impairs its ability to activate transcription. The exact binding sites are not known but bacterial two hybrid system results show that the middle part of AggR is important for Aar-AggR interaction (Santiago *et al.*, 2016).

.

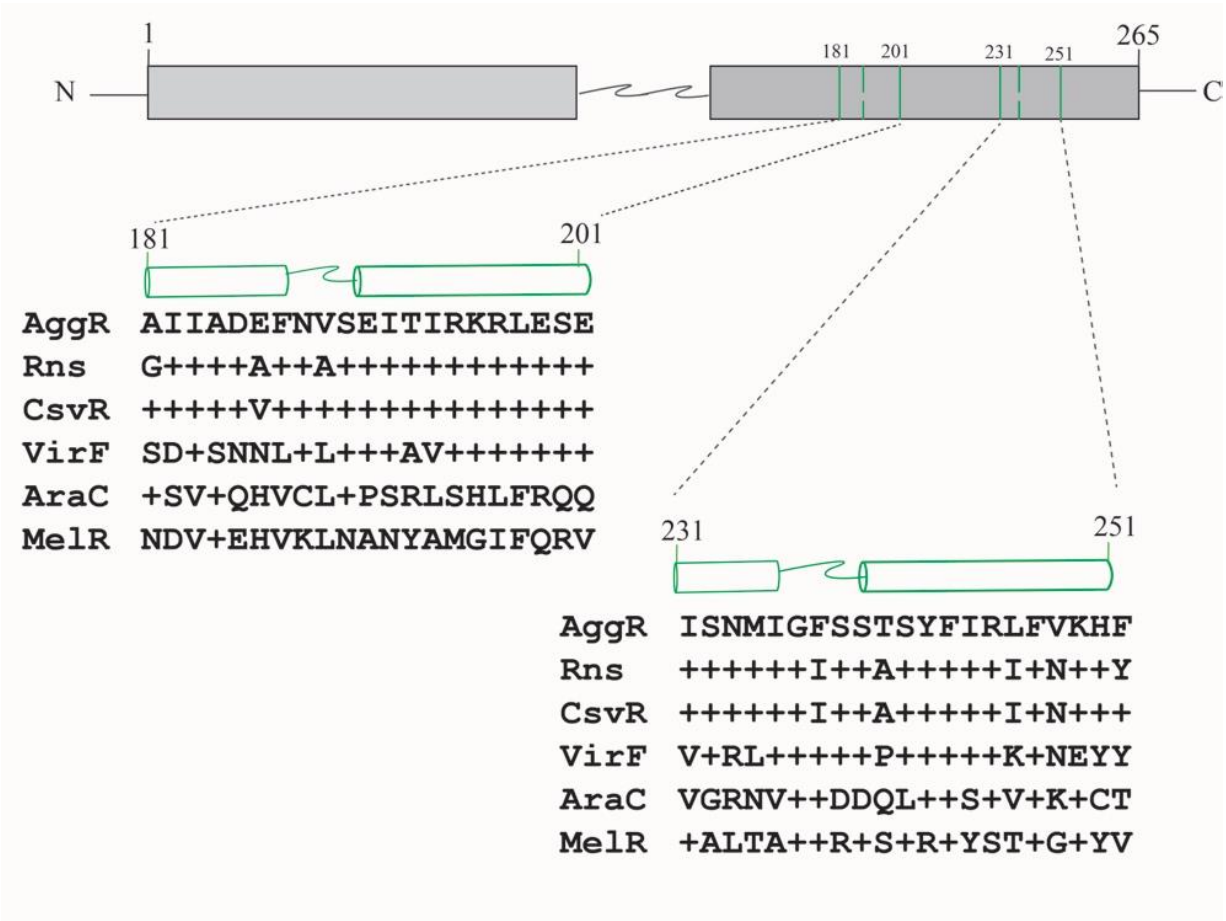


Figure 1.12 Alignment of the helix turn helix motifs of AggR with that of Rns, CsvR, VirF, AraC and MelR

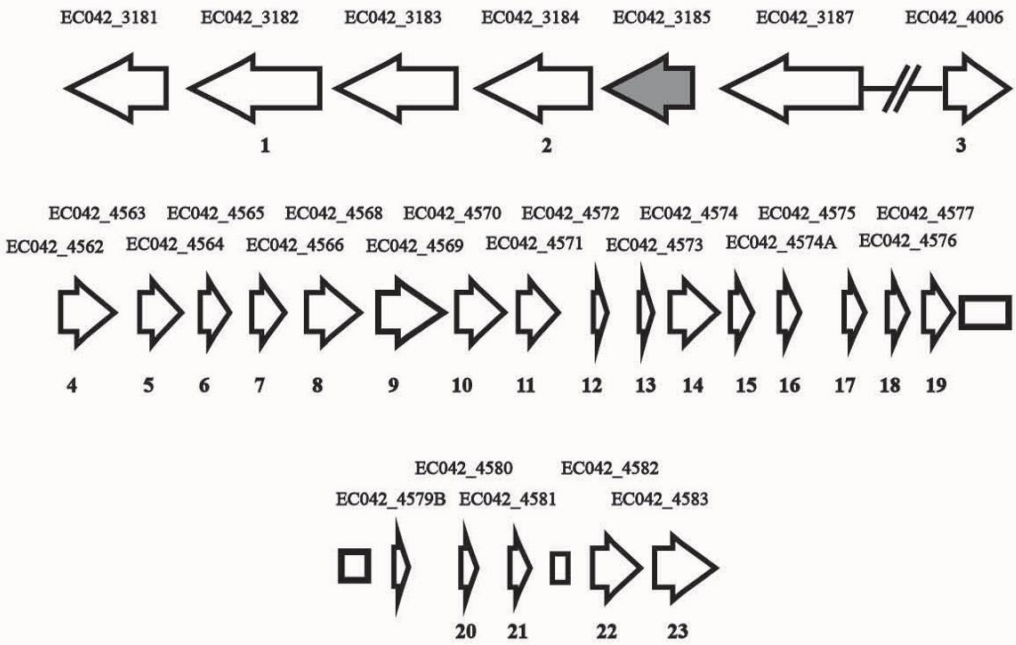
The figure shows the schematic representation of AggR as two rectangular boxes with N and C terminal domains. The green cylindrical shapes represent the HTHs of AggR. Alignment of amino acid sequences of HTHs motifs of AggR, Rns, CvsR and VirF are shown from Munson (2013). Alignment of amino acid sequences AraC and MelR to HTH of AggR is from Gallegos *et al.* (1997). The (+) sign shows the presence of same amino acid as AggR at that position.

Morin *et al.* (2013) have examined the AggR regulon in EAEC strain 042 using microarray analysis, and found that AggR regulates more than 40 genes present on pAA2, as well as on the bacterial chromosome. Most of the AggR-regulated genes identified are already identified as important virulence determinants, whilst a number of the genes encode for proteins of unknown function. As AggR regulates virulence in EAEC, it is likely that these proteins of unknown function may have a role in pathogenesis as well. Figure 1.13A and Figure 1.14A show the schematic diagrams of some of the genes on the chromosome and pAA2, respectively. Morin *et al.* (2013), determined using a microarray, that 23 AggR-regulated genes were located on the chromosome and 21 AggR-dependent genes were present on pAA2 plasmid of EAEC 042, which were verified using qPCR. The numbering below the genes represent that expression of these genes is regulated by AggR Figure 1.13A and Figure 1.14A and Figure 1.13B and Figure 1.14B contain the results of the qPCR experiment showing the fold change in decrease gene expression in the the absence of AggR on the chromosome and pAA2 plasmid, respectively.

Previous studies have provided information concerning AggR and the genes it regulates. However, there is very little information on the regulation of AggR-dependent promoters and the mechanism by which it activates transcription. Only the *aggR* gene promoter has been analysed with the AggR-binding sites (Morin *et al.*, 2010). The present study was designed with the aim of understanding AggR-dependent promoter regulation.

For this purpose, the AggR-dependent genes that are important in virulence of the EAEC prototype strain 042 were chosen in order to understand AggR-dependent regulation at these promoters. Therefore, the AAF/II and dispersin genes were selected from the pAA2 plasmid and T6SS genes from the chromosome to study AggR-dependent promoters using molecular techniques. The AggR-dependent promoter of AAF/I operon from EAEC strain 17-2 was also

A



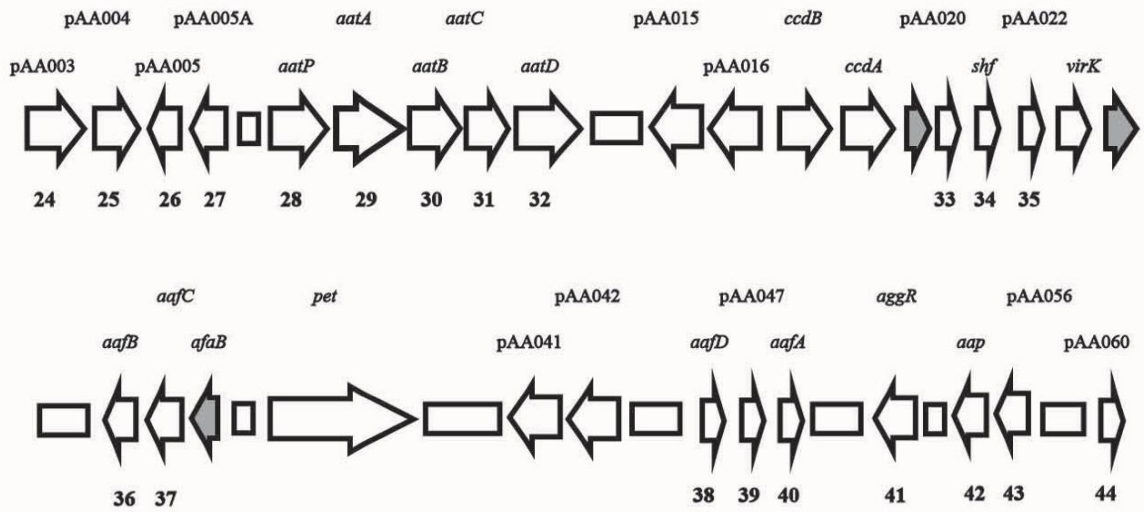
B

Serial Number	Gene Name	Fold expression	Serial Number	Gene Name	Fold expression
1	EC042_3182	4	13	EC042_4572	8
2	EC042_3184	20	14	EC042_4573	8
3	EC042_4006	4	15	EC042_4574	6
4	EC042_4562	12	16	EC042_4574A	4
5	EC042_4563	16	17	EC042_4575	4
6	EC042_4564	16	18	EC042_4576	6
7	EC042_4565	4	19	EC042_4577	6
8	EC042_4566	8	20	EC042_4580	4
9	EC042_4568	8	21	EC042_4581	6
10	EC042_4569	12	22	EC042_4582	5
11	EC042_4570	4	23	EC042_4583	5
12	EC042_4571	4			

Figure 1.13 The AggR-dependent genes on the chromosome

A. The panel shows the schematic representation of chromosomally located genes regulated by AggR. The genes are present in two loci and names are shown above each gene representation. The number below each gene represents AggR-dependent gene present on EAEC 042 chromosome.
B. The panel shows AggR-dependent expression levels of genes present on the chromosome and pAA2. The fold expression column shows the value of fold decrease in expression in absence of AggR observed by Morin *et al.* (2013).

A



B

Serial Number	Gene Name	Fold expression	Serial Number	Gene Name	Fold expression
24	pAA003	66	35	pAA022	2
25	pAA004	8	36	aafB	10
26	pAA005	16	37	aafC	10
27	pAA005a	48	38	aafD	500
28	aatP	16	39	pAA047	1000
29	aatA	128	40	aafA	10
30	aatB	32	41	aap	800
31	aatC	24	42	aggR	500
32	aatD	16	43	pAA056	6
33	pAA20	4	44	pAA060	4
34	shf	2			

Figure 1.14 The AggR-dependent genes on the pAA2 plasmid

A. The panel shows a schematic representation of genes present on pAA2 that are regulated by AggR. The gene names are shown above each gene. A number is shown below each AggR-dependent gene.
B. The panel shows AggR-dependent expression levels of genes present on the chromosome and pAA2. The fold expression column shows the value of fold decrease in expression in absence of AggR observed by Morin *et al.* (2013).

studied and the AggR-dependent model was verified by making semi-synthetic promoters.

1.11 Aims and objectives of the project

The overall aim of the project was to determine the organisation of AggR-dependent promoters and understand how the AggR-dependent genes are regulated in EAEC. Although the whole genome sequence of EAEC is available and AggR-dependent genes have been identified, the architecture of many AggR-dependent promoters has not been described. Therefore, the intended aim of this study was to choose a number of different AggR-dependent promoters from EAEC, with a view to build up a step-by-step picture of AggR-dependent regulation.

In order to do this, the key functional elements of the fimbrial promoters that control expression of AAF/II on pAA2 plasmid from EAEC 042 were determined and discussed in Chapter 3. These results gave us an understanding of the organization of AggR-dependent promoters and AggR-mediated regulation. In Chapter 4, I examined the *aap, aaiA* and semi-synthetic promoters to further understand the architecture of AggR-dependent promoters. Chapter 5 contains the details about mutational analysis of AggR protein and affect of AggR on *pet* expression.

To understand the role of AggR-dependent promoters present in other strains of EAEC, the fimbrial promoter of the EAEC strain 17-2 was studied using the same principles. It is pertinent to mention here that the AggR construct used in this study was taken from EAEC 042 and the promoter architecture of fimbrial genes of EAEC 17-2 was studied using the same principles (Chapter 6). The overall aim of the study was to determine the location of AggR-dependent promoters and the mechanism of the AggR-dependent regulation.

Chapter 2

Materials and Methods

2.1 Suppliers

All chemicals used in this study were purchased from Sigma-Aldrich, BDH and Fisher Scientific UK, unless otherwise stated. T4 DNA ligase, calf intestinal alkaline phosphatase (CIP), Phusion DNA polymerase, Q5 DNA polymerase and all restriction enzymes were obtained from New England Biolabs (NEB). Biomix red DNA polymerase, Mytaq red DNA polymerase and dNTPs were ordered from Bioline. All the deoxy-oligonucleotides were made by ALTA Biosciences (University of Birmingham). Where appropriate, solutions were autoclaved at 121°C under 15 $lbin^{-2}$ for 15 minutes or filter-sterilised using 0.2 µm pore size filters (Acrodisc®).

2.2 Bacterial growth media

2.2.1 Liquid media

All *E. coli* strains were grown in Lysogeny broth (LB) at 37°C with shaking unless otherwise stated. LB was made by mixing 10 g peptone, 5 g yeast and 5 g NaCl in one liter of distilled water (dH_2O). For the selection of plasmid-encoded resistances the medium was supplemented with the appropriate antibiotic. Super Optimal broth with Catabolite repression (SOC) medium contained 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 10 ml KCl (250 mM), 5 ml $MgCl_2$ (2M), 20 ml glucose (1M) in one liter of distilled water. Glucose was added after autoclaving. Dulbecco's modified Eagle medium (DMEM) with high glucose concentration (4.5 gl^{-1}) was purchased from Sigma Aldrich in liquid form.

2.2.2 Solid media

E. coli strains were grown on LB agar (Sigma) (15 g agar, 10 g tryptone, 5 g yeast extract and 5 g NaCl in one liter of dH_2O), Nutrient agar (Oxoid) (15 g agar, 1 g meat extract, 5 g peptone, 2 g yeast extract and 5 g NaCl in one liter of dH_2O), or MacConkey lactose agar (BD

Difco) (13 g agar, 17 g pancreatic digest of gelatin, 3 g peptones (meat and casein), 5 g NaCl, 10 g lactose, 1.5 g bile salts, 0.03 g neutral red and 0.001 g crystal violet in one liter of dH_2O) plates. Media were prepared using distilled water according to manufacturer's instructions and autoclaved. After autoclaving, the medium was cooled to between 50°C - 55°C and the appropriate antibiotics were added as required.

2.2.3 Antibiotics and other supplements

The strains carrying plasmid-encoded resistance genes were selected for, by the addition of the appropriate antibiotic to both liquid and solid media, after autoclaving. For stock solutions of antibiotics, the powder of respective antibiotics was dissolved as follow, 100 $mgml^{-1}$ of ampicillin in dH_2O, 50 $mgml^{-1}$ of kanamycin in dH_2O, 10 $mgml^{-1}$ of tetracycline in methanol (50% v/v) and 25 $mgml^{-1}$ of Chloramphenicol in methanol (100%). All the stock solutions of antibiotics were sterile filtered through 0.2 µm syringe filters (Acrodisc®) and stored at -20°C. The final concentration of the antibiotics used was 100 μgml^{-1} for ampicillin, 15 μgml^{-1} for tetracycline, 50 μgml^{-1} for chloramphenicol and 25 μgml^{-1} for kanamycin. Arabinose was used to induce protein expression using the pBAD expression system where appropriate at a concentration of 0.2% (w/v) and a stock solution 20% (w/v) was made to achieve this concentration.

2.3 Bacterial strains and plasmids

2.3.1 Bacterial strains and growth conditions

The *E. coli* K-12 strain BW25113 used throughout this study and its genotype is shown in the Table 2.1. EAEC strain 042 was also used, its genotype, and that of its variants are detailed in Table 2.1. Overnight cultures were grown by inoculating a single bacterial colony into 5 ml of LB supplemented with antibiotics (where appropriate) and incubated at 37°C with shaking for

Table 2.1 Strains used in this study

Strains	Genotype	References/ Source
BW25113	*lacI*^{q}, *rrnBT14,* Δ*lacZ*$_{WJ16,}$ *hsdR514,* Δ*araBAD*$_{AH33,}$ Δ*rhaBAD*$_{LD78}$	(Baba *et al.*, 2006)
EAEC 042	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, Diarrhoeagenic in volunteers, expresses AAF/II, biofilm positive, harbours pAA2	(Nataro *et al.*, 1995)
EAEC 042 Δ*aggR*	Wild type, prototype strain, Sm^{R}, Tet^{R}, Cm^{R}, harbours pAA2. The *aggR* gene present on pAA2 has been deleted.	(Dudley *et al.*, 2006)
EAEC DFB042	Wild type, prototype strain, Sm^{R}, Tet^{S}, Cm^{S}, expresses AAF/II, biofilm positive, harbours pAA2.	Douglas Browning
EAEC DFB042 (pBAD/*aggR*)	Wild type, prototype strain, Sm^{R}, Tet^{S}, Cm^{S}, Amp^{R} expresses AAF/II, biofilm positive, harbours pAA2. Containing pBAD/*aggR* to over express AggR.	This study
EAEC DFB042 Δ*aggR*	Wild type, prototype strain, Sm^{R}, Tet^{S}, Cm^{S}, biofilm negative, harbours pAA2. The *aggR* gene has been replaced by kanamycin resistance cassette.	This study

16-18 hours. Bacteria were sub-cultured by diluting the overnight culture 1:100 into the appropriate medium. Growth of bacterial cultures was monitored by measuring the optical density at 650 nm (OD_{650}) using a Helios Gamma Spectrophotometer (Thermo Fisher Scientific Inc.). Glycerol stocks for long term storage were made from overnight cultures by mixing glycerol stock, 850 µl of culture with 150 µl of glycerol in a cryofuge tube and the samples were stored at -80°C.

2.3.2 Plasmids

Plasmids used in this study are listed in Table 2.2.

2.4 Gel electrophoresis

2.4.1 Agarose gel electrophoresis

Agarose gels were used to resolve and purify DNA fragments and plasmids of different sizes ranging from 1 kb to 10 kb. Agarose gels contained and were run using a TBE based buffer system. The 5x TBE stock solution was prepared by dissolving 54 g Tris, 27.5 g borate and 4.65 g of EDTA in one litre of distilled water. The gels contained and were run in 0.5x TBE buffer by diluting the 5x stock with distilled water. Gels were made by mixing 15 ml molten agarose (0.8-1% (w/v) agarose in 0.5x TBE) with 1 µl of ethidium bromide (10 $mgml^{-1}$), which was poured into a gel casting plate. Wells of the desired width that could hold 10-40 µl of a sample, were created by placing a comb in vertical position into the casting tray to make sample wells. Once solidified, the comb was removed and the gel was immersed in 0.5x TBE buffer in a horizontal electrophoresis tank. DNA samples were prepared for loading by mixing one volume of loading dye (bromophenol blue (0.025% (w/v)), xylene cyanol F (0.025% (w/v)), glycerol (20% (v/v)), Tris (10 mM, pH 7.5) and EDTA (1 mM)) with 6 volumes of sample and the mixture was loaded into the wells in the gel. DNA marker from

Table 2.2 Plasmids used in this study (continued on pages 49-57)

Plasmids	Description	Source
pRW50	*oriV*, *lacZYA*, *tet*^{R}. Broad-host range, low-copy-number *lacZ* expression vector	(Lodge *et al.*, 1992)
pRW224-U9	A derivative of pRW50, which allows cloning of promoter fragments as transcriptional or translational fusions to *lacZ* carrying pUC9 linker between the EcoRI and HindIII sites.	(Islam *et al.*, 2011)
pRW225	A derivative of pRW224-U9, which allows cloning of promoter fragments using EcoRI and HindIII sites as translational fusions to *lacZ*.	(Islam *et al.*, 2011)
pRW50/*aafD*100	A derivative of pRW50 carrying the EcoRI-HindIII *aafD*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	This study
pRW50/*aafD*99	A derivative of pRW50 carrying an *aafD*99 promoter fragment constructed by deletion of 100 bp from *aafD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*98	A derivative of pRW50 carrying an *aafD*98 promoter fragment constructed by deletion of 200 bp from *aafD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*97	A derivative of pRW50 carrying an *aafD*97 promoter fragment constructed by deletion of 300 bp from *aafD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*96	A derivative of pRW50 carrying an *aafD*96 promoter fragment constructed by deletion of 320 bp from *aafD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW50/*aafD*95	A derivative of pRW50 carrying an *aafD*95 promoter fragment constructed by deletion of 340 bp, from *aafD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*94	A derivative of pRW50 carrying an *aafD*94 promoter fragment constructed by deletion of 360 bp, from *aafD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*96-90*C*	A derivative of pRW50/*aafD*96 carrying an *aafD*96-90*C* promoter fragment constructed from *aafD*96 by a substituting of T to C at position 90 and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*96-92*C*90*C*	A derivative of pRW50/*aafD*96 carrying an *aafD*96-90*C*92*C* promoter fragment constructed from *aafD*96 by substituting Ts to Cs at position 90 and 92 and cloned using an EcoRI-HindIII.	This study
pRW50/*aafD*96-65*C*	A derivative of pRW50/*aafD*96 carrying an *aafD*96-65*C* promoter fragment constructed from *aafD*96 by a substituting of an A to C at position 65 and cloned using an EcoRI-HindIII.	This study
pRW225/*aafD*1001	A derivative of pRW225, carrying *aafD*1001 promoter fragment constructed by from *aafD*100 with 10 bp deletion adjacent to HindIII site and cloned into pRW225 using an EcoRI-HindIII.	This study
pRW225/*aafD*1002	A derivative of pRW225, carrying *aafD*1002 promoter fragment constructed by from *aafD*100 with 19 bp deletion adjacent to HindIII site and cloned into pRW225 using an EcoRI-HindIII.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW224/*aafD*500	A derivative of pRW224, carrying *aafD*500 promoter fragment constructed by from *aafD*96 with 43 bp deletion adjacent to HindIII site and cloned into pRW224 using an EcoRI-HindIII.	This study
pRW50/*afaB*100	A derivative of pRW50 carrying the EcoRI-HindIII *afaB*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	This study
pRW50/*afaB*99	A derivative of pRW50 carrying an *afaB*99 promoter fragment constructed by deletion of 100 bp, from *afaB*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*afaB*98	A derivative of pRW50 carrying an *afaB*98 promoter fragment constructed by deletion of 200 bp, from *afaB*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*afaB*97	A derivative of pRW50 carrying an *afaB*97 promoter fragment constructed by deletion of 300 bp, from *afaB*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*afaB*100-320*C*318*C*	A derivative of pRW50 carrying an *afaB*100-320*C*318*C* promoter fragment constructed from *afaB*100 promoter fragment by substituting Ts to Cs at position 320 and 318 and cloned using an EcoRI-HindIII.	This study
pRW50/*afaB*100-293*C*	A derivative of pRW50 carrying an *afaB*100-293*C* promoter fragment constructed from *afaB*100 promoter fragment by substituting an A to C at position 293 and cloned using an EcoRI-HindIII.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW224/*afaB*100	A derivative of pRW224 carrying an EcoRI-HindIII *afaB*100 promoter fragment as a *lacZ* transcriptional fusion.	This study
pRW50/*afaBC*100	A derivative of pRW50 carrying the EcoRI-HindIII *afaBC100* promoter fragment cloned as a *lacZ* transcriptional fusion.	This study
pRW50/*afaBC*99	Derivative of pRW50 carrying a *afaBC*99 promoter fragment constructed from *afaBC*100 by deletion of 400 bp sequence and the promoter fragment was cloned using EcoRI and HindIII sites.	This study
pRW50/*aafA*100	A derivative of pRW50 carrying the EcoRI-HindIII *aafA*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	This study
pRW50/*aafC*100	A derivative of pRW50 carrying the EcoRI-HindIII *aafC*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	This study
pRW50/*aaiA*100	A derivative of pRW50 carrying the EcoRI-HindIII *aaiA*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	This study
pRW50/*aaiA*99	A derivative of pRW50 carrying an *aaiA*99 promoter fragment constructed by deletion of 172 bp, from *aaiA*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aaiA*98	A derivative of pRW50 carrying an *aaiA*98 promoter fragment constructed by deletion of 207 bp, from *aaiA*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW50/*aaiA*97	A derivative of pRW50 carrying an *aaiA*97 promoter fragment constructed by deletion of 228 bp, from *aaiA*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aaiA*96	A derivative of pRW50 carrying an *aaiA*96 promoter fragment constructed by deletion of 255 bp, from *aaiA*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aaiA*95	A derivative of pRW50 carrying an *aaiA*95 promoter fragment constructed by deletion of 335 bp, from *aaiA*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aaiA*98-254*C*-252*C*	A derivative of pRW50 carrying an *aaiA*98-254*C*252*C* promoter fragment constructed from *aaiA*98 promoter fragment by substituting Ts to Cs at position 254 and 252 and cloned using an EcoRI-HindIII.	This study
pRW50/*aaiA*98-226*C*	A derivative of pRW50 carrying an *aaiA*98-226*C* promoter fragment constructed from *aaiA*98 promoter fragment by substituting an A to C at position 226 and cloned using an EcoRI-HindIII.	This study
pRW50/*aap*100	A derivative of pRW50 carrying the EcoRI-HindIII *aap*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	Rita Godfrey
pRW50/*aap*100-151*C*149*C*	A derivative of pRW50 carrying an *aap*100-151*C*149*C* promoter fragment constructed from *aap*100 promoter fragment by substituting Ts to Cs at position 151 and 149 and cloned using an EcoRI-HindIII.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW50/*aap*100-124*C*	A derivative of pRW50 carrying an *aap*100-124*C* promoter fragment constructed from *aap*100 promoter fragment by substituting an A to C at position 124 and cloned using an EcoRI-HindIII.	This study
pRW50/*CC(-41.5)*	A derivative of pRW50 carrying the EcoRI-HindIII *CC(-41.5)* promoter fragment cloned as a *lacZ* transcriptional fusion.	(Savery *et al.*, 1995a)
pRW50/*DAM*20	A derivative of pRW50 carrying *DAM*20, a semi-synthetic promoter fragment constructed by transplanting a 19 bp AggR-binding motif from *aafD* promoter into *CC(-41.5)* between positions -47 to -29 from transcription start site. The AggR-binding motif spacing is 20 bp from the -10 hexamer element.	This study
pRW50/*DAM*21	A derivative of pRW50 carrying *DAM*21, a semi-synthetic promoter fragment constructed by transplanting a 19 bp AggR-binding motif from *aafD* promoter into *CC(-41.5)* and the AggR-binding motif spacing is 21 bp from the -10 hexamer element.	This study
pRW50/*DAM*22	A derivative of pRW50 carrying *DAM*22, a semi-synthetic promoter fragment constructed by transplanting a 19 bp AggR-binding motif from *aafD* promoter into *CC(-41.5)* and the AggR-binding motif spacing is 22 bp from the -10 hexamer element.	This study
pRW50/*DAM*23	A derivative of pRW50 carrying *DAM*23, a semi-synthetic promoter fragment constructed by transplanting a 19 bp AggR-binding motif from *aafD* promoter into *CC(-41.5)* and the AggR-binding motif spacing is 23 bp from the -10 hexamer element.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW50/*DAM*22-(-38*C*-36*C*)	A derivative of pRW50 carrying a *DAM*22-(-38*C*-36*C*) promoter fragment constructed from *DAM*22 promoter fragment by substituting Ts to Cs at position -38 and -36 from transcription start site and cloned using an EcoRI-HindIII.	This study
pRW50/*aggD*100	A derivative of pRW50 carrying the EcoRI-HindIII *aggD*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	Douglas Browning
pRW50/*aggD*50	A derivative of pRW50 carrying an *aggD*50 promoter fragment constructed by a deletion of 268 bp, from *aggD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aggD*49	A derivative of pRW50 carrying an *aggD*49 promoter fragment constructed by a deletion of 305 bp, from *aggD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aggD*48	A derivative of pRW50 carrying an *aggD*48 promoter fragment constructed by a deletion of 326 bp, from *aggD*100 promoter fragment, and cloned using an EcoRI-HindIII.	This study
pRW50/*aggD*49-86*C*	A derivative of pRW50 carrying an *aggD*49-86*C* promoter fragment constructed from *aggD*49 promoter fragment by substituting T to C at position 86 and cloned using an EcoRI-HindIII.	This study
pRW50/*aggD*49-60*C*	A derivative of pRW50 carrying an *aggD*49-60*C* promoter fragment constructed from *aggD*49 promoter fragment by substituting an A to C at position 60 and cloned using an EcoRI-HindIII.	This study

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pRW50/*aggD*101	A derivative of pRW50 carrying the EcoRI-HindIII *aggD*101 promoter fragment cloned as a *lacZ* transcriptional fusion.	Douglas Browning
pRW50/*aggD*102	A derivative of pRW50 carrying *aggD*102 promoter fragment, constructed by six bp deletion from *aggD*101 fragment and cloned using EcoRI and HindIII sites.	Douglas Browning
pRW50/*nlpA100*	A derivative of pRW50 carrying the EcoRI-HindIII *nlpA*100 promoter fragment cloned as a *lacZ* transcriptional fusion.	Chiara Cerrato
pDOC-K	A donor plasmid for gene doctoring carrying *sacB* gene for counter selection (*amp*^{R}*sacB*). It also contains a kanamycin resistance cassette.	(Lee *et al.*, 2009)
pDOC/*aggR*400/*aggR*401	A donor plasmid for gene doctoring with the *aggR*400 fragment cloned in homology region 1 using an EcoRI-BamHI and the *aggR*401 fragment in homology region 2 using a XhoI-SalI.	This study
pACBSR	A plasmid encoding the λ red gene and *secI* under the control of an arabinose inducible promoter for use in gene doctoring (cm^{R}).	Scarab Genomics
pBAD/*aggR*	A derivative of pBAD30 carrying *aggR* cloned using EcoRI-XbaI under the control of the *araBAD* promoter.	(Sheikh *et al.*, 2002)
pBAD/*aggR*-I14T	A derivative of pBAD/*aggR* carrying I14T substitution.	Rita Godfray
pBAD/*aggR*-N16D	A derivative of pBAD/*aggR* carrying N16D substitution.	Rita Godfray

Table 2.2 Plasmids used in this study (continued)

Plasmids	Description	Source
pBAD/*aggR*-Q230G	A derivative of pBAD/*aggR* carrying Q230G substitution.	Rita Godfray
pBAD/*aggR*-M234G	A derivative of pBAD/*aggR* carrying M234G substitution.	Rita Godfray
pBAD24	A derivative of pBAD vector series carrying restriction sites to clone the gene of interest under the control of the *araBAD* promoter and ampicillin resistance cassette.	(Guzman *et al.*, 1995)

NEB or Bioline was used in every gel to size the DNA fragments. Samples were subject to electrophoresis in 0.5x TBE tank at 80 V for 30 to 45 minutes and viewed and imaged using 300 nm UV light with an XR Gel Doc system (Bio-Rad). For DNA fragment purification, the sample were visualised by using High Performance Ultraviolet Transilluminator (Ultra Violet Products Ltd.) and DNA bands were excised for DNA purification with a QIAquick gel extraction kit (Qiagen).

2.4.2 Polyacrylamide gel electrophoresis (PAGE) of DNA

Polyacrylamide gels were used to identify and purify DNA fragments of 50 to 800 bp in length. Acrylamide gel solution of 7.5% (w/v) was made by mixing 100 ml of 5x TBE solution (section 2.4.1), 125 ml of 30% (w/v) acrylamide (National Diagnostics), 20 ml of glycerol per 500 ml. The acrylamide gel was made by adding 100 µl of freshly prepared 10% (w/v) of ammonium persulphate (APS) and 10 µl of N, N, N', N'-Tetramethylethylenediamine (TEMED) per 10 ml of 7.5% (w/v) polyacrylamide gel mix. The mixture was poured between two glass plates held apart by ~0.15 cm thick spacers. The wells were made by placing a comb at the top of the gel and the comb and bottom spacer were removed when the mixture had polymerised. The glass plates with polymerised gel were fixed into a vertical electrophoresis apparatus and the upper and lower tanks were filled with 1x TBE. DNA samples were prepared by mixing with loading dye with DNA samples in a ratio of 1:6 and loaded into washed wells in the gel. Samples were then subject to electrophoresis at 40 mA for 45 to 60 minutes. The gel was stained in a 0.5 μgml^{-1} ethidium bromide solution for 15 minutes and viewed at 300 nm using an XR Gel Doc system (Bio-Rad). For DNA fragment purification, the DNA bandds were visualised using a High Performance Ultraviolet Transilluminator (Ultra Violet Products Ltd.) and the desired DNA bands were excised for DNA electroelution and purification.

2.4.3 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein separation was carried out using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and gels were made using minigel glass plates (Bio-Rad). The SDS-PAGE gels contain two parts with different acrylamide concentrations, *i.e.* the resolving gel and the stocking gel. A 12% (w/v) resolving gel mixture (4 ml of 30% (w/v) acrylamide (National Diagnostics), 2.5 ml of lower stock solution (1.5 M Tris-HCl (pH 8.8), 0.4% (w/v) SDS, 0.001% (v/v) TEMED), 200 µl of APS (10% (w/v)) and 3.3 ml of dH_2O) was poured into assembled glass plates ¾ of the way up the glass plates. The resolving gel was left for polymerization and an isopropanol layer was added on top. After polymerization, isopropanol was tipped off and the of the gel was rinsed with dH_2O and dried with filter paper. Another 10 ml mixture of 4% (w/v) stacking gel mixture (1 ml of 30% (w/v) acrylamide (National Diagnostics), 2.5 ml of lower buffer stock solution (0.5 M Tris-HCl, pH 6.8, 0.4% (w/v) SDS, 0.002% (v/v) TEMED), 200 µl of APS (10% (w/v)) and 6.3 ml of dH_2O) was poured into the glass plates and a comb of desired well width was inserted. The stocking gel was allowed to polymerise, and then the comb was removed. The glass plates were fixed in an electrophoresis tank with running buffer (25 mM Tris-HCl, 0.1% (w/v) SDS (v/v), 190 mM glycine) in the tank. Protein samples were mixed with cracking buffer and heated for 10 minutes. The samples were centrifuged and loaded into wells along with a prestained protein ladder. The gel was electrophoresed with running buffer in the tanks (BioRed) at 120V for 1.5 to 2 hours. Protein was detected by staining gels with Coomassie blue R dye (methanol (50%v/v), acetic acid (10%v/v) and brilliant blue R (2% (w/v)) for one hour at room temperature with shaking. The gel was destained using destain solution (10% (v/v) glacial acetic acid, 40% (v/v) methanol) until the protein bands became visible.

2.5 Extraction and purification of nucleic acids

2.5.1 Phenol/chloroform extraction of DNA

In order to purify DNA after restriction enzyme digestion, phenol/chloroform extraction was used. The DNA solution was mixed with an equal volume of phenol chloroform and isoamylalcohol solution (25, 24 and 1 parts, respectively) (Sigma Aldrich), vortexed and centrifuged at 15700 g for 2 minutes. This separated the solution into two layers: the lower turbid layer of phenol cholorform and the upper clear aqueous layer, containing DNA. The upper aqueous layer was carefully transferred into a clean microfuge tube and the DNA was concentrated by ethanol precipitation as detailed below.

2.5.2 Ethanol precipitation of DNA

For ethanol precipitation, the aqueous supernatant from the phenol chloroform extraction was adjusted to 400 µl by adding sterile distilled water. Then 4 µl ($1/100^{th}$ volume) of 1M $MgCl_2$, 40 µl (1/10 volume) of 3 M sodium acetate (pH 5.2) and 888 µl (2 volumes) of ice cold absolute ethanol were added to the DNA solution and the mixture was incubated at -20°C. The microfuge tube was then centrifuged for 15 minutes at 4°C, the supernatant removed and the pellet was washed with 1 ml of ice-cold 70% (v/v) ethanol. The microfuge tube was centrifuged for another 10 minutes at 4°C and the supernatant was removed. The pellet was washed with 1 ml of ice cold 100% ethanol and centrifuged for 10 minutes at 4°C. The supernatant was again removed and the pellet was dried for 40-45 minutes in a vacuum drier. The pellet was resuspended in 20-50 µl sterile distilled water. For ethanol precipitation of single stranded DNA, the same precipitation method was used, except 3 M ammonium acetate (pH 4.8) was used instead of sodium acetate. For ethanol precipitation of RNA, the same method was also employed but 3 M sodium acetate was used at pH 7.0.

2.5.3 Extraction of DNA fragments from agarose gels

As described in section 2.4.1, DNA samples were mixed with loading dye in a ratio of 6:1 and run on 0.8% (w/v) agarose gel. DNA bands were visualised on High Performance Ultraviolet Transilluminator (Ultra Violet Products Ltd.) and excised. The excised DNA was then purified using a QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions. Extracted DNA fragments were eluted from the QIAquick columns in 50 µl of sterile distilled water.

2.5.4 Electroelution of DNA fragments from polyacrylamide gels

As described in section 2.4.2, the DNA samples were mixed with loading dye, were also run on 7.5% (w/v) polyacrylamide gels. DNA bands were visualised on High Performance Ultraviolet Transilluminator (Ultra Violet Products Ltd.) and excised from the gel. Gel slices were then placed into 6 mm dialysis tubing (Medicell International Ltd.), containing 200 µl of 0.1x TBE buffer, and the ends of tubing were sealed with dialysis clips. These samples in dialysis tubing were subject to electroelution by passing a current through them at 30 mA for 20-30 minutes. The buffer was removed from the dialysis tubing and transferred into a microfuge tube. The dialysis tubing was rinsed with 200 µl of sterile distilled water and the water was added to the same microfuge tube. The sample volume was adjusted to 400 µl and the DNA fragment was extracted from the buffer by phenol/chloroform extraction and ethanol precipitation (sections 2.5.1 and 2.5.2).

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

For preparation of plasmid DNA, a single colony was inoculated into 5 ml of LB medium containing the appropriate antibiotic. The overnight culture was transferred into a 15 ml sterile Falcon tube and centrifuged for 5 minutes at 2844 g at room temperature. The pellet was resuspended in P1 buffer and transferred into a microfuge tube. Plasmid DNA was

extracted using QIAprep Spin miniprep Kit, following the manufacturer's instructions.

2.5.6 Large-Scale preparation of plasmid DNA using QIAprep Spin maxiprep kit

For preparation of large quantities of plasmid DNA a single colony was inoculated into 250 ml of LB medium containing the appropriate antibiotic. The overnight culture (250 ml) was transferred into a centrifugation tube and the culture was centrifuged at 5400 gfor 15 minutes at 4°C. The pallet was resuspended in P1 buffer and transferred into a 50 ml falcon tube. Plasmid DNA was extracted using a QIAprep Spin maxiprep Kit, according to the manufacturer's instructions.

2.6 Transformation of *E. coli* with plasmid DNA

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

A single bacterial colony was picked and inoculated into 5 ml of LB medium containing antibiotic, where appropriate and grown overnight at 37°C with shaking. The culture was diluted 1:100 into 50 ml of fresh LB medium and incubated at 37°C with shaking until bacteria reached the mid-logarithmic phase of growth (OD_{600} = 0.3 - 0.6). The culture was then transferred to a pre-cooled 50 ml Falcon tube, kept on ice for 10 minutes and then centrifuged at 2844 g at 4°C for 15 minutes. The pellet was resuspended in 25 ml of ice cold 0.1 M $CaCl_2$ and kept on ice for a further 30 minutes. Cells were centrifugated for 15 minutes at 2844 g at 4°C and the pellet was resuspended in 3.5 ml of ice cold 0.1 M $CaCl_2$ with 10% (v/v) glycerol. Competent cells were kept on ice for upto 24 hours before being stored at -80°C in 500 µl aliquots.

2.6.2 Transformation of plasmid DNA into chemically competent cells

For transfer of plasmid DNA into competent cells, the frozen competent cells were thawed on ice. One microliter of plasmid DNA was mixed with 100 µl of competent cells on ice, and left

for 30 minutes. The cells were heat-shocked at 42°C for 2 minutes and then placed on ice for another 2 minutes. After addition of 1 ml LB medium, the cells were incubated for 1 hour at 37°C with shaking. After incubation, the cells were centrifuged for 1 minute at 15700 g rpm and the pellet resuspended in approximately 100 µl of the supernatant. The resuspended pellet was plated onto LB agar or MacConkey agar supplemented with the appropriate antibiotic(s), and incubated overnight at 37°C.

2.6.3 Preparation of electrocompetent cells

A single bacterial colony was picked and inoculated into 5 ml of LB containing antibiotic where appropriate and grown overnight at 37°C with shaking. The overnight culture was then diluted 1:100 into 50 ml of LB and incubated at 37°C with shaking until the culture reached mid-logarithmic phase (OD_{600}= 0.3 - 0.6). The culture was transferred to a pre-cooled 50 ml Falcon tube and centrifuged at 2844 g at 4°C for 15 minutes. The pellet was then resuspended in 25 ml ice-cold 10% (v/v) glycerol and this process was repeated three times to remove any traces of salt. Following the removal of all supernatant after the last centrifugation, the pellet was resuspended in 500 µl of ice-cold 10% (v/v) glycerol and the cells were kept on ice to be used on the same day.

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

For electroporation, 100 µl of electrocompetent cells were mixed with 10 µl of plasmid DNA in a pre-cooled 2 mm electroporation cuvette (Invitrogen). The mixture was left for one minute in the cuvette and cells were electroporated at 2.1 kV using a Gene Pulser (BioRad). 1 ml of pre-warmed SOC medium at 37°C was added to the cuvette and then the cells were transferred to a pre-warmed microfuge tube at 37°C. Cells were incubated at 37°C for one hour with shaking. After which they were centrifuged at 15700 g for one minute and plated at onto LB agar or MacConkey agar containing the appropriate antibiotic.

2.7 Recombinant DNA techniques

2.7.1 Polymerase chain reaction (PCR)

Phusion™ High-Fidelity DNA polymerase, Q5 DNA polymerase (New England BioLabs® Inc.), Biomix red DNA polymerase and Mytaq red DNA polymerase (Bioline) were used in the study to amplify the DNA of interest using PCR. Q5 DNA polymerase was used routinely due to its high fidelity. The DNA templates used in this study were from three sources; bacterial cell lysates, purified bacterial genomic DNA and purified bacterial plasmids.

For a typical 50 µl reaction for Q5 DNA polymerase the reaction buffer contained 10 µl of 5x Q5 buffer, 10 µl of Q5 GC buffer, 10 µl of forward primer (2.5 µM), 10 µl of reverse primer (2.5 µM), 1 µl of 0.8 mM dNTPs (Bioline), 5 µl of of template DNA (1:100 diluted), 0.5 µl of Q5 DNA polymerase (New England BioLabs® Inc.) and 3.5 µl of dH_2O. DNA amplification was performed in a thermal cycler (GeneAmp® PCR System, Applied Biosystems) and PCR cycling conditions are shown in the Table 2.3 and primers used in this study are detailed in Table 2.4. PCR products were run on agarose or polyacrylamide gels whenever the PCR product visualization was required. Biomix red DNA or Mytaq red DNA polymerase was typically used for colony PCR and error prone PCR.

2.7.2 Colony PCR

Colony PCR was used to screen for clones or confirm the presence of a DNA segment in a bacterial genome and cell lysates of candidate colonies was used as template DNA. The DNA template, was prepared by mixing a single colony off an agar plate in 100 µl of distilled water in microfuge tube which was heated at 100°C for 10 minutes. The cell debris was palleted by centrifugation of the microfuge tube and 5 µl of the supernatant was used as template in 50 µl PCR reaction. For a 50 µl reaction of Biomix Red DNA polymerase and Mytaq Red DNA

Table 2.3 PCR cycling conditions

Cycle Step	Temp.	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	10 s	
Annealing	x°C	30 s	25-30
Extension	72°C	15-30 s/1 kb amplified	
Final extension	72°C	10 minutes	1
	4°C	Hold	

Table 2.4 Oligonucleotide primers (continued on pages 67-73)

Code	Sequence (5' to 3')	Use
D10520	CCCTGCGGTGCCCCTCAAG	Anneals upstream of the EcoRI site in pRW50/ pRW224/ pRW225. Used for sequencing and amplification of inserts in this vector.
D10527	GCAGGTCGTTGAACTGAGCCTGAA ATTCAG	Anneals downstream of the HindIII site in pRW50. Used for sequencing and amplification of inserts in this vector.
D78264	GGCTGTAATGTTCTGGCATTTGGTC AGC	Anneals downstream of the HindIII site in pRW50. Used for sequencing and amplification of inserts in this vector
D19897	GGCGATTAAGTTGGGTAACGCCAG GG	Anneals downstream of the HindIII site in pRW225/ pRW224. Used for sequencing and amplification of inserts in these vectors
D49274	GGTTGGACGCCCGGCATAGTTTTTC AGCAGGTCGTTG	Anneals to pRW50 sequence downstream of the HindIII site and is used in primer extension.
D78453	GGGGGAATTCTGGTGCTTCAGGTGT GTGACATGGG	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*100, *aafD*1001 and *aafD*1002 promoter fragments.
D78454	GGGGGGAAGCTTTCCGTATTTTCAT TTTATACATTCTCC	*aafD* downstream primer containing a HindIII site for amplification of the *aafD*100, *aafD*99, *aafD*98, *aafD*97, *aafD*96, *aafD*95 *and aafD*94 promoter fragments.
D78639	GGGGGGGAATTCATGTACCGCCAA CAATGCGGG	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*99 promoter fragment.

Table 2.4 Oligonucleotide primers (continued)

Code	Sequence (5' to 3')	Use
D78640	GGGGGGGAATTCCGTAGTATTGCC AACTGAATC	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*98 promoter fragment.
D78641	GGGGGGGAATTCCACCTGATTTATT CAATAAAGTCTG	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*97 promoter fragment.
D79177	GGGGGGAATTCGTCTGCACAGTGG TGTTTATTTATC	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*96 promoter fragment.
D79178	GGGGGGAATTCTTATCTTTTTAGTA ACTTTGTTTTAAG	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*95 promoter fragment.
D79179	GGGGGGAATTCTTTTAAGTAGCATA TTAACTTAATCG	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*94 promoter fragment.
D80489	GGGGGGGAATTCGTCTGCACAGTG GTGTTTATTCACCTTTTTAGTAAC	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD96-92C90C* promoter fragment.
D80490	GGGGGGGAATTCGTCTGCACAGTG GTGTTTATTTATCTTTTTAGTAACTT TGTTTTAAGTCGCATATTAAC	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*96-65*C* promoter fragment.
D81042	GGGGGGGAATTCGTCTGCACAGTG GTnTTTATTTATCTTTTTAGTAAC	*aafD* upstream primer containing an EcoRI site for amplification of the *aafD*96-99*N* promoter fragments and n represents any base.

Table 2.4 Oligonucleotide primers (continued)

Code	Sequence (5' to 3')	Use
D78637	GGGGGGAAGCTTCATTTTATACATTCTCCATCCTCTTTAG	*aafD* downstream primer containing a HindIII site for amplification of the *aafD*1001 promoter fragment.
D78638	GGGGGGAAGCTTCATTCTCCATCCTCTTTAGAGG	*aafD* downstream primer containing a HindIII site for amplification of the *aafD*1002 promoter fragment.
D82244	GGGGGGAAGCTTTTTACGATTAAGTTAATATG	*aafD* downstream primer containing a HindIII site for amplification of the *aafD*500 promoter fragment.
D78457	GGGGGGGAATTCGACATGATAACGAATTAAGCAAGG	*afaB* upstream primer containing an EcoRI site for amplification of the *afaB*100 promoter fragment.
D78458	GGGGGGAAGCTTGGGAAAATACTCTGGAGTTGGC	*afaB* downstream primer containing a HindIII site for amplification of the *afaB*100 promoter fragment.
D79180	GGGGGGAATTCATGTGACATTCCTGCACTG	*afaB* upstream primer containing an EcoRI site for amplification of the *afaB*99 promoter fragment.
D79181	GGGGGGAATTCCCTTTCACAGGTAATGCAG	*afaB* upstream primer containing an EcoRI site for amplification of the *afaB*98 promoter fragment.
D79182	GGGGGGAATTCGATGTATTGGTAGGGGCCT	*afaB* upstream primer containing an EcoRI site for amplification of the *afaB*97 promoter fragment.

Table 2.4 Oligonucleotide primers (continued)

Code	Sequence (5' to 3')	Use
D80520	GTGTTTTTATCACCATTATGTGACA TTCCTGC	*afaB* upstream primer for megaprimer PCR amplification of the *afaB*100-320*C*318*C* promoter fragment.
D81046	CATTCCTGCACTGTCTCTTTAATAG GTGGG	*afaB* upstream primer for megaprimer PCR amplification of the *afaB*100-293*C* promoter fragment.
D78629	GGGGGAATTCAGAGTATTTTCCCTT CATCTG	*afaB* upstream primer containing an EcoRI site for amplification of the *afaBC*99 promoter fragment
D78635	GGGGGGGAATTCAAATCAACAAAC TCTAATTCTAG	*aafA* upstream primer containing an EcoRI site for amplification of the *aafA*100 promoter fragment.
D78636	GGGGGGAAGCTTTGATTTTTTTCAT GTCAACCTC	*aafA* downstream primer containing a HindIII site for amplification of the *aafA*100 promoter fragment.
D78455	GGGGGGGAATTCAGGCTGACGATA AATGGGAGG	*aafC* upstream primer containing an EcoRI site for amplification of the *aafC*100 promoter fragment.
D78456	GGGGGGAAGCTTATGTACATGTCA TCATCACTG	*aafC* downstream primer containing a HindIII site for amplification of the *aafC*100 promoter fragment.
D80569	GGGGGAATTCAAATGGGGGCGGAA TCCTAGTGTTAG	*aggD* upstream primer containing an EcoRI site for amplification of the *aggD*50 promoter fragment.
D80570	GGGGGGGAATTCATGTTAAATAAT GCTATTTTTTTAGCG	*aggD* upstream primer containing an EcoRI site for amplification of the *aggD*49 promoter fragment.

Table 2.4 Oligonucleotide primers (continued)

Code	Sequence (5' to 3')	Use
D80571	GGGGGGGAATTCTTAGCGTTATATG ATTTGAGTC	*aggD* upstream primer containing an EcoRI site for amplification of the *aggD*48 promoter fragment.
D82240	GGGGGGGAATTCATGTTAAATAAT GCTATTTTTTCAGCGTTATATG	*aggD* upstream primer containing an EcoRI site for amplification of the *aggD*49-86*C* promoter fragment.
D82312	GATTTGAGTCTTTATCTAATCAAGT TCAAG	*aggD* upstream primer for megaprimer PCR amplification of the *aggD*49-60*C* promoter fragment.
D79196	GGGGGAATTCTCCTTTGTTTTATGG ATAGTTTTCTGC	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*100 promoter fragment.
D79197	GGGGAAGCTTGTGTATTGCTCATGT TTCTATCCTAC	*aaiA* downstream primer containing a HindIII site for amplification of the *aaiA*100 promoter fragment.
D79296	GGGGGGGAATTCAGTTAGTAAATA CGAAAAATTATATAG	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*99 promoter fragment.
D81342	GGGGGGGAATTCCATTAATCAGCA AAAATGTATCACATGC	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*98 promoter fragment.
D81343	GGGGGGGAATTCCACATGCTCACTT TCTTTTTATGG	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*97 promoter fragment.
D81344	GGGGGGGAATTCCACTATATAGAA TCCATGAATATAACAAGAG	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*96 promoter fragment.

Table 2.4 Oligonucleotide primers (continued)

Code	Sequence (5' to 3')	Use
D80523	GGGGGGGAATTCCTTTCGAAAACA TCTGTGCGGTAGC	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*95 promoter fragment.
D81316	GGGGGGGAATTCCATTAATCAGCA AAAATGCACCACATGC	*aaiA* upstream primer containing an EcoRI site for amplification of the *aaiA*98-254*C*-252*C* promoter fragment.
D82313	CTTTTTATGGTCTCACTATATAGAA TCC	*aaiA* upstream primer for megaprimer PCR amplification of the *aaiA*100-226*C* promoter fragment.
D82314	GTTGCTATTTTTCACCTGGCCGCAA CTC	*aap* upstream primer for megaprimer PCR amplification of the *aap*100-151*C*149*C* promoter fragment.
D82315	CTCTTATTTATGCTCGCCTTCTAAA AGG	*aap* upstream primer for megaprimer PCR amplification of the *aap*100-124*C* promoter fragment.
D82320	GTGTTTATTTATCTTTTATCCCCTCA CTCCTG	Upstream primer for megaprimer PCR amplification of the *DAM*20 promoter fragment.
D81044	GATCAGGTAAATGGTGTTTATTTAT CTTTTATCCCCCTCACTCCTG	Upstream primer for megaprimer PCR amplification of the *DAM*21 promoter fragment.
D81045	GATCAGGTAAATGGTGTTTATTTAT CTTTTATCCCCCCTCACTCCTG	Upstream primer for megaprimer PCR amplification of the *DAM*22 promoter fragment.
D82321	GTGTTTATTTATCTTTTATCCCCCCC TCACTCCTG	Upstream primer for megaprimer PCR amplification of the *DAM*23 promoter fragment.

Table 2.4 Oligonucleotide primers (continued)

Code	Sequence (5' to 3')	Use
D82558	GATCAGGTAAATGGTGTTTATTCAC CTTTTATCCCCCCTC	Upstream primer for megaprimer PCR amplification of the *DAM*22-(-38*C*-36*C*) promoter fragment.
D80428	GGGGGGGAATTCTAATGAGCGAAA ATAGCCTG	Upstream primer containing an EcoRI site for amplification of the *aggR*400 to clone the PCR product into the donor plasmid pDOC-K
D80429	GGGGGGGGATCCATCATTCTCACAT GAAAAAAATGTC	Downstream primer containing a BamHI site for amplification of the *aggR*400 to clone the PCR product into the donor plasmid pDOC-K
D80430	GGGGGGCTCGAGAAACATGTTTCA TATCATTATTTGAG	Upstream primer containing a XhoI site for amplification of the *aggR*401 to clone the PCR product into the donor plasmid pDOC-K
D80431	GGGGGGGTCGACTATGCGCGAGGT TACCGACTGCGG	Downstream primer containing a SalI site for amplification of the *aggR*401 to clone the PCR product into the donor plasmid pDOC-K
D82322	CACCGTGTCTGGGCGATACTCAGG	Anneals to the pDOC-K plasmid upstream of the EcoRI restriction site and used for sequencing and amplification of inserts in this vector
D82323	GGAGACTGTCATACGCGTAAAACA G	Anneals to the pDOC-K plasmid downstream of the SalI restriction site. Used for sequencing and amplification of inserts in this vector

polymerase, the mixture contained 25 µl of 2x reaction mix, 10 µl of forward primer (2.5 µM), 10 µl of reverse primer (2.5 µM) (Table 2.4) and 5 µl of of template DNA (bacterial cell lysate). DNA amplification was performed in a thermal cycler (GeneAmp® PCR System, Applied Biosystems). PCR cycling conditions are shown in the Table 2.3.

2.7.3 Error prone PCR

Error prone PCR was used to introduce random mutations into the *aafD* promoter. In this PCR reaction, low fidelity Taq DNA polymerase was used. For each 50 µl reaction, containing either Biomix Red DNA polymerase or Mytaq Red DNA polymerase, 25 µl of 2x reaction mix, 10 µl of forward primer (2.5 µM), 10 µl of reverse primer (2.5 µM) (Table 2.4) and 5 µl of of template DNA (1:100 diluted) was used. The PCR cycling conditions were as in Table 2.3 except the number of PCR cycles were increased from 30 to 40.

2.7.4 Megaprimer PCR

Megaprimer PCR was used to introduce site directed mutations into the recombinant DNA fragments at positions that were not adjacent to suitable restriction sites (*e.g.* the *afaB*320*C*318*C* promoter fragment). This PCR reaction involved two separate PCR reactions, and require 3 separate primers, two primers which anneal to the ends of the DNA fragment and a third primer that annealed internally and contained the desired substitution. The first round of PCR used the internal annealing primer and one of the end primers. After the first round of PCR, the product was run on a polyacrylamide gel, and purified (sections 2.5.1 and 2.5.2). In the second round of PCR, the purified PCR product from the first round was used as one of the primers with the second end primer in the reaction to generate the full length fragment. The PCR reactions were carried out as described in section 2.7.1.

2.7.5 Restriction digestion of DNA

Both plasmid DNA and purified PCR products were digested using restriction enzymes for DNA cloning. For high copy number plasmids, 12 µl of a plasmid DNA miniprep was digested in a final volume of 60 µl using 1.5 µl of each restriction enzyme (New England BioLabs® Inc.) in 1x NEB 4 buffer or cutsmart buffer and each reaction mixture was incubated at 37°C for 3 hours. The digested vector was also incubated with 1 µl of calf intestinal alkaline phosphatase (CIP) (New England BioLabs® Inc.) for another one hour to avoid re-ligation of plasmid by removing terminal 5' phosphate groups. The digested DNA was run on a 0.8% (w/v) agarose gel and DNA bands were excised and purified using the Qiagen Gel Extraction Kit (section 2.5.3).

For low copy plasmids (*e.g.* pRW50), nine DNA minipreps were pooled together and 360 µl of this miniprep DNA was used with 40 µl of NEB 4 or cutsmart buffer and 10 µl of each restriction enzyme to make a final volume of 400 µl. The mixture was incubated at 37°C for 4 hours and to remove the terminal 5' phosphate groups, 4 µl CIP was added and incubated for another hour at 37°C. The digested vector was then purified using phenol/chloroform extraction and ethanol precipitation (sections 2.5.1 and 2.5.2).

For purified PCR products, ethanol precipitated DNA was resuspended in 48 µl of sterile distilled water and 6 µl of 10x NEB buffer 4 (or cutsmart buffer) and 3 µl of each restriction enzyme was added to make a final volume of 60 µl. The mixture was incubated at 37°C for 2.5 hours. Digested DNA fragments were run on either a polyacrylamide gel or an agarose gel and purified accordingly (sections 2.5.1, 2.5.2 and 2.5.3).

2.7.6 Ligation

The ligation reaction is a process in which an insert is melded to an appropriate vector. DNA ligation was used in this study to clone the promoter fragments or genes of interest into appropriate vectors. For ligation reactions, a final volume of 20 µl was used containing, 5-10 µl of restricted insert DNA, 3-7 µl of vector DNA, 1 µl T4 DNA ligase (New England BioLabs® Inc.), 2 µl 10x T4 DNA ligase buffer and a suitable amount of sterile dH_2O where required to make a final volume of 20 µl. Once setup, ligation mixes were initially kept on ice for 10 minutes and then incubated for at least 10 minutes at room temperature. The mixture was then placed on ice for 5 minutes before the addition of 100 µl of $CaCl_2$ competent cells. The transformation was performed as in section 2.6.2. In order to select for transformants, cells were plated onto MacConkey agar with lactose, supplemented with the appropriate antibiotic. To screen for successful candidates, transformants were grown overnight and plasmid minipreps were prepared. Plasmids were digested with appropriate restriction enzymes to excise the cloned fragment and run on 7.5% (w/v) polyacrylamide gel or 0.8% (w/v) agarose gel to check for the presence of an insert of the expected size.

2.7.7 DNA sequencing

Plasmid-to-profile sequencing was carried out by the Functional Genomics and Proteomics Laboratory, University of Birmingham, UK. For sequencing of a fragment cloned into a low copy plasmid, 7 µl of template was used with 3 µl of primer (1 µM). For a high copy number plasmid, 3 µl of template was used with 3 µl of primer (1 µM) and the volume was adjusted to 10 µl with addition of sterile dH_2O.

2.8 Strategy to locate potential AggR-binding sites in promoter fragments

DNA target sites for the binding of AggR and the closely related Rns protein have been previously investigated, using *in vivo* and *in vitro* approaches (Munson and Scott, 1999; Morin *et al.*, 2010). From these studies, it has been proposed that the potential AggR-binding site consensus sequence is 5′-AnnnnnnTATC-3′ on the forward strand and thus, 5′-GATAnnnnnnT-3′ on the reverse strand. Therefore, based on this consensus, the promoter fragments used in this work were screened for potential AggR-binding sites on both strands allowing for one mismatch to the concensus sequence. When from the deletion analysis predicted sites were found to be inessential using deletion analysis experiments, they were discounted, and potential AggR-binding sequence present in the smallest AggR-regulated fragment was analysed by mutational analysis to identify the functional site.

2.9. Cloning of promoter fragments and/or target genes

All the promoter fragments or target genes used in this study were amplified from EAEC genome using by PCR. For nested deletions, in-phase deletions or site directed mutations, the plasmid containing the starting wild type promoter fragment was used as a template. A list of all the primers used to generate each DNA fragment is shown in Table 2.4. PCR primers were designed such that the forward and reverse primers incorporate restriction enzyme sites upstream and downstream of the amplified target region, respectively. The cloning of each DNA fragment was, therefore, achieved using these restriction sites to introduce inserts into the relevant DNA plasmid.

2.9.1 Construction of pRW50 derivatives

pRW50 plasmid was prepared using both QIAprep Spin, maxiprep and miniprep kits (vector

details are in Figure 2.1) and DNA was digested with EcoRI and HindIII, treated with alkaline phosphatase and purified as in sections 2.5.1 and 2.5.2. DNA fragments were amplified, using the primers detailed in the Table 2.4, and digested with EcoRI and HindIII. Each EcoRI-HindIII promoter fragment was then ligated into pRW50 vector (sections 2.6.1 and 2.6.2).

To investigate the fimbrial operon promoters of EAEC 042, approximately, 400 bp of DNA was amplified upstream of *aafD, aafA, afaB* and *aafC* using the relevant forward and reverse primers (Figure 2.2 and Table 2.4). The customary numbering of, bases in relation to transcription start site was not followed because the transcription start sites of these promoters are not confirmed yet. The DNA fragments *afaB*100, *aafC*100, *aafD*100 and *aafA*100 are shown in Figure 2.2, were cloned into the *lacZ* expression vector pRW50 using EcoRI and HindIII sites. In order to locate the *aafD* promoter deletion analysis of *aafD*100 promoter fragment was carried out. The primers used for deletion analysis are described in the Table 2.4 and pRW50/*aafD*100 was used as template (Figure 2.3). This deletion analysis generated promoter fragments *aafD*99 to *aafD*94. Point mutations were introduced into the *aafD* promoter fragment to confirm the position of promoter elements and AggR-binding sites. This mutation analysis generated promoter fragments *aafD*96-90*C, aafD*96-92*C*90*C* and *aafD*96-65*C* (Figure 2.4).

In order to investigate *afaB* promoter in more detail, a series of nested deletions were also made using PCR with the relevant primers listed in the Table 2.4 and pRW50/*afaB*100 as a template. The DNA sequence of these *afaB* promoter fragments (*afaB*99 to *afaB*97) are detailed in Figure 2.5. Fragments were cloned into pRW50, as transcription fusions, again using EcoRI and HindIII sites. To investigate the promoter -10 hexamer element and the AggR-binding site that control expression, point mutations were introduced into the *afaB*100 promoter fragment, generating *afaB*100-293*C* and *afaB*100-320*C*318*C* (Figure 2.6). To

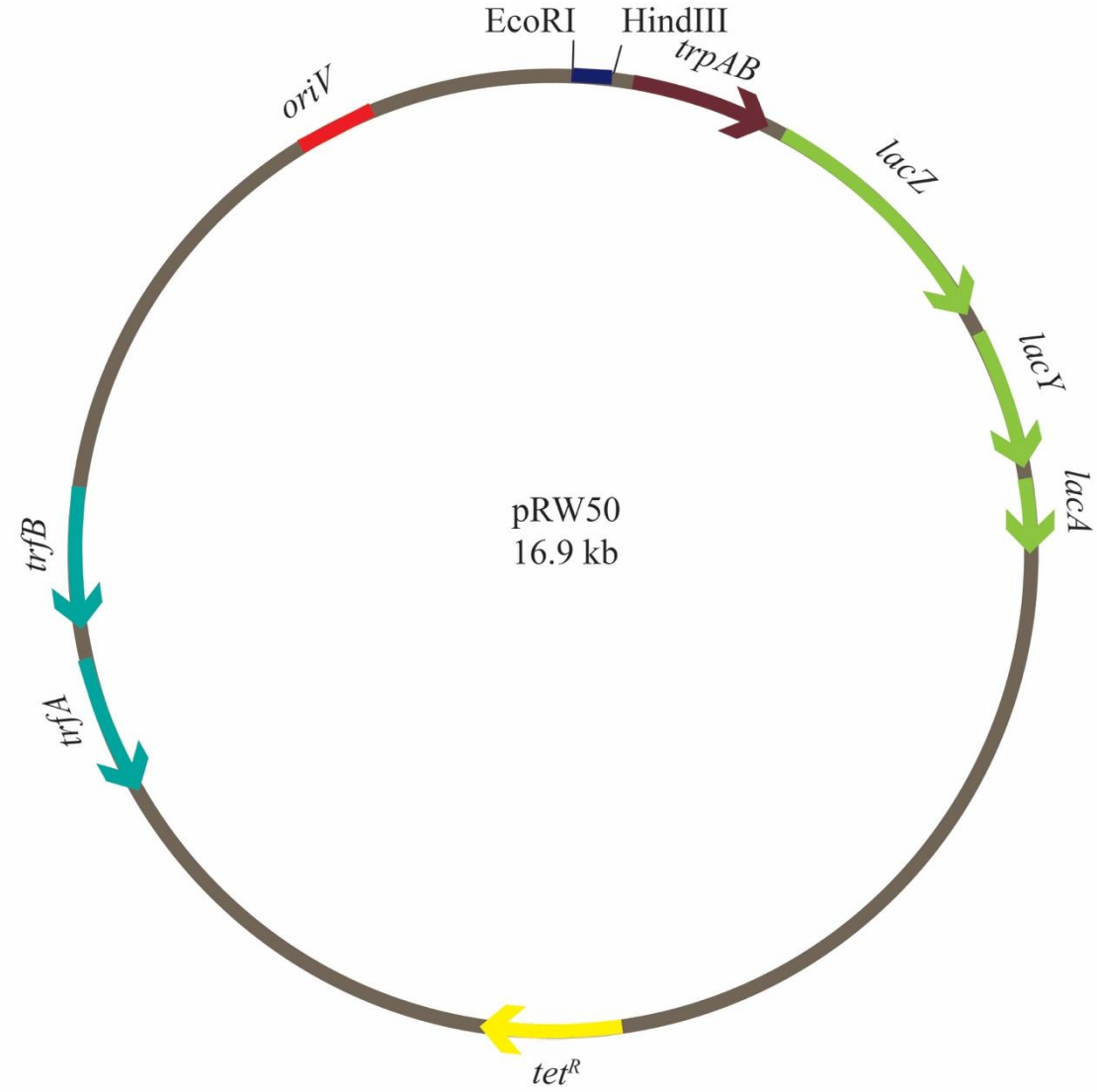
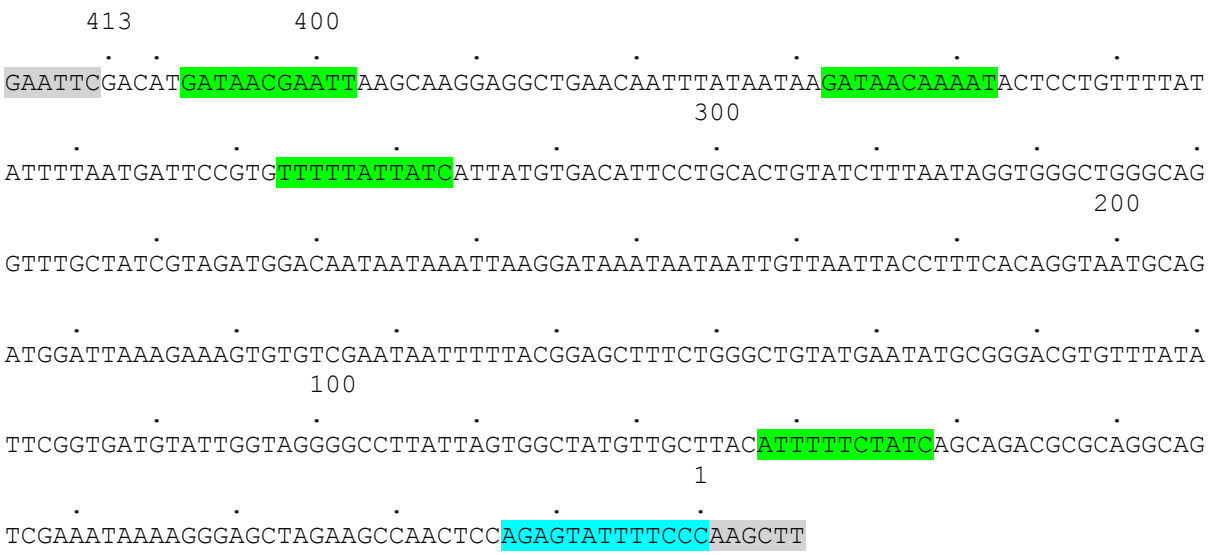


Figure 2.1 Map of plasmid pRW50

The figure shows a map of the pRW50 *lacZ* expression vector and its associated open reading frames. EcoRI and HindIII sites are located upstream of the *lacZYA* operon and are used to clone in promoter fragments for analysis. The plasmid carries tet resistance (tet^R).

A

***afaB*100 (413 bp)**



B

***aafC*100 (410 bp)**

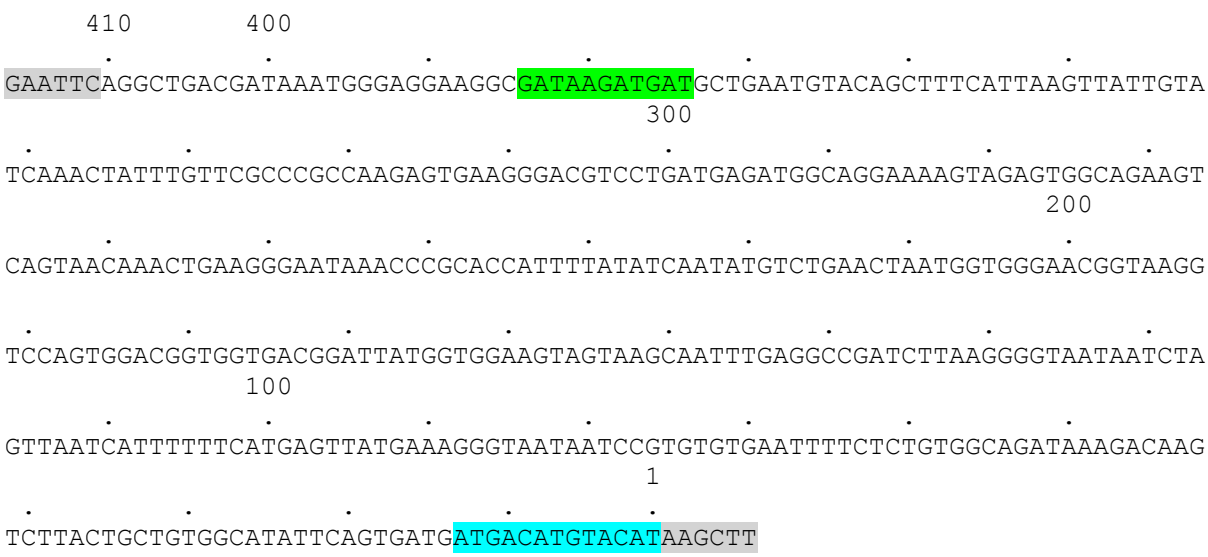


Figure 2.2 The DNA base sequence upstream of the genes encoding the AAF/II (continued)

C

***aafD*100(433 bp)**

```
     433                                400
      .  .         .         .         .         .         .         .
GAATTCGCGTAAATAGTTGGAAAAGTTCATGCAGCATGTGGATGCTTACAGCAGATGGTATAACGAGCGGCGT
                                                                    300
     .         .         .         .         .         .         .
ATAAAATTATCGCTGGGTGCAGTCAGCCCTGAAATGTACCGCCAACAATGCGGGCTGGCATAATAAAGCAGTCCA

 .         .         .         .         .         .         .         .
GGAAATCGTCCGCATCCCCTAACGGTCAAACATCGTGGCCTTGACAGCTCCCTTGGCGCGTAGTATTGCCAACTG
               200
     .         .         .         .         .         .         .
AATCCCCCGTCAATACGGTTCTCACCGTATGCTCACTGCTTACAATTGCCTGACAGTAGCAACCAACTGAGAGGA
                                        100
 .         .         .         .         .         .         .         .
TGCTATCTCACCTGATTTATTCAATAAAGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAG
                                                                    1
     .         .         .         .         .         .         .
TAGCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGAGAATGTATAAAATGAAAATACGGAAAGCTT
```

D

***aafA*100 (433 bp)**

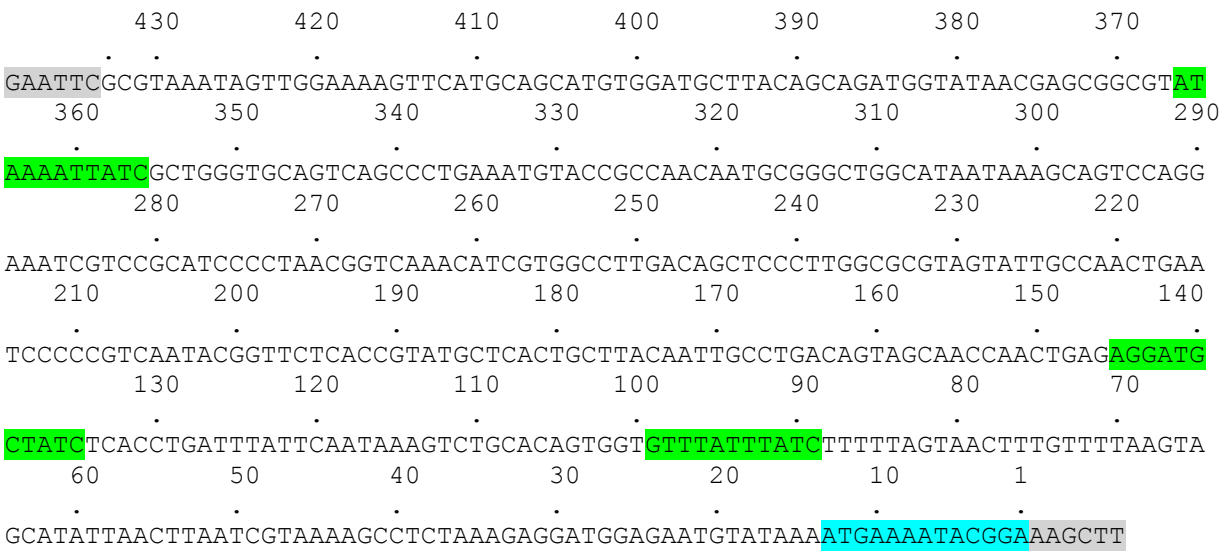
```
     433                                400
      .  .         .         .         .         .         .         .
GAATTCAAATCAACAAACTCTAATTCTAGTACAGAACGAAAAACACGCCGTATTACAGAAAATATAAACCAGTTA
                                                                  300
    .         .         .         .         .         .         .         .
CTTGGTCTATGCGGATATATAAGTGCATCATATAATACATCTTGGAATAATATCCTTCTTACTAGATTACTCATA

         .         .         .         .         .         .         .
CAACCGTAATATAATTGATTACTATGGTAACAAGAAAATCTCAAGAAACCTTATGGCGACATACATATCAAAAAA
             200
    .         .         .         .         .         .         .         .
ATAGTTTCACAGCACAAGATACTAGTATGCATGAATTATTAAGTACTGGACCACCGAAATGGCCATTCTTCTGTT
                                       100
         .         .         .         .         .         .         .
TCAGGTCAGAACGTATCGTAATGATAGAGTGTAAATAAAACGCTCTTTCTTTCTGTAGATGAATAAATATATACG
                                                                  1
    .         .         .         .         .         .         .
GAAACCACCTGTATATTTAAATGAAAAATAGATGTATTTTTAGAGGTTGACATGAAAAAAATCAAAGCTT
```

Figure 2.2 The DNA base sequence upstream of the genes encoding the AAF/II

A. The panel shows the 413 bp DNA sequence of the *afaB*100 promoter fragment.
B. The panel shows the 410 bp DNA sequence of the *aafC*100 promoter fragment.
C. The panel shows the 433 bp DNA sequence of the *aafD*100 promoter fragment.
D. The panel shows the 433 bp DNA sequence of the *aafA*100 promoter fragment.
In all panels open reading frame sequence is highlighted blue, the potential AggR-binding sites are green and EcoRI and HindIII sites are grey.

A

***aafD*100 (433 bp)**



B

***aafD*99 (333 bp)**

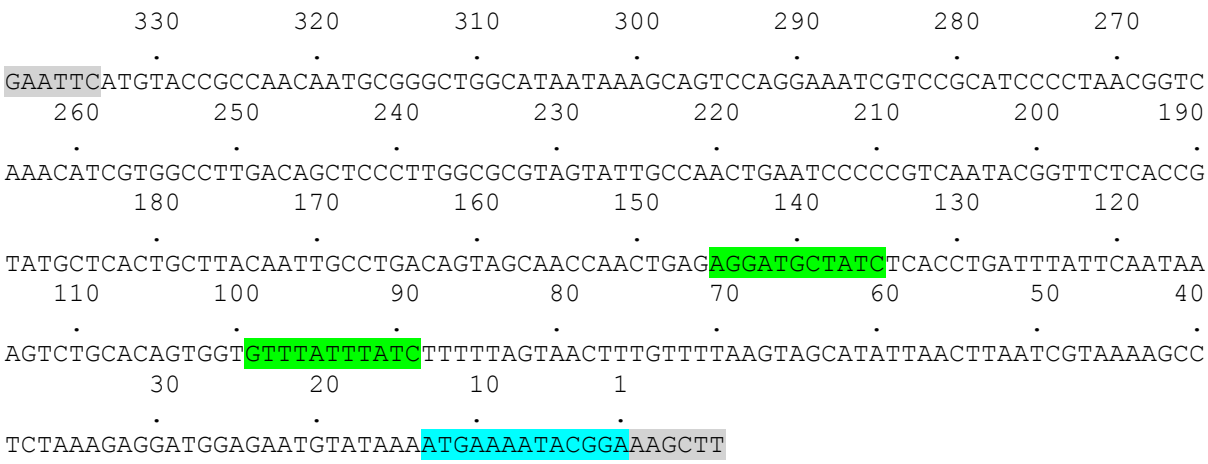


Figure 2.3 The DNA base sequence of the *aafD* nested promoter deletions (continued)

C

***aafD*98(233 bp)**

```
        230       220       210       200       190       180       170
         .         .         .         .         .         .         .
GAATTCCGTAGTATTGCCAACTGAATCCCCCGTCAATACGGTTCTCACCGTATGCTCACTGCTTACAATTGCCTG

   160       150       140       130       120       110       100       90
    .         .         .         .         .         .         .         .
ACAGTAGCAACCAACTGAGAGGATGCTATCTCACCTGATTTATTCAATAAAGTCTGCACAGTGGTGTTTATTTAT

         80        70        60        50        40        30        20
         .         .         .         .         .         .         .
CTTTTTAGTAACTTTGTTTTAAGTAGCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGAGAATGTATAA
    10        1
    .         .
AATGAAAATACGGAAAGCTT
```

D

***aafD*97 (133 bp)**

```
        130       120       110       100        90         80        70
         .         .         .         .         .         .         .
GAATTCCACCTGATTTATTCAATAAAGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAGTA
    60        50        40        30        20        10        1
    .         .         .         .         .         .         .
GCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGAGAATGTATAAAATGAAAATACGGAAAGCTT
```

E

***aafD*96 (113 bp)**

```
        110       100        90        80        70        60        50
         .         .         .         .         .         .         .
GAATTCGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAGTAGCATATTAACTTAATCGTAA
    40        30        20        10        1
    .         .         .         .         .
AAGCCTCTAAAGAGGATGGAGAATGTATAAAATGAAAATACGGAAAGCTT
```

Figure 2.3 DNA base sequence of different *aafD* promoter fragments (continued)

F

***aafD*95 (93 bp)**

```
          90         80         70         60         50         40         30
          .          .          .          .          .          .          .
GAATTCTTATCTTTTTAGTAACTTTGTTTTAAGTAGCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGA
    20         10        1
    .          .         .
GAATGTATAAAATGAAAATACGGAAAGCTT
```

G

***aafD*94 (73 bp)**

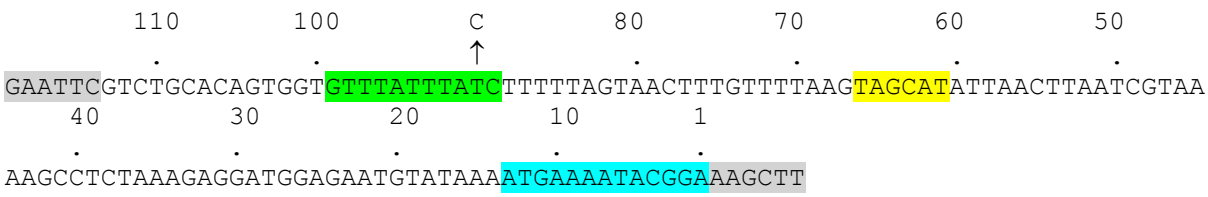
```
          70         60         50         40         30         20         10
          .          .          .          .          .          .          .
GAATTCTTTTAAGTAGCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGAGAATGTATAAAATGAAAATA
   1
   .
CGGAAAGCTT
```

Figure 2.3 DNA base sequence of different *aafD* promoter fragments

A. The panel shows DNA sequence of the 433 bp *aafD*100 promoter fragment.
B. The panel shows DNA sequence of the 333 bp *aafD*99 promoter fragment.
C. The panel shows DNA sequence of the 233 bp *aafD*98 promoter fragment.
D. The panel shows DNA sequence of the 133 bp *aafD*97 promoter fragment.
E. The panel shows DNA sequence of the 113 bp *aafD*96 promoter fragment.
F. The panel shows DNA sequence of the 93 bp *aafD*95 promoter fragment.
G. The panel shows DNA sequence of the 73 bp *aafD*94 promoter fragment.
The *aafD* open reading frame sequence is highlighted blue, the potential AggR-binding sites are green and EcoRI and HindIII sites are grey.

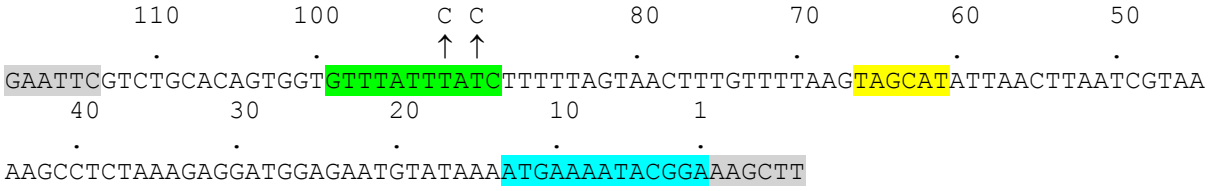
A

aafD*96-90*C



B

aafD*96-90*C*92*C



C

aafD*96-65*C

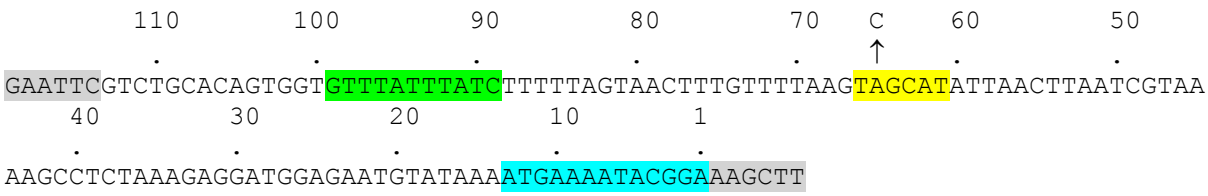


Figure 2.4 The DNA base sequence of *aafD*96 promoter fragment carrying point mutations

A. The panel shows the DNA sequence of the *aafD*96 promoter fragment with arrows that show the position of the substitution at position 90 (T to C) carried by the *aafD*96-90*C* promoter fragment.
B. The panel shows the DNA sequence of the *aafD*96 promoter fragment with arrows that show the position of the substitution at positions 90 and 92 (Ts to Cs) carried by the *aafD*96-90*C*92*C* promoter fragment.
C. The panel shows the DNA sequence of the *aafD*96 promoter fragment with arrows that show the position of the substitution at position 65 (A to C) carried by the *aafD*96-65*C* promoter fragment.
The *aafD* open reading frame sequence is highlighted blue, the potential AggR-binding sites green, the -10 hexamer element yellow and EcoRI and HindIII sites are grey.

A

***afaB*100 (413 bp)**

```
        410       400       390       380       370       360       350
         .         .         .         .         .         .         .
GAATTCGACATGATAACGAATTAAGCAAGGAGGCTGAACAATTTATAATAAGATAACAAAATACTCCTGTTTTAT
   340       330       320       310       300       290       280      270
    .         .         .         .         .         .         .        .
ATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAG
        260       250       240       230       220       210       200
         .         .         .         .         .         .         .
GTTTGCTATCGTAGATGGACAATAATAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAG
   190       180       170       160       150       140       130      120
    .         .         .         .         .         .         .        .
ATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATA
        110       100        90        80        70        60        50
         .         .         .         .         .         .         .
TTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG
    40        30        20        10        1
    .         .         .         .         .
TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCAAGCTT
```

B

***afaB*99 (313 bp)**

```
        310       300       290       280       270       260       250
         .         .         .         .         .         .         .
GAATTCATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAGGTTTGCTATCGTAGATGGACAATAA
   240       230       220       210       200       190       180      170
    .         .         .         .         .         .         .        .
TAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAGATGGATTAAAGAAAGTGTGTCGAAT
        160       150       140       130       120       110       100
         .         .         .         .         .         .         .
AATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATATTCGGTGATGTATTGGTAGGGGCCT
    90        80        70        60        50        40        30       20
    .         .         .         .         .         .         .        .
TATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAGTCGAAATAAAAGGGAGCTAGAAGCC
         10        1
         .         .
AACTCCAGAGTATTTTCCCAAGCTT
```

Figure 2.5 The DNA base sequences of the *afaB* nested promoter deletions (continued)

C

***afaB*98 (213 bp)**

```
        210       200       190       180       170       160       150
         .         .         .         .         .         .         .
GAATTCCCTTTCACAGGTAATGCAGATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTG

   140       130       120       110       100        90        80        70
    .         .         .         .         .         .         .         .
TATGAATATGCGGGACGTGTTTATATTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATT

         60        50        40        30        20        10        1
          .         .         .         .         .         .        .
TTTCTATCAGCAGACGCGCAGGCAGTCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCAAGCTT
```

D

***afaB*97(113 bp)**

```
        110       100        90        80        70        60        50
         .         .         .         .         .         .         .
GAATTCGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG
    40        30        20        10       1
     .         .         .         .        .
TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCAAGCTT
```

Figure 2.5 The DNA base sequences of the *afaB* nested promoter deletions

A. The panel shows the DNA sequence of the 413 bp *afaB*100 promoter fragment.
B. The panel shows the DNA sequence of the 313 bp *afaB*99 promoter fragment.
C. The panel shows the DNA sequence of the 213 bp *afaB*98 promoter fragment.
D. The panel shows the DNA sequence of the 113 bp *afaB*97 promoter fragment.
The *afaB* pseudogene is highlighted blue, the potential AggR-binding sites are green and EcoRI and HindIII sites are grey.

afaB*100-320*C*318*C

```
        410       400       390       380       370       360       350
         .         .         .         .         .         .         .
GAATTCGACATGATAACGAATTAAGCAAGGAGGCTGAACAATTTATAATAAGATAACAAAATACTCCTGTTTTAT
   340       330         C C       310       300       290       280      270
    .         .          ↑ ↑        .         .         .         .         .
ATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAG
        260       250       240       230       220       210       200
         .         .         .         .         .         .         .
GTTTGCTATCGTAGATGGACAATAATAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAG
   190       180       170       160       150       140       130      120
    .         .         .         .         .         .         .         .
ATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATA
        110       100       90        80        70        60        50
         .         .         .         .         .         .         .
TTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG
    40        30        20        10        1
    .         .         .         .         .
TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCAAGCTT
```

B

afaB*100-293*C

```
        410       400       390       380       370       360       350
         .         .         .         .         .         .         .
GAATTCGACATGATAACGAATTAAGCAAGGAGGCTGAACAATTTATAATAAGATAACAAAATACTCCTGTTTTAT
   340       330       320       310       300       290       280      270
    .         .         .         .         .         ↑ .         .         .
ATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAG
        260       250       240       230       220       210       200
         .         .         .         .         .         .         .
GTTTGCTATCGTAGATGGACAATAATAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAG
   190       180       170       160       150       140       130      120
    .         .         .         .         .         .         .         .
ATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATA
        110       100       90        80        70        60        50
         .         .         .         .         .         .         .
TTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG
    40        30        20        10        1
    .         .         .         .         .
TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCAAGCTT
```

Figure 2.6 The DNA base sequence of *afaB*100 promoter fragment carrying various point mutations

A. The panel shows the DNA sequence of the *afaB*100 promoter fragment and arrows indicate the base substitutions introduced into the AggR-binding site at positions 320 and 318 (Ts to Cs), used to generate the *afaB*100-320*C*318*C*.
B. The panel shows the DNA sequence of the *afaB*100 promoter fragment and arrows indicate the base substitutions introduced into the AggR-binding site at positions 293 (A to C), used to generate the *afaB*100-320*C*318*C*.
The *afaB* pseudogene sequences are highlighted blue, the potential AggR-binding sites are green, -10 hexamer elements are yellow and EcoRI and HindIII sites are grey.

determine if another promter was present downstream of the *afaB* promoter, all the sequence to *aafC* gene was cloned, using relevant primers from the Table 2.4. This generated *afaBC*100 and *afaBC*99 fragments and these were cloned into pRW50 using EcoRI-HindIII (Figure 2.7).

To examine the regulation of the fimbrial genes in EAEC strain 17-2, the DNA upstream of *aggD* was investigated. A promoter fragment from another EAEC strain 17-2 was investigated for AggR-binding site. A promoter fragment of 413 bp, named *aggD*100, was cloned into pRW50 and was acquired from Douglas Browning. In order to investigate the promoter elements and AggR-binding site, nested deletions were made using relevant primers from the Table 2.4 and pRW50/*aggD*100 as template. This generated *aggD*50 to *aggD*48 promoter fragments detailed in Figure 2.8. The point mutations were made in order to confirm the AggR-binding site and promoter elements of *aggD.* To make the site directed mutations, the relevant primers were used from Table 2.4 and pRW50/*aggD*49 was used as template to generate *aggD*49-86*C* and *aggD*49-60*C* promoter fragments (Figure 2.9).

To investigate the promoter controlling the expression of the *aai* T6SS system, the promoter fragment *aaiA*100 (479 bp) was cloned into pRW50 using EcorRI and HindIII sites and this was provided by Douglas Browning (Figure 2.10). For investigation of promoter element and AggR-binding site, a deletion analysis of *aaiA*100 fragment was performed in order to identify minimum DNA fragment with promoter activity. The primer used for deletion analysis are detailed in the Table 2.4 and pRW50/*aaiA*100 was used as template. This deletion analysis resulted in *aaiA*99 to *aaiA*95 promoter fragments (Figure 2.10). In order to confirm the AggR-binding site and the promoter elements, site directed point mutations were made using primers listed in the Table 2.4 and pRW50/*aaiA*98 as template generating *aaiA*98-254*C*252*C* and *aaiA*98-226*C* promoter fragments (Figure 2.11).

A

***afaB*100 (413 bp)**

GAATTCGACATGATAACGAATTAAGCAAGGAGGCTGAACAATTTATAATAAGATAACAAAATACTCCTGTTTTAT

ATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAG

GTTTGCTATCGTAGATGGACAATAATAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAG

ATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATA

TTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG

TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCAAGCTT

B

***afaBC*100 (1088 bp)**

GAATTCGACATGATAACGAATTAAGCAAGGAGGCTGAACAATTTATAATAAGATAACAAAATACTCCTGTTTTAT

ATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAG

GTTTGCTATCGTAGATGGACAATAATAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAG

ATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATA

TTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG

TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCTTCATCTGGGCGCAACGCGTGTGGTGTACTC

TAGCTCATTCGGAGAGGTATTGGCAGTCATTAATGATCAAAATTATTCAATGCAGGTTCAGGCTGAAGTTTATCA

GAGGACCGGAAGAGTATAGCCCCTTTTGTGGTTACTCCACTGCAATTTCGCCTTAATGGTCTGCAATCCAGTCGC

CTCCCGAATTGTCCGCACAGGAGGGGATTTTTCGGTGGACCGAGAGAGTCTTCAATGGATTTGTATAAAAGGAAT

CCCCCCCAAAGGCTGACGATAAATGGGAGGAAGGCGATAAGATGATGCTGAATGTACAGCTTTCATTAAGTTATT

GTATCAAACTATTTGTTCGCCCGCCAAGAGTGAAGGGACGTCCTGATGAGATGGCAGGAAAAGTAGAGTGGCAGA

AGTCAGTAACAAACTGAAGGGAATAAACCCGCACCATTTTATATCAATATGTCTGAACTAATGGTGGGAACGGTA

AGGTCCAGTGGACGGTGGTGACGGATTATGGTGGAAGTAGTAAGCAATTTGAGGCCGATCTTAAGGGGTAATAAT

CTAGTTAATCATTTTTTCATGAGTTATGAAAGGGTAATAATCCGTGTGTGAATTTTCTCTGTGGCAGATAAAGAC

AAGTCTTACTGCTGTGGCATATTCAGTGATGATGACATGTACATAAGCTT

Figure 2.7 The DNA sequence of the promoter fragments between *afaB*100 and *aafC* (continued)

C

***afaBC*99 (688 bp)**

GAATTCAGAGTATTTTCCCTTCATCTGGGCGCAACGCGTGTGGTGTACTCTAGCTCATTCGGAGAGGTATTGGCA

GTCATTAATGATCAAAATTATTCAATGCAGGTTCAGGCTGAAGTTTATCAGAGGACCGGAAGAGTATAGCCCCTT

TTGTGGTTACTCCACTGCAATTTCGCCTTAATGGTCTGCAATCCAGTCGCCTCCCGAATTGTCCGCACAGGAGGG

GATTTTTCGGTGGACCGAGAGAGTCTTCAATGGATTTGTATAAAAGGAATCCCCCCCAAAGGCTGACGATAAATG

GGAGGAAGGCGATAAGATGATGCTGAATGTACAGCTTTCATTAAGTTATTGTATCAAACTATTTGTTCGCCCGCC

AAGAGTGAAGGGACGTCCTGATGAGATGGCAGGAAAAGTAGAGTGGCAGAAGTCAGTAACAAACTGAAGGGAATA

AACCCGCACCATTTTATATCAATATGTCTGAACTAATGGTGGGAACGGTAAGGTCCAGTGGACGGTGGTGACGGA

TTATGGTGGAAGTAGTAAGCAATTTGAGGCCGATCTTAAGGGGTAATAATCTAGTTAATCATTTTTTCATGAGTT

ATGAAAGGGTAATAATCCGTGTGTGAATTTTCTCTGTGGCAGATAAAGACAAGTCTTACTGCTGTGGCATATTCA

GTGATGATGACATGTACATAAGCTT

D

***aafC*100 (410 bp)**

GAATTCAGGCTGACGATAAATGGGAGGAAGGCGATAAGATGATGCTGAATGTACAGCTTTCATTAAGTTATTGTA

TCAAACTATTTGTTCGCCCGCCAAGAGTGAAGGGACGTCCTGATGAGATGGCAGGAAAAGTAGAGTGGCAGAAGT

CAGTAACAAACTGAAGGGAATAAACCCGCACCATTTTATATCAATATGTCTGAACTAATGGTGGGAACGGTAAGG

TCCAGTGGACGGTGGTGACGGATTATGGTGGAAGTAGTAAGCAATTTGAGGCCGATCTTAAGGGGTAATAATCTA

GTTAATCATTTTTTCATGAGTTATGAAAGGGTAATAATCCGTGTGTGAATTTTCTCTGTGGCAGATAAAGACAAG

TCTTACTGCTGTGGCATATTCAGTGATGATGACATGTACATAAGCTT

Figure 2.7 DNA base sequences of the promoter fragments between *afaB*100 and *aafC*

A. The panel shows the DNA sequence of the 413 bp *afaB*100 promoter fragment contains *afaB* (red) and upstream of *afaB* sequence (black).
B. The panel shows the 1088 bp *afaBC*100 promoter fragment which contains the DNA sequence carried by *afaB* (red), *aafC* (blue) fragment, and the upstream of *afaB* pseudogene and *aafC* sequence (black).
C. The panel shows the 688 bp *afaBC*99 promoter fragment which contains *afaB* pseudogene sequence (red), *aafC* (blue), amd the upstream of *aafC* sequence (black).
D. The panel shows the 410 bp *aafC*100 promoter fragment, contains part of *afaB* (red), *aafC* (blue) and upstream of *aafC* sequence (black).
The potential AggR- binding sites are green and EcoRI and HindIII sites are grey.

A

***aggD*100 (413 bp)**

```
        410       400       390       380       370       360       350
         .         .         .         .         .         .         .
GAATTCTTCTGGTGCTTCAGGTGTGTGACATGGGAACTCATTCTGGATGGTTACTCTGAAAGCTCATATTCTGCC
   340       330       320       310       300       290       280      270
    .         .         .         .         .         .         .        .
ACACCCCGATTTGCAGCCTCCAGGCTGCCGTGGTTCAGGAAATCGTCCACATCCCCTTAACGGACTTCGGGGGAA
        260       250       240       230       220       210       200
         .         .         .         .         .         .         .
AACGTGTATTTTTCGTTATCCTATTTACCTCTTTCAGGGAGTTTAGTTTCCAGGATTTCCGGGACGGCCTAGCTA
   190       180       170       160       150       140       130      120
    .         .         .         .         .         .         .        .
ATAACGTGATAAATAATTATCTTTATGCGAAAAGTGTAATTTTTGGAGAAAATGGGGGCGGAATCCTAGTGTTAG
        110       100        90        80        70        60        50
         .         .         .         .         .         .         .
AAGATTGAAATATGTTAAATAATGCTATTTTTTAGCGTTATATGATTTGAGTCTTTATATAATCAAGTTCAAGT
    40        30        20        10        1
    .         .         .         .        .
TCAAGTTCAAGTTCAAGTTCAAGTGATAGCGATGAAGATTCGAAAAGCTT
```

B

***aggD*99 (145 bp)**

```
          140       130       120       110       100        90        80
           .         .         .         .         .         .         .
GAATTCAAATGGGGGCGGAATCCTAGTGTTAGAAGATTGAAATATGTTAAATAATGCTATTTTTTTAGCGTTATA
      70        60        50        40        30        20        10
       .         .         .         .         .         .         .
TGATTTGAGTCTTTATATAATCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTGATAGCGATGAAGATTCGA
1
.
AAAGCTT
```

Figure 2.8 The DNA base sequences of the *aggD* nested promoter deletions (continued)

C

***aggD*98 (108 bp)**

```
             100         90         80         70         60         50        40
              .          .          .          .          .          .         .
GAATTCATGTTAAATAATGCTATTTTTTTAGCGTTATATGATTTGAGTCTTTATATAATCAAGTTCAAGTTCAAG
          30         20         10        1
          .          .          .         .
TTCAAGTTCAAGTTCAAGTGATAGCGATGAAGATTCGAAAAGCTT
```

D

***aggD*97 (87 bp)**

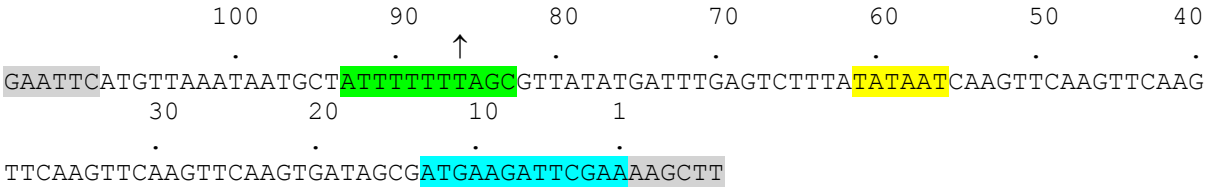
```
             80         70         60         50         40         30        20
              .          .          .          .          .          .         .
GAATTCTTAGCGTTATATGATTTGAGTCTTTATATAATCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTGA
         10        1
          .         .
TAGCGATGAAGATTCGAAAAGCTT
```

Figure 2.8 The DNA base sequences of the *aggD* nested promoter deletions

A. The panel shows the 413 bp DNA sequence of the *aggD*100 promoter fragment cloned from EAEC strain 17-2.
B. The panel shows the 145 bp DNA sequence of the *aggD*99 promoter fragment cloned from EAEC strain 17-2.
C. The panel shows the 108 bp DNA sequence of the *aggD*98 promoter fragment cloned from EAEC strain 17-2.
D. The panel shows the 87 bp DNA sequence of the *aggD*97 promoter fragment cloned from EAEC strain 17-2.
The *aggD* open reading frame sequence is highlighted blue, the potential AggR-binding sites are green and EcoRI and HindIII sites are grey.

A

aggD*98-86*C



B

aggD*98-60*C

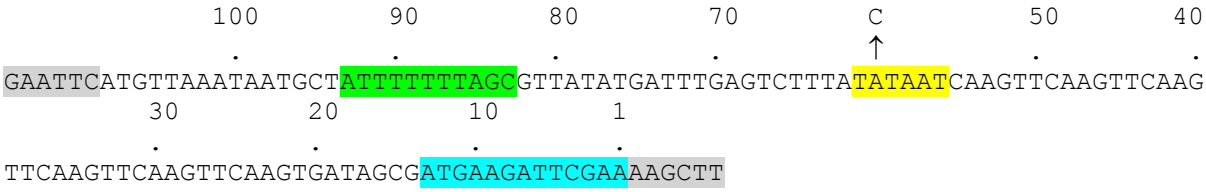


Figure 2.9 The DNA base sequences of the *aggD*98 promoter fragment carrying different point mutations

A. The panel shows the DNA sequence of the *aggD*98 promoter fragment, with the DNA substitution introduced at position 86 (T to C) used to disrupt the AggR-binding site in the *aggD*98-86*C* promoter fragment.
B. The panel shows the DNA sequence of the *aggD*98 promoter fragment, with the DNA substitution introduced at position 60 (A to C) used to disrupt the AggR-binding site in the *aggD*98-60*C* promoter fragment.
The *aggD* open reading frame sequence is highlighted blue, the AggR-binding site is green, the potential -10 hexamer element is yellow and EcoRI and HindIII sites are grey.

A

***aaiA*100 (479 bp)**

```
              470       460       450       440       430       420
               .         .         .         .         .         .
GAATTCTCCTTTGTTTTATGGATAGTTTTCTGCATCAGGCGTCGTTCTGTTCTGGGGATTGGTGTTATGATCGGC
410       400       390       380       370       360       350       340
.         .         .         .         .         .         .         .
ATCGCTCAGTCCGGTTGGTGATTTTTTCTTTTTGGCGATTGATCAGATCGCACAATCCGGGCTGAGTTCCCTCAA
    330       320       310       300       290       280       270
     .         .         .         .         .         .         .
AGTGACCTACTATTCCGCGCAGCTATTTAGTTAGTAAATACGAAAAATTATATAGAGTTATATCATTAATCAGCA
260       250       240       230       220       210       200       190
.         .         .         .         .         .         .         .
AAAATGTATCACATGCTCACTTTCTTTTTATGGTATCACTATATAGAATCCATGAATATAACAAGAGATTGATAC
    180       170       160       150       140       130       120
     .         .         .         .         .         .         .
AATCTTTTGCCAAAGATATAACTTTATATCAGATATTTTCACTTTCGAAAACATCTGTGCGGTAGCTATGGTATT
110       100        90        80        70        60        50        40
.         .         .         .         .         .         .         .
TAGGTCATATAATTATGCTATGTTCATCAATATTATGATAGTAGATTATAGTGTTCTTAATAAAAAAGAATTTAA
     30        20        10       1
     .         .         .        .
AAGCTGTTTGTAGGATAGAAACATGAGCAATACACAAGCTT
```

B

***aaiA*99 (307 bp)**

```
             300       290       280       270       260       250       240
              .         .         .         .         .         .         .
GAATTCAGTTAGTAAATACGAAAAATTATATAGAGTTATATCATTAATCAGCAAAAATGTATCACATGCTCACTT
       230       220       210       200       190       180       170
        .         .         .         .         .         .         .
TCTTTTTATGGTATCACTATATAGAATCCATGAATATAACAAGAGATTGATACAATCTTTTGCCAAAGATATAAC
  160       150       140       130       120       110       100        90
   .         .         .         .         .         .         .         .
TTTATATCAGATATTTTCACTTTCGAAAACATCTGTGCGGTAGCTATGGTATTTAGGTCATATAATTATGCTATG
        80        70        60        50        40        30        20
         .         .         .         .         .         .         .
TTCATCAATATTATGATAGTAGATTATAGTGTTCTTAATAAAAAAGAATTTAAAAGCTGTTTGTAGGATAGAAAC
   10        1
    .        .
ATGAGCAATACACAAGCTT
```

Figure 2.10 The DNA base sequences of the *aaiA* nested promoter deletions (continued)

C

***aaiA*98(272 bp)**

```
       270       260       250       240       230       220       210
        .         .         .         .         .         .         .
GAATTCCATTAATCAGCAAAAATGTATCACATGCTCACTTTCTTTTTATGGTATCACTATATAGAATCCATGAAT
  200       190       180       170       160       150       140       130
   .         .         .         .         .         .         .         .
ATAACAAGAGATTGATACAATCTTTTGCCAAAGATATAACTTTATATCAGATATTTTCACTTTCGAAAACATCTG
       120       110       100        90        80        70        60
        .         .         .         .         .         .         .
TGCGGTAGCTATGGTATTTAGGTCATATAATTATGCTATGTTCATCAATATTATGATAGTAGATTATAGTGTTCT
   50        40        30        20        10       1
    .         .         .         .         .
TAATAAAAAAGAATTTAAAAGCTGTTTGTAGGATAGAAACATGAGCAATACACAAGCTT
```

D

***aaiA*97 (251 bp)**

```
      250       240       230       220       210       200       190
       .         .         .         .         .         .         .
GAATTCCACATGCTCACTTTCTTTTTATGGTATCACTATATAGAATCCATGAATATAACAAGAGATTGATACAAT
 180       170       160       150       140       130       120       110
  .         .         .         .         .         .         .         .
CTTTTGCCAAAGATATAACTTTATATCAGATATTTTCACTTTCGAAAACATCTGTGCGGTAGCTATGGTATTTAG
      100        90        80        70        60        50        40
       .         .         .         .         .         .         .
GTCATATAATTATGCTATGTTCATCAATATTATGATAGTAGATTATAGTGTTCTTAATAAAAAAGAATTTAAAAG
  30        20        10       1
  .         .         .        .
CTGTTTGTAGGATAGAAACATGAGCAATACACAAGCTT
```

Figure 2.10 The DNA base sequences of the *aaiA* nested promoter deletion (continued)

E

***aaiA*96 (224 bp)**

```
          220        210        200        190        180        170        160
           .          .          .          .          .          .          .
GAATTCCACTATATAGAATCCATGAATATAACAAGAGATTGATACAATCTTTTGCCAAAGATATAACTTTATATC
     150        140        130        120        110        100         90
      .          .          .          .          .          .          .
AGATATTTTCACTTTCGAAAACATCTGTGCGGTAGCTATGGTATTTAGGTCATATAATTATGCTATGTTCATCAA
80         70         60         50         40         30         20         10
.          .          .          .          .          .          .          .
TATTATGATAGTAGATTATAGTGTTCTTAATAAAAAAGAATTTAAAAGCTGTTTGTAGGATAGAAACATGAGCAA
     1
     .
TACACAAGCTT
```

F

***aaiA*95 (144 bp)**

```
          140        130        120        110        100         90         80
           .          .          .          .          .          .          .
GAATTCCTTTCGAAAACATCTGTGCGGTAGCTATGGTATTTAGGTCATATAATTATGCTATGTTCATCAATATTA
     70         60         50         40         30         20         10        1
      .          .          .          .          .          .          .        .
TGATAGTAGATTATAGTGTTCTTAATAAAAAAGAATTTAAAAGCTGTTTGTAGGATAGAAACATGAGCAATACAC


AAGCTT
```

Figure 2.10 The DNA base sequences of the *aaiA* nested promoter deletion

A. The panel shows the DNA sequence of the 479 bp *aaiA*100 promoter fragment.
B. The panel shows the DNA sequence of the 307 bp *aaiA*99 promoter fragment.
C. The panel shows the DNA sequence of the 272 bp *aaiA*98 promoter fragment.
D. The panel shows the DNA sequence of the 251 bp *aaiA*97 promoter fragment.
E. The panel shows the DNA sequence of the 224 bp *aaiA*96 promoter fragment.
F. The panel shows the DNA sequence of the 144 bp *aaiA*95 promoter fragment.
The *aaiA* open reading frame sequence is highlighted blue, the potential AggR-binding sites are green and EcoRI and HindIII sites are grey.

aaiA*98-254*C*252*C

```
       270       260     C C250       240       230       220       210
        .         .      ↑ ↑ .         .         .         .         .
GAATTCCATTAATCAGCAAAAATGTATCACATGCTCACTTTCTTTTTATGGTATCACTATATAGAATCCATGAAT
  200       190       180       170       160       150       140       130
   .         .         .         .         .         .         .         .
ATAACAAGAGATTGATACAATCTTTTGCCAAAGATATAACTTTATATCAGATATTTTCACTTTCGAAAACATCTG
       120       110       100        90        80        70        60
        .         .         .         .         .         .         .
TGCGGTAGCTATGGTATTTAGGTCATATAATTATGCTATGTTCATCAATATTATGATAGTAGATTATAGTGTTCT
   50        40        30        20        10        1
   .         .         .         .         .         .
TAATAAAAAAGAATTTAAAAGCTGTTTGTAGGATAGAAACATGAGCAATACACAAGCTT
```

B

aaiA*98-226*C

```
       270       260       250       240       230  C    220       210
        .         .         .         .         .   ↑     .         .
GAATTCCATTAATCAGCAAAAATGTATCACATGCTCACTTTCTTTTTATGGTATCACTATATAGAATCCATGAAT
  200       190       180       170       160       150       140       130
   .         .         .         .         .         .         .         .
ATAACAAGAGATTGATACAATCTTTTGCCAAAGATATAACTTTATATCAGATATTTTCACTTTCGAAAACATCTG
       120       110       100        90        80        70        60
        .         .         .         .         .         .         .
TGCGGTAGCTATGGTATTTAGGTCATATAATTATGCTATGTTCATCAATATTATGATAGTAGATTATAGTGTTCT
   50        40        30        20        10        1
   .         .         .         .         .         .
TAATAAAAAAGAATTTAAAAGCTGTTTGTAGGATAGAAACATGAGCAATACACAAGCTT
```

Figure 2.11 The DNA base sequence of *aaiA*98 promoter fragments carrying various point mutations

A. The panel shows the *aaiA*98 promoter fragment with substitutions at positions 254 and 252 (Ts to Cs) marked with small arrows which disrupt the AggR-binding in the *aaiA*98-254*C*252*C* promoter fragment.
B. The panel shows the *aaiA*98 promoter fragment with substitutions at positions 226 (A to C) marked with small arrow which disrupt the -10 hexamer element in the *aaiA*98-226*C* promoter fragment.
The *aaiA* open reading frame sequence is highlighted blue, the potential AggR-binding site is green, the potential -10 hexamer element is yellow and EcoRI and HindIII sites are grey.

For the study of the dispersin (*aap*) promoter, a promoter fragment: *aap*100 (262) which had been cloned into pRW50 using the EcoRI and HindIII sites, was received from Rita Godfrey (Figure 2.12). In order to confirm the AggR-binding site and the promoter element, point mutations were made in *aap*100 promoter fragment, using relevant primers from the Table 2.4 and pRW50/*aap*100 was used as template. This generated *aap*100-151*C*149*C* and *aap*100-124*C* promoter fragments as described in the Figure 2.12.

To confirm that placing an AggR-binding site into a promoter can confer AggR-dependent regulation. The AggR-binding motif from the *aafD* promoter was introduced into the well-studied CRP-dependent semi-synthetic promoter, *CC(-41.5)* (Figure 2.13). A range of different semi-synthetic promoters were made with on AggR-binding motif at different distances from the -10 hexamer element of the promoter, generating *DAM*20 to *DAM*23 promoter fragments shown in Figure 2.13. To confirm, that the regulation observed was due to the transplanted sequence and AggR mediated activation, point mutations were also introduced into the *aafD* AggR-binding site, generating the *DAM*22-(-38*C*-36*C*) (Figure 2.14).

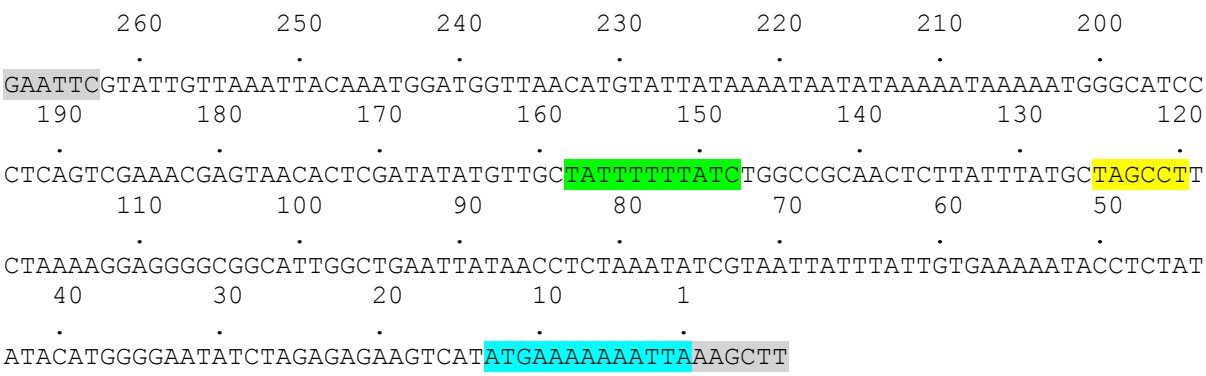
For studying mutational analysis of AggR protein, point mutations were made by substituting amino acids at position 14, 16, 230 and 234 of the AggR. To investigate, these mutations, may have affected the DNA binding ability of AggR, a known promoter fragment (*nlpA*100), repressed by AggR was used (Figure 2.15). The promoter fragment was received from Chiara and was cloned into pRW50 using EcoRI and HindIII sites.

2.9.2 Construction of pRW224-U9 and pRW225 derivatives

QIAprep Spin miniprep kits were used to prepare both pRW224-U9 and pRW225 vectors DNA (Figure 2.16 and Figure 2.17, respectively). Plasmid DNA was digested with EcoRI and

A

***aap*100 (262 bp)**



B

aap*100-151*C*149*C

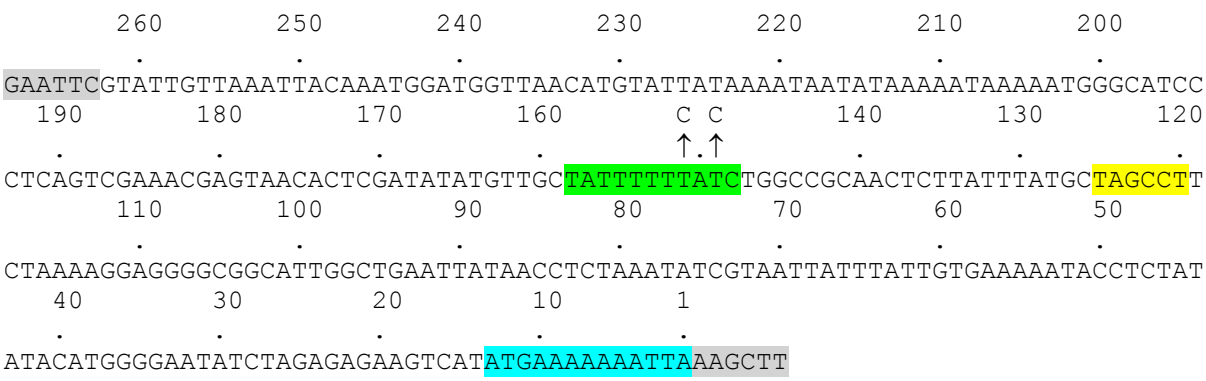


Figure 2.12 The DNA base sequences of the different *aap*100 promoter fragments (continued)

C

aap*100-124*C

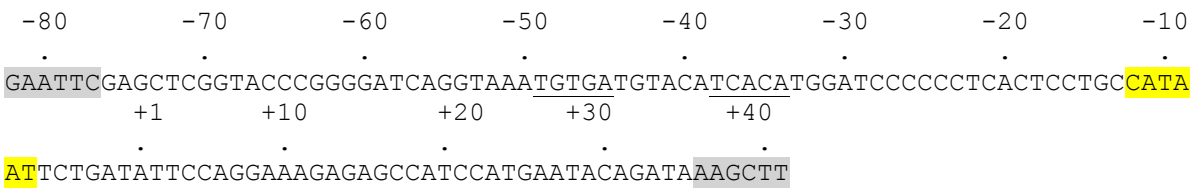
```
       260        250        240        230        220        210        200
        .          .          .          .          .          .          .
GAATTCGTATTGTTAAATTACAAATGGATGGTTAACATGTATTATAAAATAATATAAAAATAAAAATGGGCATCC
  190        180        170        160        150        140        130     C  120
   .          .          .          .          .          .          .     ↑    .
CTCAGTCGAAACGAGTAACACTCGATATATGTTGCTATTTTTTATCTGGCCGCAACTCTTATTTATGCTAGCCTT
       110        100         90         80         70         60         50
        .          .          .          .          .          .          .
CTAAAAGGAGGGGCGGCATTGGCTGAATTATAACCTCTAAATATCGTAATTATTTATTGTGAAAAATACCTCTAT
   40         30         20         10        1
   .          .          .          .         .
ATACATGGGGAATATCTAGAGAGAAGTCATATGAAAAAAATTAAAGCTT
```

Figure 2. 12 The DNA base sequences of the different *aap*100 promoter fragments

A. The panel shows the DNA sequence of the 262 bp *aap*100 promoter fragment.
B. The panel shows the *aap*100 promoter fragment with substitutions at positions 151 and 149 (Ts to Cs) marked with small arrows which disrupt the AggR-binding in the *aap*100-151*C*149*C* promoter fragment.
C. The panel shows the *aap*100 promoter fragment with substitutions at positions 124 (A to C) marked with small arrow which disrupt -10 hexamer element in the *aap*100-124*C* promoter fragment.
The *aap* open reading frame sequence is highlighted blue, the potential AggR-binding site is green, the potential -10 hexamer element is yellow and EcoRI and HindIII sites are grey.

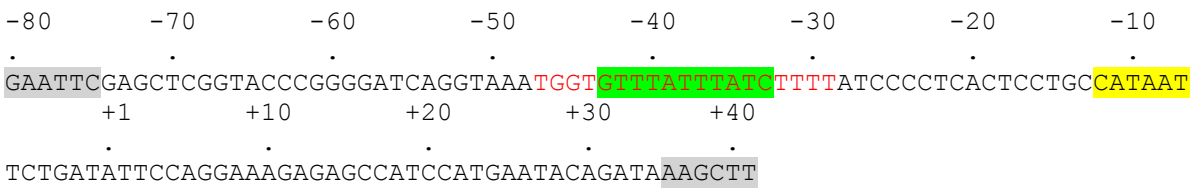
A

CC(-41.5) **(111 bp)**



B

***DAM*20 (109 bp)**



C

***DAM*21(110 bp)**

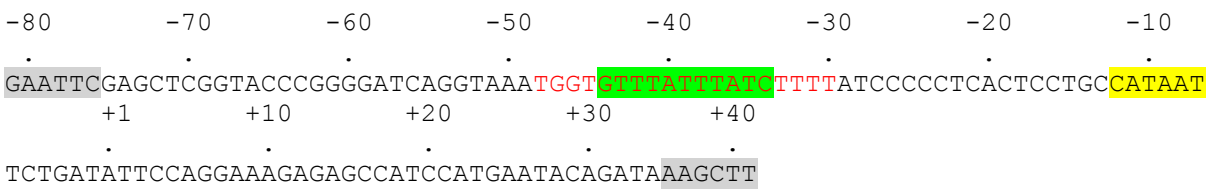


Figure 2.13 The DNA base sequences of the different semi-synthetic promoter fragments containing AggR-binding sites

D

***DAM*22 (111 bp)**

```
 -80        -70        -60        -50        -40        -30        -20        -10
  .          .          .          .          .          .          .          .
GAATTCGAGCTCGGTACCCGGGGATCAGGTAAATGGTGTTTATTTATCTTTTATCCCCCCCTCACTCCTGCCATAAT
       +1         +10        +20        +30        +40
       .          .          .          .          .
TCTGATATTCCAGGAAAGAGAGCCATCCATGAATACAGATAAAGCTT
```

E

***DAM*23 (112 bp)**

```
  -80        -70        -60        -50        -40        -30        -20        -10
   .          .          .          .          .          .          .          .
GAATTCGAGCTCGGTACCCGGGGATCAGGTAAATGGTGTTTATTTATCTTTTATCCCCCCCTCACTCCTGCCATAAT
       +1         +10        +20        +30        +40
       .          .          .          .          .
TCTGATATTCCAGGAAAGAGAGCCATCCATGAATACAGATAAAGCTT
```

Figure 2. 13 The DNA base sequences of the different semi-synthetic promoter fragments containing AggR-binding sites

A. The panel shows the DNA sequence of the *CC(-41.5)* promoter fragment from Savery *et al.* (1995b), numbered in relation to the start site of transcription (+1). The CRP-binding site centred at position -41.5 and is underlined.
B. The panel shows the sequence of the *DAM*20 promoter fragment, a variant of *CC(-41.5)* in which the AggR-binding site from *aafD* (shown in red) has been cloned 20 bp upstream of the -10 hexamer element.
C. The panel shows the sequence of the *DAM*21 promoter fragment, a variant of *CC(-41.5)* in which the AggR-binding site from *aafD* (shown in red) has been cloned 21 bp upstream of the -10 hexamer element.
D. The panel shows the sequence of the *DAM*22 promoter fragment, a variant of *CC(-41.5)* in which the AggR-binding site from *aafD* (shown in red) has been cloned 22 bp upstream of the -10 hexamer element.
E. The panel shows the sequence of the *DAM*23 promoter fragment, a variant of *CC(-41.5)* in which the AggR-binding site from *aafD* (shown in red) has been cloned 23 bp upstream of the -10 hexamer element.
The potential AggR-binding site is green, the potential -10 hexamer element is yellow and EcoRI and HindIII sites are grey.

A

***DAM*22-(-38*C*-36*C*)**

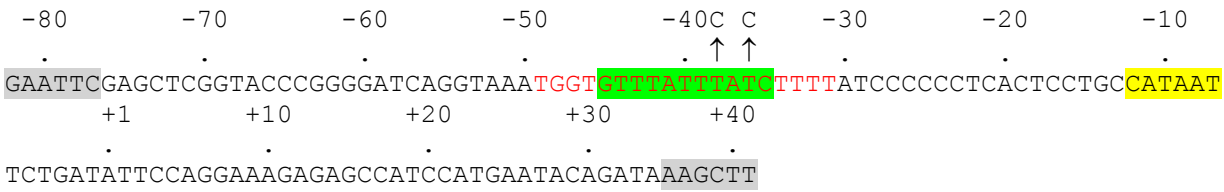


Figure 2.14 The DNA base sequence of *DAM*22 promoter fragment carrying point mutations

The figure details the DNA sequence of the DMA22 promoter fragment and shows the positions of substitutions introduced at -38 and -36 (marked with arrows and text) to disrupt the AggR-binding site and generate the *DAM*22-(-38*C*-36*C*) promoter fragment. The red colour text shows AggR-binding motif sequence cloned from *aafD*, the potential AggR-binding sites are highlighted green, the -10 hexamer is yellow and EcoRI and HindIII sites are grey.

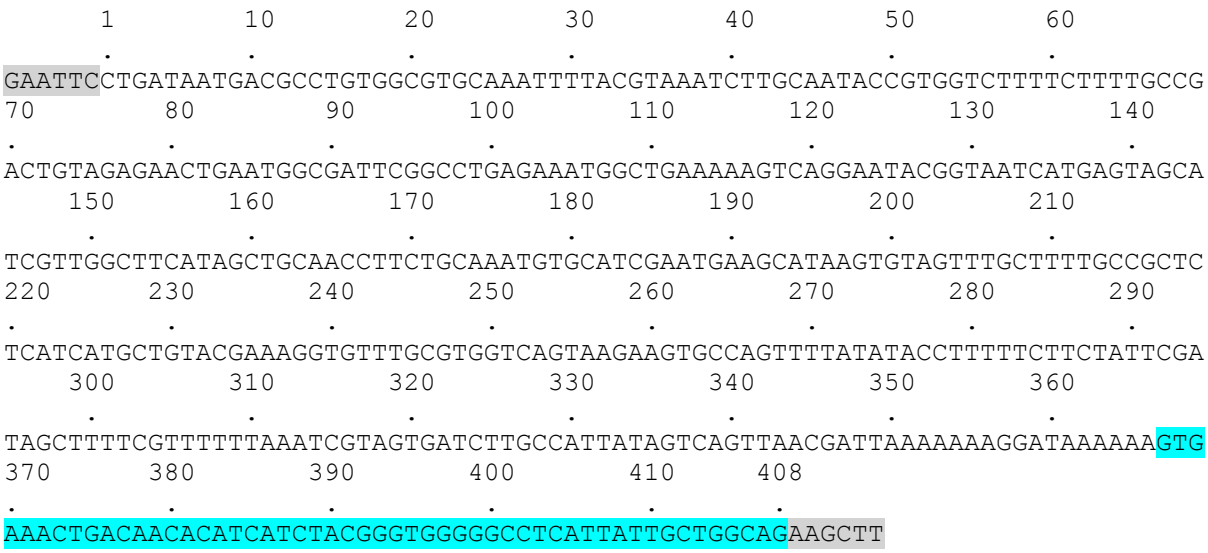
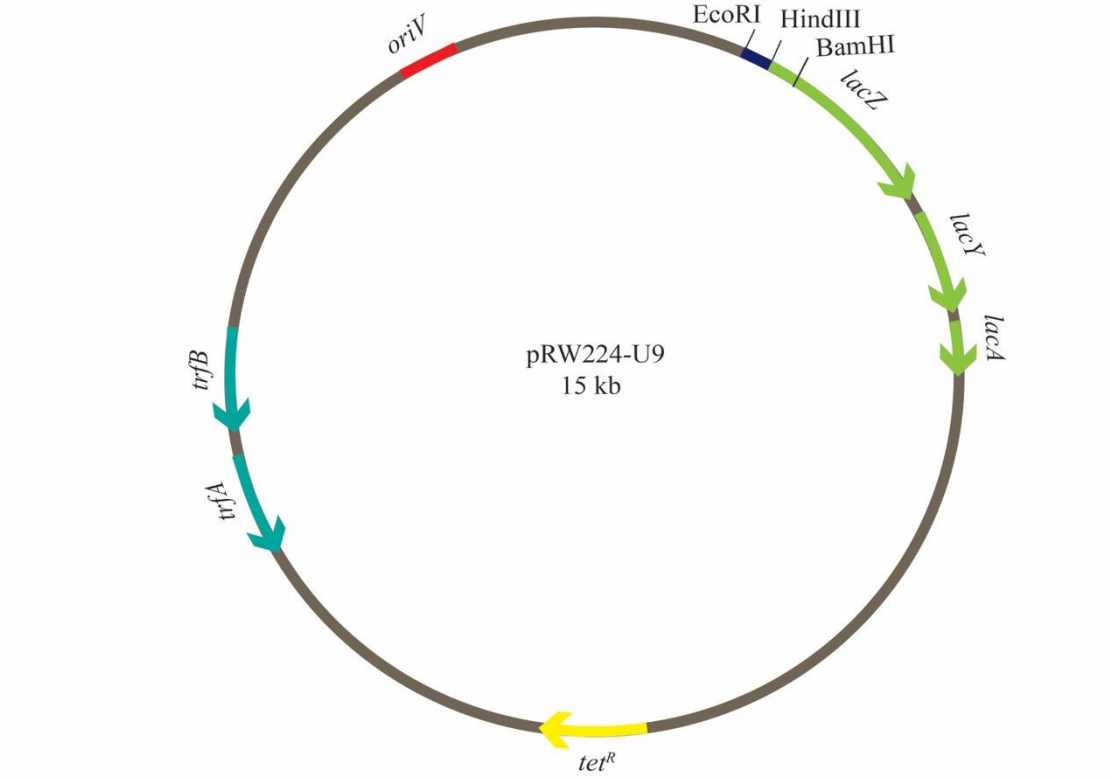


Figure 2.15 The DNA sequence of *nlpA*100 promoter fragment

The figure details the DNA sequence of the 408 bp *nlpA*100 promoter fragment. The coding sequence is highlighted blue and EcoRI and HindIII sites are grey.



EcoRI

ATTCGCGAGAGCCTTGAGTCCACGCTAGATCT**GAATTC**CCGGGGATCCGTCGACCTGCAGCC

HindIII SD sequence Translation initiation codon BamHI *lacZ*

AAGCTTAATGGAGCGAATTATGAGAGTTCTGGTTACCGCCAAGCTCCG**GGATCC**CGTCGTTT

TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC

Figure 2.16 A map of the pRW224-U9 plasmid

Plasmid pRW224-U9 is a derivative of pRW50 in which the *trpBA* region has been removed, so that either transcriptional or translational *lacZ* fusion can be generated. The sequences of the multiple cloning sites are shown in the lower part of the panel. The DNA promoter fragment can be cloned into this plasmid using either EcoRI-HindIII to generate a transcription fusion or EcoRI-BamIII to produce a translational fusion. The Shine-Dalgarno (SD) sequence and translational start site have been underlined.

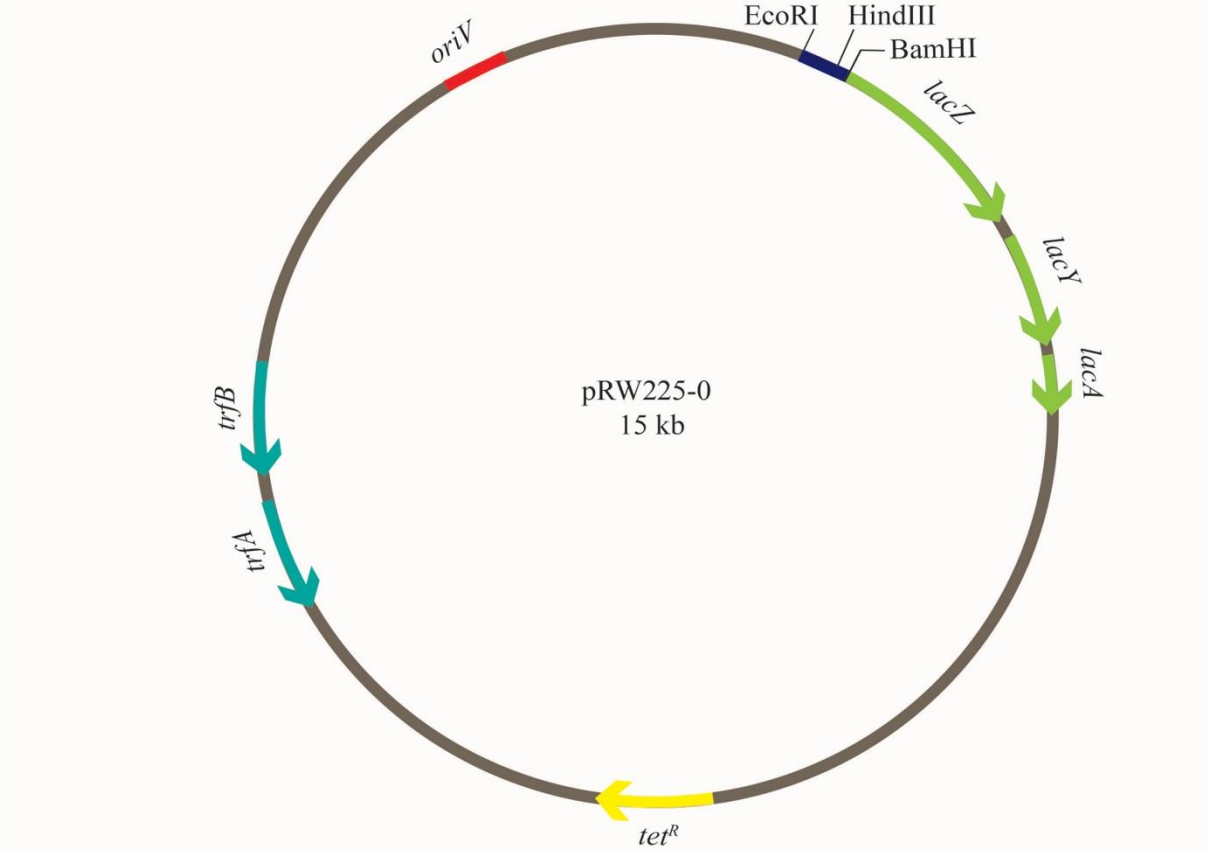


Figure 2.17 A map of plasmid pRW225

The panel shows a diagram of the pRW225 *lacZ* expression vector, a derivative of pRW224 plasmid which lacks *lac* operon translation start site. The sequences of the multiple cloning sites are shown in the lower part of the panel. EcoRI-HindIII sites are used to clone in the desired DNA promoter fragment to generate translational fusions. Note that the DNA fragment must provide its own SD and translational start site fused in-frame to *lacZ*.

HindIII, treated with CIP and purified using phenol/chloroform extraction and ethanol precipitation (sections 2.5.1 and 2.5.2). DNA fragments, of interest, were amplified using PCR and the primers detailed in Table 2.4, digested with EcoRI and HindIII and ligated into pRW224-U9 and pRW225 EcoRI-HindIII (sections 2.6.1 and 2.6.2).

pRW224-U9 is a *lacZ* expression vector that has been designed to investigate transcription regulation of cloned promoter fragments. The vector has its own Shine-Dalgarno (SD) sequence and translation start, such that transcripts produced from the upstream promoter fragment can be translated into β-galactosidase. If there is a difference in β-galactosidase produced by different promoters, it is due to a difference in the level of transcription from each promoter. For example, to investigate the strength of the *aafD* and *afaB* promoters, *aafD*500 and *afaB*100 promoter fragments were made by PCR using primers listed in Table 2.4 and cloned into pRW224-U9 (Figure 2.2 and Figure 2.18).

The pRW225 vector has been designed to investigate translation start sites such that SD sequence and translation start of the cloned promoter fragment serve as the translation start site of *lacZ* present on the vector. For the investigation of the *aafD* translation start site, the promoter fragments with different in-frame translation starts sites to *lacZ* of vector, *aafD*1001 and *aafD*1002, were generated using PCR, with the relevant primers form Table 2.4 and cloned into pRW225 vector using EcoRI-HindIII sites (Figure 2.18).

2.9.3 Construction of pDOC-K derivatives for gene doctoring of AggR

A QIAprep Spin miniprep kit was used to prepare pDOC-K vector DNA (Figure 2.19) and plasmid DNA was initially digested with EcoRI and BamHI, treated with CIP and purified by phenol/chloroform extraction and ethanol precipitation (sections 2.5.1 and 2.5.2). Two DNA sequences, present upstream and downstream of gene of interest *i.e. aggR*, were amplified

A

***aafD*1001 (423 bp)**

```
        430       420       410       400       390       380       370
      .   .       .         .         .         .         .         .
GAATTCGCGTAAATAGTTGGAAAAGTTCATGCAGCATGTGGATGCTTACAGCAGATGGTATAACGAGCGGCGTAT
   360       350       340       330       320       310       300       290
    .         .         .         .         .         .         .         .
AAAATTATCGCTGGGTGCAGTCAGCCCTGAAATGTACCGCCAACAATGCGGGCTGGCATAATAAAGCAGTCCAGG
        280       270       260       250       240       230       220
         .         .         .         .         .         .         .
AAATCGTCCGCATCCCCTAACGGTCAAACATCGTGGCCTTGACAGCTCCCTTGGCGCGTAGTATTGCCAACTGAA
   210       200       190       180       170       160       150       140
    .         .         .         .         .         .         .         .
TCCCCCGTCAATACGGTTCTCACCGTATGCTCACTGCTTACAATTGCCTGACAGTAGCAACCAACTGAGAGGATG
        130       120       110       100        90        80        70
         .         .         .         .         .         .         .
CTATCTCACCTGATTTATTCAATAAAGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAGTA
    60        50        40        30        20      11
     .         .         .         .         .       .
GCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGAGAATGTATAAAATGAAGCTT
```

B

***aafD*1002 (414 bp)**

```
        430       420       410       400       390       380       370
      .   .       .         .         .         .         .         .
GAATTCGCGTAAATAGTTGGAAAAGTTCATGCAGCATGTGGATGCTTACAGCAGATGGTATAACGAGCGGCGTAT
   360       350       340       330       320       310       300       290
    .         .         .         .         .         .         .         .
AAAATTATCGCTGGGTGCAGTCAGCCCTGAAATGTACCGCCAACAATGCGGGCTGGCATAATAAAGCAGTCCAGG
        280       270       260       250       240       230       220
         .         .         .         .         .         .         .
AAATCGTCCGCATCCCCTAACGGTCAAACATCGTGGCCTTGACAGCTCCCTTGGCGCGTAGTATTGCCAACTGAA
   210       200       190       180       170       160       150       140
    .         .         .         .         .         .         .         .
TCCCCCGTCAATACGGTTCTCACCGTATGCTCACTGCTTACAATTGCCTGACAGTAGCAACCAACTGAGAGGATG
        130       120       110       100        90        80        70
         .         .         .         .         .         .         .
CTATCTCACCTGATTTATTCAATAAAGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAGTA
    60        50        40        30        20
     .         .         .         .         .
GCATATTAACTTAATCGTAAAAGCCTCTAAAGAGGATGGAGAATGAAGCTT
```

Figure 2.18 The DNA base sequence of *aafD* promoter fragments carrying different deletions (continued)

C

***aafD*500 (70 bp)**

```
        110       100        90        80        70        60        50
         .         .         .         .         .         .         .
GAATTCGTCTGCACAGTGGTGTTTATTTATCTTTTTAGTAACTTTGTTTTAAGTAGCATATTAACTTAATCGTAA
44
.
AAAGCTT
```

Figure 2.18 DNA base sequence of *aafD* promoter fragments with different downstream deletions

A. The panel shows the DNA sequence of 423 bp *aafD*1001 promoter fragment, which is a derivative of *aafD*100 that has a 10 bp deletion adjacent to the HindIII site.
B. The panel shows the DNA sequence of 423 bp *aafD*1002 promoter fragment, which is a derivative of *aafD*100 that has a 19 bp deletion adjacent to the HindIII site.
C. The panel shows DNA sequence of the 70 bp *aafD*500 promoter fragment, which is a derivative of *aafD*96 that has a 43 bp deletion adjacent to the HindIII site.
The *aafD* open reading frame is highlighted blue, the potential AggR-binding site are green, the potential -10 hexamer element are yellow and EcoRI and HindIII sites are grey.

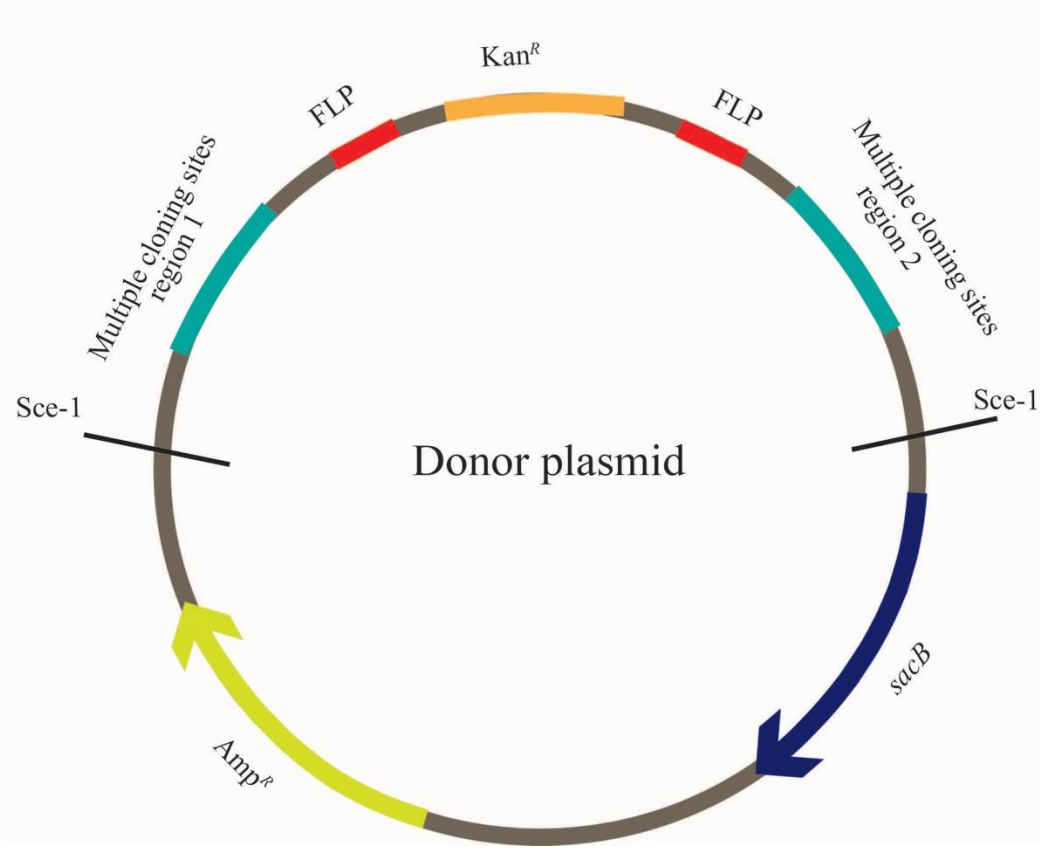


Figure 2.19 A map of the pDOC-K donor plasmid used for gene doctoring

The panel shows a diagram of the pDOC-K plasmid used to introduce deletion into EAEC. There are two multiple cloning sites regions, that were used to clone the flanking sequence of gene of interest (*e.g. aggR*). The plasmid carries kanamycin (Kan^R)and ampicillin (Amp^R) resistance genes. The *sacB* open reading frame confers sucrose sensitivity. Black line show SecI restriction sites and FLP represent flippase recombinase recognition sites used to flip out kanamycin cassette.

using PCR and named *aggR*400 and *aggR*401, respectively. The fragment *aggR*400 in Figure 2.20 was amplified using PCR and the primers detailed in Table 2.4. The purified DNA fragment was then digested with EcoRI and BamHI and ligated into pDOC-K to generate pDOC-K/*aggR*400 (sections 2.6.1 and 2.6.2).

The pDOC-K/*aggR*400 vector was then prepared and digested with XhoI and SalI, treated with CIP and purified by phenol/chloroform extraction and ethanol precipitation (sections 2.5.1 and 2.5.2). The *aggR*401 was then amplified by PCR using the primers detailed in Table 2.4 and sequence is shown in Figure 2.20. Purified DNA fragment was digested with XhoI and SalI and then ligated into the pDOC-K/*aggR*400 vector to generate pDOC-K/*aggR*400/ *aggR*401 (sections 2.6.1 and 2.6.2). The λ system was provided by plasmid pACBSR (Figure 2.21).

2.9.4 Construction of pBAD/*aggR* derivatives

In order to examine the role of AggR in *E. coli* K-12 strains, an inducible system was required to express AggR. Therefore, the vector pBAD/*aggR*, a variant of pBAD30 with *aggR* cloned under the control of the *araBAD* promoter, was acquired from J.P.Nataro (Figure 2.22) (Sheikh *et al.,* 2002). To investigate whether certain amino acids in AggR are important for promoter regulation, specific point mutations were introduced into *aggR* on pBAD/*aggR* (Figure 2.23). Two point mutations were introduced into the N-terminal region of AggR at position 14 and 16 (Figure 2.24 and Figure 2.25 respectivley) and two mutations were introduced into the C-terminal region of AggR at positions (Figure 2.26 and Figure 2.27, respectively). Therefore, *aggR* DNA fragments carrying point mutations at positions 14, 16, 230 and 234 of AggR were amplified using megaprimer PCR and each fragment was cloned into pBAD using an EcoRI-XbaI (sections 2.6.1 and 2.6.2). These pBAD/*aggR* derivatives were constructed by Rita Godfrey and the ability of these mutant AggR protein to regulate

A

***aggR*400**

```
     400
      .         .         .         .         .         .         .
GAATTCTAATGAGCGAAAATAGCCTGTTACCATCACGACAGCAGCGGGAACATAAAGGCCGAATATCGGTGAAG
                                  300
  .         .         .         .         .         .         .         .
GAAATCAACCAGCGTTGGTGTTCCGATGGCTTTGAGTTTGGTTGCGATGATAGCGAAAAACTGTGTGTTACGTT
                                                          200
        .         .         .         .         .         .         .
CTCCCTGGACTGTTGCGATCGTGAAGCCATAGACTGAGCTGCAATTAAGATACAACCCCTGTCTGGAAATAAGG

    .         .         .         .         .         .         .
GGGCAAGAATACGATAAATAATTTCCTATTGTAATTATAAGCGTAAAAATCATATCCCACATGACGATGTGGAA
         100
.         .         .         .         .         .         .         .
ATTAACAAACGTATTTTATATGAGTTAAAAATATATCTTTTTATTGATAAGAGTTAGGTCATTCTAACGCAGAT
                                  1
     .         .         .         .
TGCCTGATAAAGACATTTTTTTCATGTGAGAATGATGGATCC
```

B

***aggR*401**

```
400
      .         .         .         .         .         .         .
CTCGAGAAACATGTTTCATATCATTATTTGAGATTGCTATAAACATATTGAGATGGCTGAAGTTGGTGATGCTG
                                  300
  .         .         .         .         .         .         .         .
CCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTC
                                                          200
       .         .         .         .         .         .         .
CTGTTCAGCTACTGACGGAGTGGTGCGTAACGGCAAAAGCACTGCCGGACATCAGCGCTATCTCTGCTCTCACT

    .         .         .         .         .         .         .
GCCGTAAAACATGGCAACTGCAGTTCACTTACACCGCTTCTCAACCCGGTACGCACCAGAAAATCATTGATATG
         100
.         .         .         .         .         .         .         .
GCCATGAATGGCGTTGGCTGCCGGGCAACTGCCCGCATTATGGGCGTTGGCCTCAACACGATTTTACGTCACTT
                                  1
     .         .         .         .
AAAAAACTCAGGCCGCAGTCGGTAACCTCGCGCATAGTCGAC
```

Figure 2.20 The DNA sequence of the homology regions used to disrupt *aggR* by gene doctoring

A. The panel shows the 400 bp sequence upstream of *aggR* (cloned in multiple cloning sites of region 1) that was cloned in the pDOC-K donor plasmid in multiple cloning region 1 using EcoRI and BamHI sites, (highlighted grey).
B. The panel shows the 400 bp sequence downstream of *aggR* (cloned in multiple cloning sites of region 1) that was cloned in the pDOC-K donor plasmid in multiple cloning region 2 using XhoI and SalI sites, (highlighted grey).

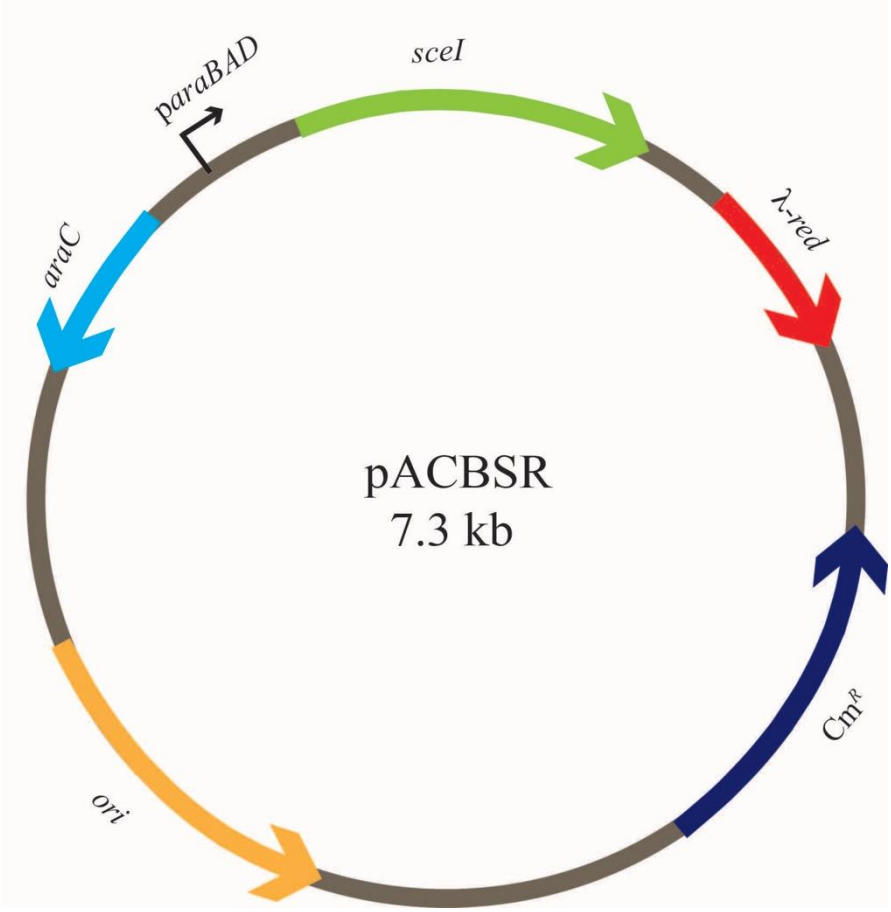


Figure 2.21 A map of the pACBSR plasmid

The panel depicts the pACBSR plasmid and shows that *sceI* and lambda (λ) red genes are under the control of *araBAD* promoter and are thus, arabinose inducible. The plasmid also contains a chloramphenicol (Cm^R) resistance gene. This plasmid expresses SceI and λ-red and is used for gene doctoring in *E. coli* strains.

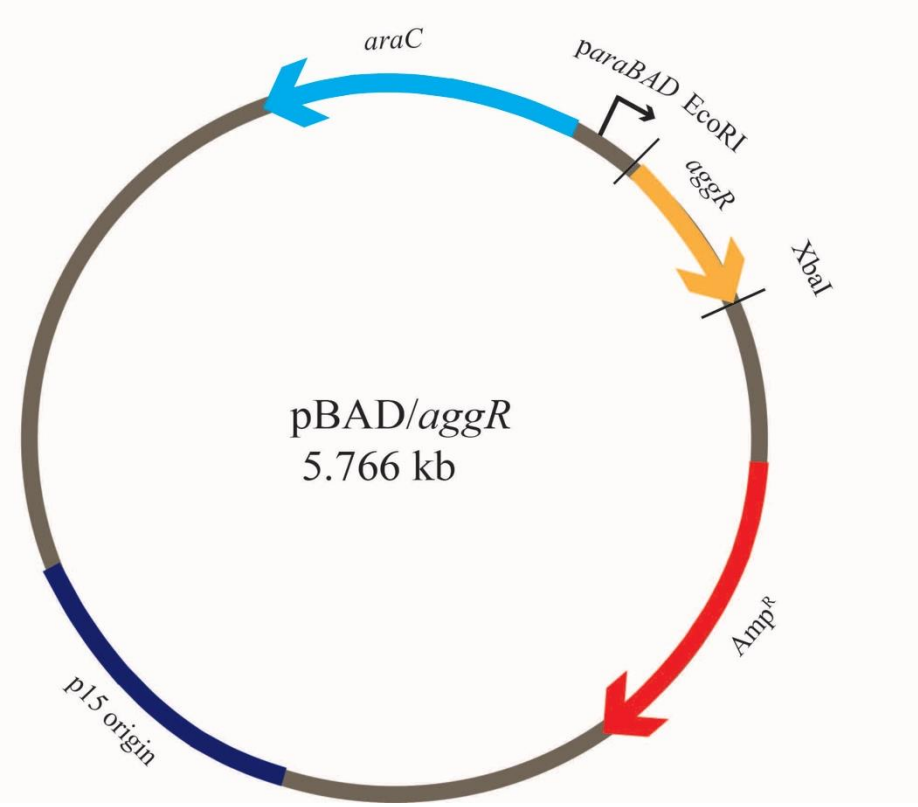


Figure 2.22 A map of the pBAD/*aggR* plasmid

The panel depicts the pBAD/*aggR* plasmid and shows that *aggR* expression is under the control of *araBAD* promoter and is inducible by arabinose. The plasmid also contains ampicillin (Amp^R) resistance gene. This plasmid expresses *aggR* and was used to study AggR-dependent promoters in *E. coli* K-12.

Figure 2.23 The sequence of *aggR* from EAEC 042

The figure shows DNA sequence of *aggR*, its leader sequence and the restriction sites, EcoRI and XbaI, used to clone this fragment into pBAD30. The start and stop codons of AggR are highlighted yellow and all the alternate codons are underlined. The letters (in blue) above each codon represent the amino acids coded by these codons. The numbers represent the numbering of amino acids in AggR. Restriction sites are highlighted grey and the sequence between EcoRI and the start codon is part of the *aggR* leader sequence.

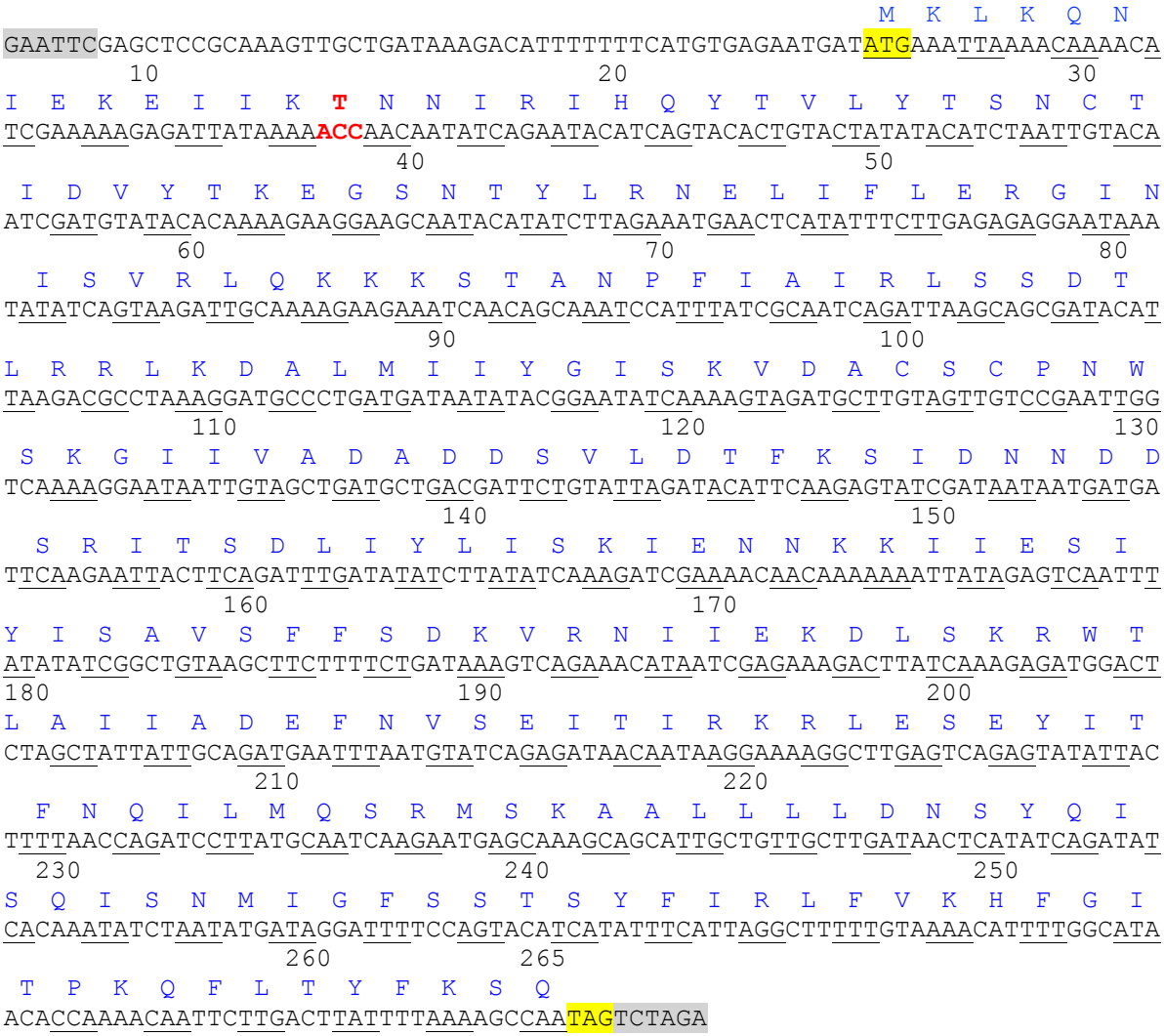


Figure 2.24 The DNA and amino acids sequence of AggR-I14T variant

The figure shows the DNA and amino acid sequence of the AggR-I14T variant, which was cloned into pBAD30 using EcoRI and XbaII restriction sites (grey). The start and stop codons of AggR are highlighted yellow and alternate codons are underlined. The amino acid sequence is shown above each codon and the amino acid numbering is above. At position 14 the codon has been changed from ATC to ACC, changing the amino acid of this position from isoleucine (I) to threonine (T) (shown red).

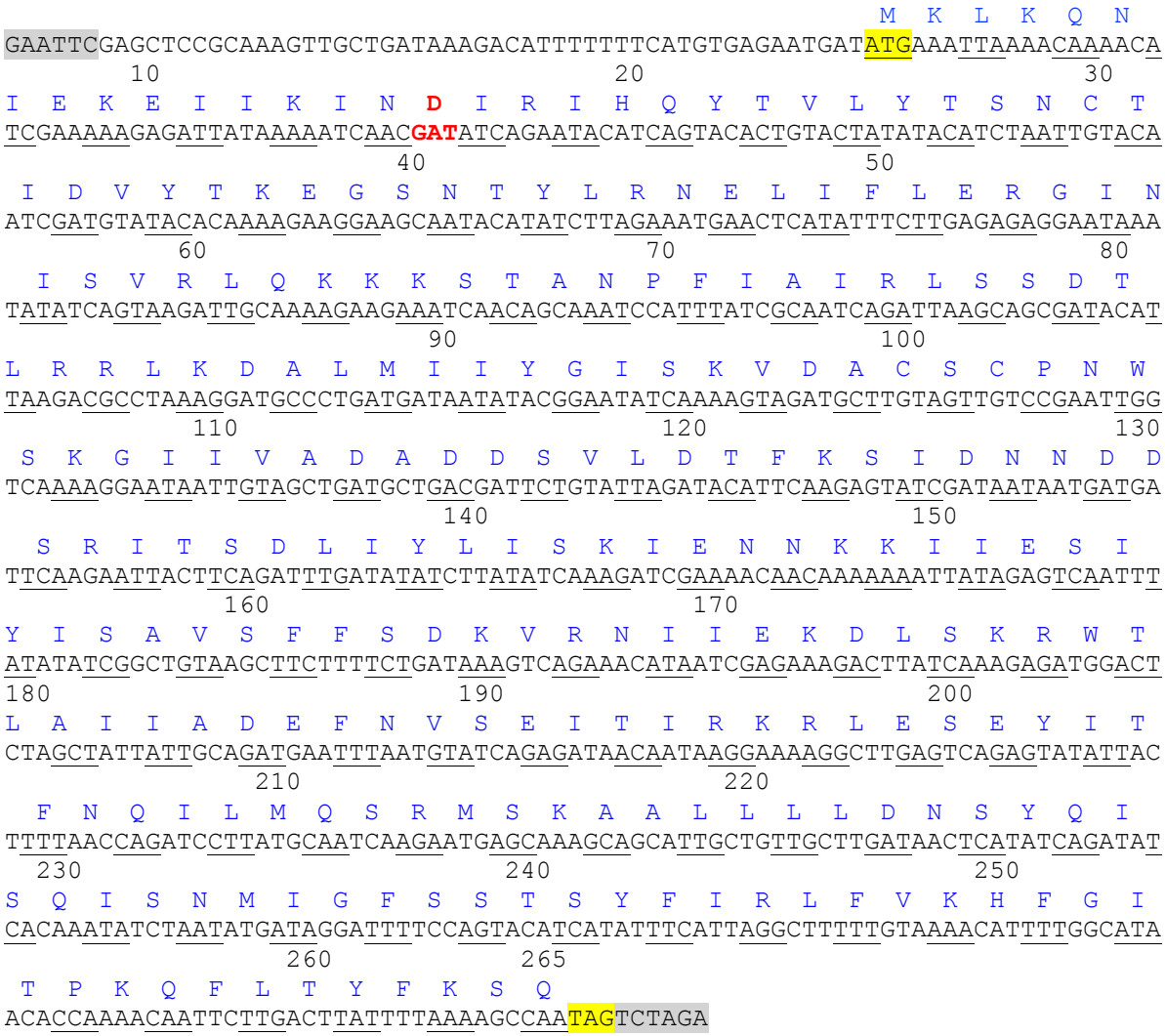


Figure 2.25 The DNA and amino acids sequence of AggR-N16D variant

The figure shows the DNA and amino acid sequence of the AggR-N16D variant, which was cloned into pBAD30 using EcoRI and XballI restriction sites (grey). The start and stop codons of AggR are highlighted yellow and alternate codons are underlined. The amino acid sequence is shown above each codon and the amino acid numbering is above. At position 16 the codon has been changed from AAT to GAT, changing the amino acid of this position from asparagine (N) to aspartic acid (D) (shown red).

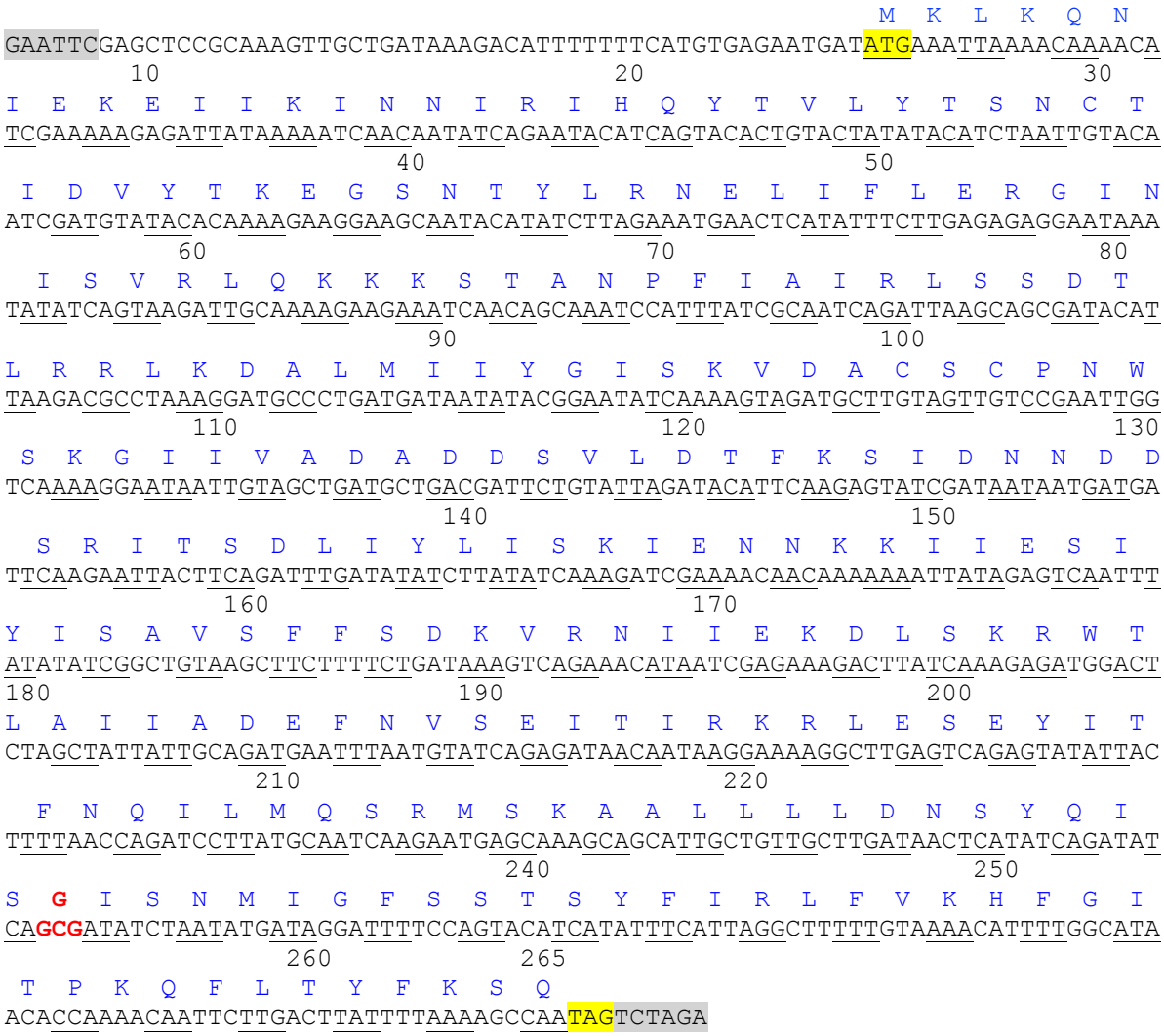


Figure 2.26 The DNA and amino acids sequence of AggR-Q230G variant

The figure shows the DNA and amino acid sequence of the AggR-Q230G variant, which was cloned into pBAD30 using EcoRI and XballI restriction sites (grey). The start and stop codons of AggR are highlighted yellow and alternate codons are underlined. The amino acid sequence is shown above each codon and the amino acid numbering is above. At position 230 the codon has been changed from CAA to GCG changing the amino acid of this position from glutamine (Q) to glycine (G) (shown red).

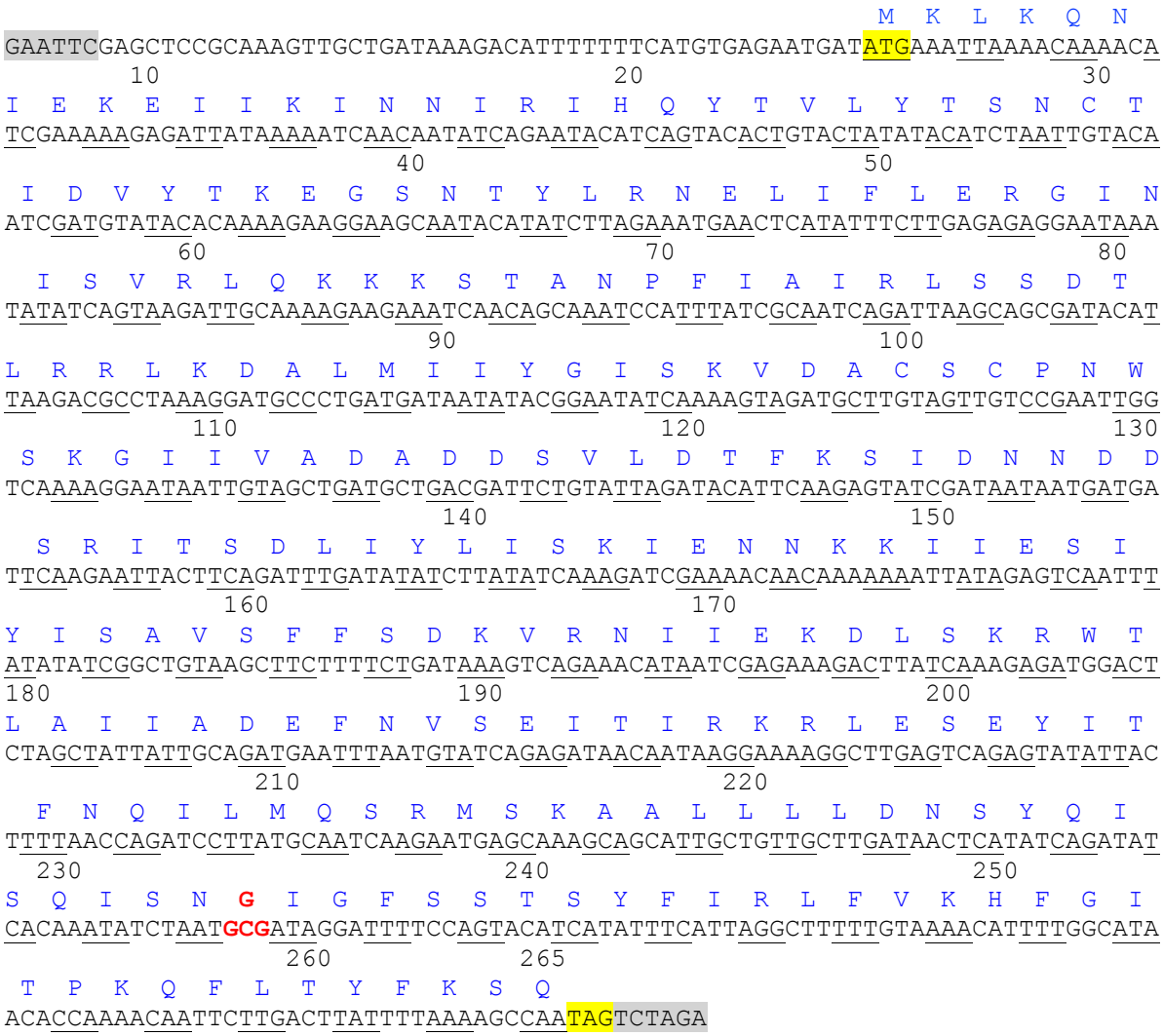


Figure 2.27 The DNA and amino acids sequence of AggR-M234G variant

The figure shows the DNA and amino acid sequence of the AggR-M234G variant, which was cloned into pBAD30 using EcoRI and XballI restriction sites (grey). The start and stop codons of AggR are highlighted yellow and alternate codons are underlined. The amino acid sequence is shown above each codon and the amino acid numbering is above. At position 234 the codon has been changed from ATG to GCG changing the amino acid of this position from methionine (M) to glycine (G) (shown red).

promoter activity was examined by myself by measuring β-galactosidase expression. The negative control used for *aggR* expression was pBAD24 plasmid (Figure 2.28).

2.10 β-galactosidase assays

In this study, β-galactosidase assays were used to measure promoter activity and in each case promoter fragments were cloned upstream of *lacZ* in the expression vectors, pRW50, pRW224 and pRW225.

2.10.1 β-galactosidase assays during exponential growth and stationary phase

Transformants, containing the desired promoter fragment in *lacZ* expression vector derivatives, were picked from a MacConkey agar plate, inoculated into 5 ml of appropriate media containing antibiotics and grown overnight at 37°C with shaking. Cultures were then diluted 1:100 in appropriate media and incubated at 37°C with shaking until mid-logarithmic phase (OD_{650}= 0.4 - 0.6). Cultures were then lysed by the addition of 3 drops of toluene and 3 drops of 1% (w/v) sodium deoxycholate and vortexed. Culture lysates were shaken at 37°C for 10 -15 minutes to evaporate the toluene. Lysates were used to assay β-galactosidase activity in Z-buffer.

Z-buffer was prepared using 0.75 g of potassium chloride (KCl), 0.25 g of magnesium sulphate 7 hydrate ($MgSO_4.7H_2O$), 8.53 g of disodium hydrogen orthophosphate (Na_2HPO_4), 4.87 g of sodium dihydrogen orthophosphate dihydrate ($NaH_2PO_4.2H_2O$) and distilled water up to 1 liter. Before use, 2.7 ml of β-mercaptoethanol and 400 mg of ortho-nitrophenol-β-D-galactopyranoside (ONPG) (a final concentration 13 mM) were added per 1 liter of Z-buffer.

The β-galactosidase assays were carried out in Z-buffer at 30°C and the reaction was started by adding 100 µl of the lysate. The mixture was incubated until a yellow colour developed. The reaction was stopped by the addition of 1 ml of 1M sodium carbonate (Na_2CO_3). The

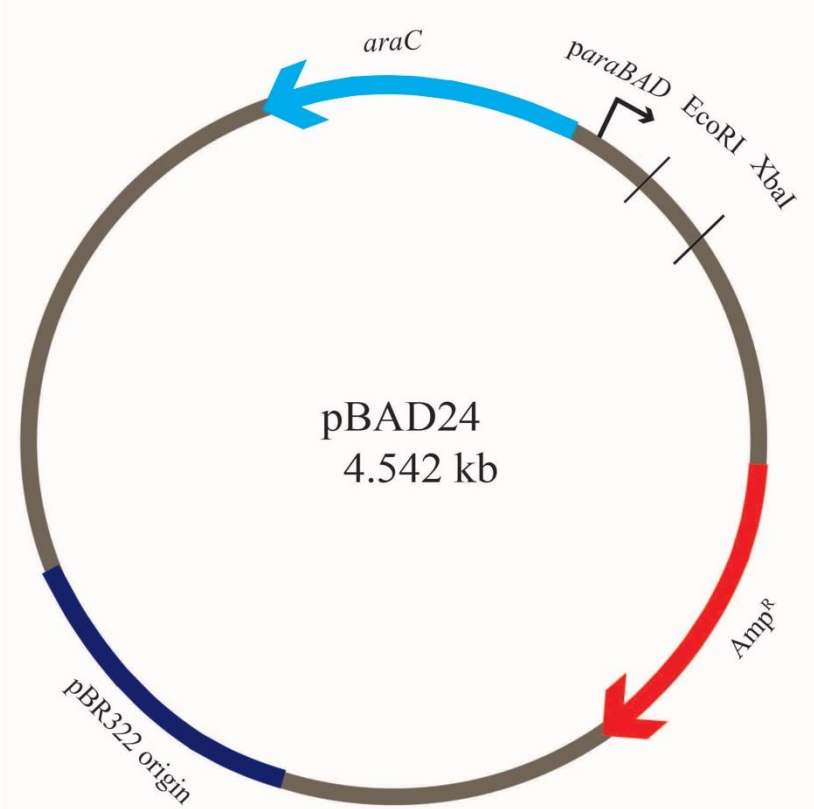


Figure 2.28 A map of the pBAD24 plasmid

The figure depicts the pBAD24 plasmid showing the location of *araC* gene, the *araBAD* promoter and multiple cloning sites. The plasmid also contains ampicillin (Amp^R) resistance gene. This plasmid is used as a negative control while studying AggR-dependent gene expression in *E. coli* K-12.

colour of the reaction was measured at OD_{420} and activity was calculated as:

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

Where,

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)

1000/4.5 = factor for conversion of OD_{420} into nmol Ortho-nitrophenyl-b-D-galactopyranoside (ONPG), based on 1 nmol ml^{-1} ONPG having an OD_{420} of 0.0045

t = incubation time (minutes)

v = volume of lysate added (ml)

The β-galactosidase activity of cells containing plasmids each carrying promoter fragment was measured using three independent transformants and the mean β-galactosidase level and standard deviation was calculated for each promoter fragment. In all assays, cells containing empty vector were used as negative control.

2.11 Primer Extension

Primer extension was carried out to find the transcription start site of various promoters cloned into the pRW50 vector. In this experiment, mRNA was isolated and used as a template to produce complementary DNA (cDNA) using a radioactive labelled primer. The length of cDNA corresponds to the transcription start site and is measured by running reactions on a gel alongside M13mp18 sequence reaction as a ladder.

2.11.1 Isolation of RNA using the QIAGEN RNeasy kit

In this experiment, the recombinant plasmid of interest was transformed into *E. coli* strain BW25113, carrying either pBAD24 or pBAD/*aggR*. Three independent transformants were used to setup over night cultures, each inoculated into 5 ml of LB medium supplemented with ampicillin 100 $mgml^{-1}$ and tetracycline 10 $mgml^{-1}$. The overnight cultures were diluted (1:50) into 10 ml of fresh LB with the antibiotics (tetracycline 10 $mgml^{-1}$ and ampicillin 100 $mgml^{-1}$) and 0.2% (w/v) arabinose and incubated at 37°C with shaking until an OD_{600} of 0.4 - 0.5. The cultures were transferred into Falcon tubes and centrifuged for 20 minutes at 2844 g. The supernatant was discarded and the pellet was resuspended in 200 µl of TE buffer containing 40 μgml^{-1} lysozyme and incubated at room temperature for 15 minutes. RNA was isolated using, QIAGEN RNeasy kit according to manufacturer's specifications and the RNA was eluted from the column in 60 µl of the elution buffer.

Any DNA contamination was removed from the RNA sample, by the addition of 1 µl of Turbo DNase (Ambion) and 3.5 µl of Turbo DNase buffer and tubes were incubated at 37°C for 30 minutes. DNA digestion was stopped by the addition of 6.1 µl of DNase inactivation reagent (Ambion) and samples were incubated at room temperature for 2 minutes with occasional mixing. Microfuge tubes were centrifuged for 1.5 minutes at 9300 g and the supernatant was transferred to a clean tube. The RNA concentration of each sample was quantified using a Nanodrop ND-1000 spectrophotometer and the RNA was stored at -80°C in aliquots, each containing 35 µg of RNA.

2.11.2 Labelling of primers

Primer D49274 was labelled with $\gamma^{32}P$ by mixing 0.5 µl of primer (100 µM), 2 µl ofpolynucleotide kinase buffer (10x), 15.5 µl of RNase free water, 1 µl of $\gamma^{32}P$ and 1 µl of T4

polynucleotide kinase (NEB) (Table 2.4). The mixture was incubated for 30 minutes at 37°C and then 10 minutes at 68°C to inactivate the enzyme. The labelled primer was stored at -20°C.

2.11.3 Sequencing reactions for primer extension

A sequencing reaction is used as a marker for primer extension reactions run on acrylamide gel and was prepared using a T7 sequencing kit (USB). The single stranded M13mp18 phage DNA provided with the kit was used as a template to prepare sequence reactions, according to manufacturer's instructions (Figure 2.29). 2 µl of the Universal primer (10 pmol) were used per reaction with 10 µl of the diluted template, and 2 µl of the annealing buffer. This was mixed and incubated at 65°C for 5 minutes, and then transferred to 37°C for 10 minutes. T7-polymerase stock (8 unitsμl^{-1}) was diluted 1:5 in dilution buffer for the labelling reaction. For each labelling reaction, 15 µl of the annealed template/primer reaction mixture, 1 µl of α-^{32}P dATP (10 $\mu Ci\mu l^{-1}$) and 2 µl of diluted T7-polymerase were incubated at room temperature for 5 minutes.

The kit contains four different ddNTPs mixes, each containing ddNTP (ddATP, ddCTP, ddGTP and ddTTP) and 2.5 µl of each was transferred to separate microfuge tubes. The sequence reactions were terminated by adding 4.5 µl of the labelling reaction into each of the four tubes, which were incubated at 37°C for 5 minutes. The reactions were stopped using 5 µl of 'STOP' solution (Formamide (Deionised) 97.5% (v/v), EDTA (10mM), Xylene cyanol (0.3% (w/v)) and bromophenol blue (0.3% (w/v))) and tubes were stored at -20°C after this step.

2.11.4 Primer Extension

The RNA aliquots (35 µg) were thawed on ice and 1 µl of labelled primer was added to each

```
1         10         20         30         40         50         60         70
.         .          .          .          .          .          .          .
GTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCT
          80         90        100        110        120        130        140
          .          .          .          .          .          .          .
CGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
         150        160        170        180        190        200        210
          .          .          .          .          .          .          .
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
         220        230        240        250        260        270        280
          .          .          .          .          .          .          .
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
         290        300        310        320        330        340        350
          .          .          .          .          .          .          .
GGAGAGGCGGTTTCCGTATTCCCCCCCACCCTCCTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGAT
         360        370        380        390        400        410        420
          .          .          .          .          .          .          .
TGCCCTTCACCCCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAA
         430        440        450        460        470        480        490
          .          .          .          .          .          .          .
ATCCTGTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGT
         500        510        520        530        540        550        560
          .          .          .          .          .          .          .
TGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAA
         570        580        590        600        610        620        630
          .          .          .          .          .          .          .
AACCGTCTATCAGGGCGATGGCCCACTCAGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGC
         640        650        660        670        680        690
          .          .          .          .          .          .
CGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCCATTTACACCTTGACCCCCAAAGC
```

Figure 2.29 A part of M13mp18 sequence

The figure shows a part of M13mp18 sequence from single stranded circular DNA. It was provided with the USB kit. The sequence shown begins with the annealing sequence of the universal primer.

aliquot. The RNA was ethanol precipitated as described in section 2.5.2. The pellet was resuspended in 30 µl of hybridization buffer (20 mM HEPES, 0.4 M NaCl, 80% (v/v) formamide), vortexed and incubated for 5 minutes at 50°C. The tube was vortexed once more, briefly centrifuged and incubated for 15 minutes at 75°C and for another 3 hours at 50°C. The RNA was ethanol precipitated as described in section 2.5.2. The pellet was resuspended in 31 µl of RNase free water and mixed with 10 µl of reverse transcriptase buffer (5x) (Promega), 1 µl of 50 mM DTT, 5 µl of 10 mM dNTPs (Bioline), 2.5 µl of AMV reverse transcriptase (Promega) and 0.6 µl of RNasin (Promega). Microfuge tubes were incubated for 60 minutes at 37°C and 10 minutes at 72°C (to inactivate the reverse transcriptase enzyme). The microfuge tubes were briefly centrifuged and 1 µl of RNase A solution (10 $mgml^{-1}$) was added to reaction to digest the RNA template. The reaction mix was then incubated for 30 minutes at 37°C.

The DNA product of the reaction was ethanol precipitated (section 2.5.2) and resuspended in 6 µl of stop solution. The primer extension product was diluted in 1:5 in stop solution and 5 µl of the diluted volume was loaded onto 6% (w/v) denatured PAGE (polyacrylamide gel electrophoresis), along with the sequence reactions from section 2.11.3. The gel was run at 60 Watts for 3 hours, dried using vacuum drier and exposed to phosphorimaging screen (Fuji) for overnight. The data was visualised by scanning the phosphorimaging screen using BioRad molecular imager FX and Quantity one software.

2.12 Gene doctoring

There are many gene manipulation techniques are available that are used for addition and deletion of certain DNA sequences in the bacterial cells. One of these techniques is “the gene doctoring technique” developed to delete or add a gene in the host cell. This technique was

been developed by Lee *et al*. (2009) for modification of the host chromosome. Gene doctoring has been used in this study to delete the *aggR* gene on plasmid pAA2 from the EAEC strain DFB042 (Table 2.1).

2.12.1 An overview of the technique

Most of the gene manipulation techniques, use the λ Red system to carry out recombinant modification. The λ Red system consists of three proteins: *exo* generates the single stranded DNA overhangs, *gam* protects the single stranded DNA and *bet* catalyse the recombination of the homologous regions. The technique was developed by Datsenko and Wanner (2000) for gene manipulation and they used linear DNA fragments, which were electroporated into bacterial cells. These DNA fragments carry an antibiotic resistance cassette and homology regions for the DNA sequence of interest. The λ Red system genes were provided by a plasmid in the host cells.

This method was further modified by Herring *et al.* (2003) who used donor plasmid instead of linear fragments of DNA. The donor plasmid contained an antibiotic resistance cassette, which was flanked by the DNA sequences of the targeted gene. This part of the plasmid was also flanked by SceI restriction sites. The genes for λ Red were provided on another plasmid in the host cells that also contain the SceI meganuclease gene required to produce linear DNA fragments from the donor plasmid. All these genes were cloned under an inducible promoter. This technique improved the efficiency as more copies of the linear DNA fragments were available for DNA recombination.

In gene doctoring, Lee *et al.* (2009) modified this technique by adding selection for successful recombinants. This method uses a *sacB* gene cassette from *Bacillus subtilis.* The product of this gene converts sucrose into levans that accumulates in periplasm of *E. coli* and it is toxic.

The cells that have *sacB* cannot grow in the presence of sucrose (Li *et al.*, 2013). In this technique, there are also two plasmids required, like in Herring *et al.* (2003). The donor plasmid consists of sequence of two homology regions around a kanamycin cassette and this whole DNA segment is located between two SceI restriction sites. In addition, the donor plasmid also has an ampicillin resistance cassette for selection and *sacB* for sucrose sensitivity. After successful recombinations, the host cell loses ampicillin resistance and becomes resistant to sucrose. The λ Red system and *sce-I* present on pACBSR, are necessary for generating linear DNA fragment, from pDOC-K/*aggR*401/*aggR*402 (Figure 2.21). The pACBSR also encodes chloramphenicol resistance.

2.12.2 Gene doctoring methodology

For gene doctoring, the donor plasmid and pACBSR were co-transformed into EAEC strain DFB042 by electroporation and transformants were selected on chloramphenicol (50 μgml^{-1}) and kanamycin (25 μgml^{-1}) LB agar plates. The transformants were re-streaked onto chloramphenicol (50 μgml^{-1}) and kanamycin (25 μgml^{-1}) LB agar plates and kanamycin (25 μgml^{-1}) and sucrose (5% (w/v)) plates simultaneously to check the system works in EAEC DFB042. A sucrose sensitive, chloramphenicol resistant and kanamycin resistant colony was inoculated into 0.5 ml of LB containing kanamycin (25 μgml^{-1}) and incubated at 37°C with aeration for ~3 hours until the culture turned turbid. The culture was centrifuged at 15700 g and washed three times with 0.5 ml of 0.1x LB to remove any traces of antibiotic. The cells were then resuspended in 0.1x LB containing 0.3% (w/v) arabinose, to induce λ Red and SceI and incubated at 37°C with shaking for 3 hours. The culture was centrifuged at 15700 g and the pellet was resuspended in 0.2 ml of 0.1x LB. This cell suspension was plated onto four LB agar plates supplemented with kanamycin (25 μgml^{-1}) and sucrose (5% (w/v)). The uncut

status of the donor plasmid was also determined by inoculating 100 µl of 1000-fold dilution of pre-induction and post-induction samples onto both LB agar and LB agar supplemented with ampicillin (100 µgml^{-1}) plates. The cells with donor plasmid that has been linearised will be sensitive to ampicillin (100 µgml^{-1}) and this indicates the efficiency of SceI digestion in producing linear DNA within the cells. All the plates were incubated at room temperature until visible colonies appeared.

2.12.3 Detection of gene doctored cells

Colonies which grew on kanamycin (25 µgml^{-1}) and sucrose (5% (w/v)) were re-streaked onto LB plates with ampicillin (100 µgml^{-1}), LB plates with sucrose (5% (w/v)) and LB plates with kanamycin (25 µgml^{-1}). Ampicillin sensitive and kanamycin resistant colonies that were not sensitive to sucrose were tested for DNA modification using colony PCR. The primers used for PCR and sequencing reactions are detailed in Table 2.4. The successful candidates were then checked for chloramphenicol (50 µgml^{-1}) sensitivity to confirm the loss of pACBSR plasmid.

2.13 Biofilm formation

Overnight cultures were inoculated from freshly grown single colonies of *E. coli,* EAEC DFB042 and EAEC DFB042Δ*aggR* in 5 ml of LB medium. The overnight cultures were diluted 1:100 in 5 ml of DMEM medium and incubated at 37°C for 1 hour with shaking and from each strain, 150 µl of the diluted culture was added to 8 wells of microtitre plate. The plate was sealed with Parafilm (TM) to avoid evaporation and incubated at 37°C for overnight. The next day, media was removed and plate was smacked on the paper towel to remove any traces of media. Then 150 µl of crystal violet was added to each well and left at 4°C for 30 minutes. The crystal violet is removed and plate was smacked on a paper towel to

remove any traces. Then, the plate is submerged into water and water was poured out into container again. It was repeated 3 times and the negative control was appearing negative. The plate was dried by smacking on a paper towel and 150 µl of ethanol/acetone solution (80 ml ethanol and 20 ml acetone) was added to each well. The plate was left on a shaker for 30 minutes and absorbance was measured at 595 nm by a Labsystems Multiskan MS plate reader (Thermo Fisher Scientific Inc).

2.14 Detection of Pet

Pet is one of the virulence factors of EAEC 042 that is secreted out of the cell. Pet production was detected for its possible AggR-dependent regulation as AggR regulates other virulence factors *e.g.* fimbriae gene expression and *pet* is present between two regions of AggR-regulated fimbrial regions.

2.14.1 Protein precipitation

Fresh isolated colonies of *E. coli* K-12, EAEC DFB042, EAEC DFB042*aggR* (pBAD/*aggR*) and EAEC DFB042Δ*aggR,* were used to set up overnight cultures. Overnight cultures were sub-cultured by 1:100 dilutions into 50 ml of DMEM supplemented with tryptone (1% (w/v)) and arabinose (0.2% (w/v)) where appropriate and incubated for different length of time to determine different time points. After completion of incubation, the flask was kept on ice for 10 minutes and culture was transferred into a precooled Falcon tube. The cultures were centrifuged using a centrifuge 5810 R (eppendrof) at 2844 g for 20 minutes at 4°C. The supernatant was filter sterilised through a 0.22 µM filter (Acrodisc® syringe filters) and transferred into appropriate centrifuge tubes for centrifugation. 5 ml of freshly prepared 100% trichloro acetic acid (TCA) was added to 25 ml each filtered supernatant.

The mixture was left overnight to precipitate at 4°C and centrifuged at 20817 g for 45 minutes

at 4°C. The pellet was resuspended in 2 ml of acetone and transferred to 2 ml microfuge tubes for subsequent steps. The mixture in acetone was centrifuged in microfuge at 20817 g for 45 minutes at 4°C and poured off supernatant. The pellet was air dried and resuspended in 100 µl of cracking buffer (10.12 ml of SDS (20% (w/v)), 20 ml of glycerol, 5 mg of bromophenol blue, made up to 92 ml by stocking gel buffer (1:10 diluted) and 8.7% (v/v) of β-mercaptoethanol). If the colour of samples turned yellow due to low pH of the solution mixture reaction, this was neutralised by addition of 5 µl of Tris-HCl (1.5 M, pH 8.0).

2.14.2 Western blot

Samples from 2.14.1 were heated at 100°C for 5 minutes and then centrifuged for 1 minute at 15700 g. The protein was loaded onto a SDS-PAGE gel along with prestained marker. The electrophoresis was carried out at 150 V until the blue dye front reached the bottom of the gel as described in section 2.4.3. After electrophoresis, the gel was soaked in transfer buffer (25 mM Tris, 190 mM glycine, 5% (v/v) methanol and 0.1% (w/v) SDS) for 10 minutes. A piece of nitrocellulose membrane (*e.g.* Hybond-ECL) of the same size as that of the gel was soaked into distilled water for 5 minutes and then in transfer buffer for another 5 minutes. Four pieces of the filter paper and the blotting pads for transfer apparatus were also soaked in transfer buffer for 5 minutes. The western blot transfer apparatus was set up by placing transfer membrane towards the anode and gel towards cathode, sandwiched between filter papers and blotting pads. The proteins were transferred to the membrane at 30 V for 1 hour using transfer buffer (at room temperature).

Following transfer, the nitrocellulose membrane was washed two times with 25 ml 1x TBS (20 mM Tris, 150 mM NaCl, pH 7.6) for 5 minutes at room temperature, with vigorous shaking. The membrane was incubated in 25 ml of blocking buffer (TBS with 0.1% (w/v)

Tween-20 and 5% (w/v) non-fat dry milk) for 1 hour at room temperature followed by three washes for 5 minutes each with 25 ml of TBS/T (TBS with 0.1% (w/v) Tween 20) with vigorous shaking.

The membrane was incubated with the primary antibody (anti-passenger domain antibodies) 1µl in 10 ml of blocking buffer with gentle agitation overnight at room temperature in a sealed bag. The membrane was rinsed with TBS/T, then washed three times for 5 minutes each with 25 ml of TBS/T with vigorous shaking. The membrane was then incubated with HRP-conjugated secondary antibody (1 µl in 10 ml blocking buffer) with gentle agitation for 1 hour at room temperature again in a sealed bag. The membrane was washed three times for 5 minutes each with ~25 ml TBS/T with vigorous shaking.

The reagents of the ECL detection system were removed from storage at 2-8 °C and transferred 1 ml of each A and B regents were into 2ml microfuge tubes to equilibrate to room temperature. The excess TBS/T drained from the membrane and placed protein side up on a sheet of cling film. Both reagents were mixed and pipetted the mixed detection reagent onto the surface of the membrane and left for 5 minutes at room temperature.

The membrane was held with a pair of tweezers and touching the edge against a tissue and the excess detection reagent drained. The membrane was wrapped into a fresh sheet of cling film with a care to avoid any air bubbles. It was placed into an X-ray cassette with protein side up and overlaid with a sheet of autoradiography film (GE Healthcare Limited). The cassette was closed and time was noted for exposure to the membrane. The film was removed and developed immediately using an x-ray developing machine CURIX 60 (AGFA). Based on the results, the next film was exposed for longer/ shorter time to get an appropriate band on the film.

Chapter 3

Organisation of the promoter regions of the aggregative adherence fimbrial genes of EAEC strain 042

3.1 Introduction

EAEC is an intestinal pathogen and there are many virulence factors that enable this pathogen to attach to the intestine and cause disease in humans. The major virulence factors include EAEC adhesins, toxins and AggR (Vila *et al.*, 2000; Morin *et al.*, 2010). The transcription factor, AggR, controls the expression of various virulence determinants in EAEC 042 (Morin *et al.*, 2013).

Aggregative adherence fimbriae II (AAF/II) are an important virulence factor that confer "the stacked brick" aggregation pattern to EAEC (Nataro *et al.*, 1987). The genes that encode these fimbriae are regulated by AggR (Morin *et al.*2010). Elias *et al.* (1999) described the arrangement of the fimbrial genes in two clusters known as Region 1 and Region 2, both on the virulence plasmid of EAEC 042 *i.e.* pAA2. Region 1 encodes the genes for chaperone (*aafD*), the fimbrial subunit (*aafA*) and the transcription regulator (*aggR*). Region 2 contains a chaperone pseudogene (*afaB*) and the genes for the usher protein (*aafC*) and fimbrial adhesin (*aafB*) (Figure 3.1). Elias *et al.* (1999) also reported that Region 1 is regulated by AggR, whilst Region 2 is not, and gave limited insight about the promoters of these genes.

The study of Morin *et al.* (2010), identified AggR-binding sites at the *aggR* promoter and these sites are similar to the binding site of another member of the AraC family, Rns. Furthermore, Morin *et al.* (2013) identified AggR-regulated genes in the EAEC 042 genome and found that the fimbrial genes *aafA, aafB, aafC* and *aafD* from Region 1 and Region 2 are regulated by AggR. However, this study did not predict the promoter elements of these genes. The work detailed in this chapter concerns the identification of the AggR-dependent promoters of the fimbrial genes of EAEC 042, present in two clusters. DNA fragments of Region 1 and Region 2 regulatory regions are described in sections 3.2 and 3.3, respectively.

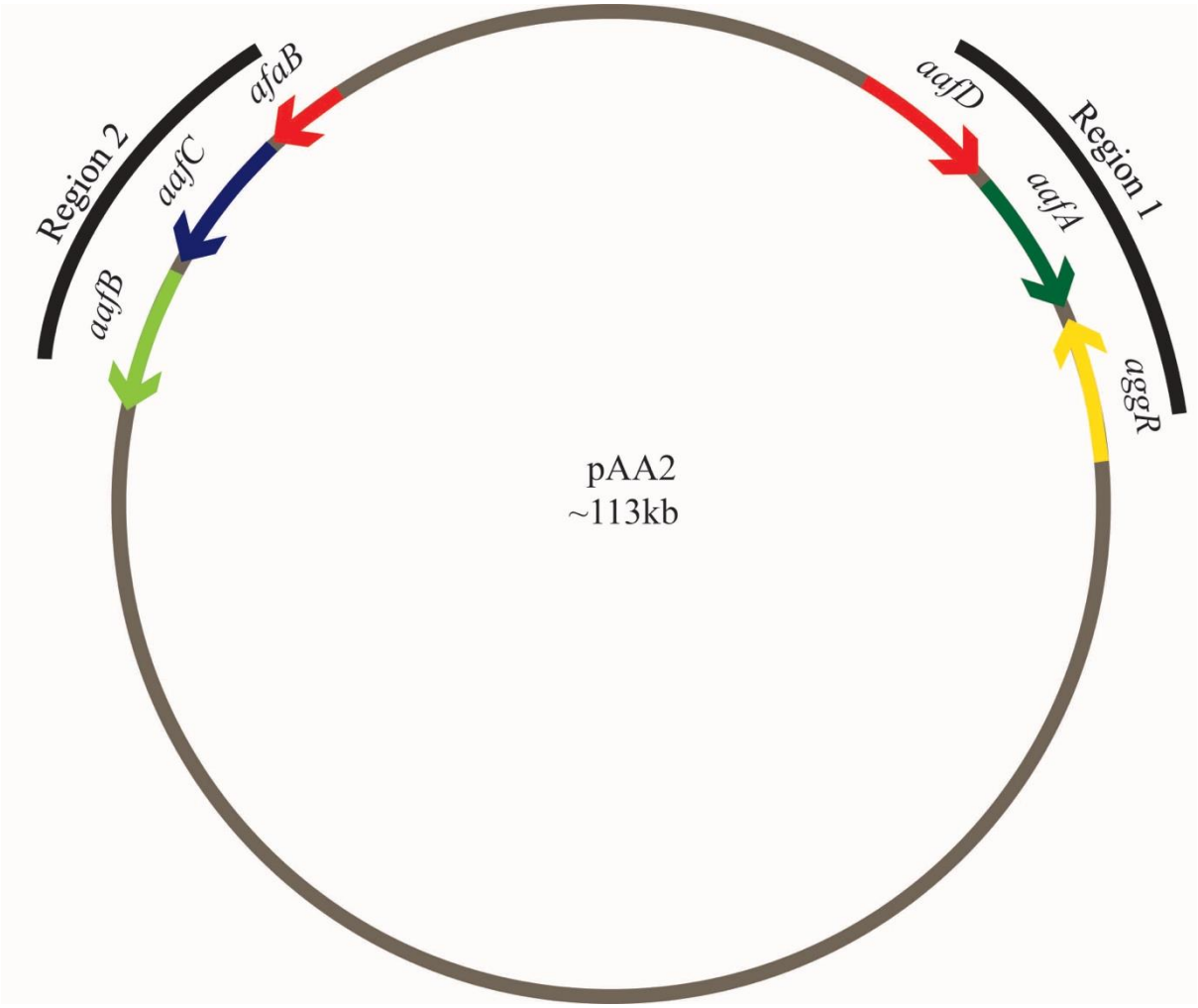


Figure 3.1 A Schematic representation of Region 1 and Region 2 of AAF/II genes on pAA2 plasmid

The figure shows a schematic diagram of the pAA2 plasmid from EAEC strain 042 and the two regions that encode the attachment adherence fimbriae (AAF/II) Region 1 encodes *aafD* (a chaperone protein), *aafA* the (fimbrial subunit) and *aggR* (a transcription activator) genes and Region 2 encodes *afaB* (a chaperone pseudogene), *aafC* (the usher protein) and *aafB* (an adhesin) genes. The diagram is not to scale and modified from Figure 1.4.

3.2 Analysis of the regulation of Region 1

3.2.1 Analysis of the upstream DNA sequence of *aafD* and *aafA*

In Region 1 of pAA2, there are two structural genes *aafD* and *aafA,* which are likely to be transcribed as an operon. In order to locate promoters, which control the expression of *aafD* and *aafA*, DNA sequences upstream of the genes in Region 1 were cloned into pRW50 (Figure 2.1). This is a vector that has a promoterless *lac* operon and any fragment cloned upstream of this operon, using the EcoRI and HindIII sites, can serve as a *lac* operon promoter. Thus, *lacZ* expression depends on the promoter activity of the cloned fragment.

For the *aafD* gene, a 433 bp DNA fragment, *aafD*100*,* was cloned into pRW50 and bioinformatic analysis showed that this fragment contains potential AggR-binding sites (Figure 3.2). The schematic diagram of the *aafD100* promoter fragment is shown in Figure 3.3A. The recombinant plasmid, pRW50/*aafD*100*,* was transformed into the *E. coli* K-12 strain BW25113 Δ*lac* carrying either pBAD/*aggR* or pBAD24 vector (Table 2.1). The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6). In the recombinant plasmid pBAD/*aggR*, *aggR* is cloned under the control of the *araBAD* promoter, so, 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids.

Results in Figure 3.3B show that during growth in LB with 0.2% (w/v) arabinose, β-galactosidase levels measured in cells containing pBAD/*aggR* with pRW50/*aafD*100 increased almost 30-fold in the presence of AggR when compared to an AggR negative control containing pBAD24 with pRW50/*aafD*100 (16593 and 524 units, respectively). This shows that there is an AggR-dependent promoter present upstream of *aafD*.

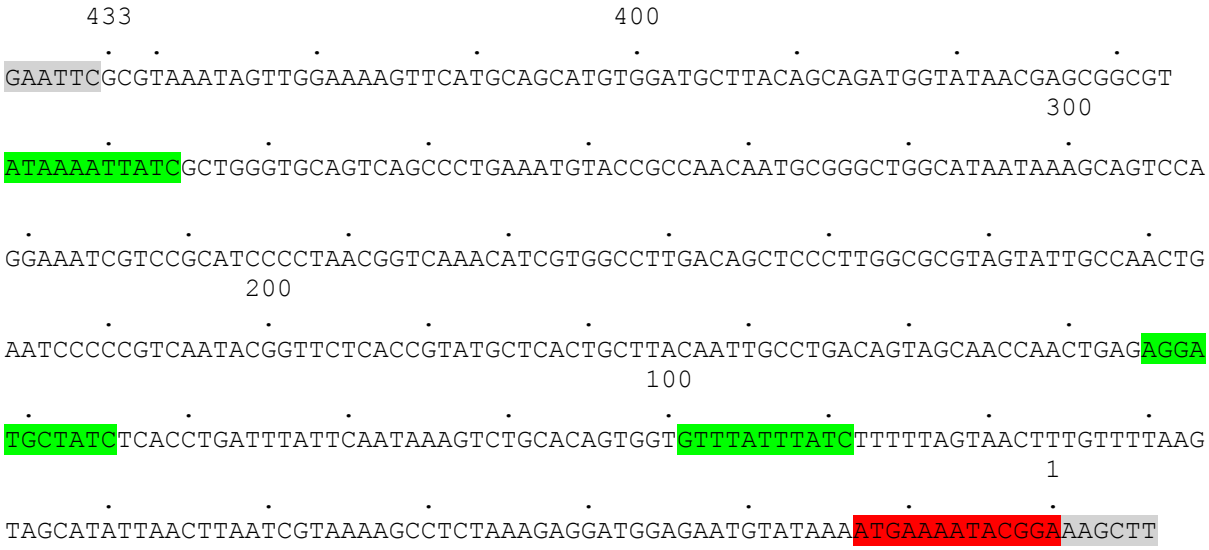


Figure 3.2 The DNA base sequence upstream of the *aafD* gene

The figure shows the 433 bp DNA sequence of the *aafD*100 promoter fragment. The open reading frame sequence is highlighted red and the potential AggR-binding sites are green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.2.

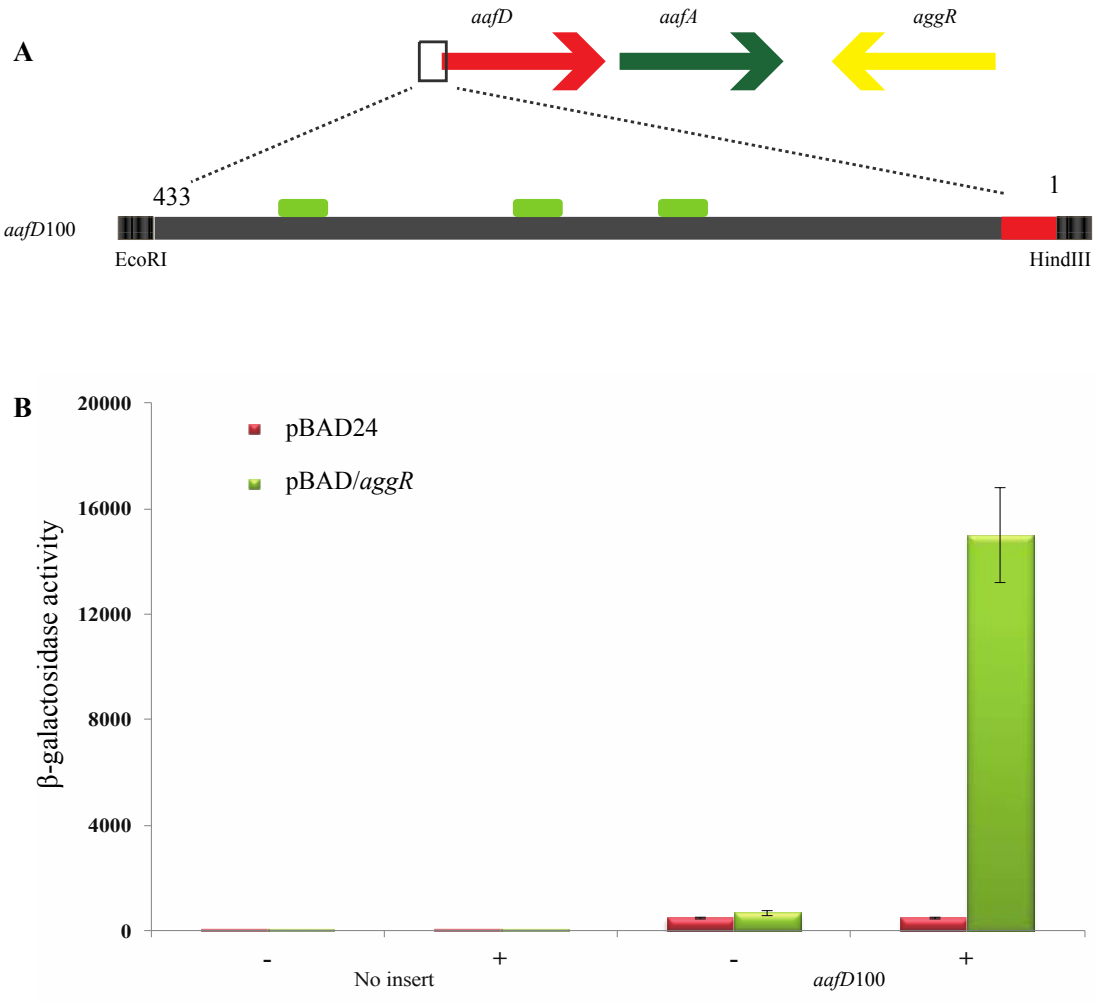


Figure 3.3 Analysis of *aafD* promoter fragment from Region 1

A. The upper line shows a schematic arrangement of the Region 1 genes of pAA2. The lower part illustrates the *aafD*100 promoter fragment. The grey bar represents the upstream sequence of *aafD* and the red part of the bar represents 13 bp sequence of *aafD* present on the promoter fragment. The potential AggR-binding sites are indicated as green boxes. The DNA sequence in this diagram is numbered above the fragment from 1 to 433 and the fragment was cloned using EcoRI and HindIII restriction sites into pRW50. This diagram is not to scale.
B. The panel illustrates β-galactosidase activity measured in the *E. coli* K-12 strain BW25113 Δ*lac*, containing the *lacZ* expression vector (pRW50) or *aafD*100 promoter fragment cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

A study by Elias *et al.* (1999) has shown that *aafA* is also an AggR-regulated gene, so, to investigate this, another fragment, *aafA*100, was cloned into pRW50. Bioinformatic analysis showed that this fragment also contains potential AggR-binding sites on the forward and the reverse strands Figure 3.4). The schematic diagram of the *aafA*100 promoter fragment is shown in Figure 3.5A. The results of β-galactosidase assays showed that cells containing pRW50/*aafA*100 with pBAD/*aggR* have similar β-galactosidase levels when compared to cells containing pRW50/*aafA*100 with pBAD24 (122 and 95 units, respectively) (Figure 3.5B). This shows, there is a very weak promoter present upstream of *aafA*, and this promoter is not regulated by AggR. As cells containing pRW50/*aafD*100 showed β-galactosidase activity that was increased by AggR induction, while cells containing pRW50/*aafA*100 showed little β-galactosidase activity, with no increase in the presence of AggR, this suggestss that an AggR-regulated promoter is present on the *aafD*100 promoter fragment and there is no AggR-regulated promoter on *aafA*100. This indicates that AggR regulates the proposed operon promoter of Region 1 and that the operon promoter is present upstream of *aafD*.

3.2.2 Identification of the minimal regulatory region of the *aafD* promoter necessary for AggR-mediated activation

To find out the minimal region of the *aafD*100 fragment fully induced by AggR, a series of nested deletions were made from the *aafD*100 fragment (*aafD*99 (333 bp) to *aafD*94 (73 bp)) and cloned into the *lacZ* expression vector pRW50 (Figure 3.6A and B). The first 3 deletions (*aafD*99 to *aafD*97) were made by successively deleting 100 bp in each nested fragment and the other 3 fragments (*aafD*96 to *aafD*94) were made by progressively deleting 20 bp in each nested fragment. The β-galactosidase activities measured from lysates of cells containing recombinant plasmids, pRW50/*aafD*99 to pRW50/*aafD*96, showed AggR-dependent

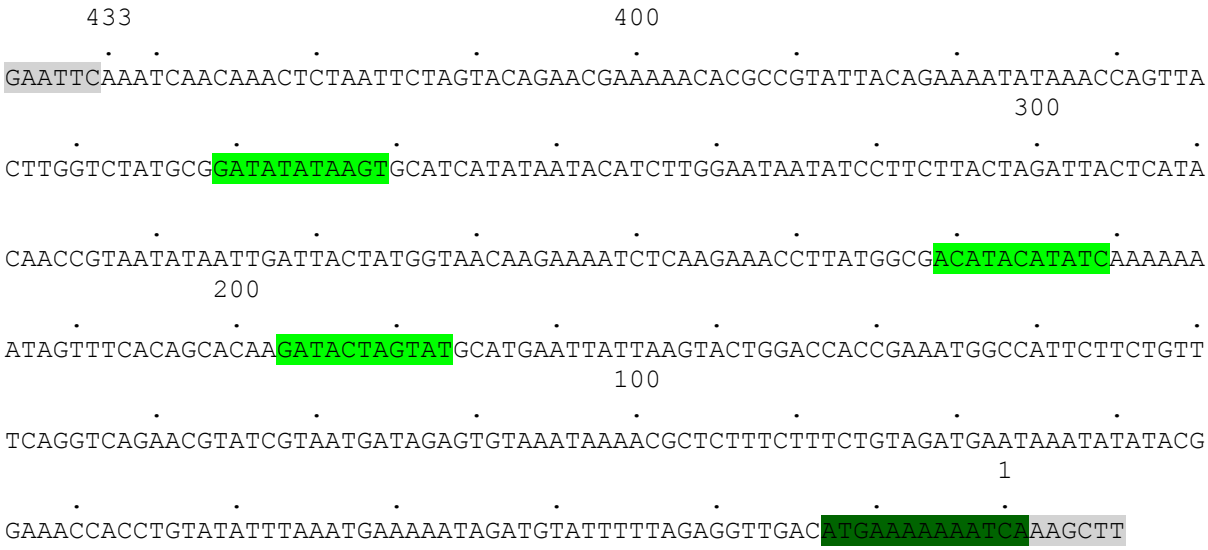


Figure 3.4 The DNA base sequence upstream of the *aafA* gene

The figure shows the 433 bp DNA sequence of the *aafA*100 promoter fragment. The open reading frame sequence is highlighted dark green and the potential AggR-binding sites are light green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.2.

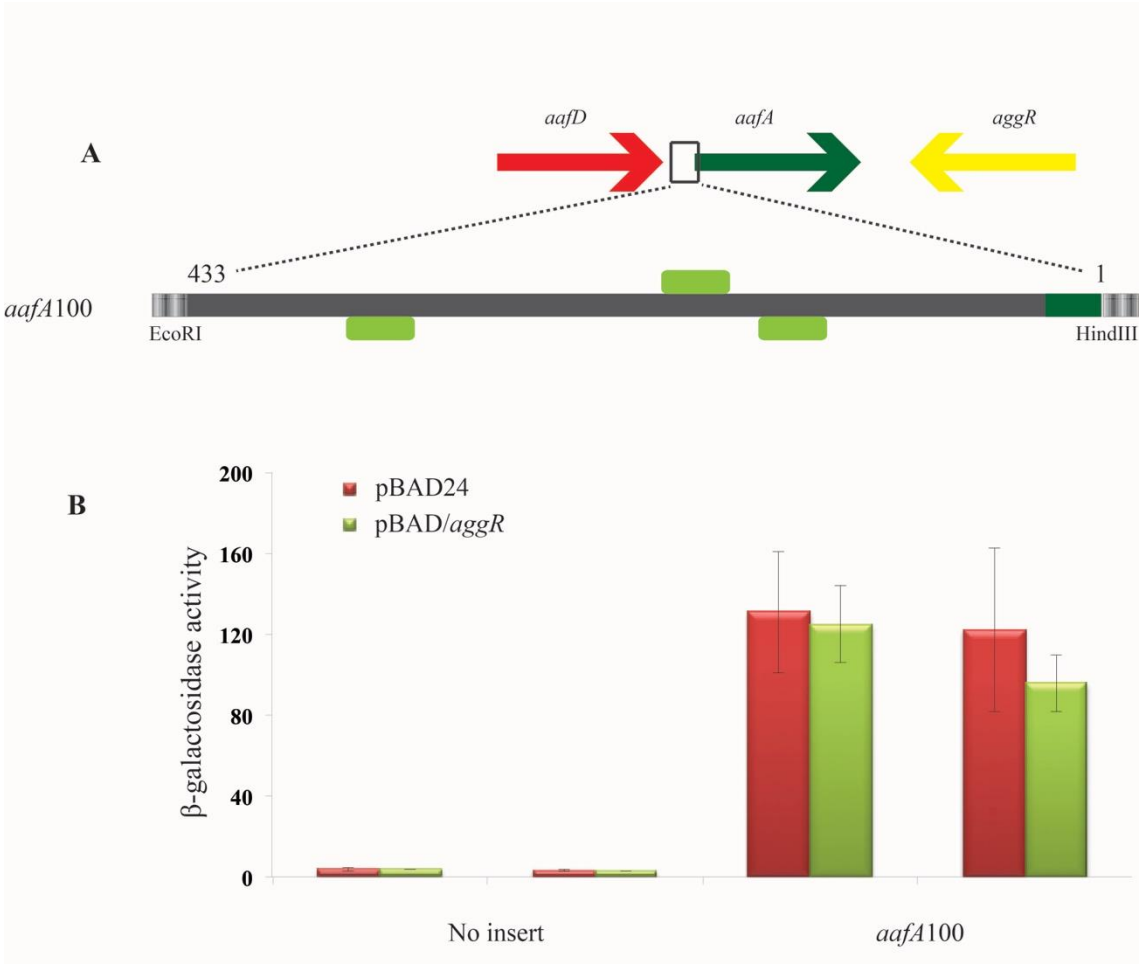


Figure 3.5 Analysis of *aafA* promoter from Region 1

A. The upper line shows the schematic arrangement of the Region 1 genes on pAA2. The lower part illustrates *aafA*100 promoter fragment. The grey bar represents upstream sequence of *aafA* and the dark green part of the bar represents 13 bp sequence of *aafA* present in the promoter fragment. The potential AggR-binding sites are illustrated as green boxes on the forward and the reverse strands. The base sequence is numbered above the fragment from 1 to 433 and the fragment was cloned using EcoRI and HindIII restriction sites into pRW50.This diagram is not to scale.

B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aafA*100 promoter fragment cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

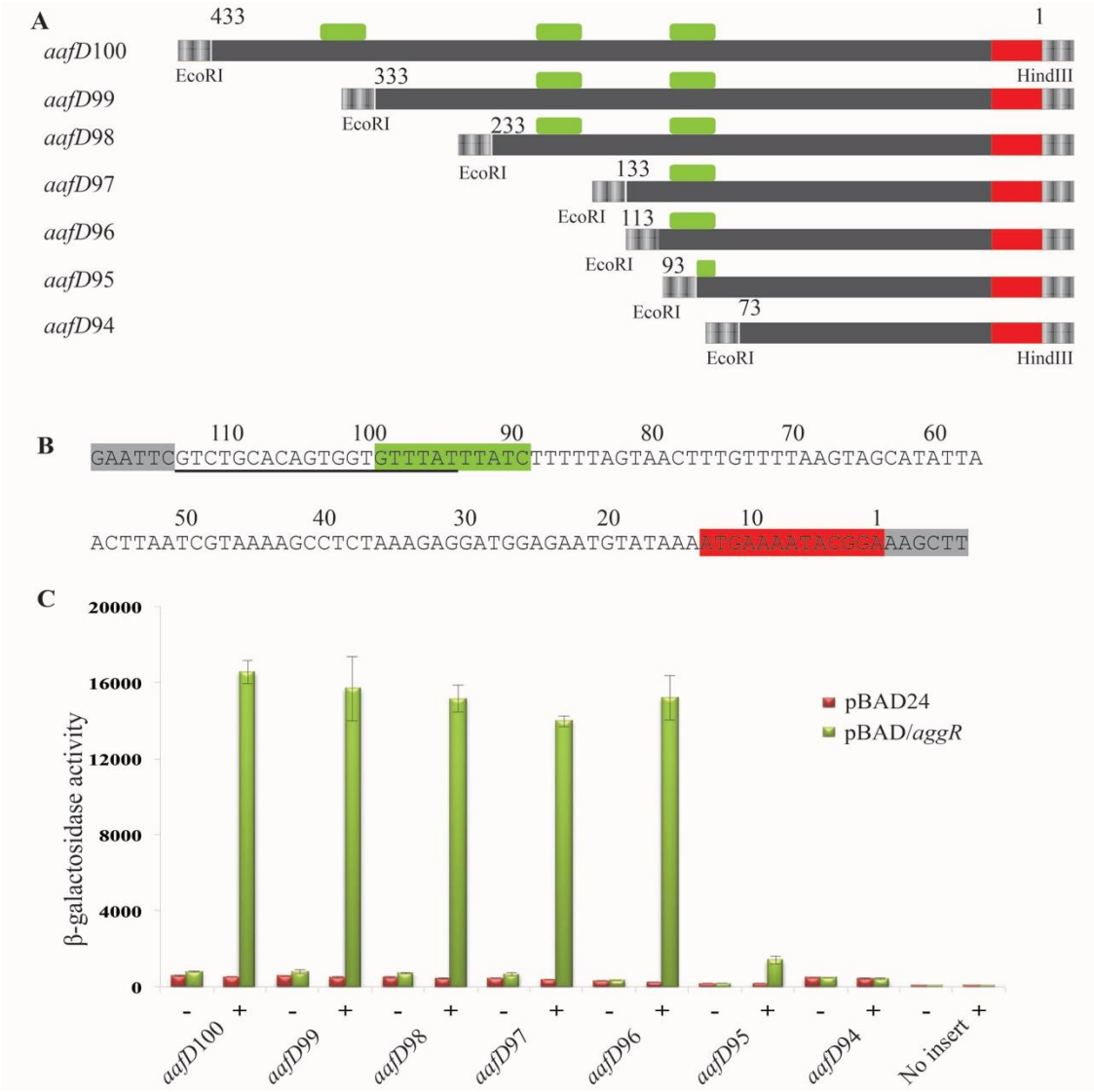


Figure 3.6 Nested deletion analysis of the *aafD* regulatory region

A. The panel illustrates *aafD*100, *aafD*99, *aafD*98, *aafD*97, *aafD*96, *aafD*95 and *aafD*94 promoter fragments. The grey bars represent upstream sequence of *aafD* and red bars represent coding sequence of *aafD*. Green boxes illustrate potential AggR-binding sites. All The fragments were cloned into pRW50 using EcoRI and HindIII restriction sites. This diagram is not to scale.

B. The panel shows the sequence of the *aafD*96 fragment. The AggR-binding site is highlighted green, restriction sites grey and the sequence of *aafD* red. The underline sequence adjacent to EcoRI site was deleted to construct the *aafD*95 promoter fragment.

C. The panel illustrates β-galactosidase activity measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aafD*100 promoter fragment derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

regulation of almost 30-fold while cells containing pRW50/*aafD*95 showed only about 4-fold increase in the presence of AggR (pBAD/*aggR*) (15192 and 1402 units, respectively) as compared to the AggR negative control (pBAD24). There was no AggR-induction observed in cells carrying pRW50/*aafD*94 (1-73 bp) (Figure 3.6B and C).

These results show that an AggR-binding sequence is located between the end points of *aafD*95 and *aafD*96. The DNA binding sequence of AggR is predicted to be the Rns-binding site (Munson, 2013; Morin *et al.*, 2010) (Figure 3.7). The AggR-binding site at the *aafD* promoter has TAT (from position 9 to 11) at the downstream end as predicted in other studies but has a G at the upstream end, instead of A which was found in the AggR-binding sites identified by Morin *et al.* (2010) (Figure 3.7). The low level of induction in cells containing pRW50/*aafD*95 is due to the AggR-binding site has been partially deleted in this fragment. The TATC part of the binding site is present on the fragment and the upstream part is deleted (Figure 3.6).

3.2.3 Identification of the AggR-binding site at the *aafD* promoter, using mutational analysis

Munson *et al.* (2010) confirmed the AggR-binding site at the *aggR* promoter by making point mutations in conserved T residues present in the downstream part of the AggR-binding site at positions 10 and 12 (Figure 3.7). To investigate the potential AggR-binding site identified at the *aafD* promoter, point mutations were made at positions 10 and 12 of the AggR-binding site (Figure 3.7). These positions of the AggR-binding site correspond to positons 90 and 92 on the *aafD*96 promoter fragment (Figure 3.8A and B). The recombinant plasmids pRW50/*aafD96-90C* and pRW50/*aafD*96-92*C*90*C* were made by substituting the Ts with Cs at positions number 90 and 92. The plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR* or pBAD24. The cells were grown in LB

Figure 3.7 Alignment of the AggR-binding site of the *aafD* regulatory region with consensus sequence

A. The panel the illustrates consensus sequence from Rns-binding logo by (Munson, 2013). The height of each letter corresponds to number of times each letter is present in the 31 Rns-binding sites identified in the ETEC genome.
B. The panel shows the sequence of the AggR-binding site identified at the *aggR* promoter of EAEC 042 (Morin *et al.*, 2010).
C. The panel shows the AggR-binding site from the *aafD* promoter sequence. The highlighted sequence shows the AggR-binding site (from position 3 to position 13) suggested to be important for AggR-binding. The flanking sequence of AggR-binding site from *aafD* is also shown.

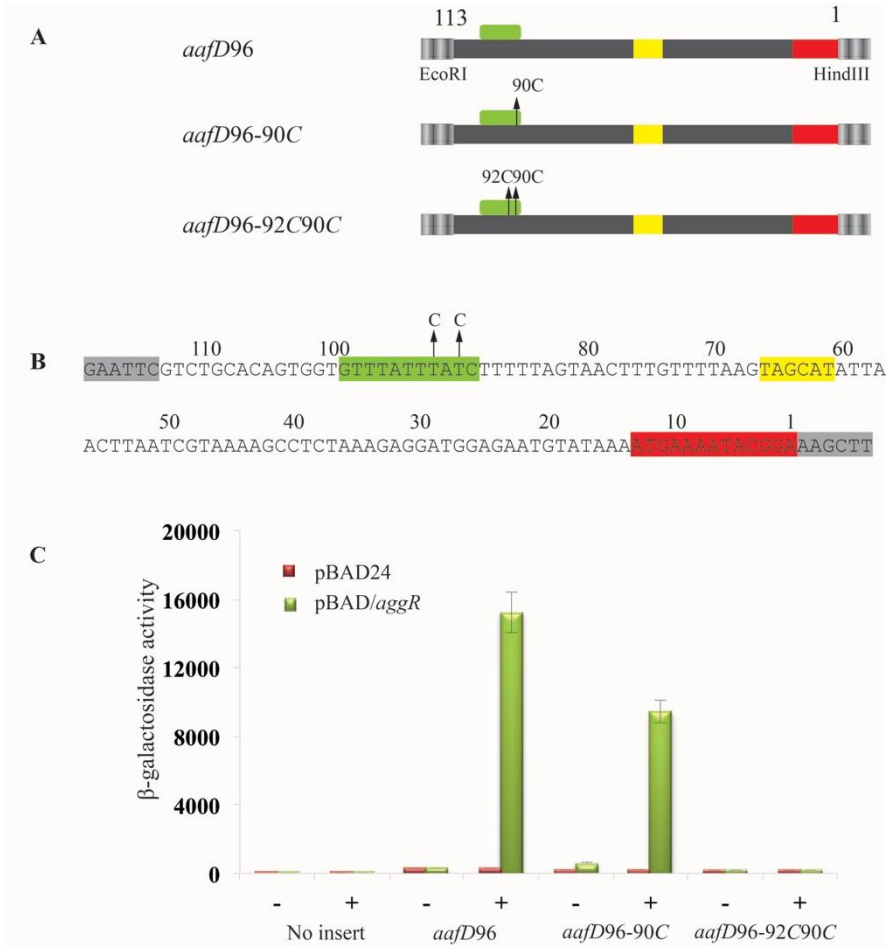


Figure 3.8 Mutational analysis of AggR-binding site at *aafD* promoter.

A. The figure illustrates *aafD*96, *aafD*96-90*C* and *aafD*96-92*C*90*C* promoter fragments. The grey bars represent upstream sequence of *aafD* and red bars represent *aafD* encoding sequence. Green boxes illustrate the potential binding sites of AggR and yellow boxes represent potential -10 hexamer element. The arrows with text indicate the positions of the point mutation and the base that has been substituted. This diagram is not to scale.
B. The panel shows the base sequence of the *aafD*96 fragment. The AggR-binding site is highlighted in green, the potential -10 element yellow, the restriction sites grey and the *aafD* sequence is red. The arrows indicate the positions of point mutations and the base that has been substituted.
C. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing *aafD* promoter derivatives cloned into the *lacZ* expression vector, pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in the presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6) with and without 0.2% (w/v) arabinose to express *aggR*. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 3.8C show that the β-galactosidase activities measured in cells containing pRW50/*aafD*96-90*C* is decreased slightly when compared to the starting plasmid pRW50/*aafD*96 (8000 units and 15500 units, respectively). No substantial AggR-dependant expression was observed in the cells containing pRW50/*aafD*96-92*C*90*C* in presence of AggR, in comparison to the AggR negative control (193units and 165 units, respectively) (Figure 3.8C). These results show that we have identified the AggR-binding site correctly and the T residues at position 90 and 92 are important for AggR-binding and activation. These results show that mutations are made in the AggR-binding site essential for AggR-dependent induction of *aafD* promoter. Moreover, these two thymine bases are likely important for AggR-binding and AggR-dependent transcription regulation.

Morin *et al.* (2010) identified the AggR-binding site sequence at the *aggR* promoter and the AggR-binding site identified at the *aafD* promoter is very similar. However, the binding site of the *aafD* promoter has G at position 3 instead of A (Figure 3.7), suggesting that other bases may be tolerated at this position. In order to identify the importance of this base, a series of mutant promoter fragments were made in which each of the 3 alternate bases were introduced at this position and the effect on AggR-dependent regulation was investigated (Figure 3.9A and B). Recombinant plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6). The cells were grown in LB with and without 0.2% (w/v) arabinose to express the *aggR*. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The results in

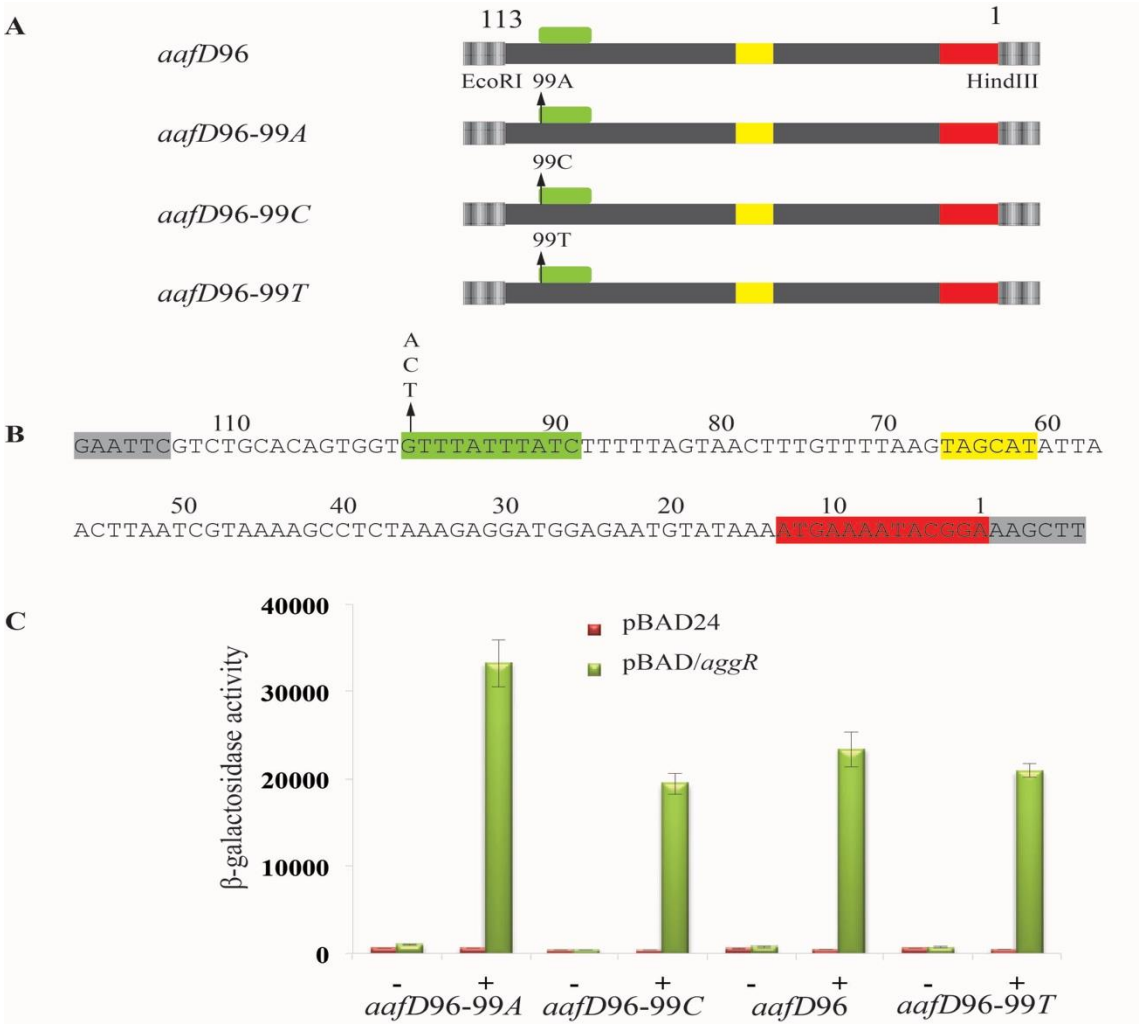


Figure 3.9 Mutational analysis of the AggR-binding site at the *aafD* promoter.

A. The figure illustrates the *aafD*96, *aafD*96-99*A*, *aafD*96-99*C* and *aafD*96-99*T* promoter fragments. The grey bars represent upstream sequence of *aafD* and red bars represent coding sequence of *aafD*. Green boxes illustrate the potential binding sites of AggR and yellow boxes represent the potential -10 hexamer element. The arrow with the text indicates the position of point mutations and the base that has been substituted. This diagram is not to scale.

B. The panel shows the base sequence of the *aafD*96 fragment. The AggR-binding site is highlighted green, the potential -10 hexamer element is yellow, the restriction sites grey and the coding sequence from *aafD* red. The arrow with the text indicates position of point mutation and the base that has been substituted to make point mutation.

C. The panel illustrates β-galactosidase activity measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing *aafD* promoter derivatives cloned into *lacZ* expression vector, pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

Figure 3.9C show that β-galactosidase activities measured in cells containing these plasmids have only two fold increase in activity in presence of AggR, when G was changed to A. Cells containing other mutants, had a similar β- galactosidase activity in presence of AggR to the starting plasmid, pRW50/*aafD*96. These results show that other bases can be tolerated at this position and suggest this pasition of the AggR-binding site may not play a crucial role in AggR-dependant promoter regulation.

3.2.4 Mapping the transcription start site of the *aafD* promoter by primer extension

To investigate the *aafD* promoter further, we carried out transcription start site mapping, in this instance using the *aafD*96 promoter fragment cloned into pRW50. The plasmid was transformed into *E. coli* K-12 strain BW25113 either containing pBAD/*aggR* or pBAD24. Cells were grown in LB medium with 0.2% (w/v) arabinose in triplicate cultures. The messenger RNA (mRNA) was collected in mid-logarithmic phase and purified. A ^{32}P labelled primer that anneals downstream of the HindIII site in pRW50 was used to prime the RNA template for DNA synthesis, resulting in extension of the primer to the position corresponding to the 5′ end of the mRNA. An RNA dependant DNA polymerase was used to copy the messenger into DNA. The transcription start site was determined from the length of the extended product. Reaction mixes after extension were loaded on to a 6% (w/v) denaturing polyacrylamide gel alongside a M13 reference genome sequence, which was used as a sequence ladder (Figure 3.10A).

Results in Figure 3.8B show that the primer extension product from the *aafD* promoter fragment produced two bands that are separated by 2 bp and correspond to the 56 and 54 bp of *aafD*96 fragment, respectively. The length of these products were calculated by comparing the bands to corresponding base of calibrated M13mp18 reference genome sequence (Figure 2.29). The two bands in the primer extension reaction can be explained by presence of two

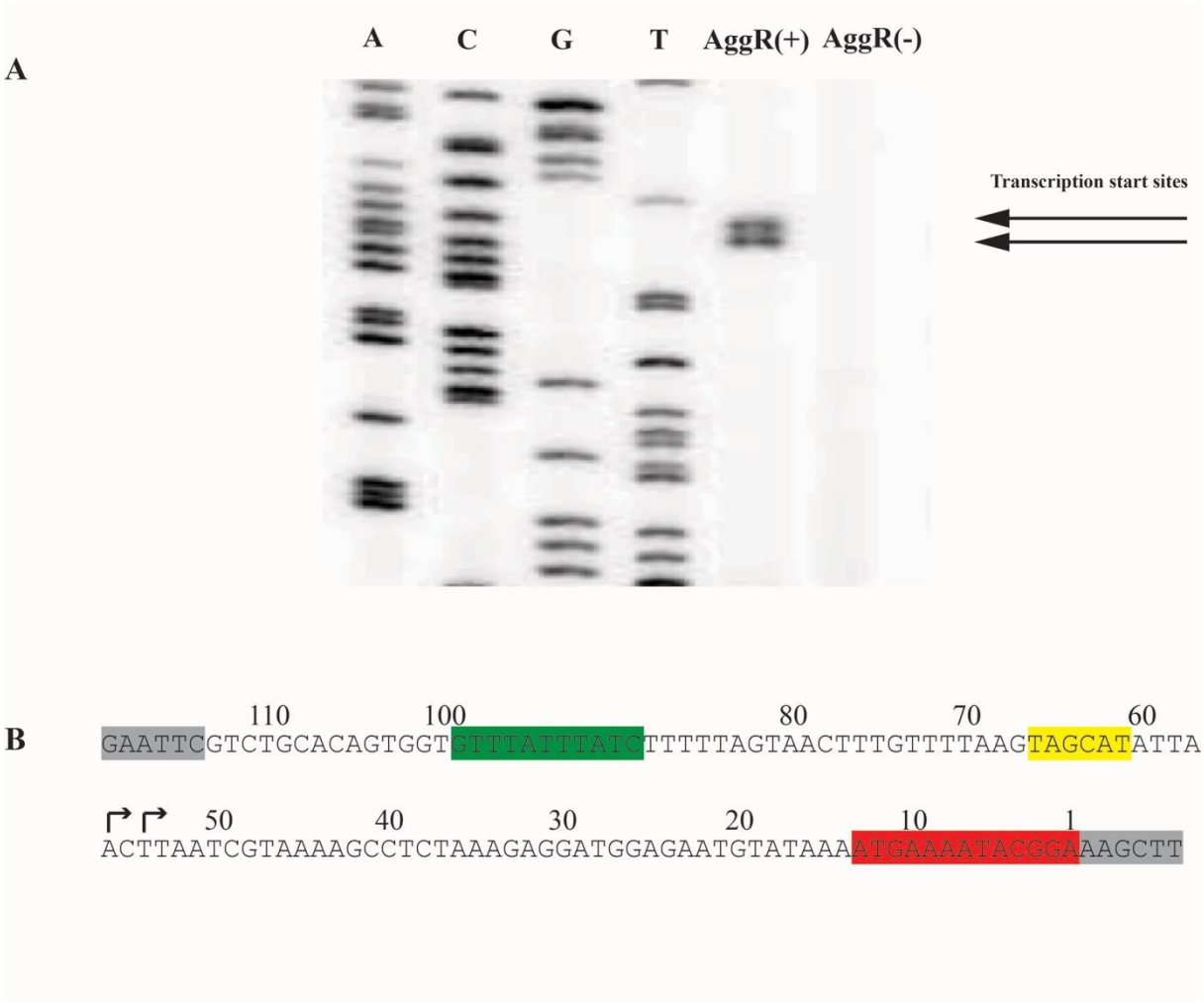


Figure 3.10 Mapping the transcription start site of the *aafD* promoter

A. The panel shows an autoradiogram of the gel run to determine the extension products from the *aafD* promoter calibrated with Phage M13 reference sequence (A, C, G and T) which serves as nucleotide ladder in this reaction. AggR (+) and AggR (-) represent presence of pBAD/*aggR*, and pBAD24, respectively. Arrows show the bands of the primer extension products in the presence of AggR. The length of these bands was calculated by using the calibration sequence.

B. The panel shows the sequence of the *aafD*96 fragment. The AggR-binding site is highlighted green, the potential -10 hexamer element yellow, the restriction sites grey and the sequence from *aafD* red. The bent arrows indicate the transcription start sites of the *aafD* promoter and are marked on the basis of the primer extension product length shown above. The M13mp18 reference sequence shown in Figure 2.29 is used to calculate the product length.

transcription start sites, the smaller product may be produced after post transcriptional processing of an original product (Figure 3.10B). Most probably the longer product reflects the transcription start site because it is at the most appropriate distance from a potential -10 element (TAGCAT). Note, there were no visible bands for AggR negative samples. This shows no transcriptional expression of *aafD* occurs in the absence of AggR (Figure 3.10) and the transcript is formed due to AggR-dependent induction of the promoter.

3.2.5 Identification of the -10 element of the *aafD* promoter using mutation analysis

A potential -10 hexamer element was identified on the basis of similarity to consensus sequence (TATAAT) and the distance from the transcription start site (Figure 3.10B). There are four out of six base matches to the consensus sequence in this potential -10 hexamer element. The potential -10 hexamer element was investigated by making point mutations at position 2 (corresponds to position 65 bp on *aafD*96 promoter fragment) and an A present at position 65 of the *aafD*96 was substituted with C to construct *aafD*96-65C promoter fragment, which was cloned into pRW50 (Figure 3.11A and B).

The plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac* carrying either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6). In the recombinant plasmid pBAD/*aggR*, *aggR* is cloned under the control of the *araBAD* promoter, so, 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 3.11C show that β-galactosidase activity measured in cells containing pRW50/*aafD*96-65*C* (contains a single bp mutation) resulted in the complete loss of AggR-dependent expression. This confirms that the -10 element has been predicted correctly and there is only one AggR-dependent *aafD* promoter.

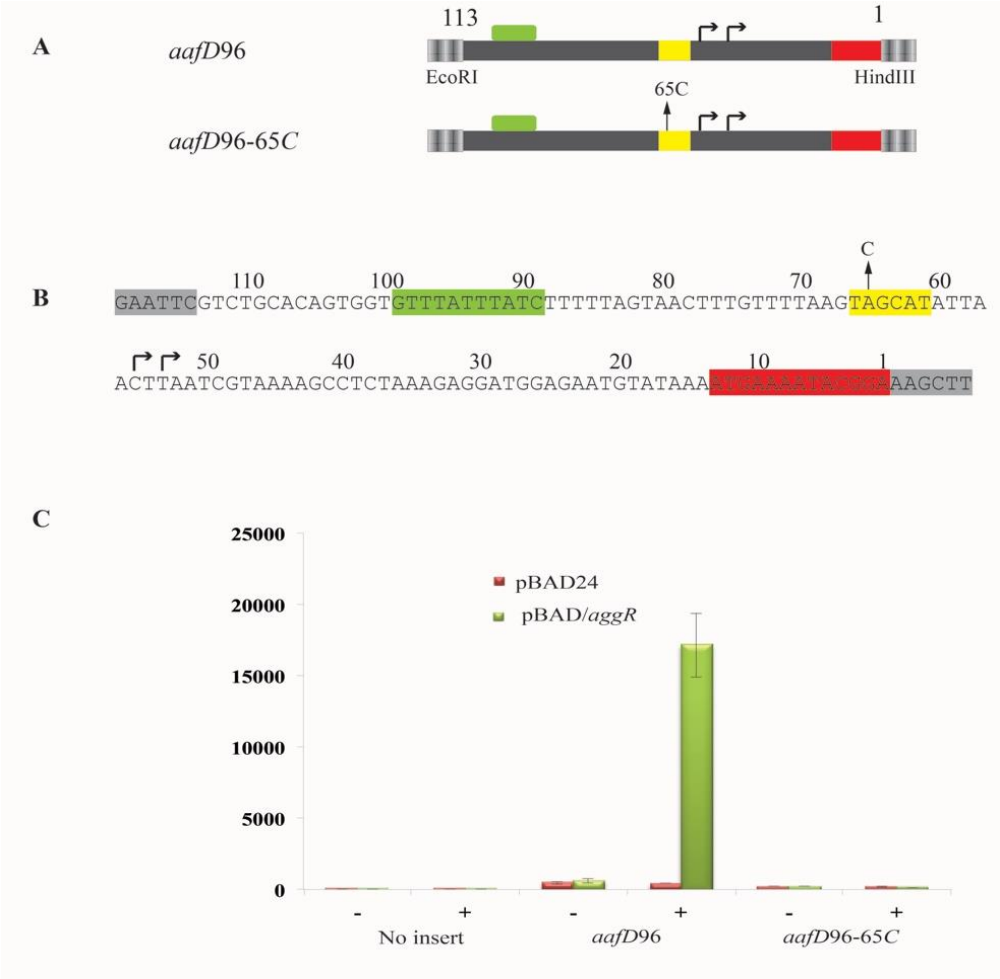


Figure 3.11 Identification of the -10 element of the *aafD* promoter

A. The figure illustrates the *aafD*96 and *aafD*96-65*C* promoter fragments The grey bars represent the upstream sequence of *aafD* and red bars represent the coding sequence of *aafD*. Green boxes illustrate the binding sites of AggR and yellow boxes represent the -10 hexamer element. The arrow indicates the position of the point mutation and the base that has been substituted. The bent arrows indicate the transcription start sites of the *aafD* promoter. This diagram is not to scale.

B. The panel shows the sequence of the *aafD*96 promoter fragment. The AggR-binding site is highlighted green, the -10 hexamer element yellow, the restriction sites grey and the sequence of *aafD* red. The arrow indicates the position of the point mutation introduced into the fragment and the base that has been substituted. The bent arrows indicate transcription start sites of the *aafD* promoter.

C The panel illustrates β-galactosidase activity measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aafD* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

3.2.6 Identification of the translation start site of the *aafD* operon reading frame

Bioinformatic analysis of the sequence upstream of *aafD* open reading frame also showed that it contains two possible start codons (in frame ATGs) for *aafD* translation. Previously, Elias *et al.* (1999) predicted the promoter and translation start site of *aafD,* whilst Chaudhuri *et al.* (2010) proposed another translation start site. Therefore, to investigate this, the *aafD*100 promoter fragment was modified to generate an in frame translational fusion. The translation of an mRNA transcript into protein usually starts on the ATG codon, and, in the *aafD* gene, two in-frame ATG codons are just 6 bp apart (Figure 3.12A and B). Elias *et al.*, (1999) predicted the upstream ATG as the translational start codon, whilst Chaudhuri *et al.* (2010) reported that the start codon for *aafD* is two codons downstream. To resolve this, two *aafD* translational fusions, *aafD*1001 and *aafD*1002*,* were made from *aafD*100 promoter fragments by deleting 10 bp and 19 bp, respectively adjacent to HindIII site (Figure 3.12A). These derivatives with either both upstream and downstream ATGs (*aafD*1001) or upstream ATG alone (*aafD*1002) were cloned into pRW225 (Figure 2.17). Plasmid pRW225 is a *lacZ* fusion vector, which lacks a translational start codon and ribosomal binding site (RBS) for *lacZ* gene. Thus, translation depends upon the start codon and RBS of the cloned fragment. Hence, β-galactosidase production indicates that translation has occurred. The results in Figure 3.12C show that the β-galactosidase level was considerably higher when the downstream ATG was fused in-frame to *lacZ i.e.* pRW225/*aafD*1001 when compared to the upstream fusion *i.e.* pRW225/*aafD*1002 (22443 units and 269 units in presence of AggR, respectively). These results indicate that the downstream ATG codon encoded by *aafD*1001 is the translation start codon for *aafD* and that the translational start site predicted by Elias *et al.* (1999) was incorrect.

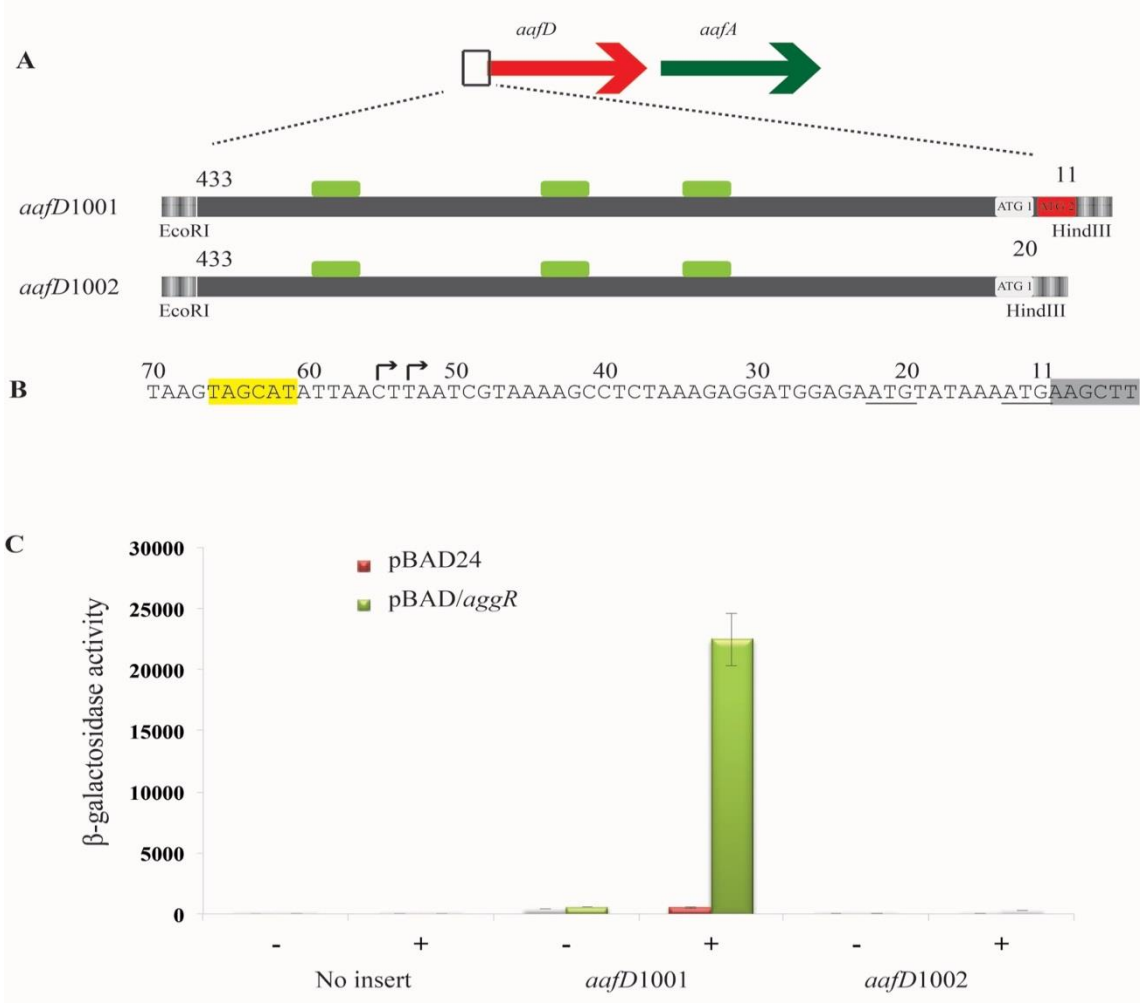


Figure 3.12 Identification of translational start site of the *aafD* open reading frame

A. The upper line shows the arrangement of the genes of Region 1 on pAA2. The lower part shows two derivatives of the *aafD*100 promoter fragment: *aafD*1001 contains the two in frame ATG (start codons) and *aafD*1002 contains only one in frame ATG codon. These promoter derivatives were cloned into pRW225 in such a way that the ATG codons from each derivative serve as translational start codons for *lacZ* on pRW225. The green boxes show potential AggR-binding sites and the red highlighted ATG is from *aafD* coding sequence. The diagram is not to scale.

B. The panel shows the sequence from 3' end of the *aafD*1001 promoter fragment that was cloned in pRW225 using HindIII sites. The grey highlighted sequence is the restriction site and yellow is the -10 hexamer element. The bent arrows indicate transcription start sites and underline sequences show the in-frame ATG codons.

C The panel illustrates β-galactosidase activity measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW225) or *aafD* promoter derivatives cloned into pRW255. The cells also carried either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

3.3 Transcriptional regulation analysis of the Region 2 of pAA2

3.3.1 Analysis of the upstream DNA sequences of *afaB* and *aafC*

Region 2 of plasmid pAA2 encodes two genes (*aafC* and *aafB*) and a pseudogene (*afaB*). The Region 2 genes and the pseudogene are likely to be transcribed as an operon and to identify the location of the operon promoter for Region 2, approximately 400 bp of DNA upstream of the *afaB* pseudogene was cloned into pRW50. This sequence carries potential AggR-binding sites, as judged by comparing to the proposed AggR-binding sequence suggested by Morin *et al.* (2010) (Figure 3.13). The schematic diagram of the *afaB*100 promoter fragment is shown in Figure 3.14A.

The recombinant plasmid, pRW50/*afaB*100 was transformed into *E. coli* K-12 strain BW25113 Δ*lac* with either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6) and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activity was determined by measuring β-galactosidase levels in lysates of cells containing the plasmid. Results in the Figure 3.14B show that β-galactosidase activity measured in cells containing pRW50/*afaB*100 increased almost 20-fold in the presence of AggR (pBAD/*aggR*) as compared to AggR negative cells (pBAD24) (558 and 25 units, respectively). These results show that there is an AggR-dependent promoter upstream of pseudogene *afaB*.

Elias *et al.* (1999) previously demonstrated that operon promoter for Region 2 was present upstream of *aafC* instead of *afaB*. To investigate this, 410 bp of DNA sequence upstream of *aafC* was cloned into pRW50 (Figure 3.15). Bioinformatics analysis also showed that this fragment carried a potential AggR-binding site on the reverse strand. The schematic diagram of the *afaC*100 promoter fragment is shown in Figure 3.16A. Plasmid pRW50/*aafC*100 was

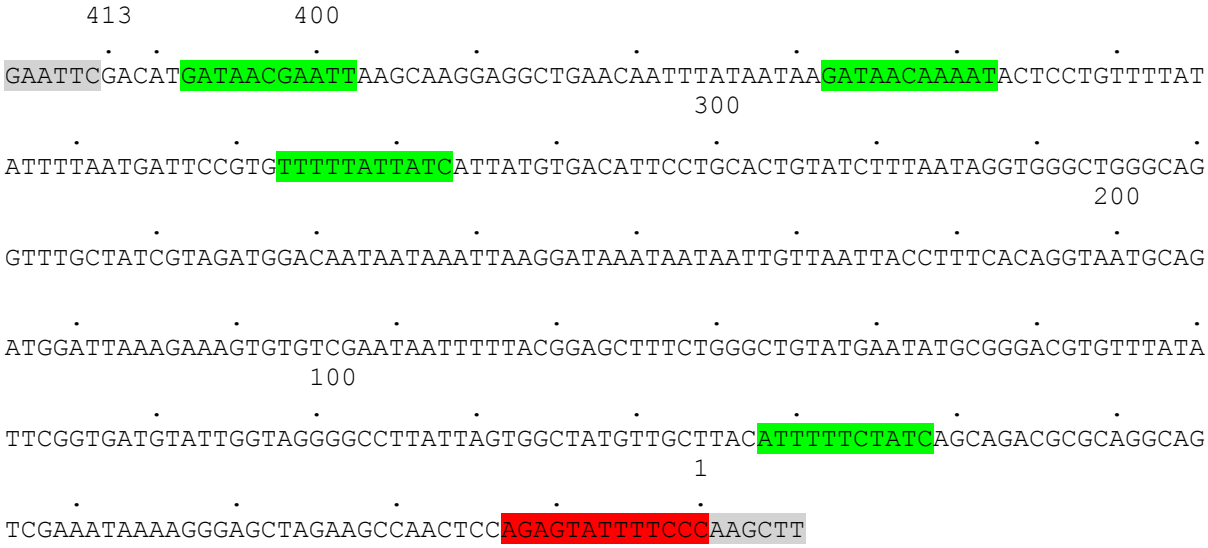


Figure 3.13 The DNA base sequence upstream of the pseudogene *afaB*

The figure shows the 413 bp DNA sequence of the *afaB*100 promoter fragment. The sequence from *afaB* pseudogene has been highlighted red and the potential AggR-binding sites are green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.2.

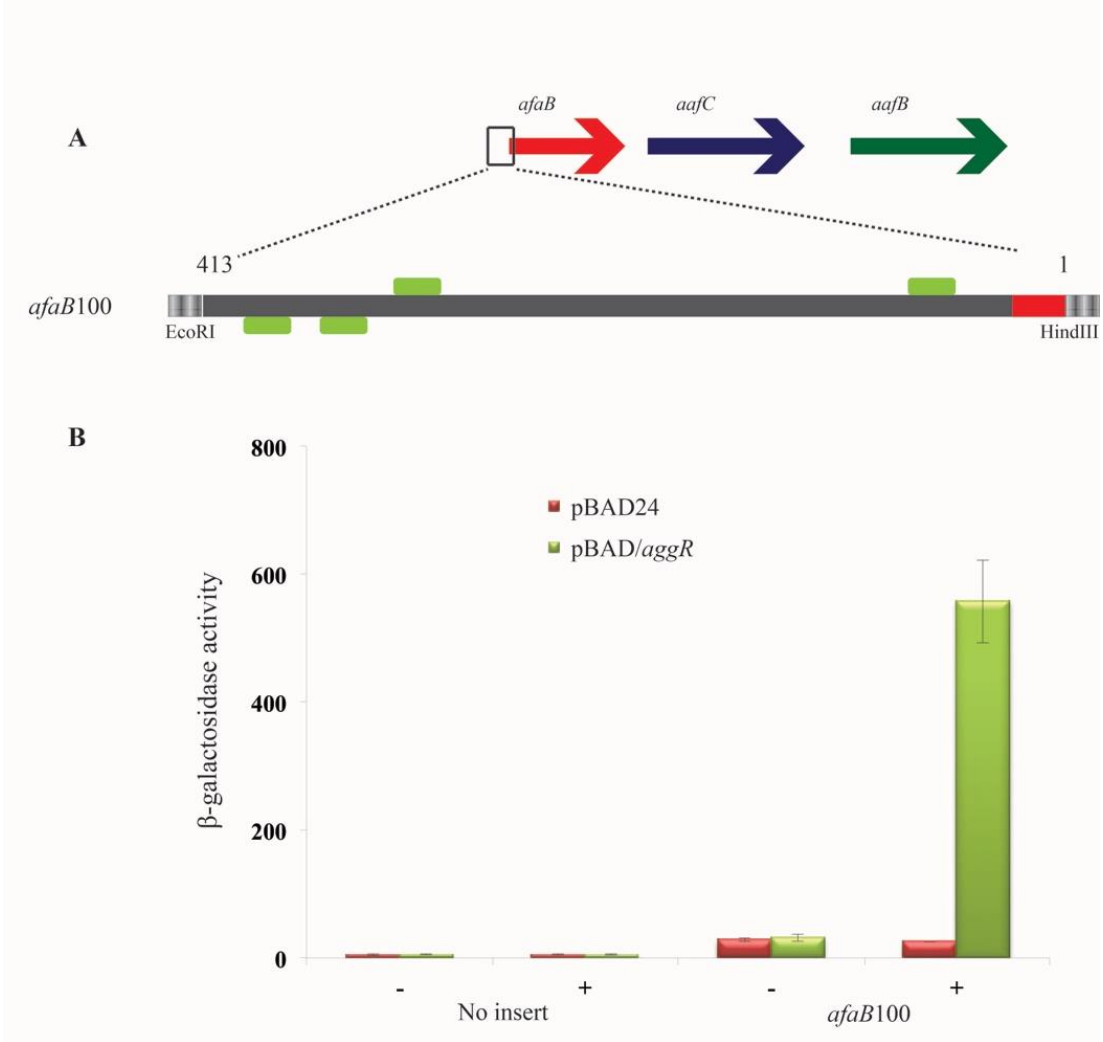


Figure 3.14 Identification of the *afaB* promoter from Region 2

A. The upper line shows the schematic arrangement of the genes on pAA2 Region 2. The lower part illustrates the *afaB*100 promoter fragment, the grey bar represents upstream sequence of *afaB* and the red bar represents 13 bp sequence of *afaB* pseudogene present in the promoter fragment. The potential AggR-binding sites are indicated as green boxes. The DNA sequence in this diagram is numbered above fragment from 1 to 413 and the fragment was cloned using EcoRI and HindIII restriction sites into pRW50. This diagram is not to scale.
B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or the *afaB*100 promoter fragment cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

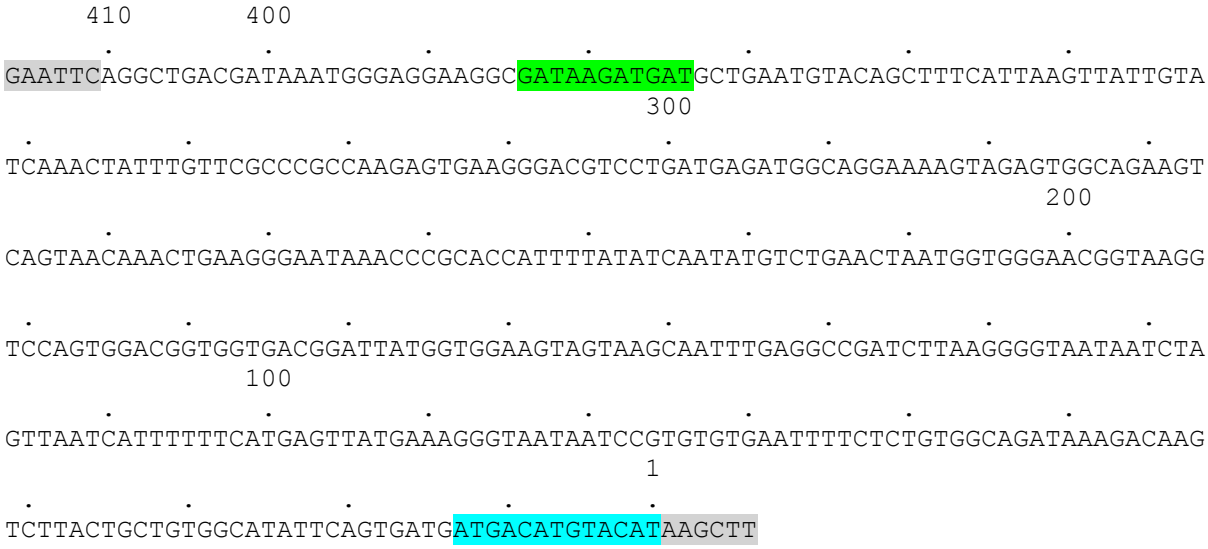


Figure 3.15 The DNA base sequence upstream of the *aafC* gene

The figure shows the 413 bp DNA sequence of the *aafC*100 promoter fragment. The open reading frame sequence is highlighted blue and the potential AggR-binding site is green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.2.

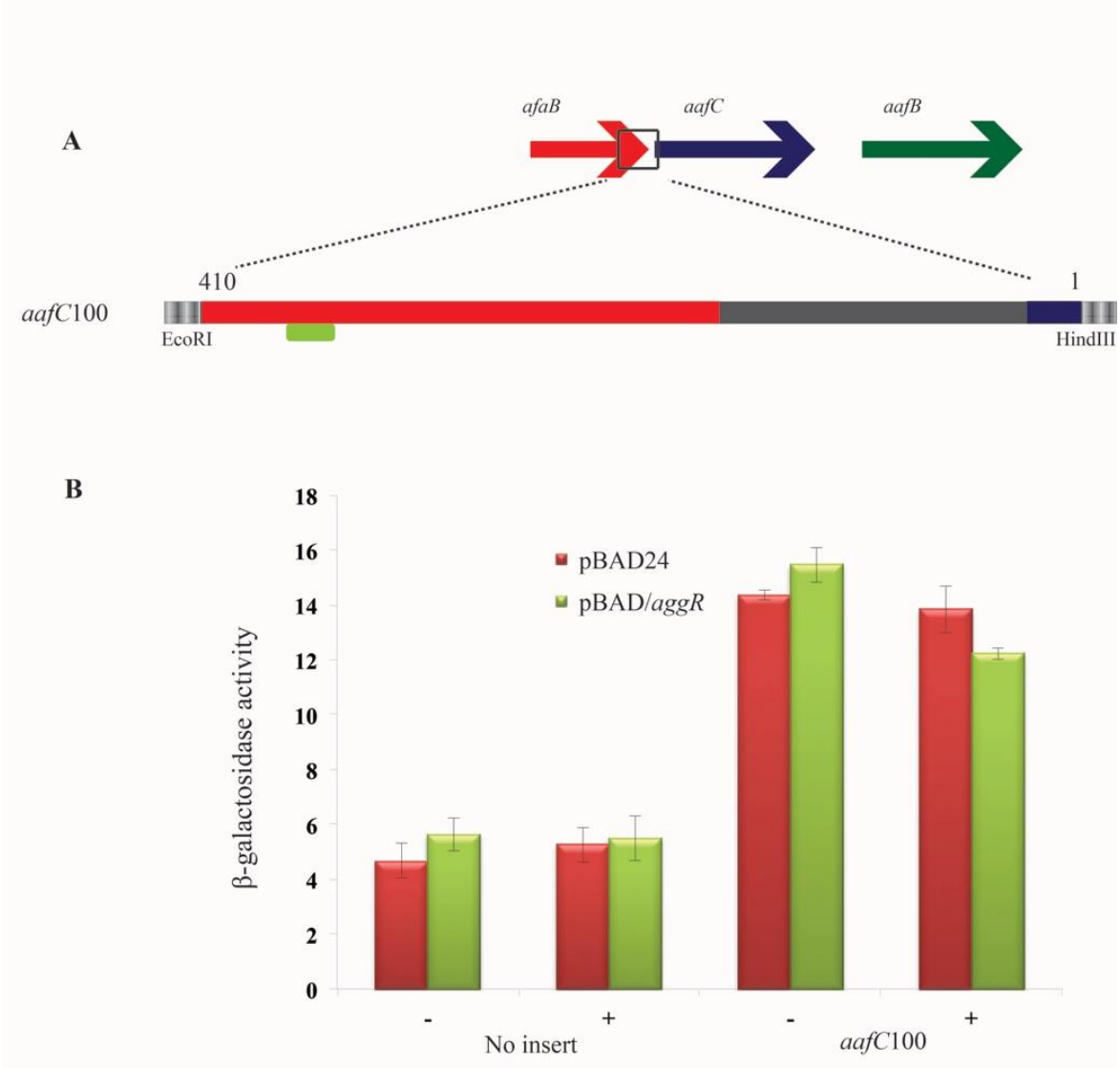


Figure 3.16 Analysis of the DNA sequence directly upstream of *aafC*

A. The upper line shows a schematic arrangement of genes on pAA2 around Region 2. The red bar represents the sequence from *afaB* pseudogene, blue bar represents *aafC* encoding sequence and grey bar represents sequence upstream of *aafC*. The potential AggR-binding site is indicated as green boxes. The DNA sequence in this diagram is numbered above fragment from 1 to 410 and the fragment was cloned using EcoRI and HindIII restriction sites into pRW50. This diagram is not to scale.
B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aafC*100 promoter fragment cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

transformed into *E. coli* K-12 strain BW25113 Δ*lac,* with either pBAD/*aggR* or pBAD24. The β-galactosidase activity measured in cells containing pRW50/*aafC*100 showed very low levels (15 units) and activity was not increased when AggR expression was induced (Figure 3.16B). This shows that there is an extremely weak promoter present upstream of *aafC* which is not AggR-regulated.

These results show that an AggR-regulated promoter is located on the *afaB*100 promoter fragment and there is no AggR-regulated promoter on the *aafC*100 promoter fragment. Thus, we conclude that an AggR-regulated promoter is present upstream of *afaB* and the proposed operon promoter of Region 2 is present upstream of *afaB* rather than of *aafC*.

3.3.2 Mapping of gene regulatory region on Region 2

The results above indicate that there is a promoter upstream of the *afaB* pseudogene, which directs transcription into Region 2 and this presumably transcribes the entire operon. This contradicts the results of Elias *et al.* (1999) which suggested that the promoter was present directly upstream of *aafC* and not *afaB*. To investigate, if there is another promoter directly upstream of *aafC*, two large promoter fragments *afaBC*100 and *afaBC*99 were cloned into pRW50. The *afaBC*100 fragment was made by cloning all the DNA sequence starting from the 5' end of the *afaB*100 promoter fragment to 3' end of *aafC*100 fragment (Figure 3.17). The other construct *afaBC*99 was made by cloning the DNA sequence starting from upstream of 5' end of *afaB* to 3' end of *aafC*100 fragment (Figure 3.18A). The recombinant plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac* with either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6) and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids.

***afaBC*100 (1088 bp)**

GAATTCGACATGATAACGAATTAAGCAAGGAGGCTGAACAATTTATAATAAGATAACAAAATACTCCTGTTTTAT

ATTTTAATGATTCCGTGTTTTTATTATCATTATGTGACATTCCTGCACTGTATCTTTAATAGGTGGGCTGGGCAG

GTTTGCTATCGTAGATGGACAATAATAAATTAAGGATAAATAATAATTGTTAATTACCTTTCACAGGTAATGCAG

ATGGATTAAAGAAAGTGTGTCGAATAATTTTTACGGAGCTTTCTGGGCTGTATGAATATGCGGGACGTGTTTATA

TTCGGTGATGTATTGGTAGGGGCCTTATTAGTGGCTATGTTGCTTACATTTTTCTATCAGCAGACGCGCAGGCAG

TCGAAATAAAAGGGAGCTAGAAGCCAACTCCAGAGTATTTTCCCTTCATCTGGGCGCAACGCGTGTGGTGTACTC

TAGCTCATTCGGAGAGGTATTGGCAGTCATTAATGATCAAAATTATTCAATGCAGGTTCAGGCTGAAGTTTATCA

GAGGACCGGAAGAGTATAGCCCCTTTTGTGGTTACTCCACTGCAATTTCGCCTTAATGGTCTGCAATCCAGTCGC

CTCCCGAATTGTCCGCACAGGAGGGGATTTTTCGGTGGACCGAGAGAGTCTTCAATGGATTTGTATAAAAGGAAT

CCCCCCCAAAGGCTGACGATAAATGGGAGGAAGGCGATAAGATGATGCTGAATGTACAGCTTTCATTAAGTTATT

GTATCAAACTATTTGTTCGCCCGCCAAGAGTGAAGGGACGTCCTGATGAGATGGCAGGAAAAGTAGAGTGGCAGA

AGTCAGTAACAAACTGAAGGGAATAAACCCGCACCATTTTATATCAATATGTCTGAACTAATGGTGGGAACGGTA

AGGTCCAGTGGACGGTGGTGACGGATTATGGTGGAAGTAGTAAGCAATTTGAGGCCGATCTTAAGGGGTAATAAT

CTAGTTAATCATTTTTTCATGAGTTATGAAAGGGTAATAATCCGTGTGTGAATTTTCTCTGTGGCAGATAAAGAC

AAGTCTTACTGCTGTGGCATATTCAGTGATGATGACATGTACATAAGCTT

Figure 3.17 The DNA base sequence upstream of the *aafC* and *afaB* pseudogene

The figure shows the 1088 bp DNA sequence of the promoter fragment containing *aafC* sequence shown in blue, upstream of *aafC* and *afaB* shown in black and *afaB* sequence shown in red. **T**he potential AggR-binding sites are green highlighted green and the restriction sites EcoRI and HindIII are grey. This figure is adapted from Figure 2.8.

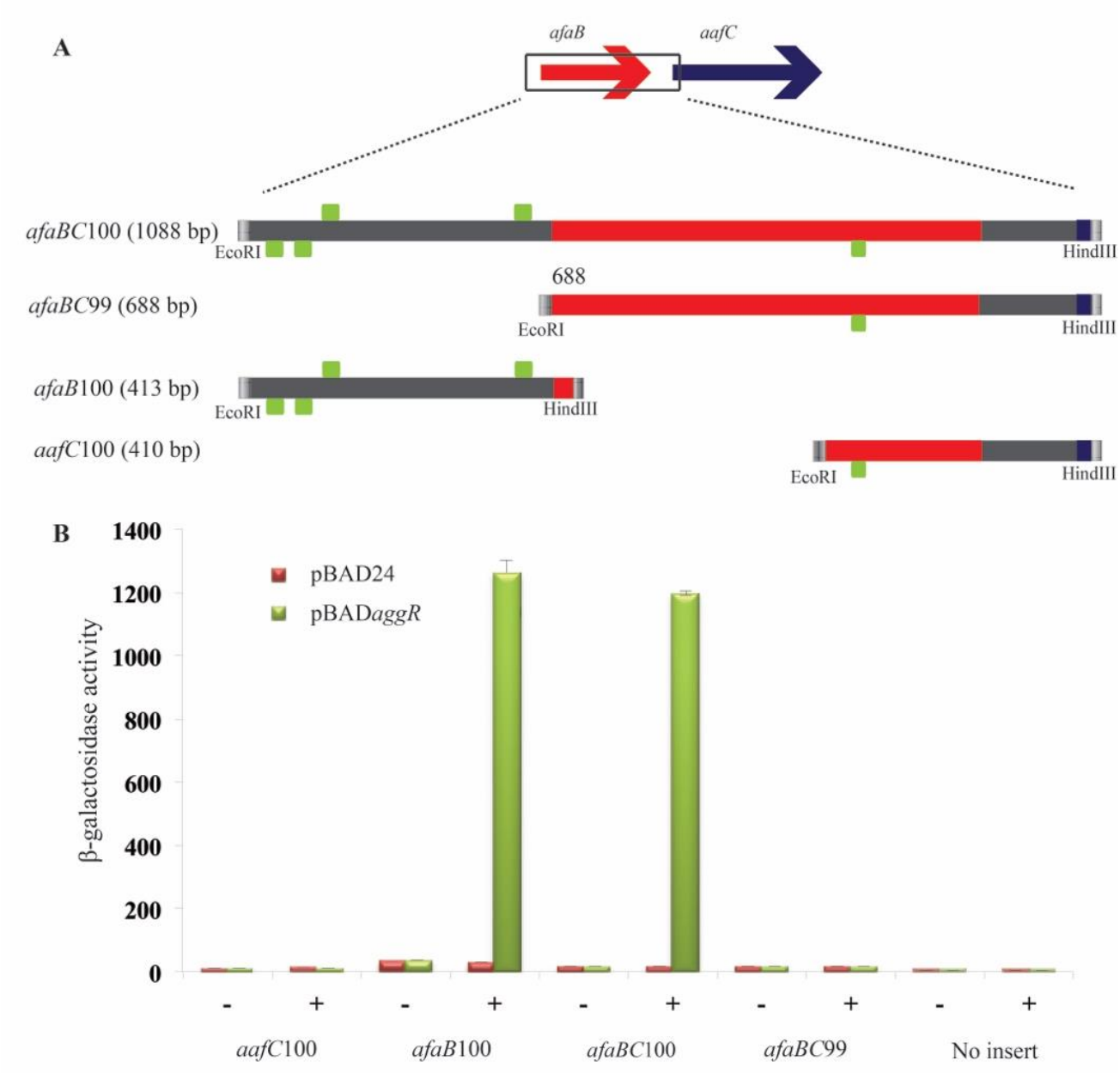


Figure 3.18 Analysis of the Region 2 operon promoter

A. The upper line shows the schematic arrangement of genes on pAA2 around Region 2. The lower part illustrates upstream sequence of *aafC* with potential AggR-binding sites as green boxes. The red bars represent the sequence from *afaB* pseudogene, blue bars represent *aafC* encoding sequence and grey bars represent sequence upstream of *afaB* pseudogene and *aafC*. These fragments were cloned using EcoRI and HindIII restriction sites into pRW50.
B. The panel illustrates the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or one of *aafC*100, *afaBC*100, *afaBC*99 and *afaB*100 promoter fragments cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

The β-galactosidase activities measured in cells containing these fragments cloned into pRW50 show that pRW50/*afaBC*99 has very low or no promoter activity and expression was not induced by AggR. In cells containing pRW50/*afaBC*100 substantial promoter activity was detected and this was AggR-inducible. Furthermore, the β-galactosidase activity observed in cells containing pRW50/*afaBC*99 was comparable to the β-galactosidase activity measured in the cells containing pRW50/*aafC*100 and similarly the level of expression measured in cells containing pRW50/*afaBC*100 was comparable to activity observed in cells containing pRW50*aafB*100 (Figure 3.18B). These results confirm that the Region 2 operon promoter is upstream of *afaB* and it is regulated by AggR, whilst there is no AggR-dependent promoter directly upstream of *aafC.*

3.3.3 Identification of the minimal regulatory region of the *afaB*100 promoter necessary for AggR-mediated activation

As AggR was found to activate the *afaB* promoter substantially, three nested deletions were made, (*afaB*99 to *afaB*97), by removing 100 bp from 5' end of the *afaB*100 promoter fragment to locate the minimal promoter fragment induced by AggR (Figure 3.19A and B). The derivative promoter fragments (*afaB*99 to *afaB*97) were then cloned into the *lacZ* expression vector pRW50. The plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac* with either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in the lysates of cells containing these plasmids. Results in Figure 3.19C show that the β-galactosidase activity measured in cells containing pRW50/*afaB*100 was increased in the presence of AggR while the β-galactosidase activity observed in cells containing pRW50/*afaB*99 and the other promoter fragment derivatives

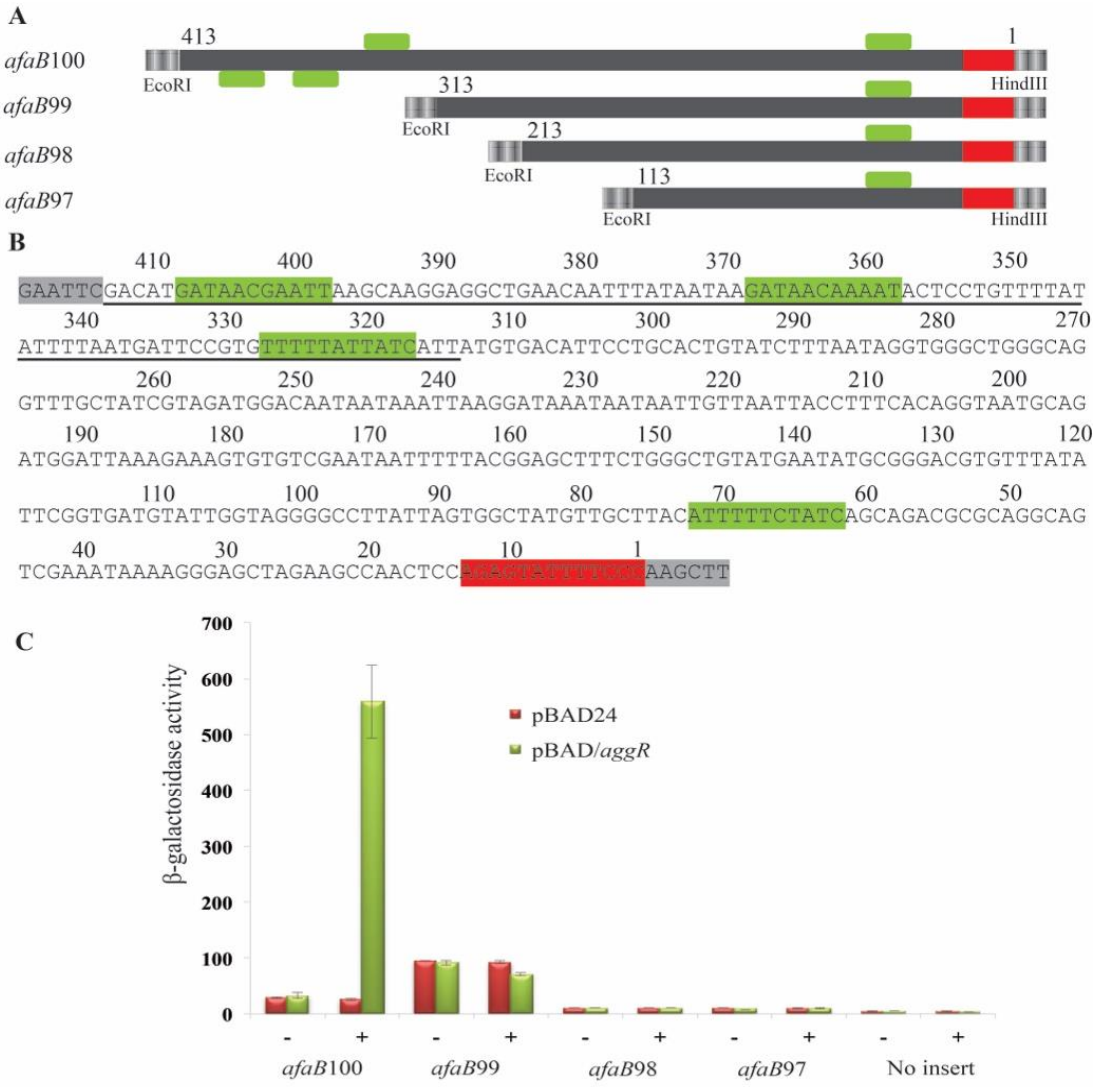


Figure 3.19 Deletion analysis of the *afaB* promoter

A. The diagram shows the *afaB*100*, afaB*99*, afaB*98 and *afaB*97 promoter fragments with the AggR-binding sites shown as green boxes on the forward and the reverse strand. The grey bar represents upstream sequence of *afaB* and the red bar represents 13 bp sequence of *afaB* pseudogene present in the promoter fragment. The base sequences are numbered from 1 to 413 on *afaB*100 promoter fragment and cloned using EcoRI and HindIII restriction sites into pRW50. This diagram is not to scale.

B. The panel shows the sequence of the *afaB*100 promoter fragment. The AggR-binding sites are highlighted green, the restriction sites grey and the sequence from *afaB* pseudogene red. The underlined sequence adjacent to EcoRI was deleted to construct the *afaB*99 fragment.

C The panel illustrates the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *afaB*100 promoter fragment derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

showed no increase in levels when AggR-expression was induced. This suggests that the AggR-binding site responsible for promoter induction has been deleted or disrupted in *afaB*99 fragment.

3.3.4 Identification of the AggR-binding site on the *afaB* promoter using mutation analysis

Sequence analysis of the *afaB* promoter sequence suggest that there were three potential AggR-binding sites between the *afaB*99 and *afaB*100 promoter fragments. There was one potential AggR-binding site on the forward strand and two potential AggR-binding sites on the reverse strand (Figure 3.19A and B). To investigate the AggR-binding site required for AggR-dependent regulation, point mutations were introduced into the conserved T residues of the AggR-binding site on the forward strand and this sequence was aligned with consensus sequence inFigure 3.20. In this *afaB*, AggR-binding site, the Ts at position 10 and 12 of AggR-binding site were substituted with Cs. These mutations in the AggR-binding site correspond to position 318 and 320 on the *afaB*100 promoter fragment and substitutions in the AggR-binding site were introduced to construct *afaB*100-320*C*318*C* fragment (Figure 3.21A and B).

The plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac* with either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The β-galactosidase activity measured in cells containing pRW50/*afaB*100-320*C*318*C* showed no increase when AggR expression was induced (Figure 3.21C). These results indicate that the AggR-binding site that has been disrupted is essential for AggR-dependent transcription activation at the *afaB* promoter.

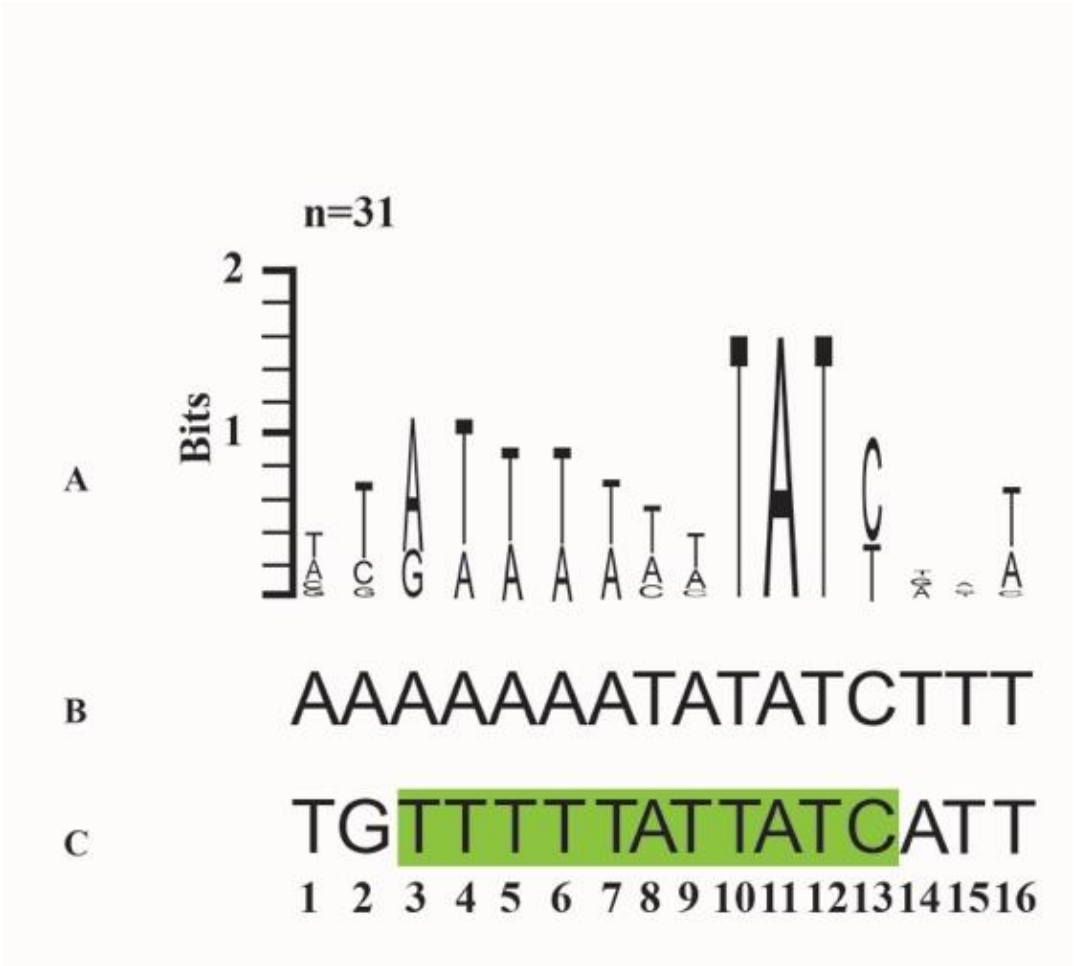


Figure 3.20 Alignment of the AggR-binding site of the *afaB* regulatory region with the consensus sequences of Rns and AggR

A. The panel illustrates consensus binding sequence for Rns, (Munson, 2013). The height of each letter corresponds to conservation of sequence in 31 Rns identified on ETEC genome.
B. The panel shows the sequence of the essential AggR-binding site identified on *aggR* promoter (Morin *et al.*, 2010).
C. The panel shows the AggR-binding sequence from *afaB* promoter. The green highlighted sequence shows the AggR-bindings sequence (from position 3 to position 13) which is suggested to be important for AggR-binding.

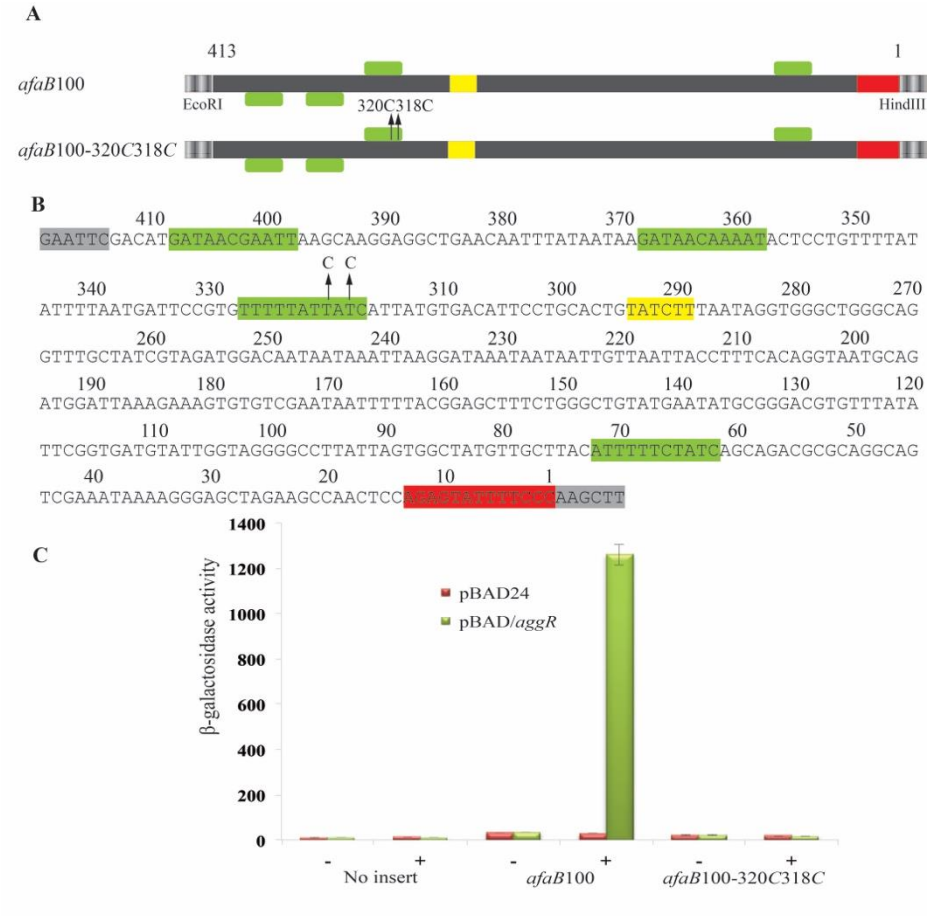


Figure 3.21 Mutational analysis of the *afaB100* promoter fragment

A. The figure illustrates the *afaB*100 and *afaB*100-320*C*318*C* promoter fragments. The grey bar represents upstream sequence of *afaB* and the red bar represents 13 bp sequence of *afaB* pseudogene present in the promoter fragment. Green boxes indicate the potential AggR-binding sites on the forward and reverse strands and the potential -10 hexamer is presented as yellow. These fragments were cloned into pRW50, using EcoRI and HindIII restriction sites. The arrows, with text, indicate the positions of point mutations and the bases that have been substituted to make these point mutations. This diagram is not to scale.
B. The panel shows the sequence of *afaB*100 fragment. The AggR-binding sites are highlighted green, the potential -10 hexamer element yellow, the restriction sites grey and the sequence from *afaB* red. The arrows with the text indicate position of the point mutations.
C The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *afaB* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

3.3.5 Mapping of the transcription start site of the *afaB* promoter by primer extension

To investigate the transcription start site of the *afaB* promoter, primer extension was used. In this method plasmid pRW50/*afaB*100 was transformed into *E. coli* K-12 cells, carrying either the pBAD/*aggR* vector or pBAD24. The cell cultures were grown in triplicate to mid-logarithmic phase in LB in the presence of 0.2% (w/v) arabinose with shaking at 37°C. The mRNA was collected in mid-logarithmic phase and purified. A ^{32}P labelled primer that anneals downstream of the HindIII site in pRW50 was used to prime DNA synthesis, resulting in the extension of the primer to the position corresponding to the 5′ end of the message, using RNA dependant DNA polymerase. The transcription start site was then determined from the length of the extended primer product.

Reaction mixes were loaded onto a 6% (w/v) denaturing polyacrylamide gel. The length of the products was determined by comparing the bands to the calibrated M13mp18 reference sequence (Figure 2.29 and Figure 3.22A). Results in Figure 3.22B show that the two bands correspond to the positions 280 and 266 on the *afaB*100 promoter fragment. These two bands could be explained by the presence of two transcription start sites or the smaller product may have been produced after the post transcription processing of the original product. In this study, only the large product is considered as a transcription start site due to its appropriate distance from the potential -10 hexamer element. There were no visible bands in the AggR negative samples indicating that no transcription has occurred in the absence of AggR.

3.3.6 Identification of the -10 element on the *afaB* promoter using mutational analysis

At the *afaB* promoter region a potential -10 hexamer element was identified on the basis of resemblance to the consensus sequence (TATAAT) and the distance from the major transcription start site seen in primer extension (Figure 3.23A and B). A potential -10 hexamer element (TATCTT) with a 4 out of 6 bp match to the consensus was present

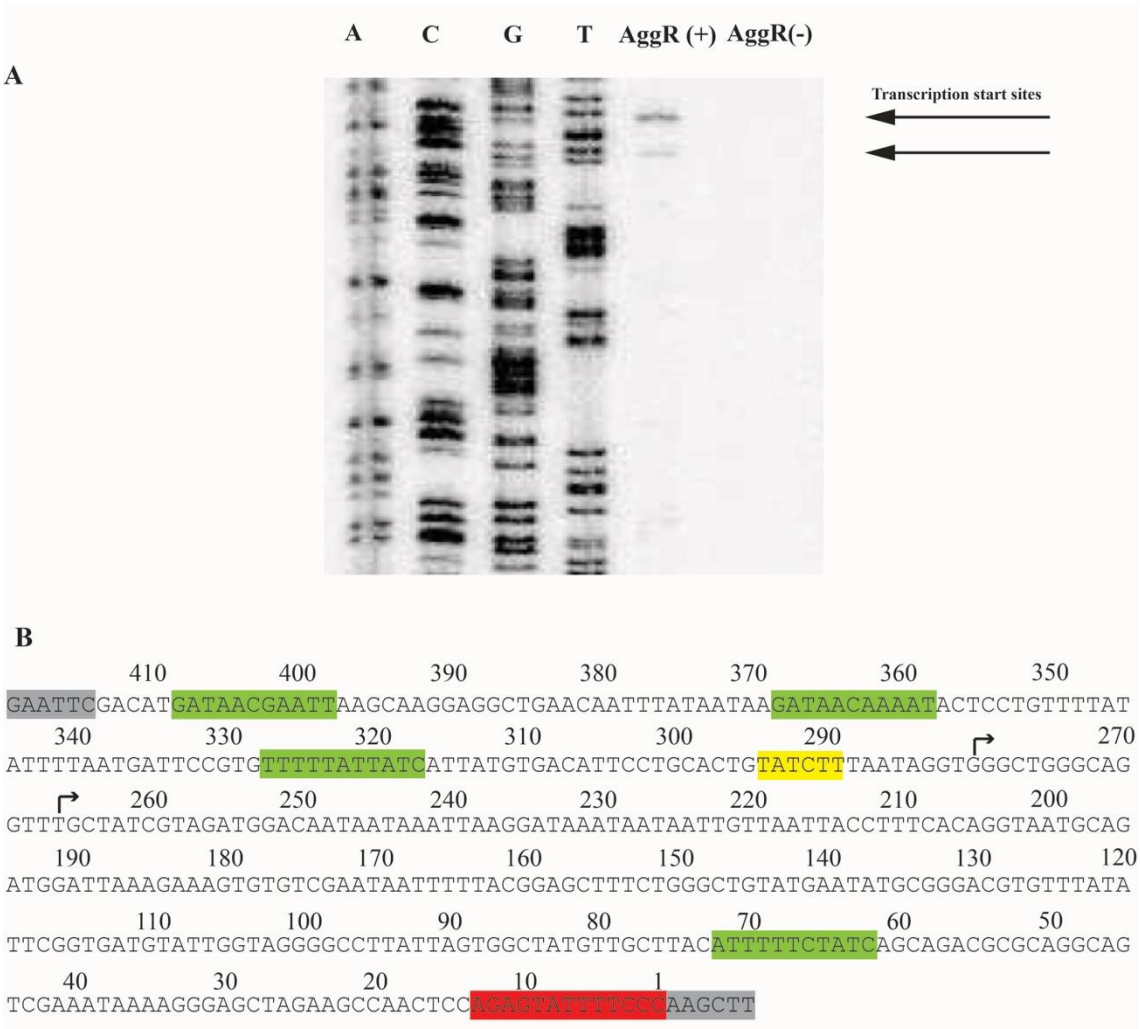


Figure 3.22 Analysis of the transcription start site of the *afaB* promoter

A. The panel shows an autoradiogram of the gel run to determine the extension product of the *afaB* transcript and is calibrated with Phage M13mp18 reference sequence. Phage M13 reference sequence serves as nucleotide ladder in this reaction (lane A, C, G and T). Arrows show the bands of the primer extension product produced in the presence of AggR. The length of these bands is calculated by corresponding base in the calibration sequence.
B. The panel shows the sequence of the *afaB*100 fragment. The AggR-binding site is highlighted green, the potential -10 hexamer element yellow, the restriction sites grey and the sequence from *afaB* pseudogene red. The bent arrows indicate the transcription start sites of *afaB* identified by primer extension product. The M13mp18 reference sequence shown in Figure 2.29 is used to calculate the product length.

Figure 3.23 Identification of -10 element of the *afaB* promoter

A. The figure illustrates the *afaB*100 and *afaB*100-293*C* promoter fragments. The grey bar represents upstream sequence of *afaB* and the red bar represents 13 bp sequence of *afaB* pseudogene present in the promoter fragment. Green boxes indicate potential AggR-binding sites on forward and reverse strands. These fragments were cloned into pRW50 using EcoRI and HindIII restriction sites. The arrow, with text, indicates positon of point mutation introduced. The bent arrows illustrate transcription start sites. This diagram is not to scale.
B. The panel shows the sequence of the *afaB*100 fragment. The AggR-binding sites are highlighted green, the potential -10 hexamer element is yellow, the restriction sites are grey and the sequence from *afaB* is red. The arrows, with the text, indicate position of the point mutation introduced. The bent arrows illustrate the transcription start sites.
C The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *afaB* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

upstream of the transcription start site. Therefore, a point mutation from A to C at position 2 of the -10 hexamer element was introduced into the *afaB*100 promoter fragment. This corresponds to position 293 on *afaB*100 promoter fragment and it resulted in the *afaB*100-293*C* promoter fragment (Figure 3.23A and. B). The plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac* carrying either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6) and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were then determined by measuring the β-galactosidase levels in lysates of cells containing these plasmids. The β-galactosidase levels measured in cells containing pRW50/*afaB*100-293*C* were lower when compared to the wild type promoter activity with this single base pair mutation (Figure 3.23C). This confirms that it is the -10 hexamer element of the *afaB* promoter and it is required for AggR-dependent expression.

3.3.7 Comparison of *aafD* and *afaB* promoters

Promoter elements of *aafD* and *afaB* have been identified in this study and β-galactosidase assays in cells containing pRW50/*aafD*96 and pRW50/*afaB*100 showed that the higher expression was observed in cells containing *aafD*96 promoter fragment in comparison to *afa*B100 promoter fragment in the presence of AggR (Figure 3.6 and Figure 3.14). However, to compare the expression levels of these promoters correctly the promoters must be cloned into a suitable vector. Therefore, to investigate, if there is any difference in promoters strength at the transcription level, *aafD* and *afaB* promoter fragments were cloned into pRW224, using EcoRI and HindIII sites. This plasmid contains its own SD and translational start codon for *lacZ* and thus, any difference in β-galactosidase activity would be completely due to transcription regulation. As the *aafD* promoter fragment already contains a translational start codon, a derivative of the promoter was made by removing 43 bp from adjacent to HindIII

site, *i.e.* the *aafD*500 fragment. As *afaB* is a pseudogene and does not contain a translational start codon, the *afaB*100 promoter fragment was cloned into pRW224 (Figure 3.24A and B). The plasmids were transformed into *E. coli* K-12 strain BW25113 Δ*lac* carrying either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD_{650}=0.4-0.6) and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The β-galactosidase levels measured in cells containing pRW224/*aafD*500 showed an 18-fold higher activity in comparison to the cells containing pRW224/*afaB*100 with AggR-induction (Figure 3.24C). These results show that the *aafD* promoter is a consdierably stronger promoter than of the *afaB* promoter.

3.4 Discussion

EAEC is a pathogen and it has many virulence factors. The first step in the disease process is the attachment of EAEC to the intestinal cells of the human gut, which is mediated by AAF. The expression of the fimbrial genes and many other virulence determinants is under the control of AggR (Elias *et al.*, 1999; Morin *et al.*, 2013). In this study, the fimbrial promoters have been located in both Region 1 and Region 2 and the AggR-binding sites important for AggR-dependent regulation, have been identified at these promoters.

This current work found that the fimbrial encoding genes in Region 1 are transcribed by a single operon promoter, present upstream of *aafD,* which is regulated by AggR. The study by Elias *et al.* (1999) had predicted an AggR-dependent promoter upstream of *aafD* but the promoter elements identified in this present work are different. In Region 2 the fimbrial genes of EAEC 042 have an unusual arrangement, with a pseudogene *afaB* upstream followed by

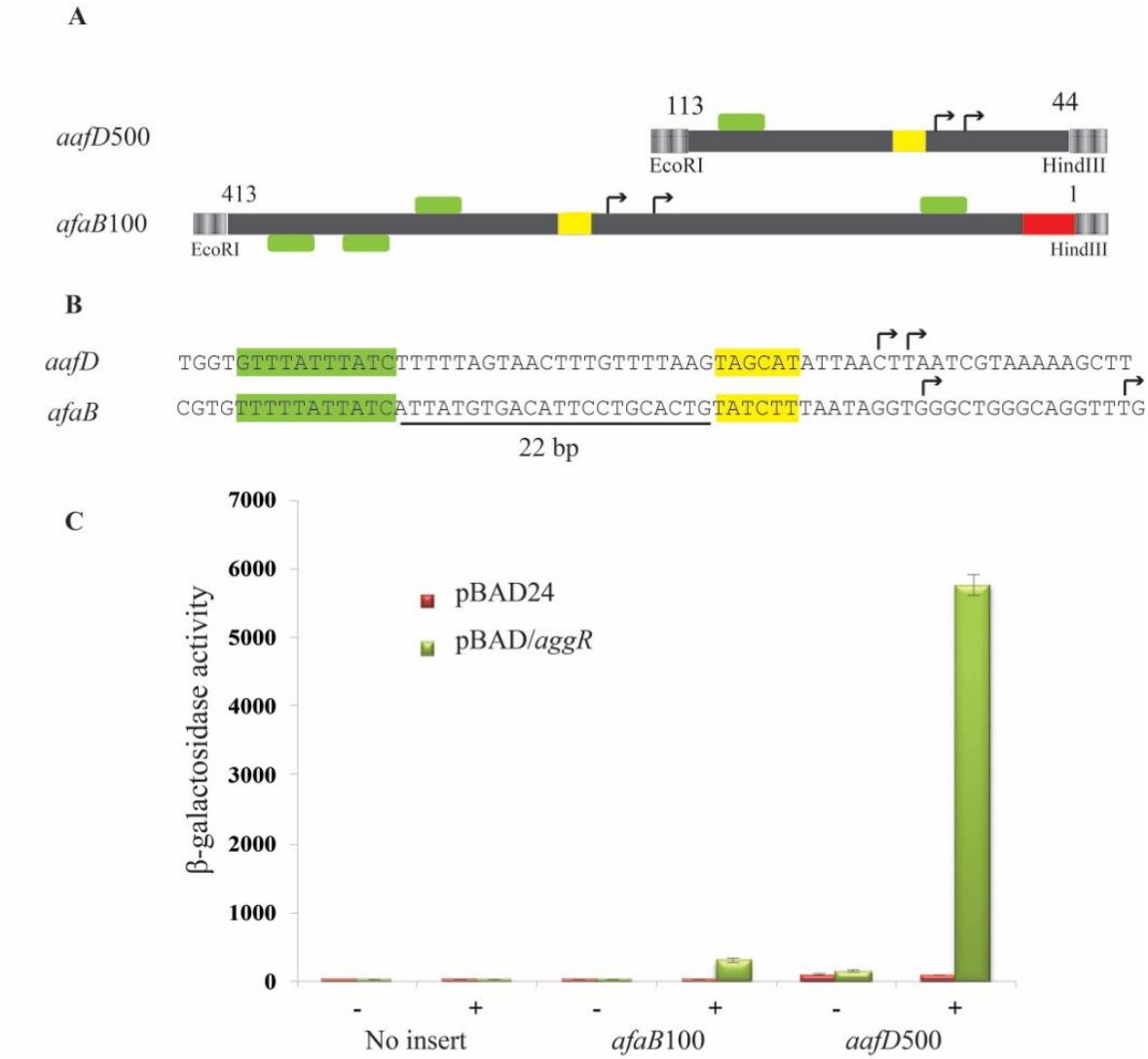


Figure 3.24 Comparison of *aafD* and *afaB* promoter transcriptional fusions

A. The figure illustrates the *aafD*500 and *afaB*100 promoter fragments. Green boxes indicate the potential AggR-binding sites on forward and reverse strands and yellow boxes represent the -10 hexamer elements. The upstream sequence of genes is shown as grey bars and red bar represent 13 bp of *afaB* pseudogene. These fragments were cloned into pRW50 using EcoRI and HindIII restriction sites. The bent arrows indicate the trancriptional start sites.

B. The panel shows the sequences of the *aafD* and *afaB* promoter elements. The AggR-binding sites are highlighted green, the -10 elements yellow and the bent arrows illustrate the transcription start sites. The distance between -10 hexamer elements and AggR-binding sites is 22 bp, marked with a line.

C The panel illustrates the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW224), *afaB*100 promoter fragment or *aafD*500 promoter fragment cloned into pRW224. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per milligram of bacterial mass.

aafC and *aafB*. The study by Elias *et al.* (1999) found that the promoter for Region 2 lies directly upstream of *aafC* rather than the *afaB* pseudogene. The current study has shown that the promoter of Region 2 is located upstream of *afaB* instead of *aafC* gene. The differential findings from this previous study might be explained by differences in the techniques and strains used to map the promoters.

The AggR-binding sites at the *aafD* and *afaB* promoters overlap with the -35 element of the promoter and there are 22 bp between the -10 hexamer element and the AggR-binding site (Figure 3.24). This result is similar to that observed by Bodero *et al.* (2007) for Rns-dependent promoters (*CS17*, *CS19* and *PCF071*), as the distance between -10 hexamer element and Rns-binding site is 22 bp and overlapping the -35 hexamer element. As no structural data is available for AggR, Rns or other closely related members of the AraC family, a little information of protein DNA interaction is available. However, Munson (2013) showed that Rns binds to one face of the DNA and its recognition HTHs bind to two adjacent major grooves (Munson, 2013). As AggR has the same number of amino acids and the sequence is conserved in the HTH region, it can be suggested that AggR binds in a similar fashion on the promoter as Rns does. The position of the AggR-binding site in relation to the other promoter elements at both *afaB* and *aafD*, suggests these promoters are Class II promoters (Browning and Busby, 2004).

The results from Figure 3.24C show that the promoter of Region 1 (*aafD*) is much stronger than the promoter of Region 2 (*afaB*). This is logical as it is physiologically required to produce more chaperone and fimbrial subunits than usher protein and adhesion subunits (Figure 1.2). To identify, the molecular basis for promoter strength, the promoter sequences were aligned and the -10 hexamer element of *afaB* promoter was found to be closer to the consensus sequence (TATAAT) than the *aafD* promoter (Figure 3.24). This does not explain

the situation and no outstanding differences were evident when the AggR-binding sites of both promoters were compared. A single base difference at position 3 of the AggR-binding site was observed between *afaB* and *aafD* promoter when compared to the consensus sequence of (Morin *et al.*, 2010). However, the difference at this location can be tolerated (Figure 3.9). The studies of Rns-dependent promoters have also shown that position 3, can also be variable (Munson and Scott, 1999; Munson, 2013). Thus, these differences do not explain the difference in promoter strength between the *aafD* and *afaB* promoters. The spacer element of the *aafD* promoter is more A-T rich than *afaB* promoter and it could be important for transcription regulation (Figure 3.24). The flanking sequences around promoter elements and the AggR-binding sites could also contribute to promoter strength (Singh *et al.*, 2011). However, it is not obvious why the strength of the *aafD* promoter is much higher than *afaB* at this present time.

Results in Figure 3.12 demonstrate that the translational start site for *aafD,* proposed by Elias *et al.* (1999) is incorrect and that another translational start site two codons downstream is in fact the correct one. This finding matches with the translational start site proposed by Chaudhuri *et al.* (2010). Moreover, the SD sequence of *aafD* is closer to the consensus sequence and it might have further contributed to the strength of *aafD* promoter to translate its transcript into functional protein.

The work detailed in this chapter has given us understanding of AggR-dependent regulation of transcription and the organisation of AggR-dependent promoter. Further characterization of AggR and its mechanism of regulation would reveal new insight into this important transcription regulation.

Chapter 4

Regulation of dispersin and T6SS gene promoters of the EAEC 042 and semi-synthetic promoters by AggR

4.1 Introduction

AggR regulates 44 genes present on the genome of EAEC 042 and in this chapter two of the operon promoters present upstream of the *aap* and *aai* genes are studied (Morin *et al.*, 2013). In the first part of this chapter (sections 4.2 and 4.3), deletion and mutational analysis of these promoters is detailed. To understand better the organisation of AggR-regulated promoters, semi-synthetic promoters were also constructed and these are discussed in the last part of the chapter.

4.2 Regulation of dispersin expression by AggR

Dispersin is an important virulence determinant and is encoded by the *aap* (anti-aggregation protein) gene. Dispersin was named, due to its ability to bind to the EAEC bacterial surface and reduce autoaggregation (Sheikh *et al.*, 2002). This reduction in autoaggregation helps in the attachment of cells to intestinal surfaces and facilitates the bacteria in spreading infection (Figure 1.5). Hence, an *aap* depleted strain of EAEC 042 was less virulent when compared to the wild type strain (Sheikh *et al.*, 2002).The *aap* gene is located upstream of *aggR* on plasmid pAA2 and it is regulated by AggR at the level of transcription (Figure 1.4). Dispersin is not confined to EAEC strains only but it is also present in other *E. coli* strains. Although, the dispersin protein sequence has been identified in different strains of *E. coli* including EAEC 042, but the promoter elements for *aap* have not been investigated to date.

4.2.1 Analysis of the *aap* DNA upstream sequence of *aap*

In order to study the promoter of *aap*, a 262 bp DNA fragment, *aap*100, was cloned into the *lacZ* expression vector, pRW50 (Figure 4.1). The schematic diagram of the *aap*100 promoter fragment is shown in Figure 4.2A. The conventional numbering system according to the transcription start site was not followed here because the transcription start site of the

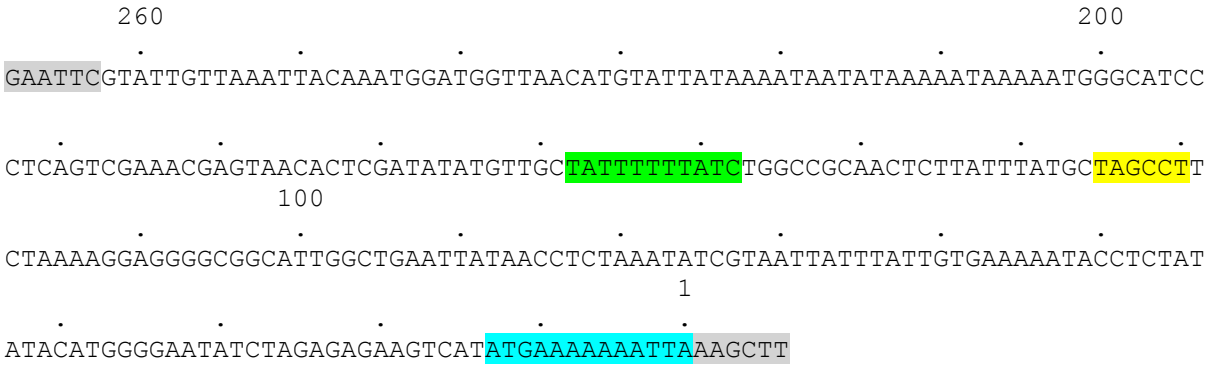


Figure 4.1 The DNA base sequence upstream of the *aap* gene

The figure shows the 262 bp DNA sequence of the *aap*100 promoter fragment. The open reading frame sequence is highlighted blue and the potential AggR-binding sites are green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.12.

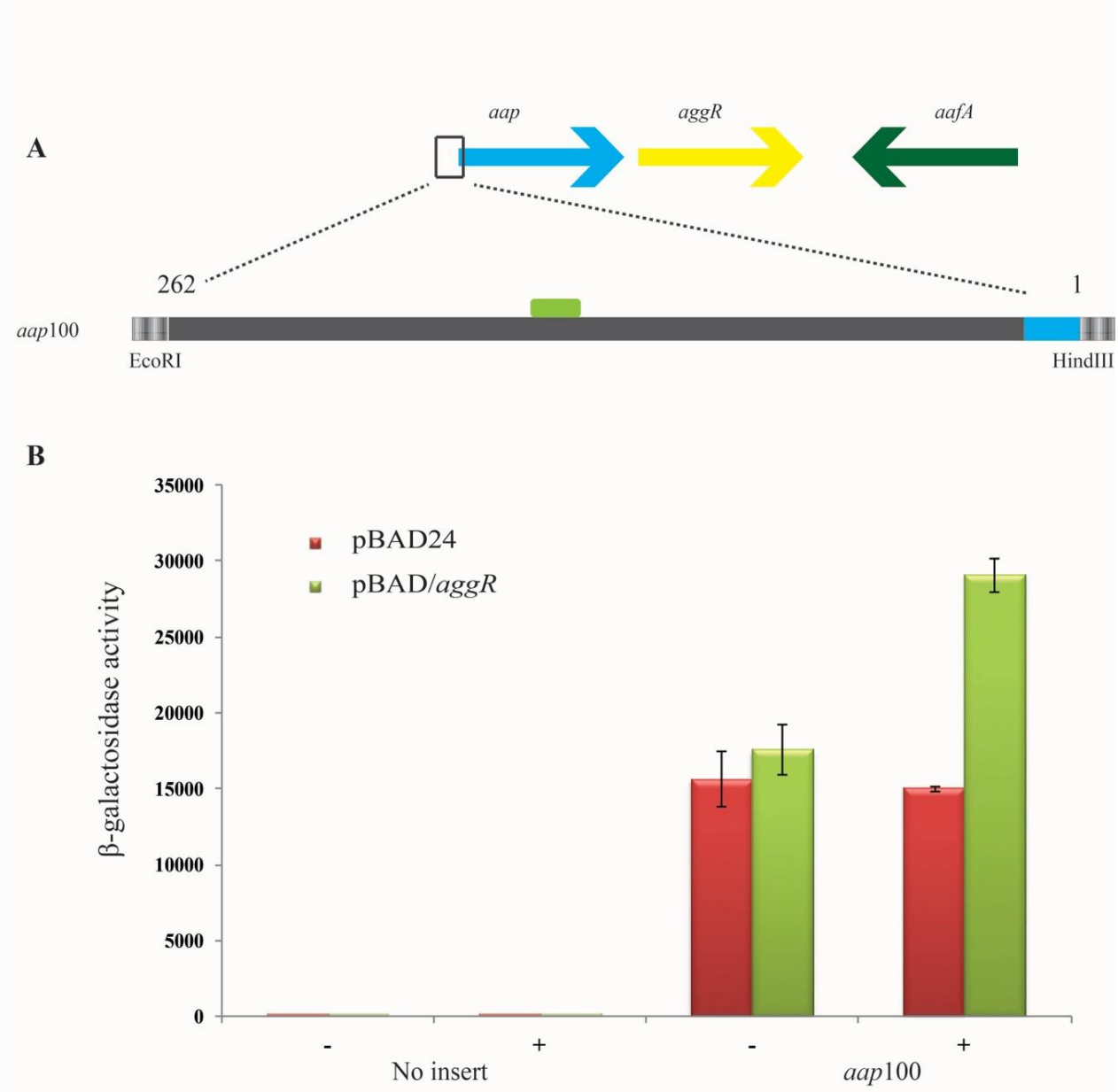


Figure 4.2 Analysis of the *aap* promoter from pAA2

A. The upper line shows the schematic arrangement of some of the genes on pAA2. The grey bar represents the upstream sequence of *aap* and the blue bar represents the *aap* coding sequence present on the promoter fragment. The potential AggR-binding site is indicated as a green box. The DNA sequence in this diagram is numbered above the fragment from 1 to 262 and the fragment was cloned into pRW50 using EcoRI and HindIII. This diagram is not to scale.
B. The panel illustrates the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the empty *lacZ* expression vector pRW50 or the *aap*100 promoter fragment cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in the presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

promoter has not been confirmed. An arbitrary numbering of the bases was chosen, starting from the base adjacent to HindIII site. Plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* containing either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to the LB medium to induce the expression of *aggR,* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The results in Figure 4.2B show that the β-galactosidase activity of the cells containing *aap*100 is increased by 2-fold in the presence of AggR-induction when compared to the AggR negative control. The results show that the *aap*100 promoter fragment carries a promoter that is AggR-regulated, and expression levels are very high suggesting that it is a very active promoter.

4.2.2 Analysis of the *aap*500 promoter fragment

The results of the *aap*100 promoter fragment showed AggR-dependent activity but the base level of expression was very high. The high β-galactosidase activity without AggR could be as a result of a very strong promoter fragment or due to presence of more than one promoters. This high level of AggR-dependent activity made it difficult to study and interpret the results of the *aap* promoter. During cloning of the *aap*100 fragment into pRW50, a derivative was cloned that had a one bp deletion at position 12 (Figure 4.3A and B) and this *aap*100Δ*T*12 was named as *aap*500. To examine the *aap*500, cloned into pRW50, the recombinant plasmid was transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR* or pBAD24. Cells were again grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Interestingly, β-galactosidase activity measured

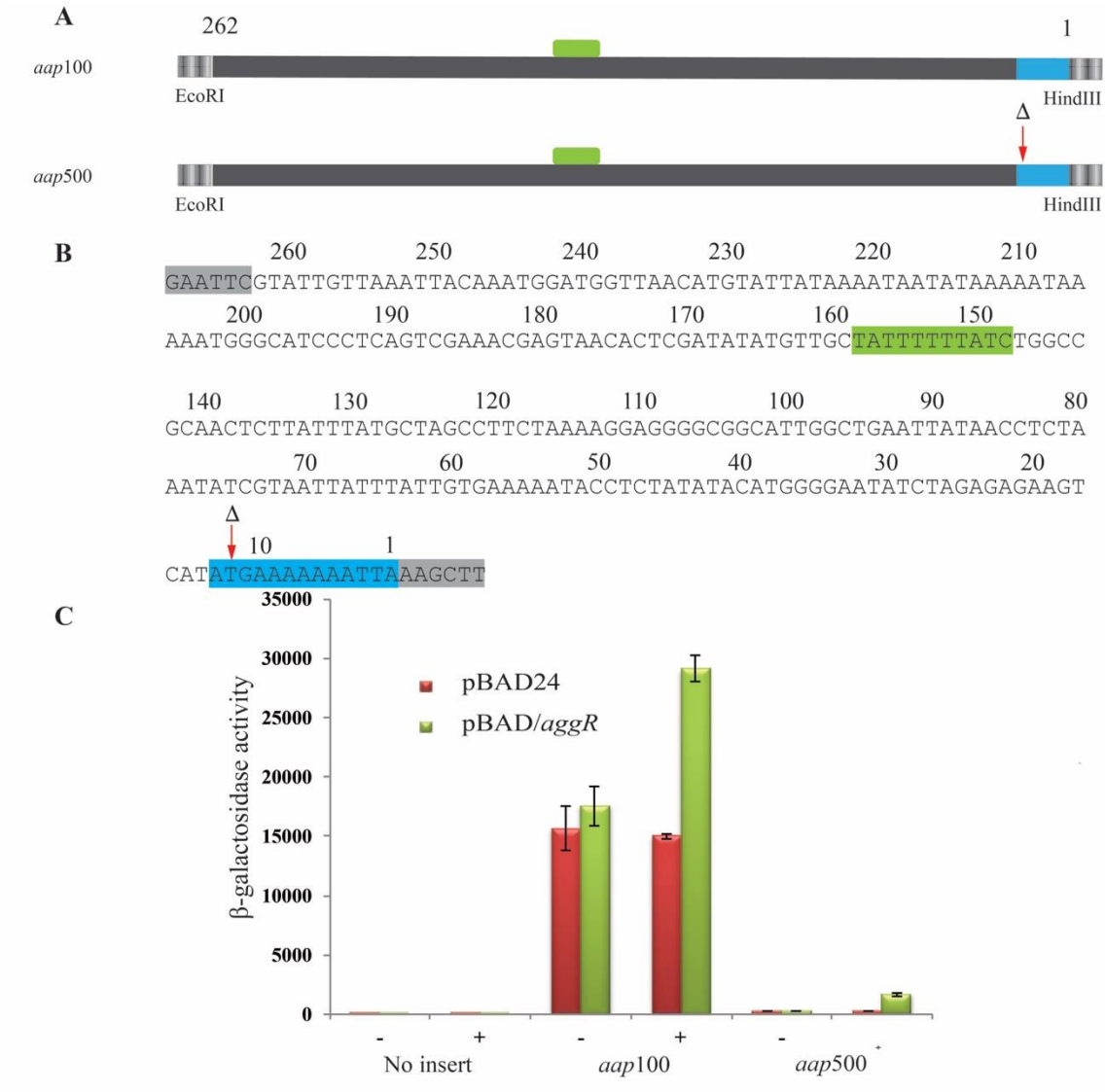


Figure 4.3 Comparison of the *aap*100 and the *aap*500 promoter fragments

A. The figure illustrates the *aap*100 and *aap*500 promoter fragments. The *aap*500 fragment is a derivative of *aap*100 in which the T at position 12 is deleted. The grey bar represents upstream sequence of *aap* and the blue bar represents *aap* open reading frame that is present on the promoter fragment. A green box illustrates the potential binding site of AggR. The red arrow with Δ indicates the position of the one base deleted in *aap*500. This diagram is not to scale.
B. The panel shows the sequence of the *aap*100 promoter fragment. The AggR-binding site is highlighted green, the restriction sites are grey and the *aap* coding sequence is blue. The red arrow with Δ indicates the position of the deletion in the *aap*500 promoter fragment.
C The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the empty *lacZ* expression vector pRW50 or *aap* promoter derivatives cloned into pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars) and were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

in cells containing pRW50/*aap*500 was much lower than the original *aap*100 fragment in the presence of AggR. Results in the Figure 4.3C also show that the β-galactosidase activity measured from lysate of cells containing pRW50/*aap*500 was increased eight-fold by AggR, when compared to AggR negative control (744 units and 90 units, respectively). The one bp deletion has shifted the reading frame and this change may had made the mRNA unstable and resulted in low β-galactosidase levels. Thus, pRW50/*aap*500 containing cells showed lower β-galactosidase levels without AggR. As the β-galactosidase levels are lower using the *aap*500 fragment, further analysis of the *aap* promoter region was carried out using this fragment.

4.2.3 Identification of the AggR-binding site on the *aap*500 promoter fragment

Bioinformatic analysis of the *aap* promoter region suggested that a potential AggR-binding sequence is present on the *aap*500 promoter fragment, which is similar to the Rns-binding weblogo by Munson (2013) and the AggR-binding site identified by Morin *et al.* (2010) (Figure 4.4). To investigate this potential AggR-binding site, point mutations were introduced at positions 10 and 12 of the AggR-binding site, corresponding to positons 151 and 149 of the *aap*500 promoter fragment (Figure 4.5A and B). The recombinant promoter fragment *aap*500-151*C*149*C* was made by substituting the Ts at positions 151 and 149 with Cs in the *aap*500 fragment. This fragment was cloned into pRW50 and recombinant plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* containing either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 4.5C show that the β-galactosidase activity measured in cells containing pRW50/*aap*500-151*C*149*C* was not

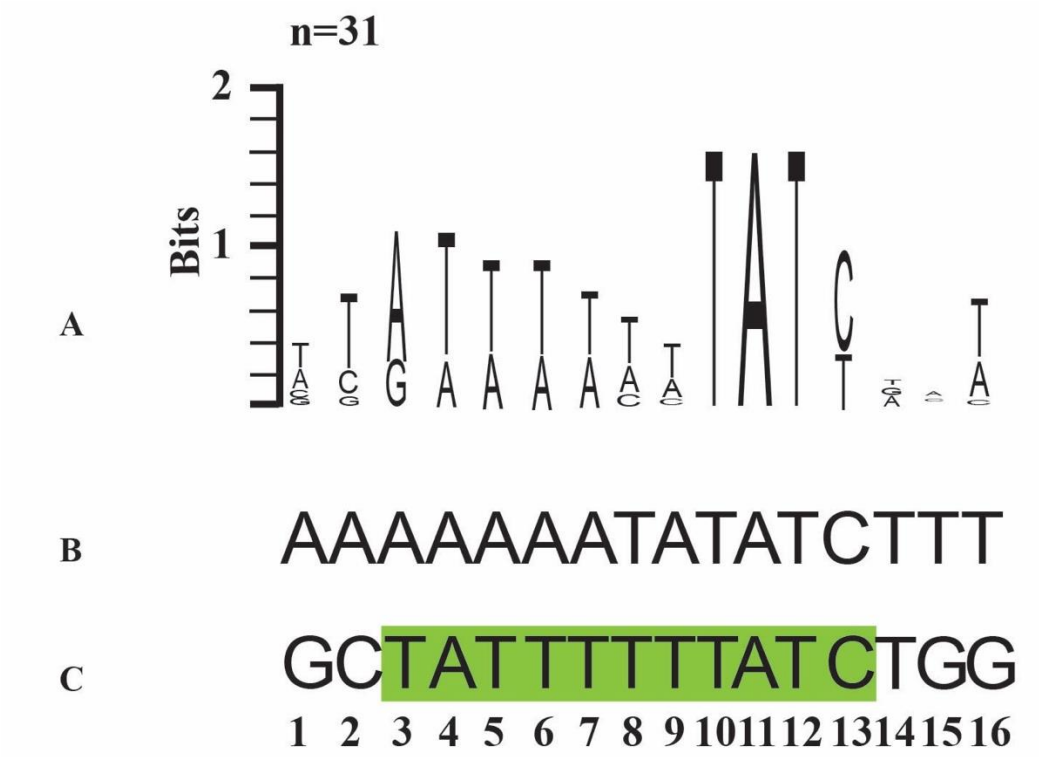


Figure 4.4 Alignment of the AggR-binding site from the *aap* regulatory region with consensus sequences of Rns and AggR

A. The panel illustrates the consensus sequence from Rns-binding logo (Munson, 2013). The height of each letter corresponds to number of times each letter is present in the 31 Rns-binding sites identified in the ETEC genome.
B. The panel shows the sequence of the essential AggR-binding site identified at the *aggR* promoter of EAEC 042 (Morin *et al.*, 2010).
C. The panel shows the proposed AggR-binding site from the *aap* promoter. The highlighted sequence shows the potential AggR-binding site (from position 3 to position 13) suggested to be important for AggR-binding. The flanking sequence of the AggR-binding site from the *aap* promoter is also shown.

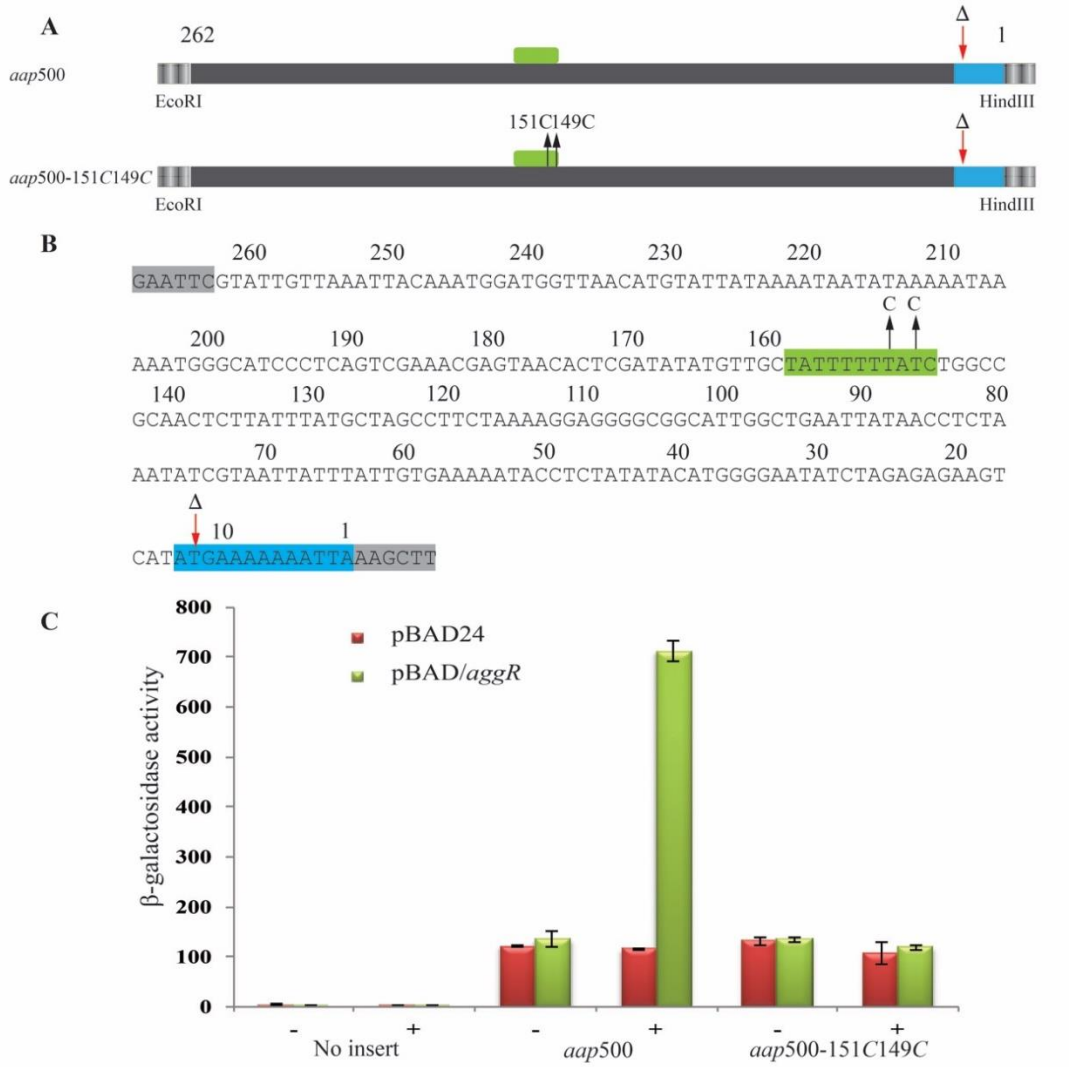


Figure 4.5 Mutational analysis of the AggR-binding site at the *aap* promoter

A. The figure illustrates the *aap*500 and *aap*500-151*C*149*C* promoter fragments. The grey bar represents upstream sequence of *aap* and the blue bar represents the 13 bp of the *aap* coding sequence present on the promoter fragment. The red arrow with Δ indicates the position of the base deleted from the *aap*100 fragment. Green boxes illustrate the potential AggR-binding site. The black arrows, indicate the positions of the point mutations introduced into the *aap*500 fragment. This diagram is not to scale.

B. The panel shows the nucleotide base sequence of the *aap*100 fragment. The AggR-binding site is highlighted green, the restriction sites are grey and the *aap* coding sequence is blue. The red arrow with Δ indicates the position of the base deleted in the *aap*500 fragment. The black arrows indicate positions of the mutations introduced into the *aap*500151*C*149*C* promoter fragment.

C. The panel shows the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aap* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

increased when AggR-expression was induced. These results indicate that this AggR-binding site has been disrupted at the *aap* promoter and is essential for AggR-dependent transcription activation.

4.2.4 Identification of the -10 hexamer element of the *aap* promoter

Inspection of the *aap*500 DNA sequence identified a potential -10 hexamer element 22 bp downstream of the AggR-binding site, which resembled the consensus sequence TATAAT. This potential -10 element has a three out of six base match to the consensus sequence and a TG motif (extended -10 hexamer element) present one bp upstream of the -10 hexamer element. To investigate this promoter element further, a point mutation A to C was introduced at position 2 of -10 hexamer element, corresponding to position 124 of the *aap*500 promoter fragment, resulting in the *aap*500-124*C* promoter fragment (Figure 4.6A and B). This fragment was cloned into pRW50 and recombinant plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* harbouring either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 4.6C show that the β-galactosidase levels measured from lysate of cells containing pRW50/*aap*500-124*C* was much lower than the cells carrying the starting *aap*500 fragment. This suggests that this substitution has disrupted the -10 hexamer element of the *aap* promoter. It is interesting to note that although the *aap* promoter carries an extended -10 hexamer element, disruption of the -10 hexamer element is sufficient to abolish expression.

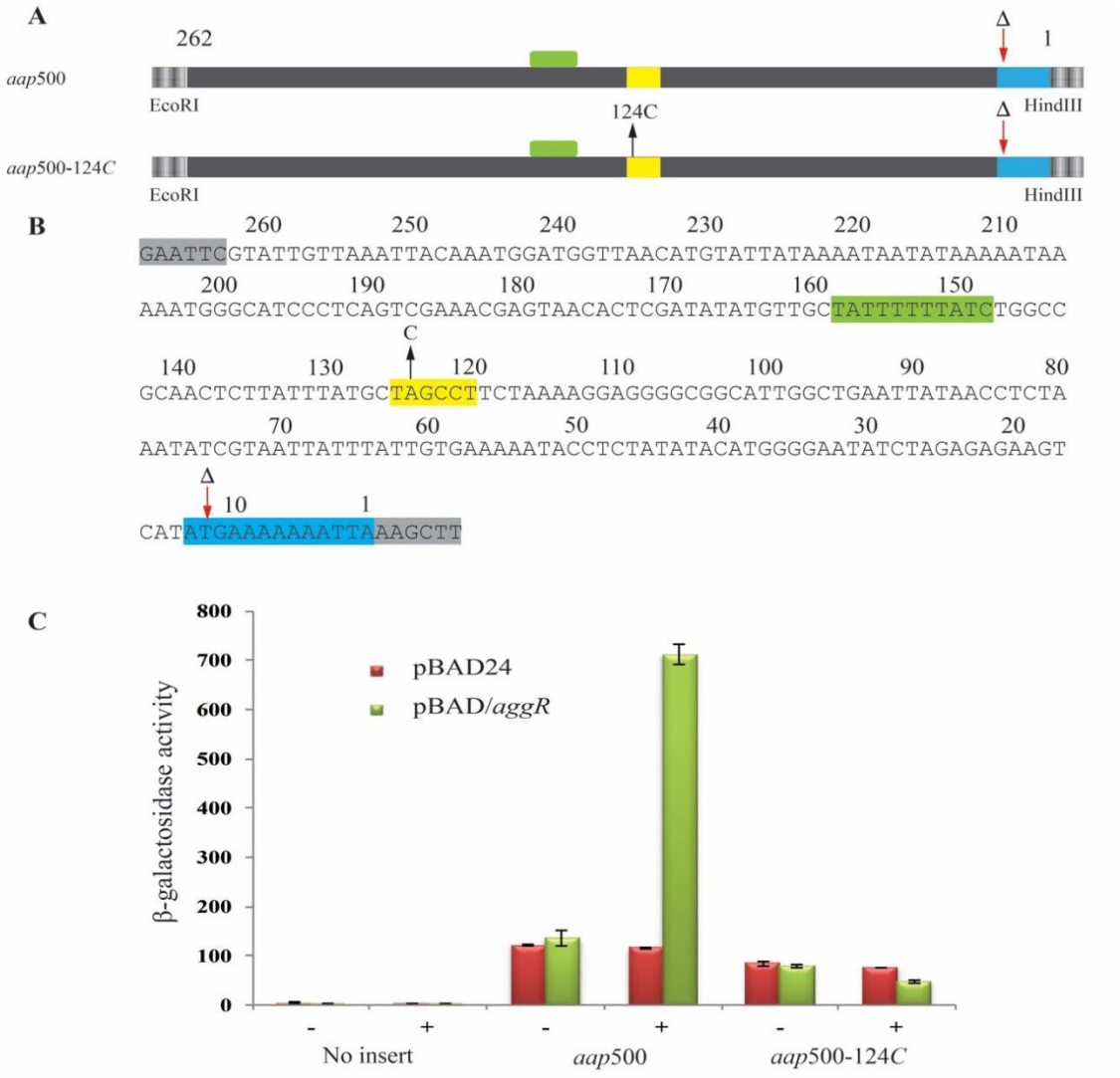


Figure 4.6 Identification of the -10 hexamer element of the *aap* promoter

A. The figure illustrates the *aap*500 and *aap*500-124*C* promoter fragments. The grey bar represents the upstream sequence of *aap* and the blue bar represents the 13 bp of the *aap* coding sequence present on the promoter fragment. The red arrow with Δ indicates the position of the base deleted during *aap*500 construction. Green boxes illustrate the potential binding site of AggR, yellow boxes the -10 hexamer element and the *aap* sequence is in blue. The black arrow indicates the position of point mutation introduced. This diagram is not to scale.

B. The panel shows the sequence of the *aap*100 fragment. The AggR-binding site is highlighted green, the -10 element yellow, restriction sites grey and the *aap* sequence is blue. The red arrow with Δ indicates the position of the base deleted in the *aap*500 promoter fragment. The black arrow with text indicates the position of 124C point mutation.

C The panel illustrates the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or the *aap* promoter derivatives cloned into pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

4.3 Transcription regulation of the EAEC 042 chromosomally encoded T6SS operon

AggR-regulated genes are present on the chromosome of EAEC strain 042 as well as on the pAA2 plasmid. EAEC 042 has two operons that encode for T6SS and they are named as *sci-1* and *sci-2*. Expression of *sci-1* is not affected by the presence of AggR but a study by Morin *et al.* (2010), using microarray methods found that genes located in the *sci-2* region are AggR-regulated. Dudley *et al.* (2006) also identified an AggR-regulated operon of 25 genes (*aaiA* to *aaiY*) present in the *sci-2* region on the EAEC genome. The later study used proteomic and genomic techniques to investigate the AggR-dependent operon present on the chromosome and used a promoter fragment of 766 bp length (carrying 466 bp upstream of the translation start site of *aaiA* gene and 300 bp of the coding sequence) to study the AggR-dependent promoter. Although, they identified the approximate location of the *aaiA* promoter, but they did not investigate the promoter elements.

4.3.1 Analysis of the *aaiA* upstream DNA sequence

In order to understand the AggR-dependent regulation of the T6SS expression from EAEC 042, the *aaiA*100 promoter fragment (carrying 466 bp upstream of *aaiA* and 13 bp of the coding sequence of *aaiA*) was cloned into the *lacZ* expression vector, pRW50 (Figure 4.7). The schematic diagram of the *aaiA*100 promoter fragment is shown in Figure 4.8A. The promoter fragment was constructed using the same upstream sequence of *aaiA* used by Dudley *et al.* (2006) but they had included 300 bp downstream sequence from the translation start site of *aaiA* and cloned this fragment into a different *lacZ* expression vector pRS551. While in the present study, only 13 bp was included downstream of the translational start site of *aaiA* and cloned into pRW50. Bioinformatic analysis indicated that on the *aaiA*100

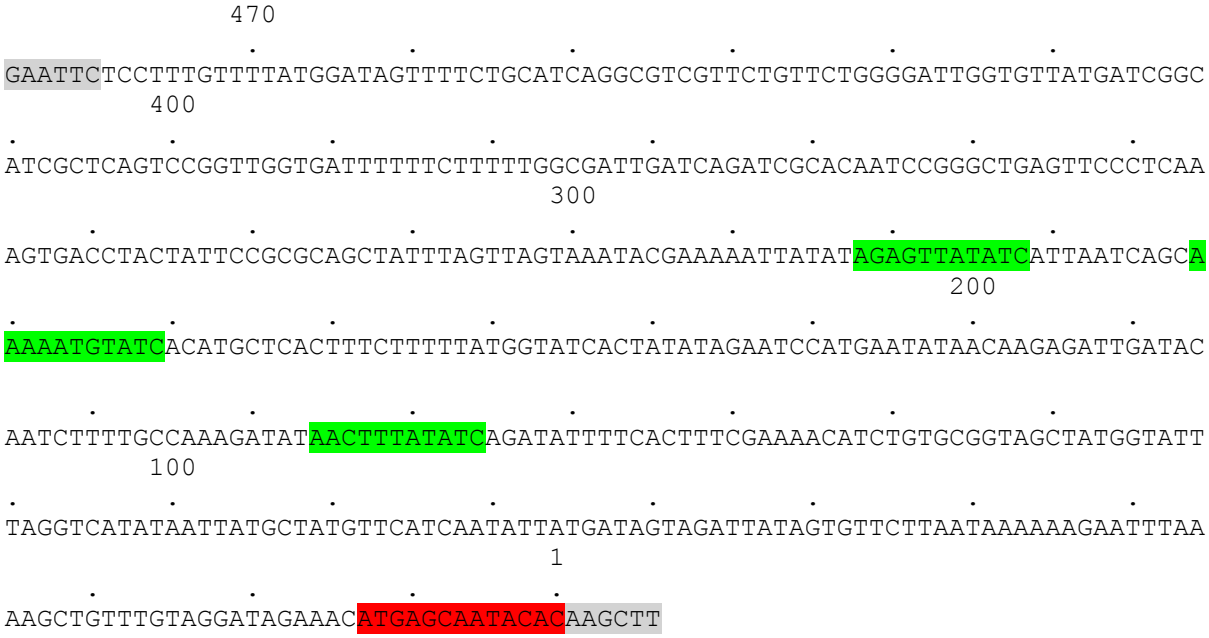


Figure 4.7 The DNA base sequence upstream of the *aaiA* gene

The figure shows the 479 bp DNA sequence of the *aaiA*100 promoter fragment. The open reading frame sequence is highlighted red and the potential AggR-binding sites are green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.10.

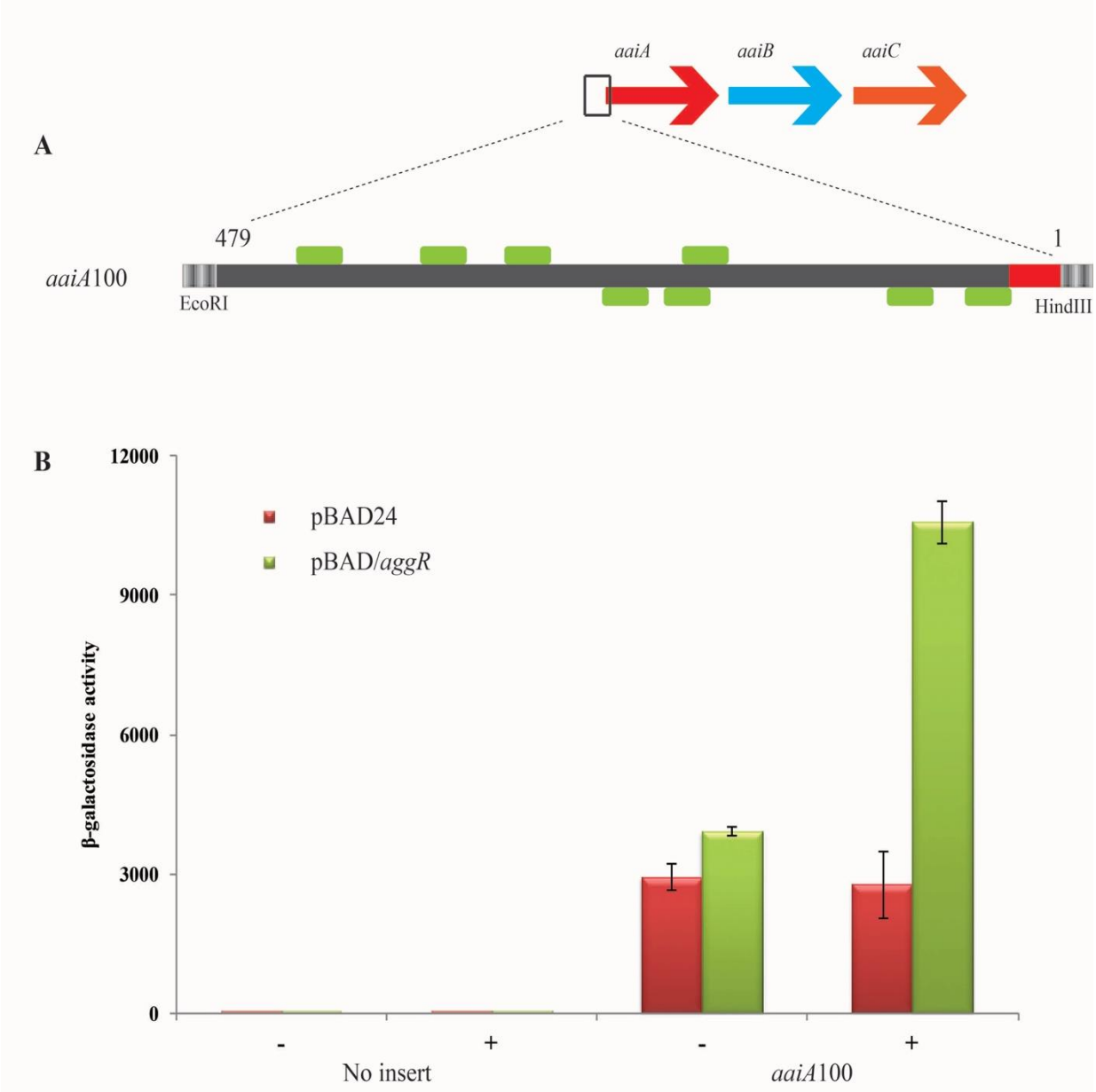


Figure 4.8 Analysis of the *aaiA* promoter from EAEC 042

A. The upper line illustrates the *aaiA* gene from the chromosome of EAEC 042 and the lower part represents *aaiA*100 promoter fragment. The grey bar represents the upstream DNA sequence of *aaiA* and the red bar represents the 13 bp of the *aaiA* coding sequence present on the fragment. Potential AggR-binding sites are indicated as green boxes. The DNA sequence in this diagram is numbered above fragment from 1 to 479 and the fragment was cloned into pRW50 using EcoRI and HindIII sites. This diagram is not to scale.
B. The panel illustrates β-galactosidase activity measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50) or the *aaiA*100 promoter fragment cloned into pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

promoter fragment, there are four potential AggR-binding sites on the forward strand and four potential sites on the reverse strand (Figure 4.8A). Plasmids were transformed into the Δ*lac E. coli* K-12 strain, BW25113, carrying either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase, and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were then determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 4.8B show that the β-galactosidase levels observed in cells containing pRW50/*aaiA*100 is increased 4-fold by AggR-induction in comparison to the AggR negative control. This shows that the *aaiA*100 promoter fragment carries the *aaiA* promoter and that the promoter is activated by AggR. Note, there is considerable AggR-independent activity without AggR-induction, which this could be due to the presence of a strong promoter or more than one promoters on the *aaiA*100 promoter fragment.

4.3.2 Identification of the minimal regulatory region of the *aaiA*100 promoter necessary for AggR-mediated activation

To find the minimal regulatory region of the *aaiA*100 fragment that is induced by AggR, a series of nested deletion mutants were made from the *aaiA*100 fragment (*i.e.* the *aaiA*99 (307-bp) to *aaiA*95 (144-bp) promoter fragments) which were all cloned into the *lacZ* expression vector pRW50 (Figure 4.9A and B). Plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac* containing either pBAD/*aggR* or pBAD24. Cells were again grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 4.9C show that the β-galactosidase activity measured in cells containing

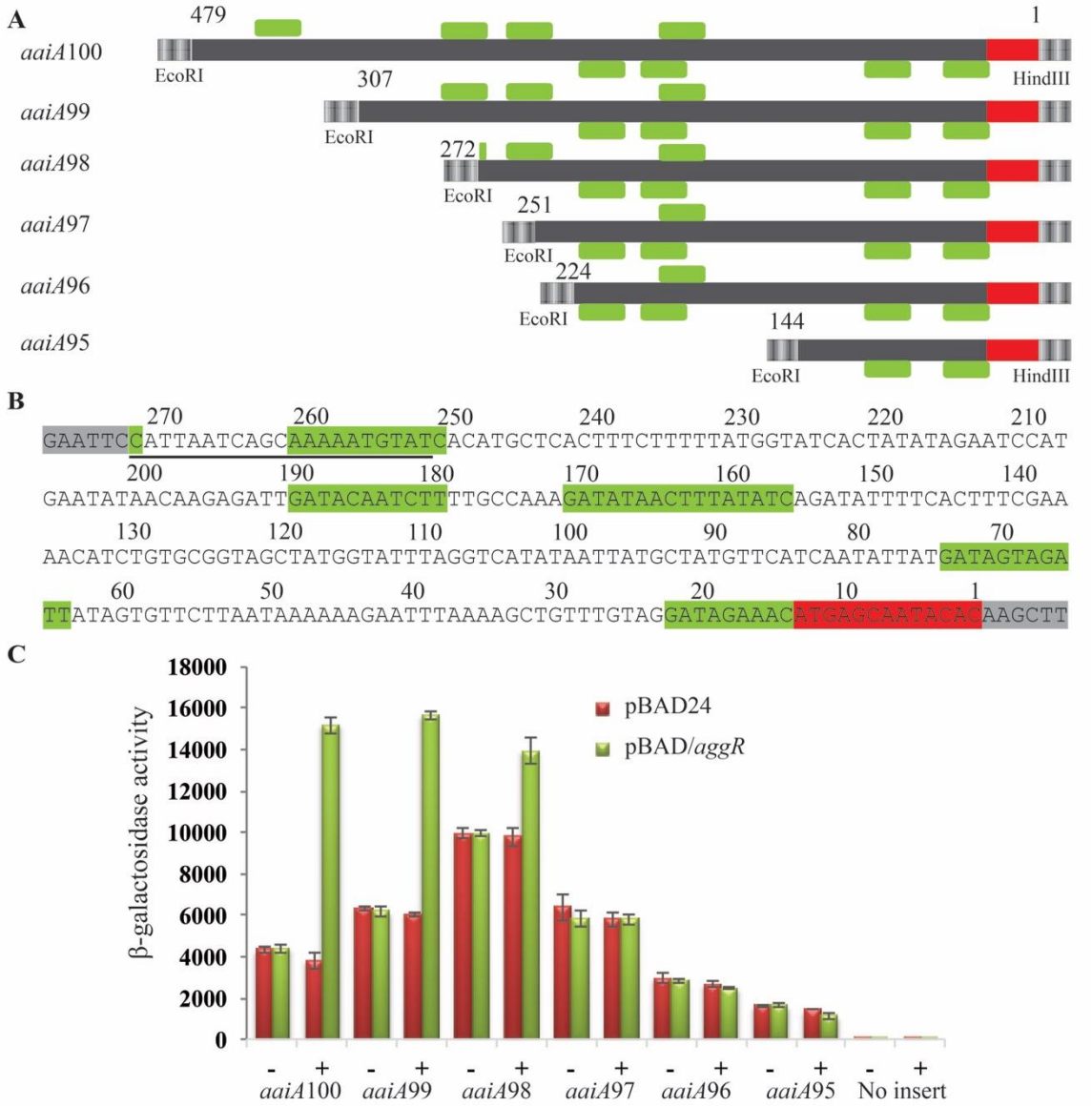


Figure 4.9 Deletion analysis of the *aaiA* regulatory region

A. The panel illustrates the *aaiA*100, *aaiA*99, *aaiA*98, *aaiA*97, *aaiA*96 and *aaiA*95 promoter fragments. The grey bar represents upstream sequence of *aaiA* and the red bar represents the 13 bp of the *aaiA* coding sequence present on each promoter fragment. Green boxes illustrate the potential binding sites of AggR. All the fragments were cloned into pRW50 using EcoRI and HindIII sites. This diagram is not to scale.

B. The panel shows the sequence of *aaiA*98 fragment and the potential AggR-binding sites are highlighted green, restriction sites are grey and the *aaiA* sequence is red. The underline sequence adjacent to EcoRI site was deleted to construct the *aaiA*97 promoter fragment.

C. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or *aaiA*100 promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

pRW50/*aaiA*99 and pRW50/*aaiA*98 is increased by AggR while for pRW50/*aaiA*97, and smaller deletions, no increase was detected in comparison to the AggR negative control. These results show that essential AggR-binding site, required for AggR-dependent regulation is present on *aaiA*98 but it has been deleted or disrupted on *aaiA*97. It is also important to note that the AggR-independent activity of cells containing pRW50/*aaiA*99 and pRW50/*aaiA*98 is higher than the cells containing pRW50/*aaiA*100 (Figure 4.9C). This could be the result of deletion of some repressor element during this deletion analysis in *aaiA*99 and *aaiA*98. Furthermore, high β-galactosidase activity was observed in cells containing pRW50/*aaiA*97 or smaller promoter fragments cloned into pRW50, so there is possibility of presence of other promoter that is still present.

4.3.3 Identification of the AggR-binding sites at the *aaiA* promoter using mutational analysis

The results suggest that an essential AggR-binding sequence is present on the *aaiA*98 promoter fragment and, that it is disrupted or deleted on the *aaiA*97 promoter fragment (Figure 4.9A and B). This potential AggR-binding site (from 261 bp to 251 bp) is similar to Rns-binding weblogo proposed by Munson, 2013 and the AggR-binding site identified by Morin *et al.* (2010) (Figure 4.10). To investigate this, point mutations were introduced into positions 10 and 12 of the AggR-binding site shown in Figure 4.10, that correspond to positions 254 and 252 on the *aaiA*98 promoter fragment (Figure 4.11A and B). The recombinant promoter fragment *aaiA*98-254*C*252*C* was made by substituting the Ts at positions number 254 and 252 with Cs and was cloned into pRW50. Plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac* containing either pBAD/*aggR* or pBAD24. Cells were again grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to induce the expression of *aggR* where

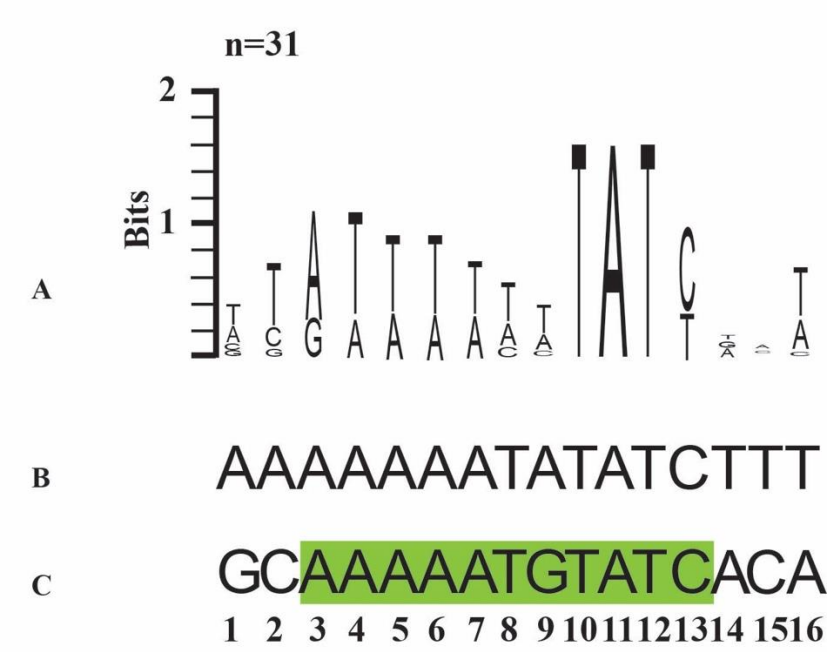


Figure 4.10 Alignment of the AggR-binding site from the *aaiA* regulatory region with the consensus sequences of Rns and AggR

A. The panel illustrates the consensus sequence from the Rns-binding logo (Munson, 2013). The height of each letter corresponds to number of times each letter is present in the 31 Rns-binding sites identified in the ETEC genome.
B. The panel shows the sequence of the essential AggR-binding site identified at the *aggR* promoter of EAEC 042 (Morin *et al.*, 2010).
C. The panel shows the AggR-binding site from the *aaiA* promoter sequence. The highlighted sequence shows the potential AggR-binding site (from position 3 to position 13) suggested to be important for AggR-binding. The flanking sequence of AggR-binding site from the *aaiA* regulatory region is also shown.

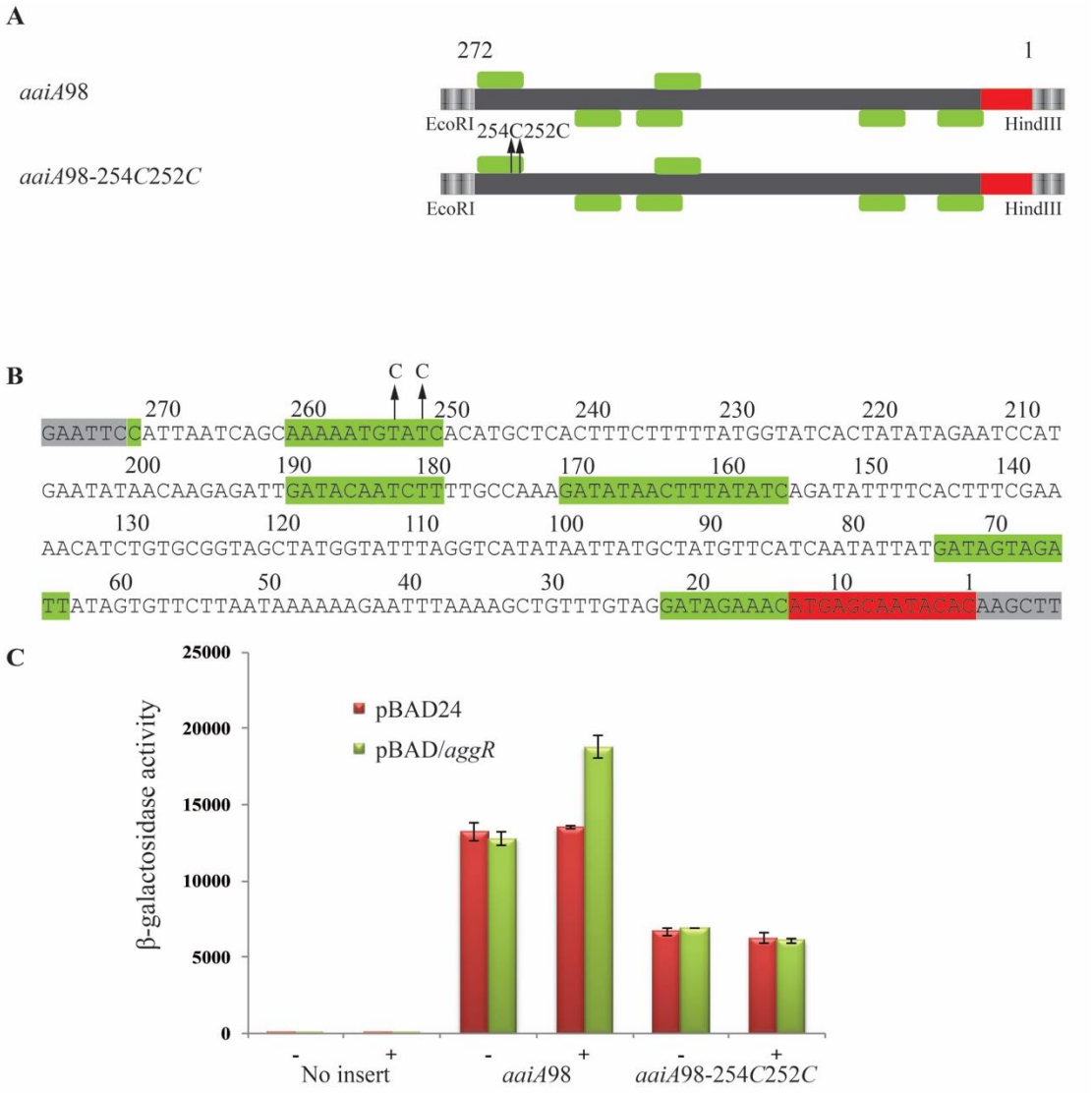


Figure 4.11 Mutational analysis of the AggR-binding site at the *aaiA* promoter

A. The figure illustrates the *aaiA*98 and *aaiA*98-254*C*252*C* promoter fragments. The grey bar represents the upstream sequence of *aaiA* and the red bar represents the 13 bp of the *aaiA* of coding sequence present on the promoter fragment. Green boxes illustrate potential binding sites of AggR. The arrows indicate the positions of point mutations introduced into the fragment. This diagram is not to scale.

B. The panel shows the base sequence of the *aaiA*98 fragment. The AggR-binding sites are highlighted green, restriction sites are grey and *aaiA* sequence red. The arrows, indicate positions of point mutations introduced into the *aaiA*98 promoter fragment.

C. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50, *aaiA* promoter derivative cloned into pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars) and were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 4.11C show that the β-galactosidase activity measured from lysates of cells containing pRW50/*aaiA*98-254*C*252*C*, did not increase when compared to the AggR negative control. These results show that the AggR-binding site has been disrupted on the *aaiA*98-254*C*252*C* promoter fragment. Moreover, there is a decrease in AggR-independent activity for *aaiA*98-254*C*252*C,* indicating that these mutations may have disrupted promoter elements, important for *aaiA* expression.

4.3.4 Analysis of the -10 hexamer element of the *aaiA* promoter

A potential -10 hexamer element (TATCAC) was identified 23 bp downstream of the AggR-binding site, which resembles the consensus sequence, TATAAT. In this case, four out of the six bases matched the consensus sequence of the -10 hexamer element and it also has an extended -10 element (an upstream TG motif). To investigate this promoter element, a point mutation was introduced at position 2 of -10 hexamer element, corresponding to position 226 of the *aaiA*98 promoter fragment, resulting in the *aaiA*98-226*C* fragment (Figure 4.12A and B). This fragment was cloned into pRW50 and transformed into the *E. coli* K-12 strain BW25113 Δ*lac* harbouring either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were then determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in the Figure 4.12C show that the β-galactosidase levels measured in cells containing pRW50/*aaiA*98-226*C* showed only a small decrease in promoter activity, indicating that this mutation had a little effect on the promoter. This could be explained by the presence of the TG motif on *aaiA*98 promoter fragment as this could compensate for the mutation introduced. However, due to time constraints, it was not possible to explore this further.

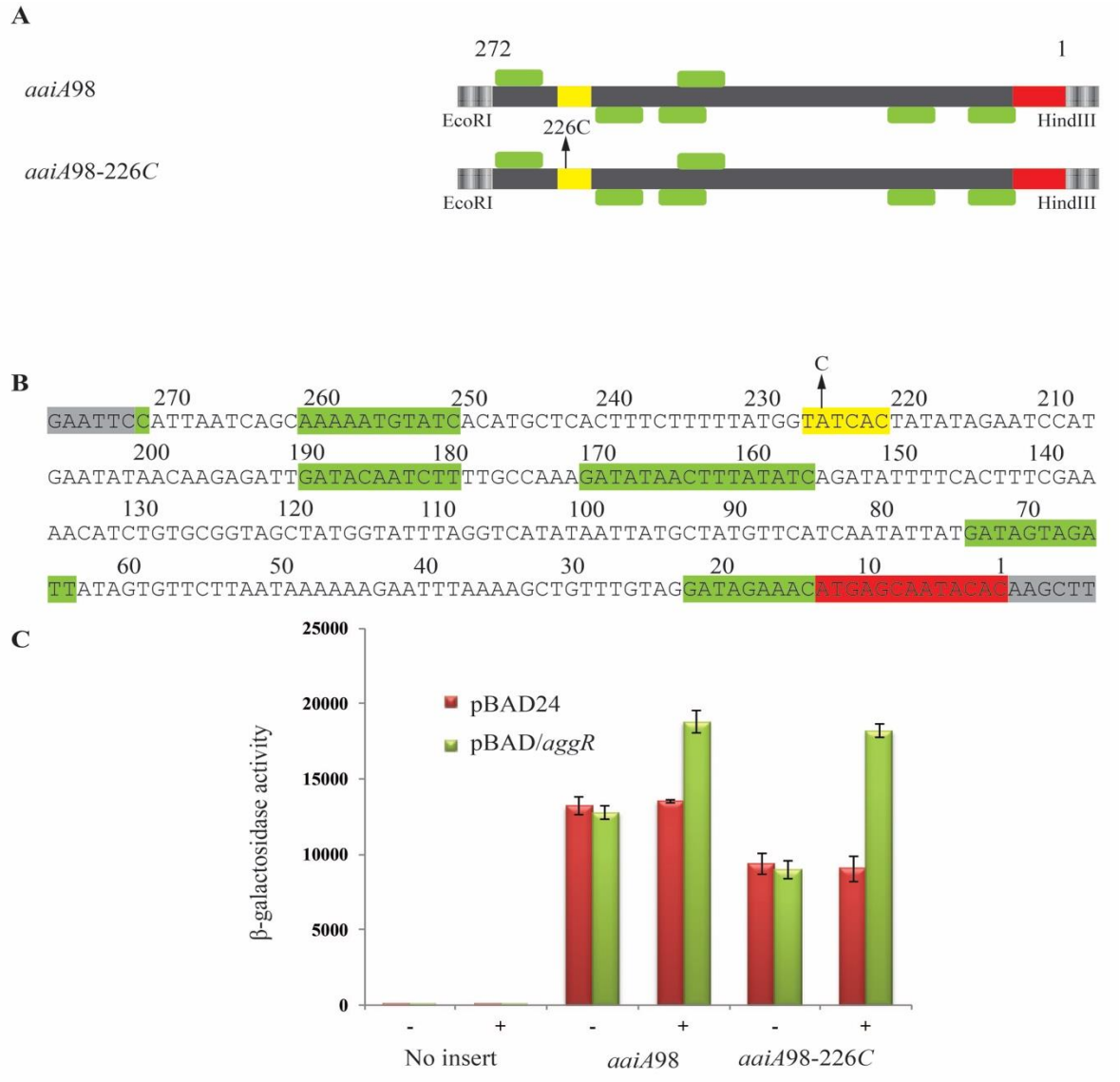


Figure 4.12 Identification of the -10 hexamer element of the *aaiA* promoter

A. The figure illustrates the *aaiA*98 and *aaiA*98-226*C* promoter fragments. The grey bar represents the upstream sequence of *aaiA* and the red bar represents the 13 bp sequence of the *aaiA* coding sequence present on the promoter fragment. Green boxes illustrate potential binding sites of AggR and yellow boxes represent the -10 hexamer elements. The arrow with text indicates position of the point mutation introduced in the *aaiA*98-226*C* promoter fragment.

B. The panel shows the sequence of the *aaiA*98 promoter fragment. The AggR-binding sites are highlighted green, the -10 hexamer element is yellow, the restriction sites are grey and the *aaiA* coding sequence is red. The arrow indicates position of point mutation introduced into the *aaiA*98-226*C* promoter fragment.

C The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or *aaiA* promoter derivatives cloned into pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars) and were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

4.4 Construction of AggR-dependent semi-synthetic promoters

For further understanding of AggR-binding sites and AggR-dependent promoters, a series of semi-synthetic promoters were constructed. A well studied CRP-dependent semi-synthetic promoter was used as starting promoter fragment in this study (*i.e. CC(-41.5)*) (Figure 4.13). The AggR-binding site sequence from *aafD* promoter was cloned into *CC(-41.5)* to generate a series of semi-synthetic promoters. To understand promoter architecture at AggR-dependent promoters, the fragments used in this study were aligned.

4.4.1 Alignment of the AggR-binding sites and the -10 hexamer elements from AggR-regulated promoters

In this project, a number of AggR-dependent promoters from EAEC 042 have been investigated with important AggR-binding sites identified at the AAF/II, T6SS and dispersin promoters. These AggR-binding sites are in general very similar but a few differences in important regions have been noted. The differences in AggR-dependent promoters that have been investigated also have different levels of AggR-dependent expression, with some promoters displaying considerable AggR-independent activity. To highlight the differences and similarities between the AggR-binding sites and promoter elements, these promoter sequences were aligned (Figure 4.14A). Furthermore, to find an AggR consensus sequence, the AggR-binding sequences, including flanking bases, were copied into the weblogo website software and a weblogo was generated (Figure 4.14B). In each case a 16 bp sequence was used, which is the same fragment length used by Munson *et al.* (2001) for Rns. The sequences from the *aafD, afaB, aap* and *aaiA* promoters from this study and the *aggR* promoter Morin *et al.* (2010) were copied into the input box of http://weblogo.berkeley.edu to generate the weblogo (Figure 4.14B) (Crooks *et al.*, 2004). The height of each letter corresponds to the number of times, each letter is present in the AggR-binding sites at a certain position. The

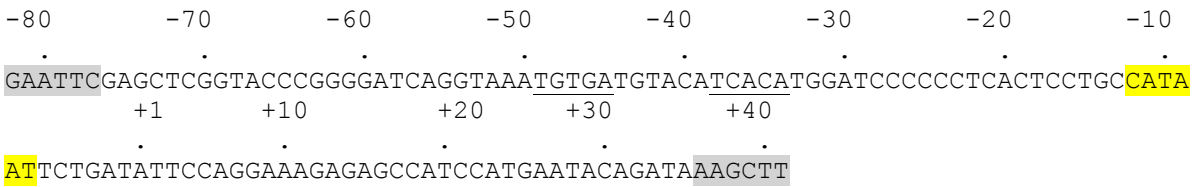


Figure 4.13 The promoter sequence of *CC(-41.5)* used to make semi-synthetic promoter fragments containing AggR-binding sites

The panel shows the DNA sequence of the *CC(-41.5)* promoter fragment from Savery *et al.* (1995b), numbered in relation to the start site of transcription (+1). The CRP-binding site centred at position -41.5 and is underlined. The potential -10 hexamer element is yellow and EcoRI and HindIII sites are grey.

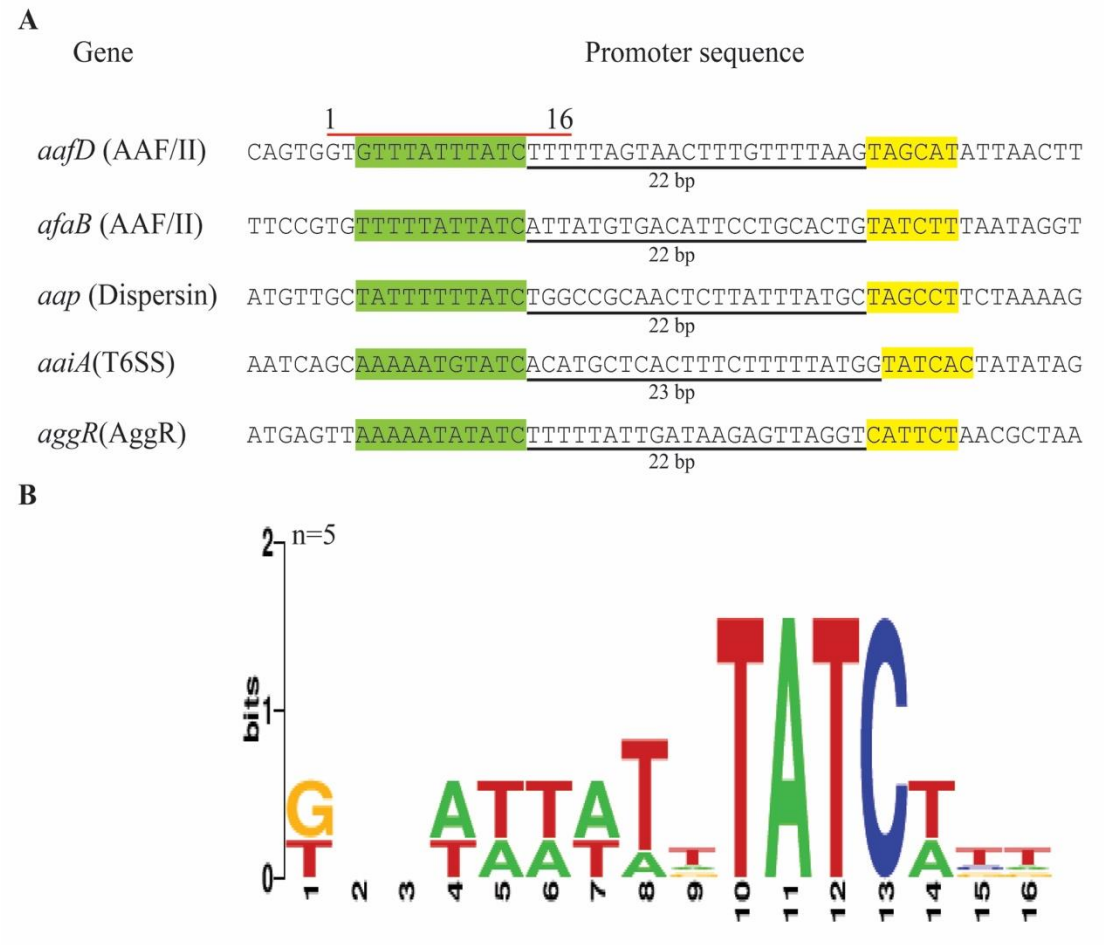


Figure 4.14 Alignment of DNA sequences from AggR-dependent promoters

A. The panel illustrates the sequences of AggR-dependent promoters investigated in this study. The AggR-binding sites are highlighted green and the -10 hexamer elements yellow. The underline sequences show the distance between AggR-binding sites and -10 hexamer elements. The red line on the top of the sequences with the numbers from 1 to 16 represents the part of promoter used to make the weblogo shown in panel B.

B. The panel illustrates the consensus sequence for AggR. The n represents the number of sequences used to create the weblogo (n=5). 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 five AggR-binding sites used to generate this weblogo (Crooks *et al.*, 2004).

absence of a sequence at certain position represents high variability of bases at this position. In order to find a more reliable consensus sequence, more AggR-binding sites are required, however, this weblogo gives us an idea about the AggR-binding motif.

4.4.2 Analysis of semi-synthetic AggR-dependent promoters

The alignment of promoter sequences in Figure 4.14A shows that AggR-binding sites are generally located 22 or 23 bp upstream from the -10 hexamer elements.. In order to investigate, if an AggR-binding site can be transplanted into a promoter that is not normally AggR-binding site and its flanking sequence from the *aafD* promoter (Figure 4.15A). The *aafD* AggR-binding site was cloned at different distances from the -10 hexamer element (*i.e.* ranging from 20 bp to 23 bp) and the resultant promoter fragments named *DAM*20 to *DAM*23. These fragments were cloned into the *lacZ* expression vector pRW50. Plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac* also carrying, either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The β-galactosidase activity measured in cells containing *CC(-41.5)* and *DAM*20 showed no increase in the presence of AggR. However, the β-galactosidase activity observed in cells containing *DAM*21, *DAM*22 and *DAM*23 showed a 4, 8 and 2-fold increase in expression levels, respectively, when compare to the AggR negative control (Figure 4.15B). This shows that the AggR-binding site can be transplanted into a promoter to make it AggR-dependent and that the distance between the AggR-binding site and the -10 hexamer element is important for AggR-dependent activity. The results also show that a spacing of 22 bp is optimal for AggR-dependent induction. Interestingly, increasing the spacing to 23 bp as in the *aaiA* promoter increases the basal promoter activity

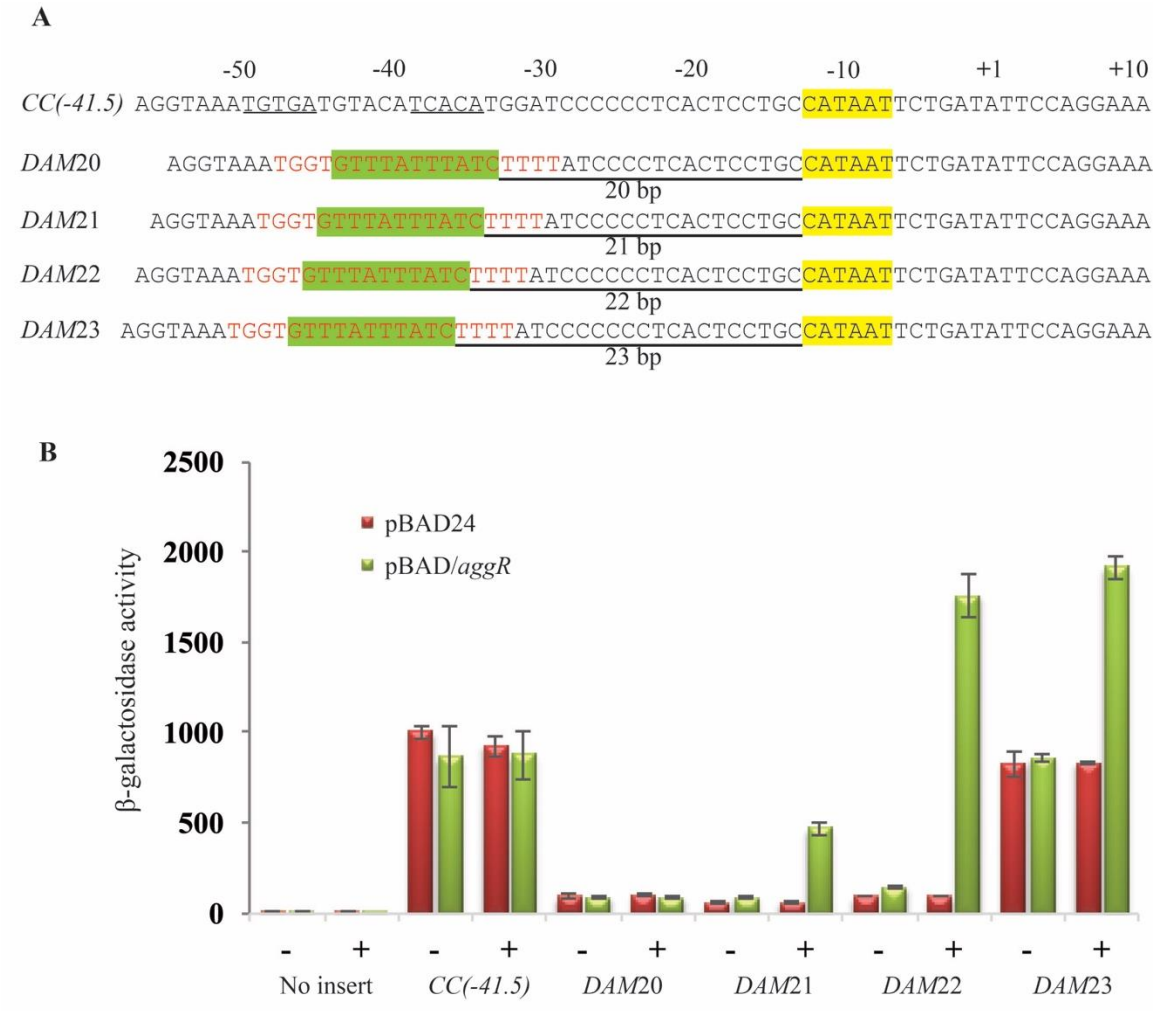


Figure 4.15 Transplantation of the *aafD* AggR-binding site into the *CC(-41.5)* promoter fragment

A. The figure shows part of the DNA sequence of the *CC(-41.5)* promoter fragment and the *DAM20* to *DAM23* promoters. In the *DAM20* to *DAM23* promoters the AggR-binding site has been transplanted at different distance from the -10 hexamer elements (20 bp to 23 bp, respectively). In *CC(-41.5)* underling represents the CRP-binding half-sites and the red colour text of AggR-binding site and flanking sequence shows the sequence transplanted from the *aafD* promoter into *CC(-41.5)*. Green boxes indicate the AggR-binding site and yellow boxes indicate the -10 hexamer elements. Fragments were cloned into pRW50 using EcoRI and HindIII.
B. The panel illustrates the β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50, *CC(-41.5),* or *DAM* promoter derivatives cloned into pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

and reduces the AggR-dependent induction by 2 fold (Figure 4.15B and Figure 4.9C).

4.4.3 Disruption of the AggR-binding site at the *DAM*22 promoter

The β-galactosidase activity results in Figure 4.15 show that the *aafD* promoter AggR-binding site can be transplanted into an AggR-independent promoter to change its regulation. To confirm that the *aafD* AggR-binding site was responsible for this regulation, the AggR-binding site in the *DAM*22 promoter fragment was mutated by introducing point mutations into the conserved Ts within the site (Figure 4.16A). These positions correspond to positons -38 and -36 on the *DAM*22 promoter fragment and this generated the *DAM*22-(-38*C*-36*C*) fragment, which was cloned into pRW50. The recombinant plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR* or pBAD24. Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in the Figure 4.16B show that the β-galactosidase activity measured in cells containing pRW50/*DAM*22-(-38*C*-36*C*) showed no increase when AggR-expression was induced. These results confirm that the *aafD* AggR-binding site is essential for AggR-dependent activation of the *DAM*22 promoter fragment.

4.5 Discussion

AggR regulates many virulence genes, for example, *aap* which is located on the pAA2 plasmid. The results of this study have shown that expression from the *aap*100 promoter fragment is increased in the presence of AggR and this is consistent with other studies (Morin *et al.*, 2013). The essential AggR-binding site at the *aap* promoter differs from the consensus at position 3 from the Rns-binding sequence, suggesting it is not optimal (Figure 4.4). As

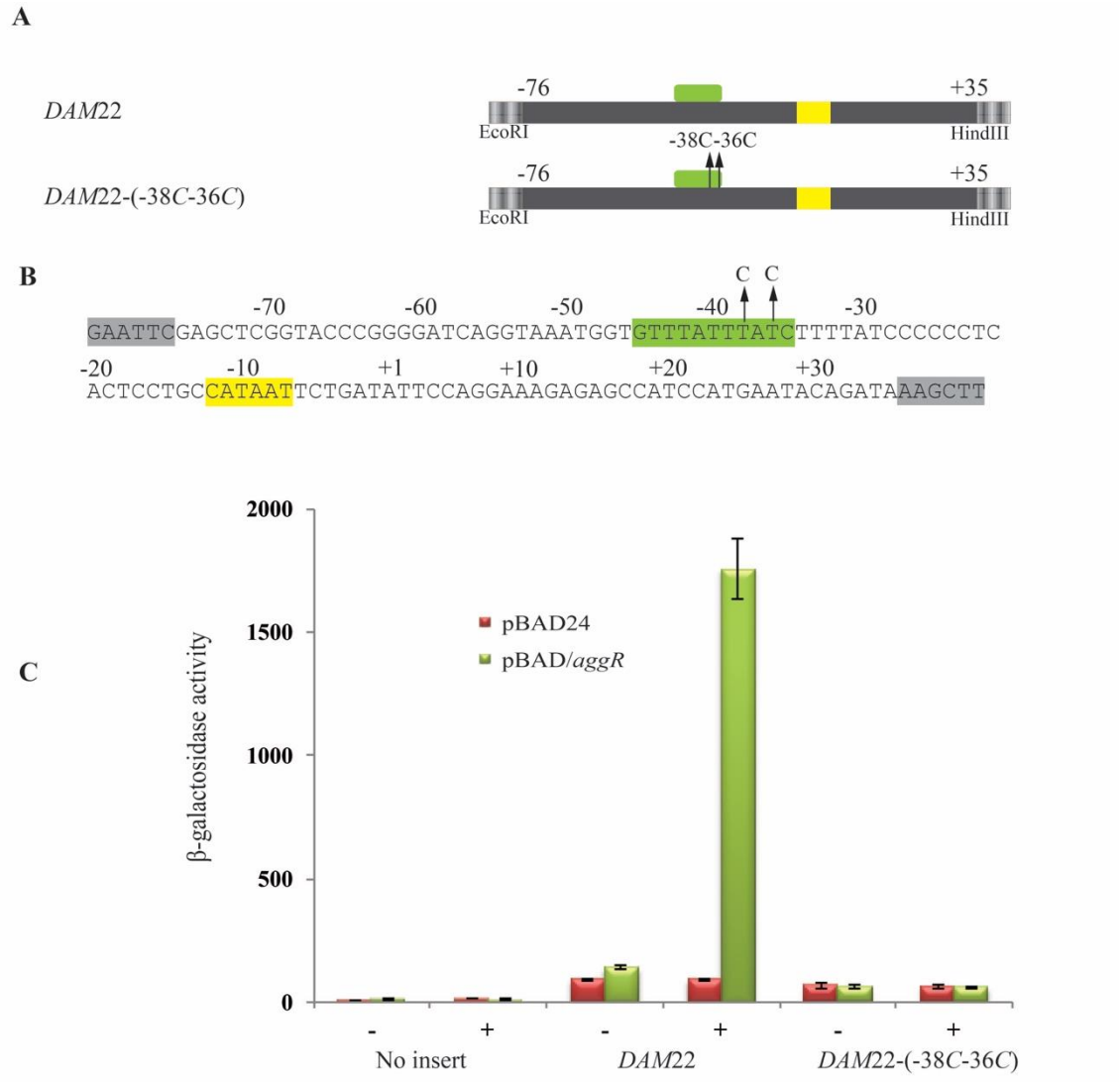


Figure 4.16 Mutational analysis of the AggR-binding site at the *DAM*22 promoter fragment

A. The figure illustrates the *DAM*22 and *DAM*22-(-38*C*-36*C*) promoter fragments. Green boxes illustrate the binding sites of AggR and yellow boxes represent the -10 element. The arrows, with text, indicate the positions of the point mutations introduced into *DAM22* promoter. This diagram is not to scale.

B. The panel shows the nucleotide base sequence of the *DAM*22 fragment. The AggR-binding site is highlighted green, the potential -10 element is yellow and restriction sites are grey. The arrow indicates position of point mutations introduced into *DAM*22 to generate the *DAM*22-(-38*C*-36*C*) promoter fragment.

C. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or *DAM22* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

lacZ expression from the *aap*100 promoter fragment was extremely high, the results were not very conclusive. For this reason, a derivative of the *aap*100 promoter fragment, *i.e.* *aap*100Δ12*T* (*aap*500) was used in this study to understand the *aap* promoter in more detail. The distance between the -10 hexamer element and the AggR-binding site is 22 bp, which is optimal and likely contributes to the high expression levels of the promoter. During infection, high levels of expression of *aap* is needed by EAEC as dispersin forms a coat around the bacterium. It is important to mention that dispersin is secreted by the system encoded in a separate operon (*i.e. aatPABCD* operon), and this arrangement ensures that there is high level production of dispersin and that the optimal number of its secretory system proteins are produced.

Similarly, the data from the *aaiA* operon analysis shows that an AggR-dependent promoter lies upstream of *aaiA*. The AggR-binding site identified at the *aaiA* promoter is similar to the Rns- binding weblogo and the *aggR* promoter AggR-binding site at conserved positions. The β-galactosidase levels measured from cell lysates containing pRW50/*aaiA*100 without induction were very high and this could be either due to the presence of the TG motif of the promoter or the presence of an additional AggR-independent promoter. In the presence of AggR, there was only a four-fold induction, which is less than that observed in the previous study by Dudley *et al.* (2006) that was an 8-fold increase. The promoter fragment used by Dudley *et al.* (2006), possessed the same DNA upstream of the translation start site, but had an additional 287 bp downstream in comparison to the fragment used in this study. Thus, this difference in expression could be explained by the difference in strains, plasmids or constructs used. Another important finding to note was, that AggR-induction was reduced from 4 fold to 2 fold when the promoter fragment was truncated to 272 bp (*i.e.* the *aaiA*98 promoter fragment) due to the increase in basal activity of the *aaiA*98 promoter fragment

(Figure 4.9). This increase in basal level could be a result of the deletion of a repressor element.

Results in this study show that AggR-binding sites overlap the -35 hexamer elements of the promoters it regulates and this suggests that AggR activates transcription by a Class II mechanism (Browning and Busby, 2004).The alignment of all AggR-binding sites and the promoter elements of the genes investigated in this study shows the variations in the AggR-binding site sequences, the -10 hexamer elements, the distance between AggR-binding sites and the -10 hexamer elements (Figure 4.14). To identify the appropriate distance between the -10 hexamer element and AggR-binding sites using semi-synthetic promoters, I have shown that 22 bp is the optimal distance for AggR-dependent activation (Figure 4.15). A similar observation has been found by Bodero *et al.* (2008), for Rns-dependent promoters.

This study has given an insight into the organisation of AggR-dependent promoters and AggR-binding sites. The mutational analysis of AggR has increased our understanding of AggR-dependent activation and the semi-synthetic promoters have enhanced our understanding about AggR-binding site positioning from the promoters.

Chapter 5

Mutational analysis of AggR and effects of AggR on *pet* expression in EAEC 042

5.1 Introduction

AggR is the master virulence regulator of EAEC strains and it regulates a number of promoters. In order to understand the organisation of AggR-regulated promoters, a mutational analysis of AggR was carried out. AggR mutants were studied *in-vivo* by measuring activities of both semi-synthetic and AAF/II promoters. In the first part of the chapter, I have described the activation by AggR mutants on AggR-dependent promoters. In the second part of this chapter, I have detailed a study on *pet* (plasmid encoded toxin) expression and how AggR may affect *pet* expression.

5.2 Mutational Analysis of the AggR protein

The results in this study indicate that, at target promoters, AggR binds near the -35 hexamer element, and so it is likely that it interacts with the RNAP directly, as other AraC family members (Grainger *et al.*, 2004). Results in the Figure 4.15B shows that the optimal distance between the -10 hexamer element and an AggR-binding site is 22 bp and suggest that AggR-regulated promoters can be categorised as Class II promoters (Browning and Busby, 2004). AggR is closely related to other members of the AraC family, particularly Rns, which has the same number of amino acids and has considerable sequence similarity to AggR. Both transcription factors have two HTHs for DNA binding, present towards the C-terminal of each protein, and this HTH arrangement is conserved in AraC family members, playing an important role in AggR-dependent regulation (Figure 1.12) (Munson, 2013). Basturea *et al.* (2008), showed that the N-terminal domain of Rns is also important in Rns-dependent regulation at ETEC promoters. In this work, they introduced many substitutions into Rns and found that, at position 14, an isoleucine to threonine (I14T) substitution, and at position 16, an asparagine to aspartic acid (N16D) substitution, was important for Rns-dependent regulation. AggR also has isoleucine and asparagine at position 14 and 16, respectively (Figure 5.1). To

Figure 5.1 Alignment of AggR and Rns sequence

The figure shows amino acid sequence of AggR and Rns. Standard abbreviations are used to represent amino acids. The black text represents the AggR sequence and blue text represents Rns. Amino acids are numbered from the N-terminus from 1 to 265. The symbols below show the similarity between the two proteins; (*) represents same amino acid at that position, (:) indicates amino acids have similar biochemical properties, (.) indicate that amino acids have somewhat similar biochemical properties and space means the two amino acids have different biochemical properties.

investigate if these amino acid residues are important in AggR-dependent regulation, substitutions I14T and N16D were introduced into AggR. In parallel, two versions of AggR carrying substitutions at C-terminal regions, were constructed following a study on MelR (Grainger *et al.*, 2004). The study on MelR showed that aspartic acid and threonine amino acids in the C-terminal region of MelR interact directly with σ_4 domain of the RNAP, and the alignment of MelR with AggR shows these positions correspond to 230 and 234 in AggR (Figure 5.2). To investigate whether AggR interacts with RNAP via same position as MelR, glutamine at position 230 and methionine at position 234 in AggR were substituted with glycine. The AggR mutants were cloned into pBAD30 using EcoRI and XbaI restriction sites, placing them under the control of the *araBAD* promoter.

The introduction of substitutions into proteins can be detrimental to their stability. In order to find out if any of the substitutions in AggR affected the stability of the protein, the ability of AggR derivatives to repress and activate the transcription of a number of promoters was examined. Bodero *et al.* (2007) found that expression of *nlpA* (which encodes an inner membrane protein present in ETEC and EAEC) is repressed by Rns and AggR. Using *in vitro* experiments, it was shown that Rns binds to the promoter fragment, around the transcription start site, and sterically blocks RNAP from binding to the promoter. To confirm that substitutions introduced into AggR in this study do not alter DNA binding, the regulation of *nlpA* was studied. A 408 bp promoter fragment, *nlpA*100, (carrying 356 bp upstream of *nlpA* and 52 bp of the *nlpA* coding sequence) was cloned into the *lacZ* expression vector pRW50 using EcoRI and HindIII. Plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* also carrying either pBAD/*aggR*, pBAD/*aggR*-I14T, pBAD/*aggR*-N16D, pBAD/*aggR*-Q230G, pBAD/*aggR*-M234G or pBAD24. The cells were grown in LB medium at 37°C with shaking until mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to

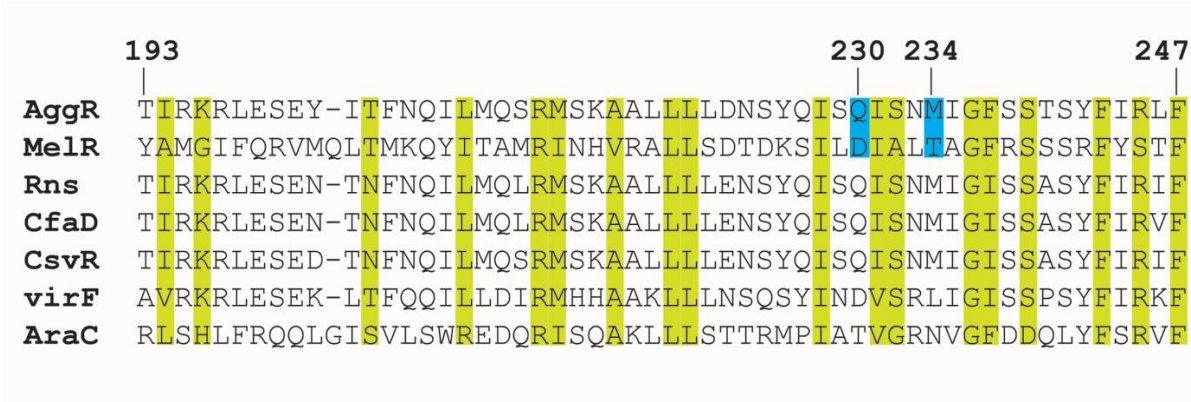


Figure 5.2 Multiple alignment of proteins belonging to the AraC/XylS family

The figure shows part of aligned sequence of AggR with some of the AraC/XylS family members. The numbering shown at in the top row is according to AggR amino acids. Green highlighted letters represent conserved amino acids in AggR according to Gallegos *et al.* (1997) and all aligned sequences at this position are also highlighted. Blue highlighted amino acids of MelR found important for interaction with σ_4 domain (Grainger *et al.*, 2004).Blue highlighted amino acids of AggR represent the positions of point substitutions in AggR.

induce the expression of *aggR*. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in the Figure 5.3A shows that the β-galactosidase levels measured in plasmids containing pRW50/*nlpA*100 decreased almost 2 fold in the presence of wild type AggR as well as all cells carrying the AggR mutants constructs. These results argue that all the AggR mutants are stable and can bind to DNA sequences effectively to repress expression from the *nlpA* promoter.

Since each mutant version of AggR represses the *nlpA* promoter, the ability of each AggR-mutation to activate transcription was studied using the *aafD*96 and *afaB*100 promoter fragments, and the semi-synthetic promoter fragment, *DAM*22. All the promoter fragments were cloned into pRW50 and plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR*-I14T, pBAD/*aggR*-N16D, pBAD/*aggR*-Q230G, pBAD/*aggR*-M234G or pBAD24 plasmids. Results shown in Figure 5.3B show that the β-galactosidase levels measured in cells containing pRW50/*DAM*22 showed similar increase in promoter activity for the pBAD/*aggR*, pBAD/*aggR*-Q230G and pBAD/*aggR*-M234G constructs when compared to AggR negative control. This indicates that AggR Q230G and M234G substitutions did not affect the ability of AggR to activate transcription, even though this region is part of the AggR HTH. However, little β-galactosidase activity was observed in cells carrying pBAD/*aggR*-I14T and pBAD/*aggR*-N16D when compared to cells containing pBAD/*aggR*. This finding highlights the importance of N-terminal amino acids in AggR-dependent activation.

Results in the Figure 5.4A and B show the β-galactosidase levels measured from cells lysates containing pRW50/*aafD*96 or pRW50/*afaB*100 carrying either wild type or mutant AggR constructs. As for the *aafD*96 and *afaB*100 promoter activity substantially decreased in the

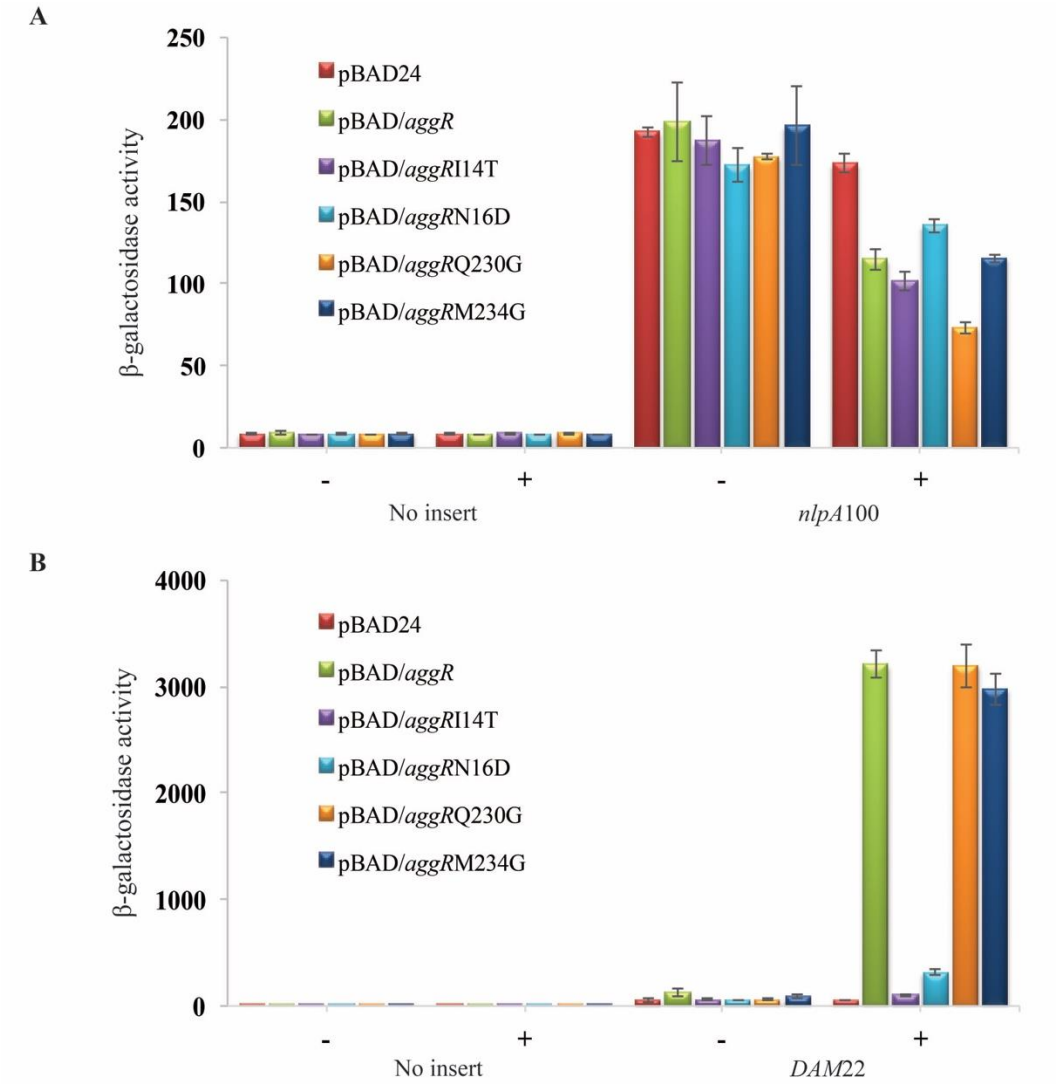


Figure 5.3 AggR-dependent regulation of the *nlpA* and *DAM*22 promoters

A. The panel shows β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or the *nlpA*100 promoter fragment cloned into pRW50. The cells also carried either pBAD/*aggR* (green bars), pBAD/*aggR*-I14T(purple bars), pBAD/*aggR*-N16D (light blue bars), pBAD/*aggR*-Q230G (orange bars), pBAD/*aggR*-M234G (blue bars) or pBAD24 (red bars) plasmids. Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or the *DAM*22 promoter cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars), pBAD/*aggR*-I14T(purple bars), pBAD/*aggR*-N16D (light blue bars), pBAD/*aggR*-Q230G (orange bars), pBAD/*aggR*-M234G (blue bars) or pBAD24 (red bars) plasmids. Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

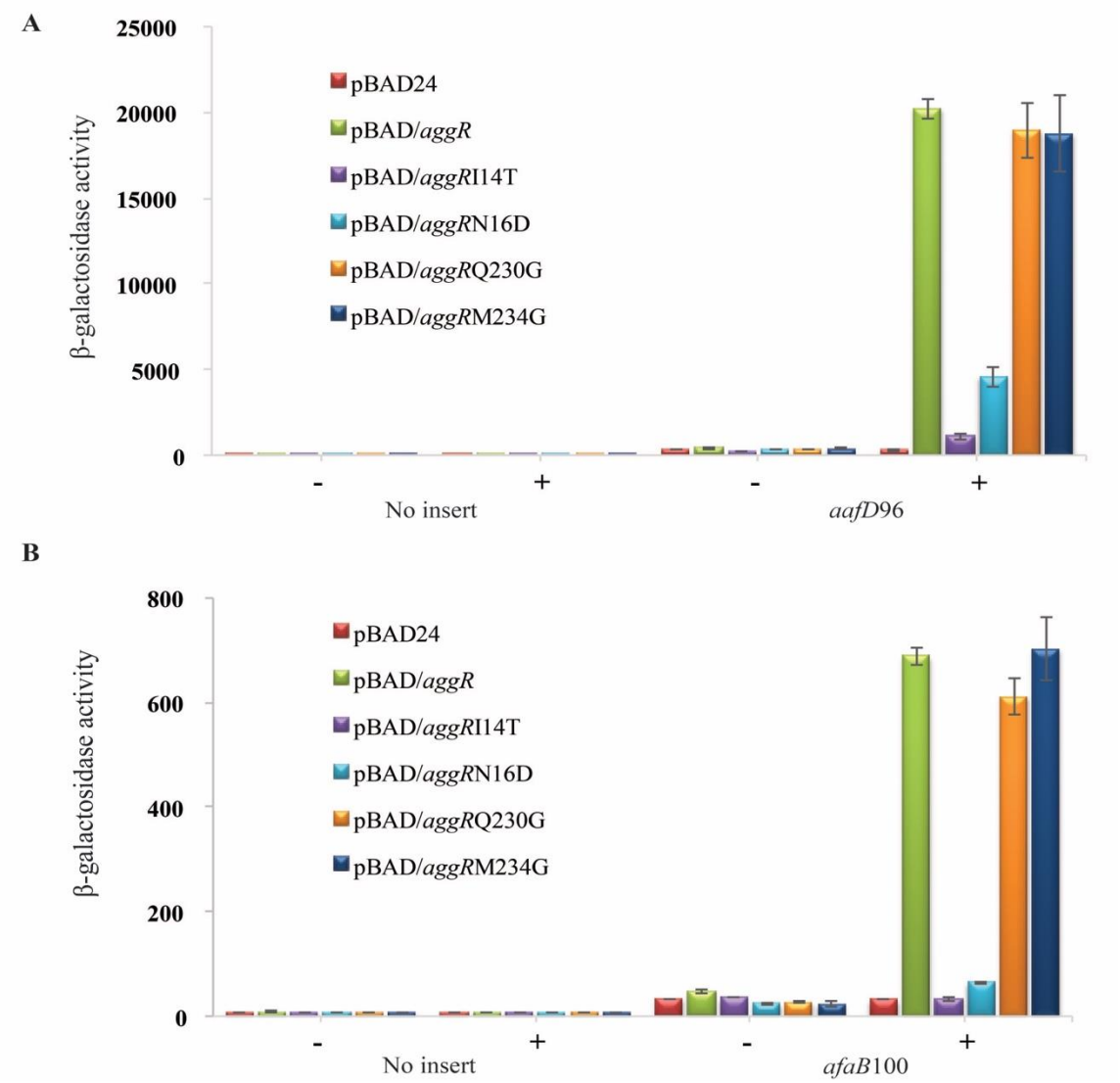


Figure 5.4 AggR-dependent regulation of the *aafD* and *afaB* promoters

A. The panel shows β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or the *aafD*96 promoter fragment cloned into pRW50. Cells also contained either pBAD/*aggR* (green bars), pBAD/*aggR*-I14T (purple bars), pBAD/*aggR*-N16D (light blue bars), pBAD/*aggR*-Q230G (orange bars), pBAD/*aggR*-M234G (blue bars) or pBAD24 (red bars) plasmids. Cells were grown in LB medium in the presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

B. The panel shows β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector pRW50 or the *afaB*100 promoter fragment cloned into pRW50. Cells also contained either pBAD/*aggR* (green bars), pBAD/*aggR*-I14T (purple bars), pBAD/*aggR*-N16D (light blue bars), pBAD/*aggR*-Q230G (orange bars), pBAD/*aggR*-M234G (blue bars) or pBAD24 (red bars) plasmids. Cells were grown in LB medium in the presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

presence of pBAD/*aggR*-I14T and pBAD/*aggR*-N16D when compared to cells containing pBAD/*aggR*. This finding again demonstrates the importance of N-terminal domain of AggR in transcription regulation. The β-galactosidase levels measured in cells containing pRW50/*aafD*96 or pRW50/*afaB*100 showed very little decrease in the presence of pBAD/*aggR*-Q230G and pBAD/*aggR*-M234G, when compared to the cells containing pBAD/*aggR* (Figure 5.4).

At all three promoters, the AggR mutations showed similar trends in their ability to activate transcription with the N-terminal mutations displaying a defect in activation, while mutations at C-terminal positions did not have an effect. These results show that N-terminal amino acids are important for AggR-dependent regulation, while the Q230G and M234G of C-terminal amino acids are not.

5.3 Investigation of *pet* expression

Pet is one of the important toxins secreted by EAEC 042 and plays an essential role in pathogenesis. The *pet* gene is present on the pAA2 plasmid, between Region 1 and Region 2 encoding AAF/II in EAEC 042 (Figure 1.4). Both regions are regulated by AggR but the *pet* gene is known to be regulated by global transcription regulators, CRP and FIS. The location of *pet* between two AggR-regulated operons could be important for gene regulation, or it could be a random insertion of gene at this location. This is why I began to investigate potential AggR effects on *pet* expression.

5.3.1 Investigation of the *pet* promoter fragment

In order to investigate any effects of AggR, the *pet* regulatory region was investigated (Rossiter *et al.*, 2011). A promoter fragment with the *pet* upstream sequence, *AER1* cloned into pRW50 using EcoRI and HindIII sites, was acquired from Douglas Browning and its

sequence is shown in Figure 5.5A. The recombinant plasmid pRW50/*AER1* was transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR* or pBAD24. Cell were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The β-galactosidase levels in cells containing pBAD/*aggR* and pRW50/*AER*1 showed no substantial difference, when compared to cells containing pBAD24 as AggR negative control and pRW50/*AER*1 (Figure 5.5B). These results show that the promoter fragment, *AER1* is not regulated by AggR. However, AggR-binding sites are located on the pAA2 plasmid, present both upstream (*afaB*) and downstream (*aafD*) of *pet* at ~1.8 kb and ~6.5 kb, respectively (Figure 1.4). These binding sites might come into play if DNA looping or coiling was involved. To investigate this, an experiment was designed to measure expression in EAEC 042 in presence and absence of AggR.

5.3.2 Deletion of *aggR* from EAEC strain DFB042

For the investigation of possible effects of AggR on *pet* expression, a comparative approach was used. I deleted the *aggR* gene from the EAEC strain DFB042 to construct EAEC DFB042Δ*aggR* (Table 2.1). The gene doctoring technique was used for deletion (section 2.12), and the deletion was confirmed by a phenotype characteristic, the ability to form a biofilm in the presence of AggR (section 2.13).

For biofilm formation, overnight cultures of EAEC DFB042, EAEC DFB042Δ*aggR*, and *E. coli* K-12 were diluted 1:100 in 5 ml of DMEM media with a high glucose concentration (Sigma Aldrich). The cultures were incubated at 37°C for one hour and then each strain was added to 8 wells (A to H) of a 96-well microtitre plate. The microtitre plate was sealed with paraffin film to avoid dehydration and incubated at 37°C for overnight. Biofilms were stained

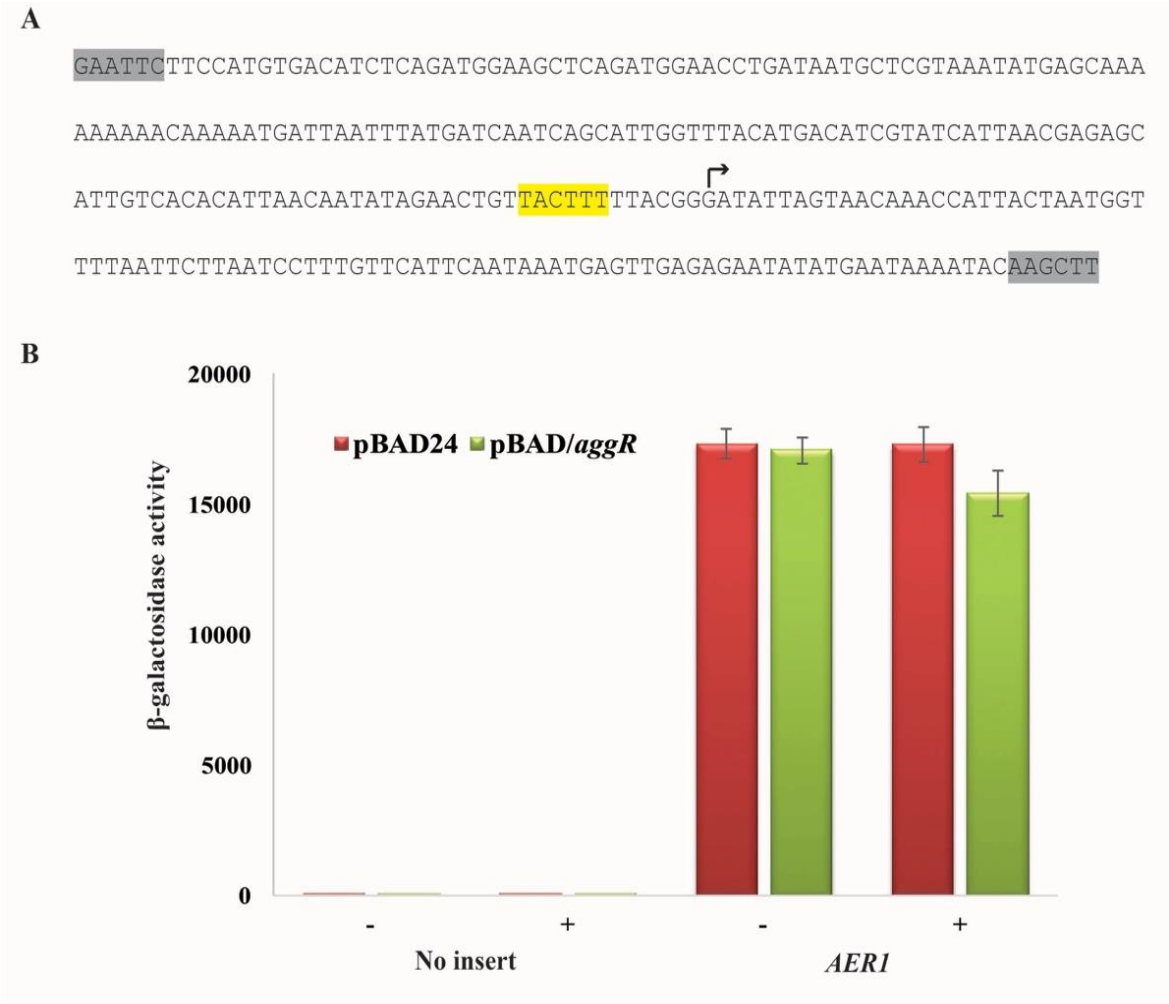


Figure 5.5 Measurement of AggR-dependent activity of *AER1* promoter fragment

A. The panel shows the base sequence of the *AER1* fragment (Rossiter *et al.,* 2011). The -10 hexamer element is highlighted yellow and the restriction sites grey. The bent arrow indicates the position of the transcription start site.
B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing *AER1* promoter cloned into *lacZ* expression vector, pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

with 0.1% (w/v) crystal violet, rinsed with water to remove excess stain. Then ethanol/acetone was added in each well of microtitre plate to dissolve the dye retained by biofilm. The absorbance of blue dye dissolved in ethanol/acetone was measured using 595 nm filter in Labsystems Multiskan MS plate reader (Thermo Fisher Scientific Inc.) and it corresponded to biofilm formation phenotype of the strain.

The results of the plate showed that there was dark blue colour in the wells containing, EAEC DFB042 and there was faint blue colour in the wells containing DMEM medium only, *E. coli* K-12 and EAEC DFB042Δ*aggR* (Figure 5.6A). When the absorbance of the solutions, in the microtitre plate wells, was measured at 595 nm using Labsystems Multiskan MS plate reader (Thermo Fisher Scientific Inc.), EAEC DFB042 containing wells showed high absorbance when compared to absorbance measured from the wells containing DMEM medium, *E. coli* K-12 and EAEC DFB042Δ*aggR* (Figure 5.6B). These results showed that biofilm is only formed in wells containing EAEC DFB042 and as AggR plays an important role in biofilm formation, so it proves that *aggR* has been deleted in EAEC DFB042Δ*aggR*.

5.3.3 Detection of *pet* expression

To investigate the effect of AggR on *pet* expression, two EAEC strains DFB042 and EAEC DFB042Δ*aggR* were compared using western blotting. The overnight cultures of EAEC DFB042, EAEC DFB042Δ*aggR* and *E. coli* K-12 were diluted 1:100 in 50 ml of DMEM medium with high glucose concentration and incubated at 37°C for 6 hours. The cultures were transferred into 50 ml Falcon tubes, cells were harvested and supernatant was filtered using 0.2 μm syringe filters. The supernatant was acid precipitated and electrophoresed by SDS-PAGE. Proteins were transferred to nitrocellulose membrane and detected using anti-passenger domain antibodies raised in rabbit. The results of the western blotting in Figure 5.7A showed that there are bands present, indicating the presence of a protein of ~104 kda,

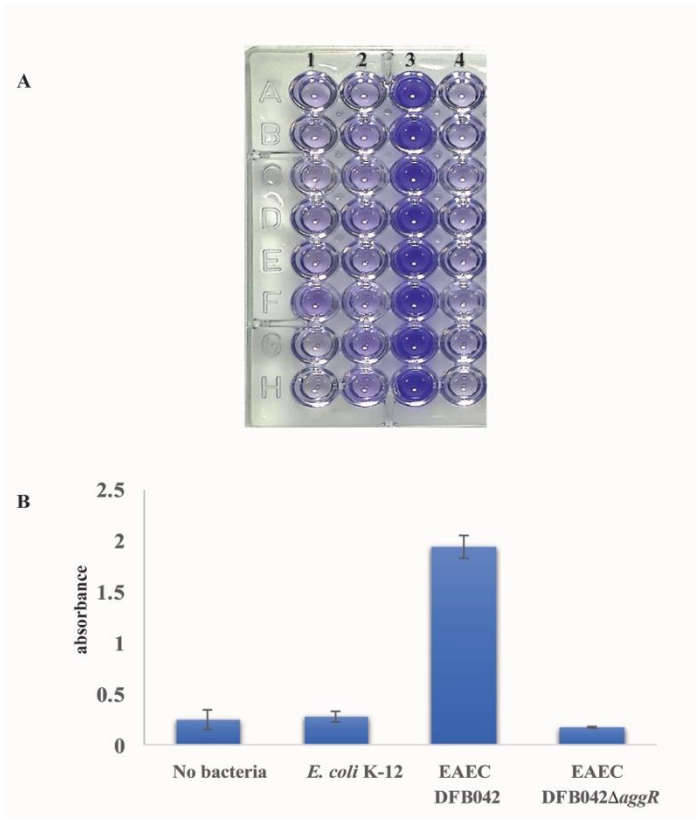


Figure 5.6 Biofilm formation by EAEC strain DFB042

A. The panel shows part of 96 well microtiter plate, each column of the plate contains technical repeats of same sample. Column 1 contains only DMEM, coulmn 2 contains *E. coli* K-12 in DMEM, column 3 has EAEC strain DFB042 and column 4 contains EAEC DFB042Δ*aggR*. The blue colour development corresponds to biofilm formed, and stained with crystal violet.
B. The bar chart represents the absorbance of blue colour of microtiter plate columns shown in panel 1, at 595 nm using plate reader. The error bars represent the difference in technical repeats in one column.

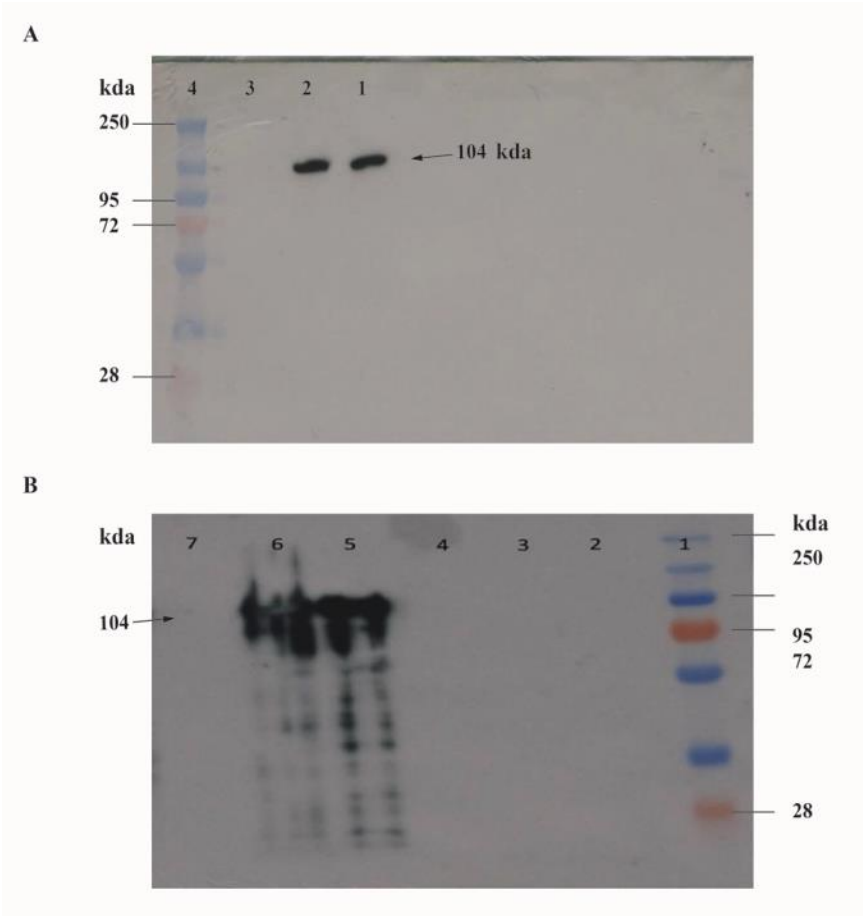


Figure 5.7 *pet* expression in different *E . coli* strains

A. The panel shows western blot scan detected using antipassenger domain antibodies for Pet. For this experiment, the bacterial strains were grown in DMEM medium at 37°C for 6 hours. Lane 1, 2 and 3 contained proteins from supernatants of EAEC DFB042, EAEC DFB042Δ*aggR, E. coli* K-12, respectively. Lane 4 shows the prestained protein ladder and band sizes are labelled.
B. The panel shows western blot scan detected using antipassenger domain antibodies for Pet, when the bacterial strains are grown in DMEM medium supplemented with tryptone at 37°C. Lane 1 shows the prestained protein ladder and bands sizes are labelled. Lane 2, 3 and 4 show no detectable bands and these contained proteins from supernatants of EAEC DFB042, EAEC DFB042Δ*aggR* and EAEC DFB042 (pBAD/*aggR*), respectively grown in DMEM medium supplemented with tryptone and arabinose (0.2% (w/v)) at 37°C for 2 hours. Lane 5 and 6 show bands while lane 7 shows no detectable band. These contained proteins from supernatants of EAEC DFB042, EAEC DFB042Δ*aggR* and EAEC DFB042 (pBAD/*aggR*), respectively grown in DMEM medium supplemented with tryptone and arabinose (0.2% (w/v)) at 37°C for 6 hours.

and it is detected using anti-passenger domain antibodies of *pet*, suggests, it is Pet. It was detected in EAEC DFB042, EAEC DFB042Δ*aggR* and no band was detected for *E. coli* K-12 (Figure 5.7A). This experiment showed the specificity of antibodies but no difference was observed in *pet* expression in both strains EAEC DFB042 and EAEC DFB042Δ*aggR*.

Moreover, when EAEC DFB042 was grown in DMEM medium, low level of *pet* expression was observed. To increase *pet* expression in DMEM medium, it was supplemented with tryptone (Betancourt-Sanchez and Navarro-Garcia, 2009). So, another experiment was designed using DMEM medium supplemented with tryptone to express *pet* in EAEC DFB042 and EAEC DFB042Δ*aggR*. In this experiment, another strain EAEC DFB042 with (pBAD/*aggR*) was also used to overexpress AggR. The overnight cultures of EAEC DFB042, EAEC DFB042Δ*aggR* and EAEC DFB042 with (pBAD/*aggR*) were diluted 1:100 in 50 ml of DMEM medium with high glucose concentration, and supplemented with tryptone and 0.2% (w/v) arabinose. For each of the 3 strains mentioned above, 2 set of cultures diluted and incubated at 37°C with shaking. One set of cultures was incubated for 2 hours and the other was incubated for 6 hours. After incubation, the cultures were transferred into 50 ml Falcon tubes and centrifuged at 2844 g for 15 minutes. The supernatant of each culture was filtered after centrifugation, acid precipitated, and run on SDS-PAGE. The protein was transferred to nitrocellulose membrane and detected using anti-passenger domain antibodies.

The results of western blotting showed that there were no detectable bands seen, in the supernatant of cultures, when the supernantant was analysed after 2 hours of incubation (Figure 5.7B). While the *pet* expression was detected in the supernatant of EAEC DFB042 and EAEC DFB042Δ*aggR* cultures incubated for 6 hours at 37°C, but no *pet* expression was seen for EAEC DFB042(pBAD/*aggR*) (Figure 5.7B). These results suggest that the over expression of AggR in EAEC DFB042(pBAD/*aggR*) has inhibited *pet* expression. This shows

that AggR has a repressive effect on *pet* expression, and high AggR concentration is important for *pet* repression. The major bands detected were thick and degraded protein can be seen all the way down in the gel. This could be due to breakdown of Pet during acid precipitation.

These findings need to be validated with more sensitive techniques*, e.g.* quantitative real time PCR and also making more EAEC DFB042 mutants, like deletion of confirmed AggR-binding sites present upstream (*afaB* promoter) and downstream (*aafD* promoter) of *pet* that are possibly involved in repression. I could not validate the results by alternate methods, due to time constraints.

5.4 Discussion

AggR is one of AraC family members and it regulates many virulence determinants of EAEC strains. AggR-dependent activation of promoters was studied in this chapter by introducing point mutations in the N-terminal and C-terminal domains of AggR. The N-terminal domain mutations were reported to be important for Rns-dependent activation of promoters in ETEC (Basturea *et al.*, 2008). The results of present study also showed a similar trend for AggR-dependent promoter activation, as AggR-dependent promoters are not activated with AggR-I14T and AggR-N16D but these mutants can still repress expression of *nlpA*100. The lack of activation by mutants indicate that these amino acids are important for AggR-dependent activation and repression of *nlpA*100 shows that these mutants are stable proteins. The N-terminal domain of AraC family members are usually important for dimerization, however, there are conflicting opinions, as whether AggR exists as a homo-dimer. Thus, it is unclear of the N-terminal domain of AggR interacts with RNAP on AggR-activated promoters or if it is involved in dimerization. A study by Grainger *et al.* (2004), showed the

C-terminal domain of MelR is important but the results of present study showed that the mutations in corresponding positions in AggR, showed no effect on AggR-dependent activation (Figure 5.2 and 5.3B). These results indicate that AggR might not interact in the same manner as MelR or these proteins have different folding patterns that might have affected the positon of interacting amino acids.

The expression of *pet* is effected by over expression of AggR in EAEC DFB042 (Figure 5.7). This suggests that AggR plays a role in *pet* expression, and *pet* location between two AggR-regulated regions has some physiological importance. This finding is not mentioned by Morin *et al.* (2013), who found that 44 genes of EAEC 042 are regulated by AggR. The possible explanation is, AggR expresses when EAEC 042 is grown in DMEM medium and does not express when the bacterium is grown in LB medium, whilst *pet* is exactly the opposite to AggR. The *pet* toxin expresses when EAEC 042 is grown in LB medium and does not express well when grow in DMEM medium (Betancourt-Sanchez and Navarro-Garcia, 2009). So the effect of AggR, on *pet* expression, might have been masked due to this fact.

This study has increased our understanding of AggR-dependent activation and the study on *pet* expression has depicted that AggR might has more crucial rule in EAEC virulence than predicted form previous studies.

Chapter 6

Organisation of the AggR-regulated fimbrial promoter in EAEC strain 17-2

6.1 Introduction

AggR is the transcription factor that regulates fimbrial gene expression in many strains of EAEC (Zamboni *et al.,* 2004). The attachment adherence fimbriae I (AAF/I) is found in EAEC strain 17-2. The AAF/I is formed by a chaperone usher system and the genes that encode this system in EAEC 17-2 are present on the pAA plasmid in two clusters, *i.e.* Region 1 and Region 2 (Figure 6.1). The genes that encode structural components of AAF/I are present in Region 1 and the transcription regulator, AggR that regulates this system is present in Region 2 (Savarino *et al.,*1994). The aggregation pattern of EAEC 17-2 was disrupted by deleting either the genes that encode the chaperone usher system or AggR itself. These results proved that fimbrial genes expression requires AggR, as well as the structural proteins encoding genes, but to date no study has investigated the AggR-dependent promoters from EAEC 17-2 in detail. The transcription factor AggR, from both strains EAEC 042 and EAEC 17-2 is similar with only four amino acid differences (Figure 6.2). In this study, the *aggD* promoter from EAEC 17-2 was studied in *E. coli* K-12 strain BW25113, and AggR from EAEC 042 was used to induce the AggR-dependent regulation.

EAEC strains have been found in foodborne outbreaks and the strains found in outbreaks also have AAF/I. Analysis of the promoter region upstream of *aggD* from EAEC 17-2 had shown that there is a run of six 5'-TCAAGT-3' hexamer repeats present (Savarino *et al.*, 1999). Moreover, the strain of EAEC that was involved in the foodborne outbreaks in Europe has AAF/I and the upstream sequence had greater number of 5' -TCAAGT-3' repeats. This finding of the *aggD* promoter made it important to look into the comparative activity of the promoter regions from different strains with different number of repeats that might be important for gene expression and it will be discussed in this chapter.

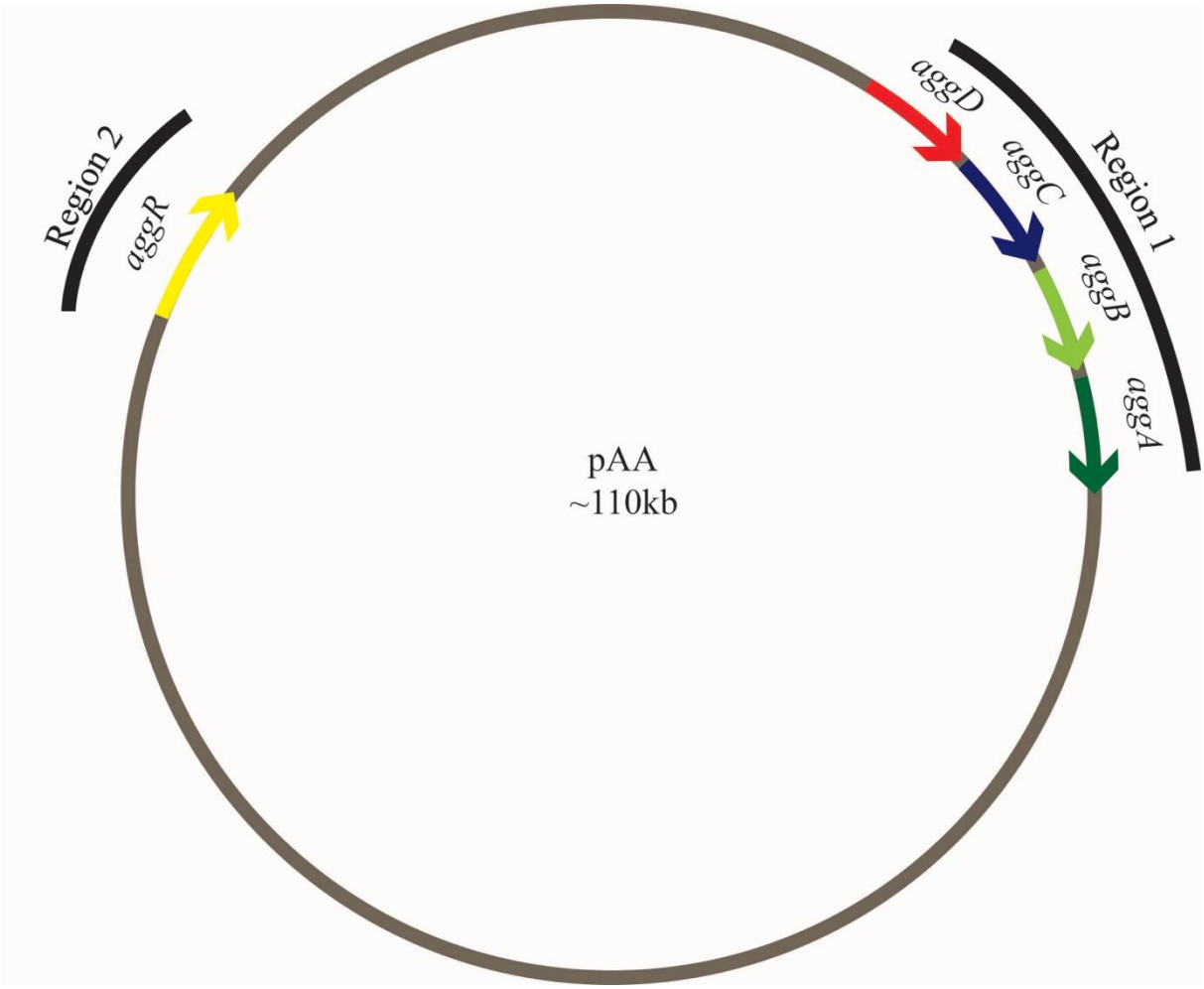


Figure 6.1 A schematic representation of Region 1 and Region 2 on the pAA plasmid from EAEC 17-2

The figure shows a schematic diagram of the pAA plasmid from EAEC strain 17-2 and the two regions that are important for the expression of the attachment adherence fimbriae type I (AAF/I). Region 1 encodes *aafD* (the chaperone protein), *aafC* (the usher protein) *aafB* (the fimbrial adhesin) and *aafA* (the fimbrial subunit) genes and Region 2 encodes *aggR*, transcription activator. This diagram is not to scale and modified from Figure 1.3.

Figure 6.2 Alignment of AggR from EAEC 042 and EAEC 17-2

The figure shows amino acid sequence of AggR from EAEC 042 and EAEC 17-2. Standard abbreviations are used to represent amino acid. The black colour alphabets represent AggR sequence from EAEC 042 and blue colour alphabets represent AggR sequence from EAEC 17-2. Amino acids are numbered from N-terminal to C-terminal from 1 to 265 and shown above black alphabets. The symbols below blue alphabets show the similarity between two proteins; (*) represents the same amino acid at that position, (:) represents amino acid replaced has similar biochemical properties, (.) represent amino acid replaced has somewhat similar biochemical properties and space means the two amino acids have different biochemical properties.

In the previous chapters, I have discussed, how AggR regulates the fimbrial genes in EAEC 042 and in this chapter, fimbrial gene regulation of EAEC 17-2 will be discussed. AggR also regulates the other fimbrial systems AAF/III in EAEC strain 55989 and AAF/IV in EAEC C1010-00 and the promoters of these systems are also predicted and aligned with AAF/I and AAF/II promoters in the last part of this chapter.

6.2 Analysis of the regulation of Region 1 on pAA from EAEC 17-2

6.2.1 Analysis of the DNA sequence upstream of the *aggDCBA* operon

Region 1 encodes for the structural genes of the EAEC 17-2 fimbriae system and Region 2 encodes for the transcription regulator, AggR (Figure 6.1) (Nataro *et al*., 1992). It has been proposed that the AggR-regulated promoter of Region 1 is upstream of *aggD* and all the structural genes are transcribed as a single operon. To investigate this, a 413 bp DNA fragment, *aggD*100, which carries the DNA upstream sequence of *aggD*, was cloned into the *lacZ* expression vector, pRW50. Bioinformatic analysis indicated that there are two potential AggR-binding sites on the forward strand and one potential AggR-binding site on the reverse strand of the *aggD*100 promoter fragment (Figure 6.3). The schematic diagram of the *aggD100* promoter fragment is shown in Figure 6.4A.

The *aggD*100 promoter fragment was cloned into pRW50 and the recombinant plasmid pRW50/*aggD*100 was transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* carrying either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. During growth with 0.2% (w/v) arabinose, β-galactosidase levels in cells containing pBAD/*aggR* and pRW50/*aggD*100

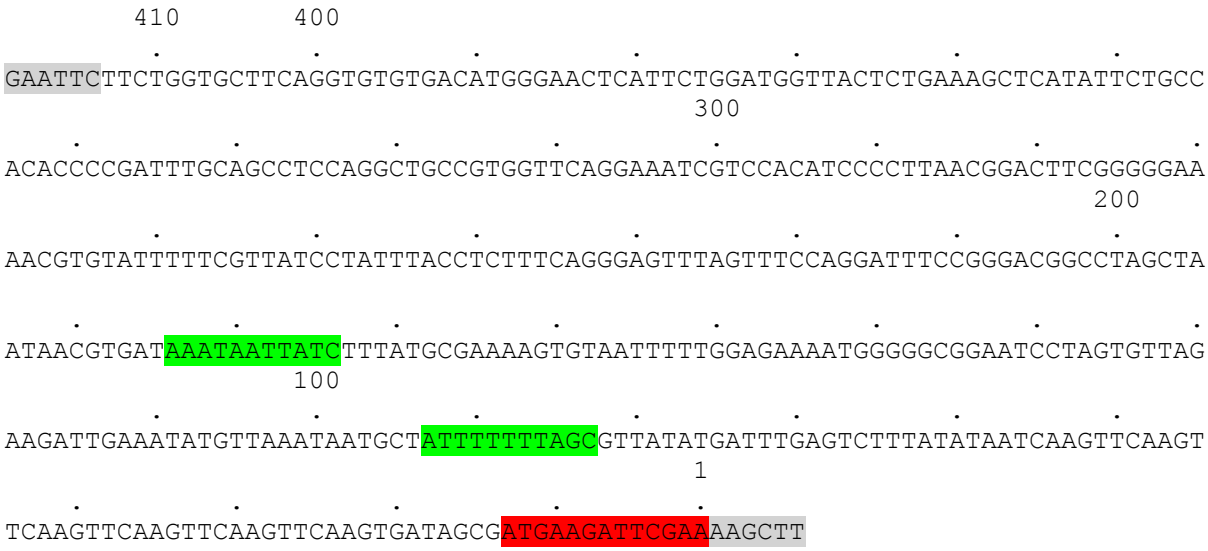


Figure 6.3 The DNA base sequence upstream of the *aggD* gene

The figure shows the 413 bp DNA sequence of the *aggD*100 promoter fragment cloned from EAEC strain 17-2. The open reading frame sequence is highlighted red and the potential AggR-binding sites are green. The restriction sites EcoRI and HindIII are highlighted grey. This figure is adapted from Figure 2.8.

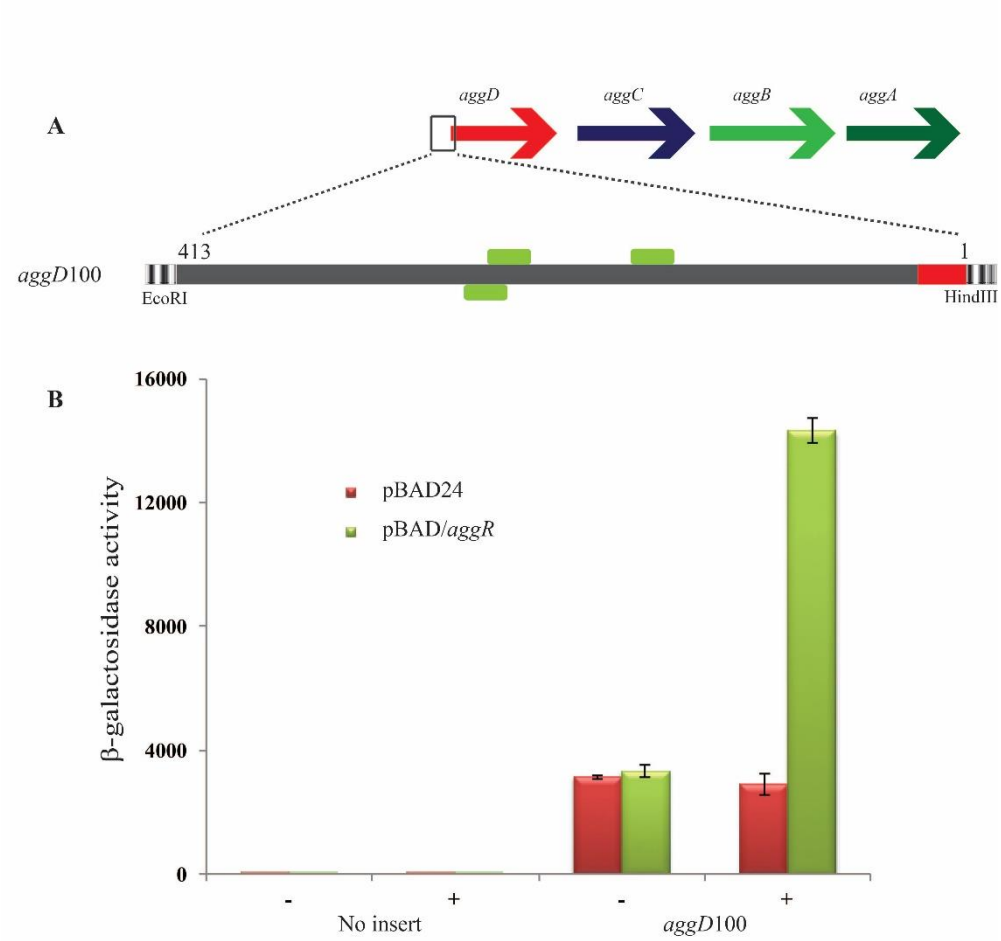


Figure 6.4 Analysis of the *aggD* promoter from Region 1 of pAA

A. The upper line shows the schematic arrangement of the fimbrial genes of pAA plasmid from EAEC 17-2 from the Region 1. The lower part illustrates *aggD*100 promoter fragment. The grey bar represents upstream sequence of *aggD* and the red part of the bar represents 13 bp sequence of *aggD* coding sequence present on the fragment. Potential AggR-binding sites are indicated as green boxes. The DNA sequence in this diagram is numbered above fragment from 1 to 413 and the fragment was cloned using EcoRI and HindIII restriction sites into pRW50. This diagram is not to scale.
B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aggD*100 promoter fragment cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

increased almost 5-fold, when compared to cells containing pBAD24 as the AggR minus control (Figure 6.4B). This shows that the *aggD*100 fragment contains an AggR-regulated promoter, and this AggR-regulated promoter is present upstream of *aggD*.

6.2.2 Identification of the minimal regulatory region of the *aggD* promoter necessary for AggR-mediated activation

To investigate the AggR-regulated promoter upstream of *aggD* in Region 1 of pAA plasmid, three nested deletions were made, (*aggD*99 to *aggD*97), by deleting sequence from the upstream of the *aggD*100 promoter fragment (Figure 6.5A and B). Once more, each promoter fragment (*aggD*99 to *aggD*97) was cloned into the *lacZ* expression vector pRW50. Plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* containing either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 6.5C show that the β-galactosidase activities measured from lysates of cells containing *aggD*100, *aggD*99 and *aggD*98 cloned into pRW50 were increased 5-fold when AggR was induced. However, in cells containing pRW50/*aggD*97 no increase was observed when AggR was induced. These results suggest that the AggR-binding site or the promoter itself has been deleted or disrupted in the *aggD*97 fragment whilst it is intact in *aggD*98 fragment.

The results of deletion analysis also showed a decrease in promoter activity in the absence of AggR as the number of bases are reduced. This may be due to the presence of second promoter that has been deleted or may be disruption of UP-element sequence in smaller fragments.

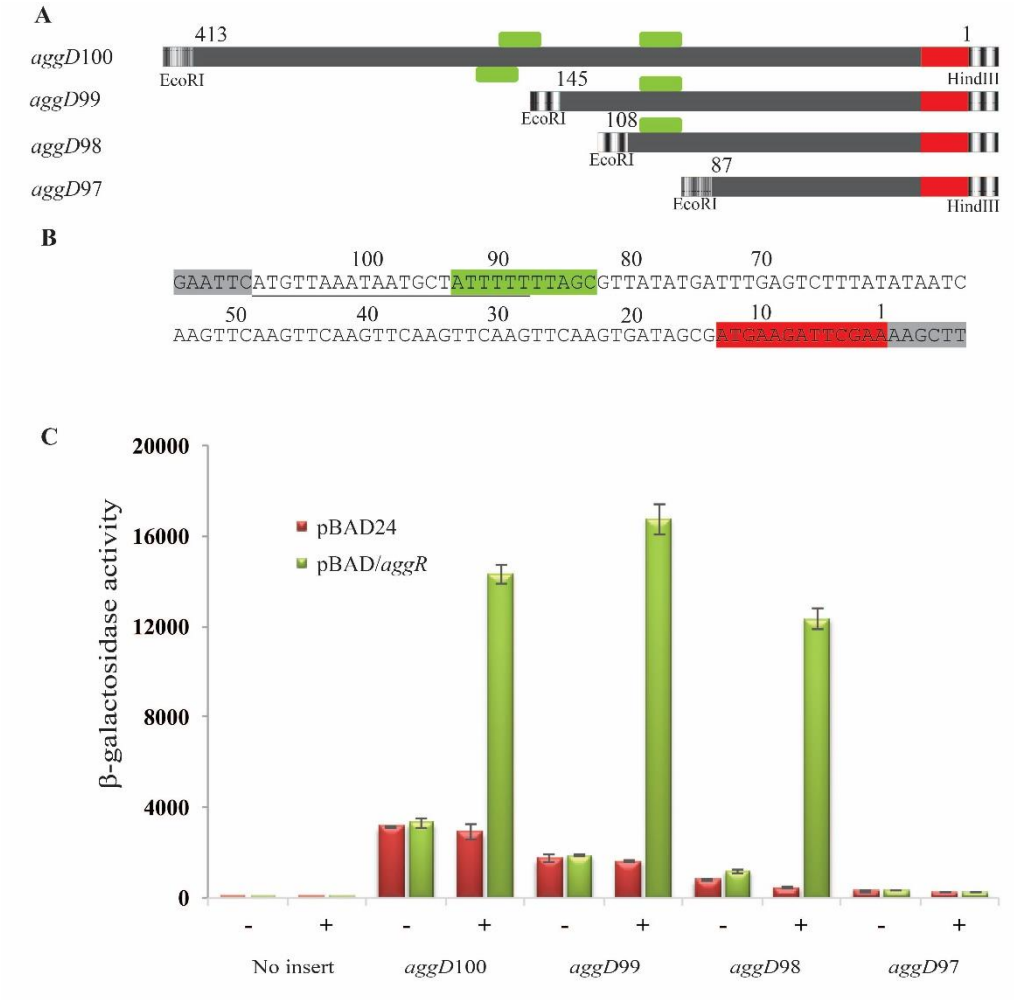


Figure 6.5 Deletion analysis of the *aggD* regulatory region

A. The panel illustrates the *aggD*100*, aggD*99*, aggD*98 and *aggD*97 promoter fragments. The grey bar represents upstream sequence of *aggD* and the red part of the bar represents 13 bp sequence of *aggD* coding region present on the fragment. Green boxes illustrate potential binding sites for AggR. All the fragments were cloned into pRW50 using EcoRI and HindIII restriction sites. The diagram is not to scale.

B. The panel shows the sequence of *aggD*98 fragment and the AggR-binding site is highlighted green, restriction sites are grey and *aggD* sequence red. The underline sequence adjacent to EcoRI site was deleted to in the *aggD*97 promoter fragment.

C. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aggD*100 promoter fragment derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

6.2.3 Identification of an essential AggR-binding site at the *aggD* promoter using mutational analysis

Inspection of the DNA sequence upstream of *aggD* suggests that an essential AggR-binding sequence is present on the *aggD*98 promoter fragment, which has been disrupted in the *aggD*97 promoter fragment. This AggR-binding site is similar to the Rns-binding weblogo prepared by Munson (2013) and the AggR-binding site identified by Morin *et al.* (2010), with an upstream A at position 3, a downstream G at position 12, instead of a T, as found at other sites (Figure 6.6) (Morin *et al.,* 2010; Munson, 2013).

Previously, AggR-binding sites have been confirmed by introducing point mutations at the conserved T residues present in the downstream part of the AggR-binding site (Morin *et al.*, 2010). To investigate this potential AggR-binding site in more detail, a point mutation was introduced at position 10 of the binding site (Figure 6.6). This corresponds to position 86 on the *aggD*98 promoter fragment (Figure 6.7A and B) and the recombinant *aggD*98-86*C* fragment was made by substituting the T with C at position 86. The *aggD*98-86*C* fragment was cloned into pRW50 and plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* containing either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 6.7C showed that the β-galactosidase activity measured in cells containing pRW50/*aggD*98-86*C,* does not increase when AggR expression was induced. This indicates, that this site, is essential for AggR-dependent transcription activation at the *aggD* promoter.

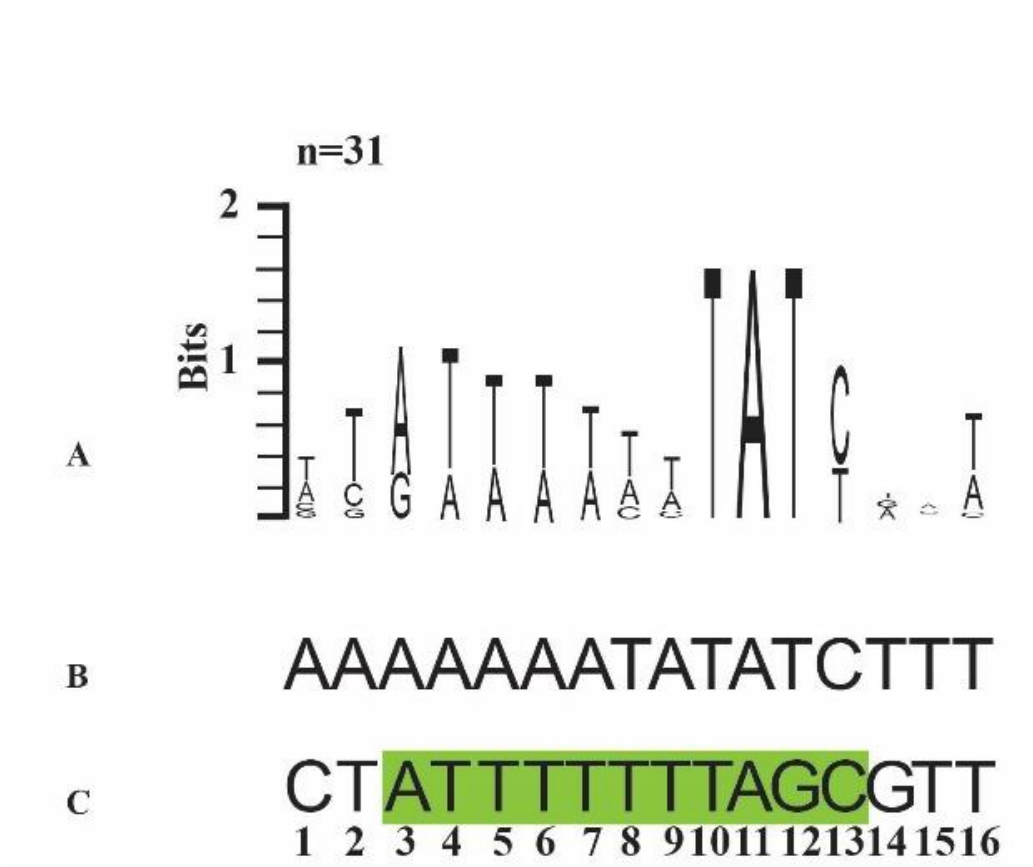


Figure 6.6 Alignment of the AggR-binding site from the *aggD* regulatory region with the consensus sequences of Rns and AggR

A. The panel illustrates consensus sequence from the Rns-binding logo devised by (Munson, 2013). The height of each letter corresponds to no of times each letter is present in the 31 Rns-binding sites identified in the ETEC genome.
B. The panel shows the sequence of the essential AggR-binding site identified at the *aggR* promoter of EAEC 042 (Morin *et al.*, 2010).
C. The panel shows the proposed AggR-binding site from the *aggD* promoter. The highlighted sequence shows the potential AggR-binding site (from position 3 to position 13) suggested to be important for AggR-binding and activation. The flanking sequence of AggR-binding site from *aggD* also shown.

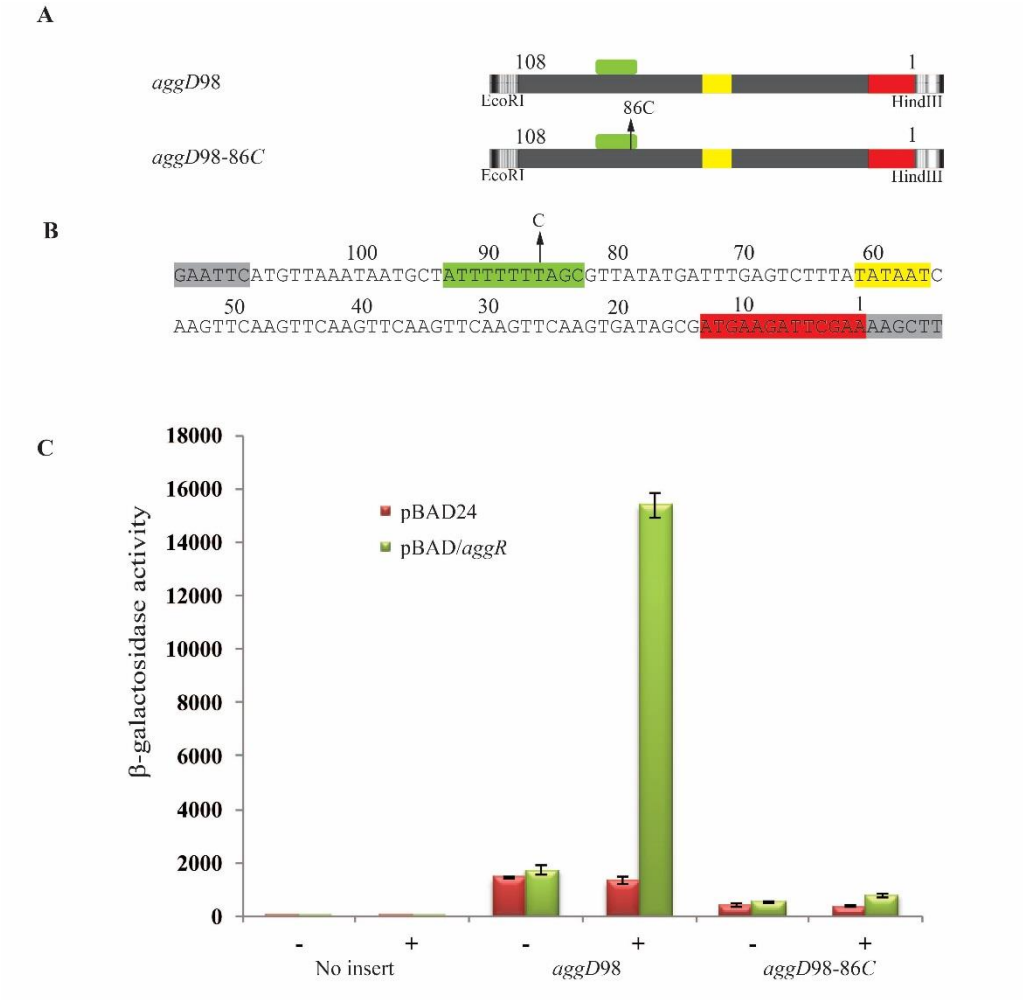


Figure 6.7 Mutational analysis of the AggR-binding site at the *aggD* promoter

A. The figure illustrates the *aggD*98 and *aggD*98-86*C* promoter fragments. The grey bars represent upstream sequence of *aggD* and the red part of the bars represent 13 bp sequence of *aggD* coding sequence present on the promoter fragment. Green boxes illustrate potential binding sites of AggR and yellow boxes represent the proposed -10 element. The arrow, indicates the position of point mutation.
B. The panel shows the base sequence of the *aggD*98 fragment. The AggR-binding site is highlighted green, the potential -10 hexamer element yellow, the restriction sites grey and the *aggD* coding sequence red. The arrow with text indicates the position of the point mutation introduced into *aggD98* promoter fragment.
C. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing pRW50 or *aggD* promoter derivatives cloned into *lacZ* expression vector, pRW50. Cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activity was measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

6.2.4 Identification of the -10 hexamer element on the *aggD* promoter by mutation analysis

A potential -10 hexamer element was identified 21 bp downstream of the AggR-binding site based on its resemblance to the -10 hexamer consensus sequence (TATAAT). A point mutation (A to C) at position 2 of the -10 hexamer element was introduced which corresponds to position 60 on *aggD*98 promoter fragment and resulted in the *aggD*98-60*C* promoter fragment (Figure 6.8A and B).

Plasmids were transformed into the Δ*lacZ E. coli* K-12 strain BW25113, carrying either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. Results in Figure 6.8C show that little β-galactosidase activity could be detected in cells containing pRW50/*aggD*98-60*C,* confirming that this is the -10 hexamer element of the *aggD* promoter. The basal activity of promoter has decreased substantially, indicating that promoter elements have been disrupted.

6.3 Comparison of the *aggD* promoter from two EAEC strains

Promoter analysis of the upstream sequence of *aggD* had shown that there was a run of six 5' -TCAAGT-3' hexamer repeats (Savarino *et al.,* 1994). Interestingly, the strain that was involved in the foodborne outbreaks in Europe also had AAF/I and the upstream sequence of *aggD* had a higher number of 5'-TCAAGT-3' hexamer repeats. To investigate the role these repeats in AggR-dependent promoter regulation, constructs with different number of repeats were made to study promoter regulation.

In this study, the starting promoter *aggD*99 from EAEC 17-2 that has six of these hexamer

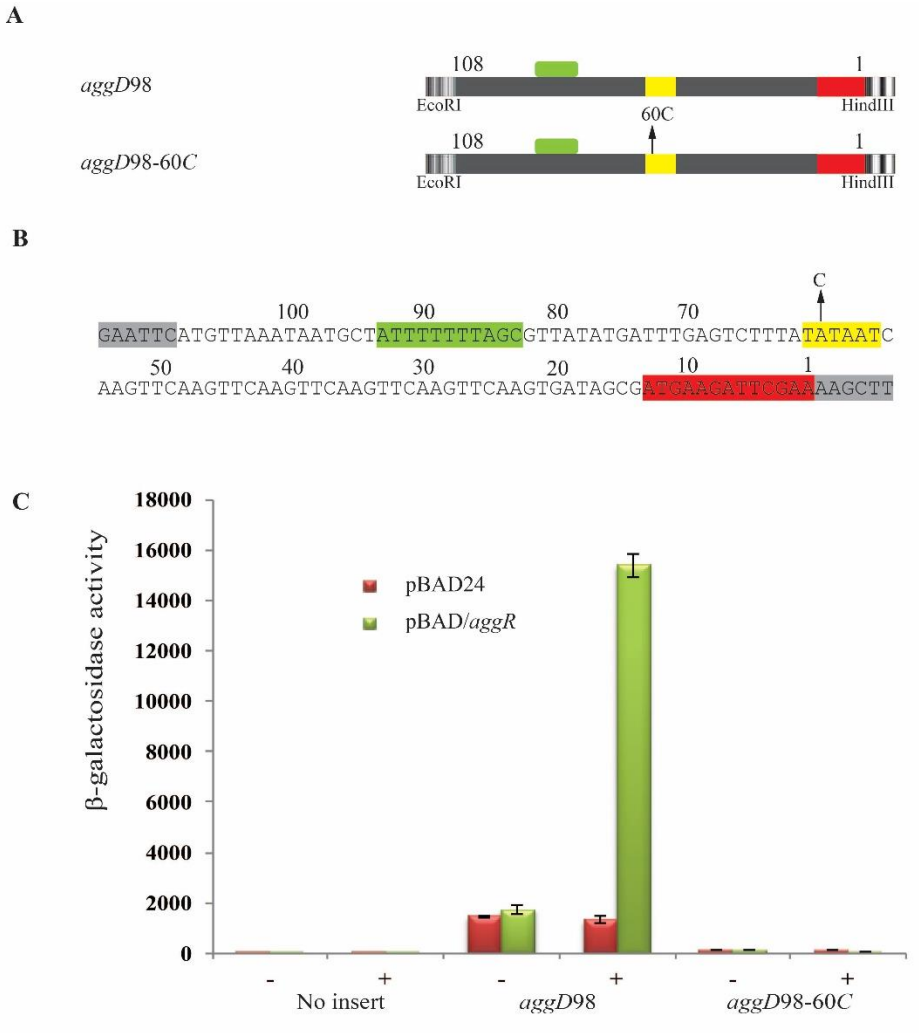


Figure 6.8 Identification of the -10 hexamer element of the *aggD* promoter

A. The figure illustrates the *aggD*98 and *aggD*98-60*C* promoter fragments. The grey bars represent the upstream sequence of *aggD* and the red part of the bars represent the 13 bp sequence of the *aggD* coding sequence present on the promoter fragment. Green boxes illustrate potential binding sites of AggR and yellow boxes represent the -10 hexamer element. The arrow, with text indicates position of point mutation.

B. The panel shows the sequence of the *aggD*98 fragment, the AggR-binding site is highlighted green, the -10 hexamer element yellow, the restriction sites grey and *aggD* sequence red. The arrow, with text indicates position of point mutation introduced into the *aggD98* promoter fragment and the base that has been substituted to make point mutation.

C The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aggD* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

repeats in the promoter. An *aggD* promoter based on the sequence form EAEC strain C227-11, that has 15 repeats was synthesised by Life Technologies ^{TM} and it is labelled as *aggD*101 (Figure 6.9A). A derivative of *aggD*101 was made by mega-primer PCR (section 2.4.3) with 14 repeats and denoted *aggD*102.

The *aggD* fragments were cloned into pRW50 and plasmids were transformed into the *E. coli* K-12 strain BW25113 Δ*lac,* containing either pBAD/*aggR* or pBAD24. The cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase and 0.2% (w/v) arabinose was added to LB medium to induce the expression of *aggR* where appropriate. Promoter activities were determined by measuring β-galactosidase levels in lysates of cells containing these plasmids. The results show that during growth in LB with 0.2% (w/v) arabinose, β-galactosidase levels in cells containing pBAD/*aggR* with pRW50/*aggD*99, pRW50/*aggD*101 or pRW50/*aggD*102 increased almost 7-fold in the presence of AggR, when compared to cells containing pBAD24 as AggR negative control (Figure 6.9B). The β-galactosidase activity showed that the number of repeats in the different promoters do not affect promoter regulation in this experiment. So, it indicates that these repeats may not be playing any critical role in gene regulation.

6.4 Alignment of AggR-dependent AAF promoters from different strains

Five AAF systems (AAF/I-AAF/V) have been identified on the basis of fimbrial antigen differences, present in different strains of EAEC. AAF/I and AAF/II are present in EAEC 17-2 and EAEC 042, respectively, and the promoters of these strains have been discussed earlier. AAF/III and AAF/IV are present in EAEC 55989 and EAEC C1010-00, respectively, and the fimbrial promoters sequences have been predicted on the basis of results of this study. AAF/V present in strain EAEC C338-14 and I could not find annotated *aggD* sequence for this

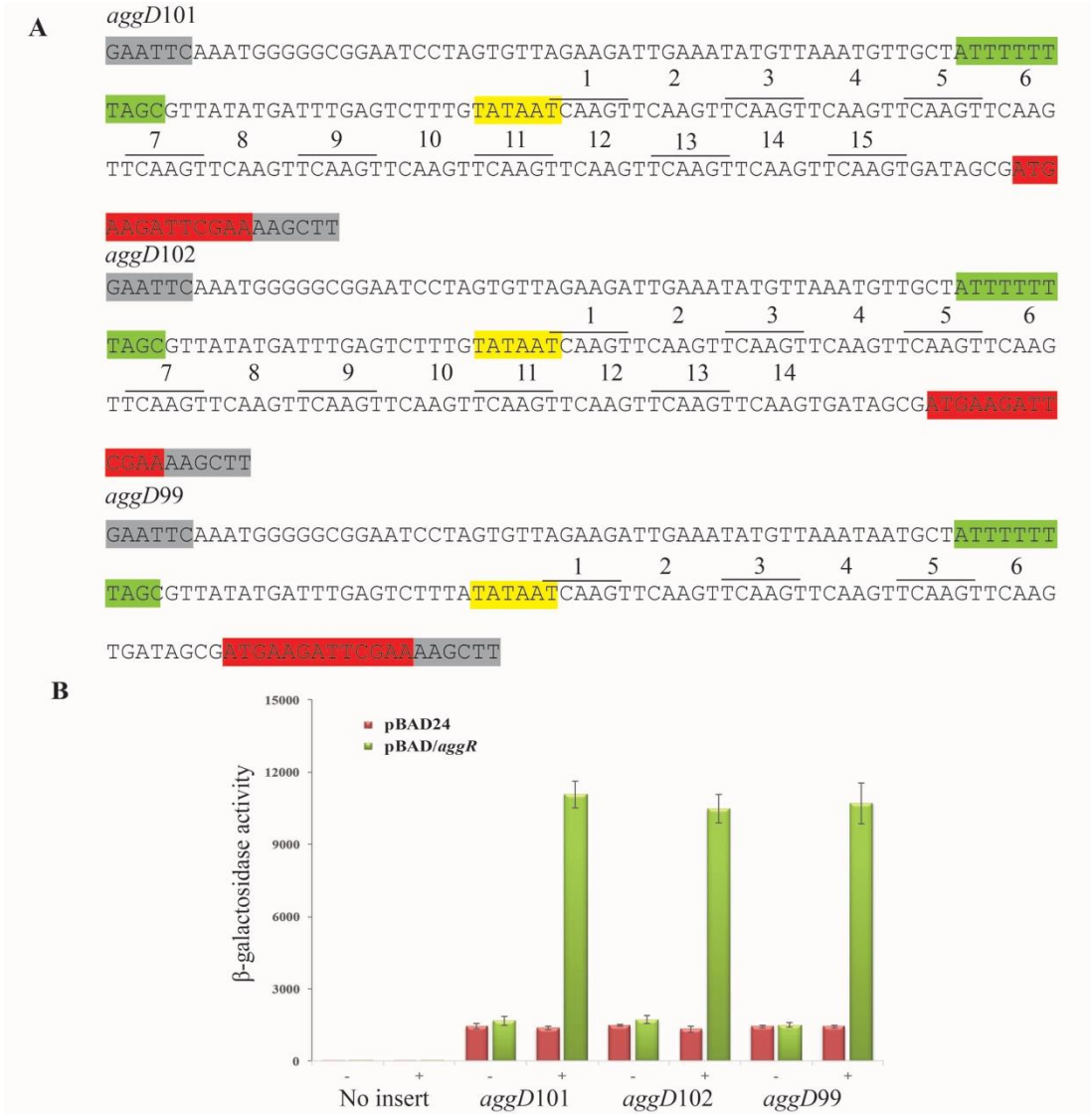


Figure 6.9 Comparison of the *aggD* promoter from two EAEC strains

A. The sequence of *aggD*101 and *aggD*102 is representing the *aggD* promoter fragment derivatives from EAEC C227-11 strain and *aggD*99 is representing the *aggD* promoter fragment from EAEC 17-2. There are difference number of 5' -TCAAGT-3' repeats are different in each fragment and numbering represents the number of this hexamer repeats in each fragment. The AggR-binding site is highlighted green, the -10 hexamer elements yellow, the restriction sites grey and the *aggD* coding sequence red.

B. The panel illustrates β-galactosidase activities measured in the *E. coli* K-12 strain BW25113 Δ*lac,* containing the *lacZ* expression vector (pRW50) or *aggD* promoter derivatives cloned into pRW50. The cells also carry either pBAD/*aggR* (green bars) or pBAD24 (red bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% (w/v) arabinose. β-galactosidase activities were measured as nmol of ONPG hydrolysed per minute per mg of bacterial mass.

strain. An alignment of the fimbrial promoter sequences of AAF/I-AAF/IV is shown in Figure 6.10. The AggR-dependent promoter has been predicted on the basis of AggR-binding site as mentioned by Morin *et al.* (2010). AggR-binding site has been predicted on the basis of TATC with presence of A at the upstream at mentioned position (Figure 6.6). The -10 hexamer element have been predicted on the basis of appropriate distance from the AggR-binding site and there was exact match to the consensus the -10 hexamer element. The -10 hexamer element of *agg3D* is present at 21 bp distance from AggR-binding site and the -10 hexamer element of *agg4D* is present at 22 bp distance from AggR-binding site. The exact match to consensus -10 hexamer elements and exact match of the AggR-binding site (on the basis of Rns weblogo and AggR study) indicate the presence of a strong promoters of *agg3D* and *agg4D.*

6.5 Discussion

The four structural genes of chaperone usher system for EAEC 17-2 are present in in Region 1 on pAA and these four genes are located in an operon (Figure 6.1) (Savarino *et al.,*1994). The promoter for this operon is present upstream of *aggD*, and the -10 hexamer element of the promoter matches in all six bp with the consensus sequence but the AggR-binding site has shown one bp variation in the conserved region at position 12 (Figure 6.6). Moreover, the distance between the AggR-binding site and the -10 hexamer element is 21 bp, whilst the optimum distance, found in a previous study as well as in other promoters of present study, between the -10 element and AggR-binding site is 22 bp. The variation of the nucleotide seperates AggR-binding site from consensus sequence and the distance variation between AggR-binding site and -10 hexamer element could be important to neutralise the effect of variation in the consensus spacer region. Another explanation for the slight difference in the consensus binding site sequences for Rns and AggR might be due to a difference in the amino

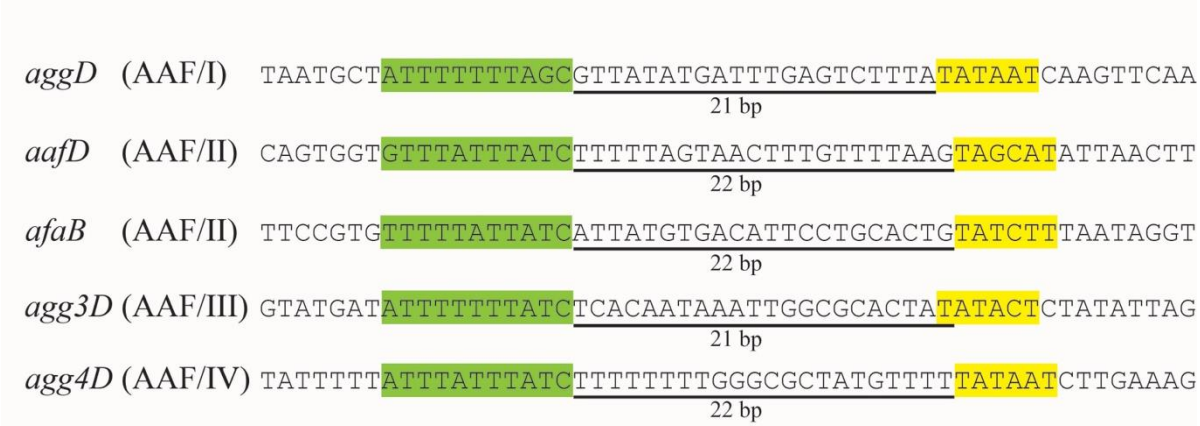


Figure 6.10 Alignment of AAF fimbrial promoters

The figure shows fimbrial promoters from AAF/I to AAF/IV. The AggR-binding sites are highlighted green, the -10 hexamer element yellow, the spacer distance between AggR-binding sites and -10 hexamer elements are marked by black line and distance is mentioned as bp numbers. *agg3D* and *agg4D* promoters have been predicted and are not supported by experimental results.

acid of HTHs of each protein (Munson, 2013). This can be investigated by introducing point mutations at certain positions of AggR binging site to identify the tolerance of different bases and an increasing in the distance between -10 hexamer element and AggR-binding site to 22 to study the effects of these changes on promoter regulation. The AggR-binding site has also shown a variation at position 3 from the consensus sequence of AggR and Rns-binding sequence (Figure 6.6). However, the results of the mutational analysis in Chaper 3on *aafD* promoter AggR-binding site has shown that any base at this position can be tolerated.

AAF are divided into 5 systems AAF/I to AAF/V on the basis of fimbrial subunit homology (Prager *et al.,* 2014). The fimbrial systems have similarities like the genes for these systems are present on plasmids, transcription is regulated by AggR, fimbriae are synthesised by the chaperone usher system and have conserved sequence of chaperon and usher proteins (Elias *et al.,* 1999, Prager *et al.,* 2014).). The alignment of structural operon promoters of four AAF systems has given an insight about promoter structure and prediction about AggR-binding sites. The AggR-binding sites overlap the -35 hexamer elements of the promoters. The location of AggR-binding sites on these promoters suggest that it is a Class II promoter (Browning and Busby, 2004). The alignment of four AAF promoters with AggR-binding sites has also shown the appropriate distance between the -10 hexamer element and the AggR-binding sites is 21bp and 22 bp (Figure 6.10). The results from another study have shown that the distance between Rns-binding site and -10 element is 22 bp for Rns dependent promoters (Bodero *et al.,* 2008). This difference in spacer distance in AAF/I and AAF/III could have come up during evolution.

AAF/I is more important to discuss as it has been reported in foodborne outbreaks pathogens. The *aggD* promoters has a run of 5' -TCAAGT-3' hexamer repeats and these repeats occur in higher number in pathogens from outbreak strains (Figure 6.9). The promoter fragments with

different number of repeats were studied by observing the β-galactosidase activities from cells containing these promoters. The results did not show substantial difference in the β-galactosidase activities measured from cells containing these promoters cloned in *lacZ* expression vector. These findings indicate either there is no role of these repeats in the promoter regulation or the technique I have used may not be sensitive enough to point out the difference in gene expression by these promoters.

This experiment can be repeated in EAEC strains instead of *E. coli* K-12 strain and more sensitive techniques like qPCR may be used to measure the gene expression. This study has given insight of organisation of AggR-dependent promoters and the AggR-binding sites present in AAF promoters.

Chapter 7
Final Discussion

EAEC has been investigated intensively in the last two decades due to its pathogenicity in humans, with researchers studying the epidemiology, physiology, protein structure, genome sequence and genetic expression of EAEC. A large amount of data has been gathered about EAEC from pathogenesis to antibiotic susceptibility but only a few studies have tried to understand gene expression in EAEC. In this study, I have examined the various mechanisms of gene expression in EAEC strains, particularly focusing on the transcriptional regulator, AggR. Thus, the study of AggR-regulated genes has enhanced the understanding of AggR-dependent promoters and helped me to suggest a model for AggR-dependent promoter activation, *i.e.* AggR-dependent promoters are regulated by a Class II transcriptional mechanism (Figure 7.1) (Browning and Busby, 2004).

7.1 AggR-dependent promoters

EAEC 042 is one of the first EAEC strains that was shown to be pathogenic in volunteer studies (Nataro *et al.*, 1995). Infection starts with attachment of bacteria to the gut epithelium by means of fimbriae and the genes encoding these fimbriae are located on the large attachment adherence plasmids found in many EAEC strains (Ito *et al.*, 2014). There are four genes that are generally involved in formation of fimbriae in EAEC 042 and these are present in two regions on pAA2; Region 1 and Region 2 (Elias *et al.*, 1999). In EAEC strain 17-2, all four genes required for fimbriae production are present in Region 1 on pAA (Nataro *et al.*, 1994). Although, in these two strains the organisation of the fimbrial genes is different, the transcription of these genes is regulated in the similar manner, *i.e.* the promoters upstream of each encoding region are regulated by the AggR transcription factor (section 3.2, 3.3 and 6.2).

By analysing the upstream sequence of each operon, which encodes the fimbrial genes, I found that an AggR-dependent promoter is present that initiates transcription of each operon.

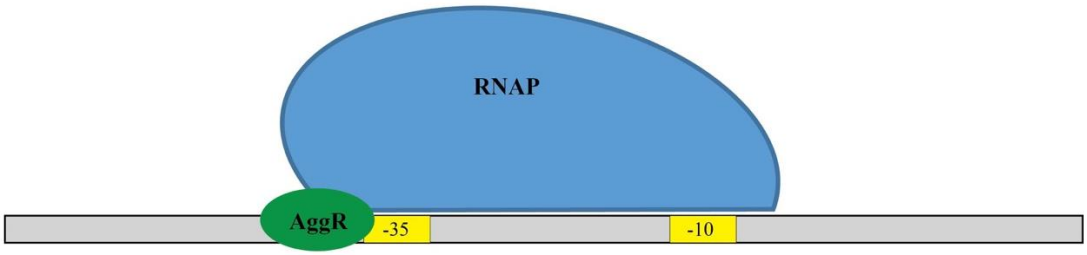


Figure 7.1 A model of AggR-dependent promoter activation

The figure illustrates an AggR-dependent promoter model, with the gray bar represents DNA double helix, the -35 and -10 hexamer elements are shown yellow, the blue structure represents RNAP and AggR is shown in green.

This finding, that fimbrial genes are regulated by AggR, is consistant with the finding that other AraC family members, *e.g.* Rns/CfaD, RegA, PerA, and Virf, regulate the fimbrial genes present in ETEC, *Citrobacter rodentium,* EPEC and *Shigella flexneri,* respectively (Santiago *et al.*, 2014; Tobe *et al.*, 1993). An important finding in this study is with regard to the fimbrial operon promoter of Region 2 EAEC 042 (section 3.3). Elias *et al.* (1999) previously showed that Region 2 in EAEC 042 contains the *afaB*, *aafC* and *aafB* genes, and suggested that the promoter was located upstream of *aafC* gene, but not the upstream of the *afaB* pseudogene (Figure 1.4). Moreover, they suggested that Region 2 genes were not regulated by AggR and that only the genes of Region 1 are AggR-dependent. The results of this present study indicate that the operon promoter is present upstream of *afaB* and it is regulated by AggR. The regulation of Region 2 by AggR also indicates that *aafC* and *aafB* expression is tightly controlled in a similar manner to that of *aafD* and *aafA* making it possible for the bacterium to regulate the chaperone usher system using a single transcription regulator. In EAEC strain 17-2, an AggR-dependent promoter is also present upstream of *aggD,* indicating that the chaperon usher system in this EAEC strain is regulated by the same system confirming AggR as the master transcription regulator of EAEC strains.

In section 3.3.7, of this study results showed that the promoter for Region 1 of plasmid pAA2 in EAEC 042 is stronger than the promoter upstream of Region 2. This difference is important, because for a chaperon usher system more chaperon and fimbrial subunits are required than usher and adhesins subunits. By controlling the expression of these subunits at the level of transcription in EAEC 042 the bacterium is able to produce the required amounts of each fimbrial protein. In EAEC 17-2, as all four genes are transcribed from a single AggR-dependent promoter, the different expression levels for these genes, might be regulated at the level of translation.

As AggR is the master virulence regulator of EAEC, it also regulates other virulence genes in EAEC 042, in addition to fimbrial genes (Morin *et al.*, 2013). Thus, I also investigated the upstream sequence of *aap* encoded on plasmid pAA2 and *aaiA,* present on the chromosome (sections 4.2 and 4.3). Strong AggR-regulated promoters were identified upstream of both genes, in line with the studies of Sheikh *et al.* (2002) and Dudley *et al.* (2006). On both pAA and pAA2, *aap* is present upstream of *aggR* whilst its secretion system (*aatPABCD*) is encoded by a separate operon which is also regulated by AggR. The most probable explanation for this secretion system being in a separate operon is that dispersin (Aap) is needed in large amounts and is regulated by a strong promoter (Figure 4.2), while its secretion is likely to only require a few secretion system proteins (Figure 1.5). Thus, they are encoded at separate loci but regulated by AggR-dependent promoters so secretion of dispersin occurs in coherent manner. The *aaiA* promoter is an operon promoter that regulates the whole operon encoding for the T6SS encoded in *sci-2* PAI in EAEC 042. It is also a strong promoter and the production of various T6SS proteins in different amount might be controlled at the level of translation in EAEC.

Many of the *aap* and *aaiA* promoter fragments used in this study possess relatively high AggR-independent expression levels indicating the presence of either strong promoters or the presence of more than one promoters on these fragments. A TG-motif is present on both promoters upstream of each -10 hexamer element, and so the high AggR-independent expression levels seen for these promoters could be attributed to the presence of this TG-motif (Figure 4.14). The presence of a second promoter is also possible as the deletion analysis of the *aaiA* promoter first showed an increase and then a gradual decrease in the activity even in the absence of AggR (Figure 4.9). In this study, I have focused on AggR-

dependent promoters, so the presence of alternate promoters has been acknowledged but was not investigated further.

7.2 AggR-binding sites at EAEC promoters

There is little information available on AggR-binding sites but other AraC family members, such as Rns has been studied in detail and an Rns binding site consensus sequence has been proposed (Munson, 2013). AggR is very similar to Rns on amino acid level, and it has been shown to induce transcription regulation of Rns-dependent promoters *in vivo* in ETEC, suggesting structural and functional similarity between the two proteins (Bodero *et al.*, 2007). A study on the *aggR* EAEC 042 promoters showed that AggR-binding sites are conserved at position 3 and 10-13 similar to Rns weblogo (Figure 4.14) (Morin *et al.*, 2010). In this study, I have used deletion and mutational analysis to define the essential AggR-binding sites on the *aafD, afaB, aap, aaiA* and *aggD* promoters. The AggR-binding sites at these promoters are conserved in the downstream part of the sequence of the AggR-binding motif, as in Rns but the upstream base at position 3 is variable (Figure 4.14). Thus a mutational analysis of the *aafD* promoter AggR-binding site at position 3 was carried out to determine the most appropriate base for this position (Figure 3.7). Results showed that AggR-dependent activation occurs if any base is present at this position but an A was optimal (Figure 3.9). These finding suggest that this position is less important in AggR-binding.

The central sequence of AggR-binding sites is highly AT rich and this seems to an play important role in AggR-binding. The *aaiA* promoter AggR-binding site has G at position 9, and promoter is activity increased by only 2 fold in the presence of AggR. These findings suggest that the sequence between conserved bases (4 to 9) may directly interact with the AggR, or DNA folding or bending may have effected due to presence of C or G base. The

weblogo of Rns has shown, a C or G base can be present at position 8 or 9, but the probability is quite low (Munson, 2013).

A weblogo was generated using the five AggR-binding sequences from EAEC 042 *aafD, afaB, aap* and *aaiA* promoters (this study) and *aggR* promoter (Morin *et al.*, 2010). The resulting logo showed conservation at positions 10 to13 whilst position 3 showed variability. As the AggR weblogo was generated using only 5 AggR-binding sequences, more sequences would be required for generating a weblogo with more confidence. The weblogo of AggR (n=5) is similar to Rns weblogo (n=31), except positions 2 and 3 in AggR weblogo are highly variable. All Rns-binding sites have A or G at position 3, but AggR does not have any specific base at this position. This difference could be explained, as there are few amino acids difference in the HTHs of both proteins, so each protein might have slightly different preferences for DNA-binding sites (Figure 5.1).

7.3 Understanding the principles of AggR-dependent regulation

The aim of the study was to understand the mechanism of AggR-dependent regulation at AggR-dependent promoters. The AggR-binding sites were identified and the promoter elements were located for different genes in this study and an AggR-binding site weblogo was generated. The AggR-binding site from the *aafD* promoter was transplanted into the well studied CRP-dependent promoter *CC(-41.5)* at different distances from the -10 element, indicating that the most appropriate spacing between the -10 hexamer and the AggR-binding was 22 bp (Figure 4.15). As, the AggR-binding site overlaps with the potential -35 elements of AggR-dependent promoters this suggests that AggR regulates promoters by a Class II mechanism. The distance between the AggR-binding site and the -10 hexamer elements at *aafD, afaB* and *aap* promoters was 22 bp. Bodero *et al.* (2007) showed that at Rns-dependent

promoters, there is a 22 bp spacing between the -10 hexamer element and the Rns-binding sites, suggesting that AggR and Rns regulate by the same mechanism.

The semi-synthetic promoters were constructed by transplanting the AggR-binding motif at various distances from the -10 hexamer elements and AggR-dependent activity was found to decrease by addition or deletion of one bp from optimal length of 22 bp. This shows, that by decreasing or increasing distance of AggR-binding site, the AggR position is twisted along with DNA helix and the appropriate alignment cannot be made between AggR and RNAP.

Another gene, *pet* is important in EAEC 042 pathogenesis but its expression is inhibited by AggR over expression from pBAD/*aggR* in EAEC 042 (Figure 5.7). This findng is supported by the fact that Pet is needed after attachment in pathogenesis (Hebbelstrup Jensen *et al.*, 2014). AggR represses *pet* expression in the beginning and then this repression might release in later stage of pathogenesis (Santiago *et al.*, 2016) .

Analysis of the AggR protein has shown that substitutions at N-terminal positions 14 and 16 affect the ability of AggR to activate at various promoters (Figure 5.3 and Figure 5.4). These results suggest that AggR-dependent activity involves the N-terminus and position 14 and 16 are important for AggR-dependent activation. The substitutions might have resulted in a change in the folding or structure of these proteins that may have resulted in a change in the interaction with RNAP. These results are in agreement with a mutational study of Rns and their effect on Rns-dependent promoters (Basturea *et al.*, 2008).

7.4 Concluding remarks

AggR is a master transcription regulator of EAEC pathogenesis and it regulates promoters from both the chromosome and the large virulence plasmid. This study has enhanced our

understanding of AggR-dependent promoter organisation, AggR-binding sites and AggR-dependent regulation with the following novel findings;

- The promoter elements at the *aafD, afaB, aap* and *aggD* promoters, which were located on virulence plasmids (pAA and pAA2) identified and those of the chromosome *aaiA* locus, were also identified.
- Pet expression is inhibited by over expression of AggR in EAEC strain DFB042.
- AggR-binding sites were identified and compared with those at the *aggR* promoter, as well as, with the Rns-binding logo.
- AggR-dependent promoters are regulated by a Class II mechanism.
- The distance between the AggR-binding motif and the -10 hexamer element is very important for AggR-dependent regulation. A 22 bp distance between AggR-binding site and -10 hexamer element is optimal.
- AggR from one strain can activate AggR-dependent promoters from another EAEC strain.
- The N-terminus of AggR is important for AggR-dependent activation.

7.5 Future Work

This project has generated some exciting results that could be investigated further to enhance our understanding of AggR-dependent activation, specifically:

- The investigation of transcription regulation at the dispersin promoter (*aap*) and its secretion system (*aatPABCD*)in the EAEC strain 042.
- The AggR-dependent expression from chromosomal promoter fragment, *aaiA100* could also be investigated to identify the multiple promoter elements.

- The rules for AggR-dependent promoter regulation uncovered by this study could be used to investigate the promoters of other attachment adherence fimbriae (AAF) encoding genes present in different EAEC strains.
- The experiments from this study were carried out in *E. coli* K-12 strain, they could be repeated in EAEC strains, though it is likely that they will be similar.
- *In-vivo* results could also be supported by *in-vitro* experiments such as, band shifts and DNase footprinting to investigate the interaction of AggR with promoter DNA.
- An extensive mutagenesis of AggR itself could be helpful in understanding the interaction of AggR with promoter DNA and RNAP.
- To uncover the AggR-dependent transcriptome of various EAEC strains, RNA-seq and ChIP-seq could be employed, leading to a better understanding of AggR-dependent regulation across the whole genome.
- Understanding the factors that switch on AggR is also essential as AggR is the master virulence regulator.
- Regulation of AggR has important clinical applications as it could be a potential target for anti-infective chemical inhibitors (Koppolu *et al.*, 2013).

The above work will enhance our understanding about AggR-dependent transcription regulation of in EAEC in particular and this can provide a model to investigate the transcription regulation in pathogens in general.

References

Adachi, J. A., Jiang, Z. D., Mathewson, J. J., Verenkar, M. P., Thompson, S., Martinez-Sandoval, F.*, et al.* (2001) Enteroaggregative *Escherichia coli* as a major etiologic agent in traveler's diarrhea in 3 regions of the world. Clin Infect Dis **32:** 1706-9.

Aschtgen, M. S., Bernard, C. S., De Bentzmann, S., Lloubes, R. and Cascales, E. (2008) SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. J Bacteriol **190:** 7523-31.

Aschtgen, M. S., Gavioli, M., Dessen, A., Lloubes, R. and Cascales, E. (2010) The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall. Mol Microbiol **75:** 886-99.

Aslani, M. M., Alikhani, M. Y., Zavari, A., Yousefi, R. and Zamani, A. R. (2011) Characterization of enteroaggregative *Escherichia coli* (EAEC) clinical isolates and their antibiotic resistance pattern. Int J Infect Dis **15:** e136-9.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M.*, et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2:** 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**.

Balleza, E., Lopez-Bojorquez, L. N., Martinez-Antonio, A., Resendis-Antonio, O., Lozada-Chavez, I., Balderas-Martinez, Y. I.*, et al.* (2009) Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol Rev **33:** 133-51.

Banerjee, S., Chalissery, J., Bandey, I. and Sen, R. (2006) Rho-dependent transcription termination: more questions than answers. J Microbiol **44:** 11-22.

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:** 2279-85.

Basu, R. S., Warner, B. A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C.*, et al.* (2014) Structural Basis of Transcription Initiation by Bacterial RNA Polymerase holoenzyme. J Biol Chem **289:** 24549–59.

Beatty, C. M., Browning, D. F., Busby, S. J. and Wolfe, A. J. (2003) Cyclic AMP receptor protein-dependent activation of the *Escherichia coli* acsP2 promoter by a synergistic class III mechanism. J Bacteriol **185:** 5148-57.

Betancourt-Sanchez, M. and Navarro-Garcia, F. (2009) Pet secretion, internalization and induction of cell death during infection of epithelial cells by enteroaggregative *Escherichia coli*. Microbiology **155:** 2895-906.

Bhan, M. K., Khoshoo, V., Sommerfelt, H., Raj, P., Sazawal, S. and Srivastava, R. (1989) Enteroaggregative *Escherichia coli* and Salmonella associated with nondysenteric persistent diarrhea. Pediatr Infect Dis J **8:** 499-502.

Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. and Ebright, R. H. (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell **78:** 889-96.

Blumer, C., Kleefeld, A., Lehnen, D., Heintz, M., Dobrindt, U., Nagy, G.*, et al.* (2005) Regulation of type 1 fimbriae synthesis and biofilm formation by the transcriptional regulator LrhA of *Escherichia coli*. Microbiology **151:** 3287-98.

Blyn, L. B., Braaten, B. A. and Low, D. A. (1990) Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. EMBO J **9:** 4045-54.

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:** 1627-32.

Boll, E. J., Struve, C., Sander, A., Demma, Z., Nataro, J. P., Mccormick, B. A.*, et al.* (2012) The fimbriae of enteroaggregative *Escherichia coli* induce epithelial inflammation in vitro and in a human intestinal xenograft model. J Infect Dis **206:** 714-22.

Brennan, C. A., Dombroski, A. J. and Platt, T. (1987) Transcription termination factor rho is an RNA-DNA helicase. Cell **48:** 945-52.

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

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

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

Bruant, G., Gousset, N., Quentin, R. and Rosenau, A. (2002) Fimbrial *ghf* gene cluster of genital strains of Haemophilus spp. Infect Immun **70:** 5438-45.

Brunner, M. and Bujard, H. (1987) Promoter recognition and promoter strength in the *Escherichia coli* system. EMBO J **6:** 3139-44.

Burmann, B. M. and Rosch, P. (2011) The role of *E. coli* Nus-factors in transcription regulation and transcription:translation coupling: From structure to mechanism. Transcription **2:** 130-134.

Busby, S. and Ebright, R. H. (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell **79:** 743-6.

Busby, S. and Ebright, R. H. (1999) Transcription activation by catabolite activator protein (CAP). J Mol Biol **293:** 199-213.

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O.*, et al.* (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell **9:** 527-39.

Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K. L. and Threlfall, E. J. (2005) Identification of plasmids by PCR-based replicon typing. J Microbiol Methods **63:** 219-28.

Cascales, E. (2008) The type VI secretion toolkit. EMBO Rep **9:** 735-41.

Chattaway, M. A., Harris, R., Jenkins, C., Tam, C., Coia, J. E., Gray, J.*, et al.* (2013) Investigating the link between the presence of enteroaggregative *Escherichia coli* and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009. Euro Surveill **18**.

Chaudhuri, R. R., Sebaihia, M., Hobman, J. L., Webber, M. A., Leyton, D. L., Goldberg, M. D.*, et al.* (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. PLoS One **5:** e8801.

Clements, A., Young, J. C., Constantinou, N. and Frankel, G. (2012) Infection strategies of enteric pathogenic *Escherichia coli*. Gut Microbes **3:** 71-87.

Cook, H. and Ussery, D. W. (2013) Sigma factors in a thousand *E. coli* genomes. Environ Microbiol **15:** 3121-9.

Cravioto, A., Tello, A., Navarro, A., Ruiz, J., Villafan, H., Uribe, F.*, et al.* (1991) Association of *Escherichia coli* HEp-2 adherence patterns with type and duration of diarrhoea. Lancet **337:** 262-4.

Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004) WebLogo: a sequence logo generator. Genome Res **14:** 1188-90.

Czeczulin, J. R., Balepur, S., Hicks, S., Phillips, A., Hall, R., Kothary, M. H.*, et al.* (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. Infect Immun **65:** 4135-45.

Datsenko, K. A. and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A **97:** 6640-5.

Dehaseth, P. L. and Helmann, J. D. (1995) Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol Microbiol **16:** 817-24.

Dhiman, A. and Schleif, R. (2000) Recognition of overlapping nucleotides by AraC and the sigma subunit of RNA polymerase. J Bacteriol **182:** 5076-81.

Domingo Meza-Aguilar, J., Fromme, P., Torres-Larios, A., Mendoza-Hernandez, G., Hernandez-Chinas, U., Arreguin-Espinosa De Los Monteros, R. A.*, et al.* (2014) X-ray crystal structure of the passenger domain of plasmid encoded toxin(Pet), an autotransporter enterotoxin from enteroaggregative *Escherichia coli* (EAEC). Biochem Biophys Res Commun **445:** 439-44.

Driessen, A. J., Fekkes, P. and Van Der Wolk, J. P. (1998) The Sec system. Curr Opin Microbiol **1:** 216-22.

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:** 1267-82.

Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Pehau-Arnaudet, G., Aschtgen, M. S.*, et al.* (2015) Biogenesis and structure of a type VI secretion membrane core complex. Nature **523:** 555-60.

Ebright, R. H. and Busby, S. (1995) The *Escherichia coli* RNA polymerase alpha subunit: structure and function. Curr Opin Genet Dev **5:** 197-203.

Echeverria, P., Serichantalerg, O., Changchawalit, S., Baudry, B., Levine, M. M., Orskov, F., *et al.* (1992) Tissue culture-adherent *Escherichia coli* in infantile diarrhea. J Infect Dis **165:** 141-3.

Egan, S. M. (2002) Growing repertoire of AraC/XylS activators. J Bacteriol **184:** 5529-32.

El-Hajj, Z. W. and Newman, E. B. (2015) An *Escherichia coli* mutant that makes exceptionally long cells. J Bacteriol **197:** 1507-14.

Elias, W. P., Jr., 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*. J Bacteriol **181:** 1779-85.

Ellinger, T., Behnke, D., Knaus, R., Bujard, H. and Gralla, J. D. (1994) Context-dependent effects of upstream A-tracts. Stimulation or inhibition of *Escherichia coli* promoter function. J Mol Biol **239:** 466-75.

Emody, L., Kerenyi, M. and Nagy, G. (2003) Virulence factors of uropathogenic *Escherichia coli*. Int J Antimicrob Agents **22:** 29-33.

Eslava, C., Navarro-Garcia, F., Czeczulin, J. R., Henderson, I. R., Cravioto, A. and Nataro, J. P. (1998) Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. Infect Immun **66:** 3155-63.

Feklistov, A. (2013) RNA polymerase: in search of promoters. Ann N Y Acad Sci **1293:** 25-32.

Feklistov, A. and Darst, S. A. (2009) Promoter recognition by bacterial alternative sigma factors: the price of high selectivity? Genes Dev **23:** 2371-5.

Feklistov, A. and Darst, S. A. (2011) Structural basis for promoter -10 element recognition by the bacterial RNA polymerase sigma subunit. Cell **147:** 1257-69.

Feklistov, A., Sharon, B. D., Darst, S. A. and Gross, C. A. (2014) Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol **68:** 357-76.

Feng, Y., Zhang, Y. and Ebright, R. H. (2016) Structural basis of transcription activation. Science **352:** 1330-3.

Franca, F. L., Wells, T. J., Browning, D. F., Nogueira, R. T., Sarges, F. S., Pereira, A. C.*, et al.* (2013) Genotypic and phenotypic characterisation of enteroaggregative *Escherichia coli* from children in Rio de Janeiro, Brazil. PLoS One **8:** e69971.

Frank, C., Faber, M. S., Askar, M., Bernard, H., Fruth, A., Gilsdorf, A.*, et al.* (2011) Large and ongoing outbreak of haemolytic uraemic syndrome, Germany, May 2011. Euro Surveill **16**.

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. Curr Microbiol **56:** 474-80.

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:** 393-410.

Gomes, T. A., Blake, P. A. and Trabulsi, L. R. (1989) Prevalence of *Escherichia coli* strains with localized, diffuse, and aggregative adherence to HeLa cells in infants with diarrhea and matched controls. J Clin Microbiol **27:** 266-9.

Gorke, B. and Stulke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol **6:** 613-24.

Gourse, R. L., Ross, W. and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Mol Microbiol **37:** 687-95.

Grad, Y. H., Lipsitch, M., Feldgarden, M., Arachchi, H. M., Cerqueira, G. C., Fitzgerald, M.*, et al.* (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A **109:** 3065-70.

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:** 1297-309.

Gummesson, B., Lovmar, M. and Nystrom, T. (2013) A proximal promoter element required for positive transcriptional control by guanosine tetraphosphate and DksA protein during the stringent response. J Biol Chem **288:** 21055-64.

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. J Bacteriol **177:** 4121-30.

Hacker, J. and Blum-Oehler, G. (2007) In appreciation of Theodor Escherich. Nature Reviews Microbiology **5:** 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:** 1749-53.

Harada, T., Hiroi, M., Kawamori, F., Furusawa, A., Ohata, K., Sugiyama, K.*, et al.* (2007) A food poisoning diarrhea outbreak caused by enteroaggregative *Escherichia coli* serogroup O126:H27 in Shizuoka, Japan. Jpn J Infect Dis **60:** 154-5.

Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T. and Gerdes, K. (2015) Recent functional insights into the role of (p)ppGpp in bacterial physiology. Nat Rev Microbiol **13:** 298-309.

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:** 614-30.

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. and Ala'aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story. Microbiol Mol Biol Rev **68:** 692-744.

Herring, C. D., Glasner, J. D. and Blattner, F. R. (2003) Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. Gene **311:** 153-63.

Herzog, K., Engeler Dusel, J., Hugentobler, M., Beutin, L., Sagesser, G., Stephan, R.*, et al.* (2014) Diarrheagenic enteroaggregative *Escherichia coli* causing urinary tract infection and bacteremia leading to sepsis. Infection **42:** 441-4.

Hitchens, T. K., Zhan, Y., Richardson, L. V., Richardson, J. P. and Rule, G. S. (2006) Sequence-specific interactions in the RNA-binding domain of *Escherichia coli* transcription termination factor Rho. J Biol Chem **281:** 33697-703.

Huang, D. B., Mohanty, A., Dupont, H. L., Okhuysen, P. C. and Chiang, T. (2006a) A review of an emerging enteric pathogen: enteroaggregative *Escherichia coli*. J Med Microbiol **55:** 1303-11.

Huang, D. B., Nataro, J. P., Dupont, H. L., Kamat, P. P., Mhatre, A. D., Okhuysen, P. C.*, et al.* (2006b) Enteroaggregative *Escherichia coli* is a cause of acute diarrheal illness: a meta-analysis. Clin Infect Dis **43:** 556-63.

Huang, D. B., Okhuysen, P. C., Jiang, Z. D. and Dupont, H. L. (2004) Enteroaggregative *Escherichia coli*: an emerging enteric pathogen. Am J Gastroenterol **99:** 383-9.

Iguchi, A., Thomson, N. R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I. R.*, et al.* (2009) Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J Bacteriol **191:** 347-54.

Imuta, N., Nishi, J., Tokuda, K., Fujiyama, R., Manago, K., Iwashita, M.*, et al.* (2008) The *Escherichia coli* efflux pump TolC promotes aggregation of enteroaggregative *E. coli* 042. Infect Immun **76:** 1247-56.

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:** 681-6.

Ito, K., Matsushita, S., Yamazaki, M., Moriya, K., Kurazono, T., Hiruta, N.*, et al.* (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. Microbiol Immunol **58:** 467–73.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol **3:** 318-56.

Jenkins, C., Van Ijperen, C., Dudley, E. G., Chart, H., Willshaw, G. A., Cheasty, T.*, et al.* (2005) Use of a microarray to assess the distribution of plasmid and chromosomal virulence genes in strains of enteroaggregative *Escherichia coli*. FEMS Microbiol Lett **253:** 119-24.

Johnson, J. R. (1991) Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev **4:** 80-128.

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:** 1144-7.

Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004) Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:** 123-40.

Kariisa, A. T., Grube, A. and Tamayo, R. (2015) Two nucleotide second messengers regulate the production of the Vibrio cholerae colonization factor GbpA. BMC Microbiol **15:** 166.

Kaur, P., Chakraborti, A. and Asea, A. (2010) Enteroaggregative *Escherichia coli*: An emerging enteric food borne pathogen. Interdiscip Perspect Infect Dis **2010:** 254159.

Klemm, P. (1985) Fimbrial adhesions of *Escherichia coli*. Rev Infect Dis **7:** 321-40.

Koppolu, V., Osaka, I., Skredenske, J. M., Kettle, B., Hefty, P. S., Li, J.*, et al.* (2013) Small-molecule inhibitor of the *Shigella flexneri* master virulence regulator VirF. Infect Immun **81:** 4220-31.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A.*, et al.* (2000) A structural model of transcription elongation. Science **289:** 619-25.

Kuhnert, P., Nicolet, J. and Frey, J. (1995) Rapid and accurate identification of *Escherichia coli* K-12 strains. Appl Environ Microbiol **61:** 4135-9.

Lee, D. J., Bingle, L. E., Heurlier, K., Pallen, M. J., Penn, C. W., Busby, S. J.*, et al.* (2009) Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. BMC Microbiol **9:** 252.

Lee, D. J., Minchin, S. D. and Busby, S. J. (2012) Activating transcription in bacteria. Annu Rev Microbiol **66:** 125-52.

Li, R., Zhang, Q., Li, J. and Shi, H. (2016) Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. Nucleic Acids Res **44:** 2554-63.

Li, X. T., Thomason, L. C., Sawitzke, J. A., Costantino, N. and Court, D. L. (2013) Positive and negative selection using the t*etA-sacB* cassette: recombineering and P1 transduction in *Escherichia coli*. Nucleic Acids Res **41:** e204.

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:** 271-6.

Lombard, J. (2014) Once upon a time the cell membranes: 175 years of cell boundary research. Biol Direct **9:** 32.

Lossi, N. S., Manoli, E., Forster, A., Dajani, R., Pape, T., Freemont, P.*, et al.* (2013) The HsiB1C1 (TssB-TssC) complex of the *Pseudomonas aeruginosa* type VI secretion system forms a bacteriophage tail sheathlike structure. J Biol Chem **288:** 7536-48.

Maeda, H., Fujita, N. and Ishihama, A. (2000) Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res **28:** 3497-503.

Mahon, V., Fagan, R. P. and Smith, S. G. (2012) Snap denaturation reveals dimerization by AraC-like protein Rns. Biochimie **94:** 2058-61.

Marques, S., Manzanera, M., Gonzalez-Perez, M. M., Gallegos, M. T. and Ramos, J. L. (1999) The XylS-dependent Pm promoter is transcribed *in vivo* by RNA polymerase with sigma 32 or sigma 38 depending on the growth phase. Mol Microbiol **31:** 1105-13.

Martinez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol **6:** 482-9.

Mathew, R. and Chatterji, D. (2006) The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol **14:** 450-5.

Mathewson, J. J., Johnson, P. C., Dupont, H. L., Satterwhite, T. K. and Winsor, D. K. (1986) Pathogenicity of enteroadherent *Escherichia coli* in adult volunteers. J Infect Dis **154:** 524-7.

Miroslavova, N. S. and Busby, S. J. (2006) Investigations of the modular structure of bacterial promoters. Biochem Soc Symp**:** 1-10.

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:** 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:** 344-55.

Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonjia, E., Ebright, Y. W. and Ebright, R. H. (2001) Translocation of sigma70 with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell **106:** 453-63.

Muller-Hill, B. (1998) Some repressors of bacterial transcription. Curr Opin Microbiol **1:** 145-51.

Mulligan, M. E., Brosius, J. and Mcclure, W. R. (1985) Characterization in vitro of the effect of spacer length on the activity of *Escherichia coli* RNA polymerase at the TAC promoter. J Biol Chem **260:** 3529-38.

Munson, G. P. (2013) Virulence regulons of enterotoxigenic *Escherichia coli*. Immunol Res **57:** 229-36.

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:** 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:** 2110-7.

Munson, G. P. and Scott, J. R. (2000) Rns, a virulence regulator within the AraC family, requires binding sites upstream and downstream of its own promoter to function as an activator. Mol Microbiol **36:** 1391-402.

Murakami, K. S. and Darst, S. A. (2003) Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13:** 31-9.

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. and Darst, S. A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science **296:** 1285-90.

Murdoch, S. L., Trunk, K., English, G., Fritsch, M. J., Pourkarimi, E. and Coulthurst, S. J. (2011) The opportunistic pathogen *Serratia marcescens* utilizes type VI secretion to target bacterial competitors. J Bacteriol **193:** 6057-69.

Murphree, D., Froehlich, B. and Scott, J. R. (1997) Transcriptional control of genes encoding CS1 pili: negative regulation by a silencer and positive regulation by Rns. J Bacteriol **179:** 5736-43.

Nataro, J. P. (2005) Enteroaggregative *Escherichia coli* pathogenesis. Curr Opin Gastroenterol **21:** 4-8.

Nataro, J. P., Deng, Y., Cookson, S., Cravioto, A., Savarino, S. J., Guers, L. D., *et al.* (1995) Heterogeneity of enteroaggregative *Escherichia coli* virulence demonstrated in volunteers. J Infect Dis **171:** 465-8.

Nataro, J. P. and Kaper, J. B. (1998) Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11:** 142-201.

Nataro, J. P., Kaper, J. B., Robins-Browne, R., Prado, V., Vial, P. and Levine, M. M. (1987) Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells. Pediatr Infect Dis J **6:** 829-31.

Nataro, J. P., Mai, V., Johnson, J., Blackwelder, W. C., Heimer, R., Tirrell, S., *et al.* (2006) Diarrheagenic *Escherichia coli* infection in Baltimore, Maryland, and New Haven, Connecticut. Clin Infect Dis **43:** 402-7.

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. Infect Immun **61:** 1126-31.

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*. J Bacteriol **176:** 4691-9.

Navarro-Garcia, F. and Elias, W. P. (2011) Autotransporters and virulence of enteroaggregative *E. coli*. Gut Microbes **2:** 13-24.

Nedialkov, Y. A., Opron, K., Assaf, F., Artsimovitch, I., Kireeva, M. L., Kashlev, M., *et al.* (2013) The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. Biochim Biophys Acta **1829:** 187-98.

Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., *et al.* (2003) The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. J Biol Chem **278:** 45680-9.

Nudler, E. (2012) RNA polymerase backtracking in gene regulation and genome instability. Cell **149:** 1438-45.

Olesen, B., Scheutz, F., Andersen, R. L., Menard, M., Boisen, N., Johnston, B., *et al.* (2012) Enteroaggregative *Escherichia coli* O78:H10, the cause of an outbreak of urinary tract infection. J Clin Microbiol **50:** 3703-11.

Paul, B. J., Berkmen, M. B. and Gourse, R. L. (2005) DksA potentiates direct activation of amino acid promoters by ppGpp. Proc Natl Acad Sci U S A **102:** 7823-8.

Perederina, A., Svetlov, V., Vassylyeva, M. N., Tahirov, T. H., Yokoyama, S., Artsimovitch, I*., et al.* (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. Cell **118:** 297-309.

Perez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res **28:** 1838-47.

Piatek, R., Zalewska, B., Bury, K. and Kur, J. (2005) The chaperone-usher pathway of bacterial adhesin biogenesis -- from molecular mechanism to strategies of anti-bacterial prevention and modern vaccine design. Acta Biochim Pol **52:** 639-46.

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:** e95379.

Pukatzki, S., Mcauley, S. B. and Miyata, S. T. (2009) The type VI secretion system: translocation of effectors and effector-domains. Curr Opin Microbiol **12:** 11-7.

Rivetti, C., Guthold, M. and Bustamante, C. (1999) Wrapping of DNA around the *E. coli* RNA polymerase open promoter complex. EMBO J **18:** 4464-75.

Robb, N. C., Cordes, T., Hwang, L. C., Gryte, K., Duchi, D., Craggs, T. D*., et al.* (2013) The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. J Mol Biol **425:** 875-85.

Roberts, J. W., Shankar, S. and Filter, J. J. (2008) RNA polymerase elongation factors. Annu Rev Microbiol **62:** 211-33.

Ross, W., Aiyar, S. E., Salomon, J. and Gourse, R. L. (1998) *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. J Bacteriol **180:** 5375-83.

Rossiter, A. E., Browning, D. F., Leyton, D. L., Johnson, M. D., Godfrey, R. E., Wardius, C. A*., et al.* (2011) Transcription of the plasmid-encoded toxin gene from enteroaggregative *Escherichia coli* is regulated by a novel co-activation mechanism involving CRP and Fis. Mol Microbiol **81:** 179-91.

Rossiter, A. E., Godfrey, R. E., Connolly, J. A., Busby, S. J., Henderson, I. R. and Browning, D. F. (2015) Expression of different bacterial cytotoxins is controlled by two global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism. Biochem J **466:** 323-35.

Ruiz, R. C., Melo, K. C., Rossato, S. S., Barbosa, C. M., Correa, L. M., Elias, W. P.*, et al.* (2014) Atypical enteropathogenic *Escherichia coli* secretes plasmid encoded toxin. Biomed Res Int **2014:** 896235.

Saecker, R. M., Record, M. T., Jr. and Dehaseth, P. L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol **412:** 754-71.

Sana, T. G., Hachani, A., Bucior, I., Soscia, C., Garvis, S., Termine, E.*, et al.* (2012) The second type VI secretion system of *Pseudomonas aeruginosa* strain PAO1 is regulated by quorum sensing and Fur and modulates internalization in epithelial cells. J Biol Chem **287:** 27095-105.

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 Pathog **10:** e1004153.

Santiago, A. E., Yan, M. B., Tran, M., Wright, N., Luzader, D. H., Kendall, M. M.*, et al.* (2016) A large family of anti-activators accompanying XylS/AraC family regulatory proteins. Mol Microbiol **101:** 314-32.

Sarantuya, J., Nishi, J., Wakimoto, N., Erdene, S., Nataro, J. P., Sheikh, J.*, et al.* (2004) Typical enteroaggregative *Escherichia coli* is the most prevalent pathotype among *E. coli* strains causing diarrhea in Mongolian children. J Clin Microbiol **42:** 133-9.

Savarino, S. J., Fox, P., Deng, Y. and Nataro, J. P. (1994) Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. J Bacteriol **176:** 4949-57.

Savery, N. J., Rhodius, V. A., Wing, H. J. and Busby, S. J. (1995a) Transcription activation at *Escherichia coli* promoters dependent on the cyclic AMP receptor protein: effects of binding sequences for the RNA polymerase alpha-subunit. Biochem J **309:** 77-83.

Savery, N. J., Rhodius, V. A., Wing, H. J. and Busby, S. J. (1995b) Transcription activation at *Escherichia coli* promoters dependent on the cyclic AMP receptor protein: effects of binding sequences for the RNA polymerase alpha-subunit. Biochem J **309 (Pt 1):** 77-83.

Schleif, R. (2010) AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. FEMS Microbiol Rev **34:** 779-96.

Seedorff, J. and Schleif, R. (2011) Active role of the interdomain linker of AraC. J Bacteriol **193:** 5737-46.

Servin, A. L. (2005) Pathogenesis of Afa/Dr diffusely adhering *Escherichia coli*. Clin Microbiol Rev **18:** 264-92.

Shalom, G., Shaw, J. G. and Thomas, M. S. (2007) *In vivo* expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. Microbiology **153:** 2689-99.

Sheikh, J., Czeczulin, J. R., Harrington, S., Hicks, S., Henderson, I. R., Le Bouguenec, C.*, et al.* (2002) A novel dispersin protein in enteroaggregative *Escherichia coli*. J Clin Invest **110:** 1329-37.

Silhavy, T. J., Kahne, D. and Walker, S. (2010) The bacterial cell envelope. Cold Spring Harb Perspect Biol **2:** a000414.

Silverman, J. M., Austin, L. S., Hsu, F., Hicks, K. G., Hood, R. D. and Mougous, J. D. (2011) Separate inputs modulate phosphorylation-dependent and -independent type VI secretion activation. Mol Microbiol **82:** 1277-90.

Singh, S. S., Typas, A., Hengge, R. and Grainger, D. C. (2011) *Escherichia coli* sigma 70 senses sequence and conformation of the promoter spacer region. Nucleic Acids Res **39:** 5109-18.

Skordalakes, E. and Berger, J. M. (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell **114:** 135-46.

Stoyanov, J. V., Hobman, J. L. and Brown, N. L. (2001) CueR (YbbI) of *Escherichia coli* is a MerR family regulator controlling expression of the copper exporter CopA. Mol Microbiol **39:** 502-11.

Swint-Kruse, L. and Matthews, K. S. (2009) Allostery in the LacI/GalR family: variations on a theme. Curr Opin Microbiol **12:** 129-37.

Tagami, H. and Aiba, H. (1999) An inactive open complex mediated by an UP element at *Escherichia coli* promoters. Proc Natl Acad Sci U S A **96:** 7202-7.

Tecon, R. and Or, D. (2016) Bacterial flagellar motility on hydrated rough surfaces controlled by aqueous film thickness and connectedness. Sci Rep **6:** 19409.

Tenover, F. C. (2006) Mechanisms of antimicrobial resistance in bacteria. Am J Infect Control **34:** S3-10; discussion S64-73.

Tobe, T., Yoshikawa, M., Mizuno, T. and Sasakawa, C. (1993) Transcriptional control of the invasion regulatory gene *virB* of *Shigella flexneri*: activation by *virF* and repression by H-NS. J Bacteriol **175:** 6142-9.

Tseng, T. T., Tyler, B. M. and Setubal, J. C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol **9** S2.

Valentin-Hansen, P., Sogaard-Andersen, L. and Pedersen, H. (1996) A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. Mol Microbiol **20:** 461-6.

Vila, J., Vargas, M., Henderson, I. R., Gascon, J. and Nataro, J. P. (2000) Enteroaggregative *Escherichia coli* virulence factors in traveler's diarrhea strains. J Infect Dis **182:** 1780-3.

Wilson, A., Evans, J., Chart, H., Cheasty, T., Wheeler, J. G., Tompkins, D.*, et al.* (2001) Characterisation of strains of enteroaggregative *Escherichia coli* isolated during the infectious intestinal disease study in England. Eur J Epidemiol **17:** 1125-30.

Zamboni, A., Fabbricotti, S. H., Fagundes-Neto, U. and Scaletsky, I. C. (2004) Enteroaggregative *Escherichia coli* virulence factors are found to be associated with infantile diarrhea in Brazil. J Clin Microbiol **42:** 1058-63.

Zoued, A., Durand, E., Brunet, Y. R., Spinelli, S., Douzi, B., Guzzo, M.*, et al.* (2016) Priming and polymerization of a bacterial contractile tail structure. Nature **531:** 59-63.

Zuo, Y. and Steitz, T. A. (2015) Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. Mol Cell **58:** 534-40.