Distant regulation of the *Escherichia coli* recombinase gene, *fimB*

A thesis submitted to the University of Kent for the degree of PhD in the faculty of Science, Technology and Medical Studies

Simon Friar

2006

Department of Biosciences

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or Institute of learning.

S Friar

22/09/06

Acknowledgements

I would really like to thank Ian. He has been an enthusiastic and supportive supervisor, and has encouraged me to be as independent as possible. I would also like to thank Bali for her constant help and support in the lab. And Lizanne - you are simply brilliant! Also Mike Geeves and Jim Bloor, too, as members of my monitoring committee. And my parents, for putting up with me and offering me support.

I could never miss out Hazel, too. Simply put, you would not be reading this now if not for her help and support.

List of Abbreviations

A	absorbance
bp	base pair(s)
AT	adenine and thymine
CAM	chloramphenicol
CAP/ CRP	catabolite activator/ repression protein
CNS	central nervous system
CTAB	hexadecyltrimethylammonium bromide
(d)H_20	(deionised-) water
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
EDTA	ethylenediaminetetraacetic acid
FIS	factor for inversion stimulation
g	(Acceleration due to) gravity (rcf)
Glyc	glycerol
GlcNAc	*N*-acetyl-D-glucosamine
H-NS	histone-like nucleoid structuring protein
IHF	integration host factor
IL-6	interleukin-6
IL-8	interleukin-8
IRL	left inverted repeat
IRR	right inverted repeat
Kan	kanamycin
Kb	kilo-base pair(s)
LB	lysogeny broth (also known as Luria-Bertani broth)
LPS	lipopolysaccharide
Lrp	leucine-responsive regulatory protein
MOPS	3-(N-Morpholino)propanesulphonic acid (free acid)
Neu5Ac	*N*-acetylneuraminic acid
OD	optical density
ONPG	2-nitrophenol-β-D-galactopyranoside
ORF	open reading frame

PCR	polymerase chain reaction
PMN	polymorphonuclear leukocytes
(p)ppGpp	guanosine tetra-/ penta-phosphate
PTS	phosphotransferase system
rcf	relative centrifugal force
RD MOPS	rich defined MOPS (medium)
Rm	replacement mutation
RNAP	RNA polymerase
rpm	revolutions per minute
RT	room temperature
rut	Rho utilisation (site)
SDS	sodium dodecyl sulphate
SIDD	stress induced duplex destabilization
Tet	tetracycline
TNFα	tumour necrosis factor α
Tricine	*N*-[Tris-(hydroxymethyl)-methyl]-glycine
Tris	tris (hydroxymethyl) aminomethane
UK	United Kingdom
UPEC	uropathogenic *E. coli*
USA	Unites States of America
°C	degrees Celsius

Table of Contents

Abstract

Escherichia coli is the archetypal member of the Enterobacteriaceae, and for many decades has been utilised as a model organism for the study of microbial genetics, biochemistry and molecular biology. The majority of isolates belong to the intestinal microflora of healthy humans and other mammals, although there are a number of patho-adapted isolates that cause both intestinal and extraintestinal diseases. Adherence to host cell receptors is a strategy employed by both commensal and pathogenic bacteria alike, and this is commonly achieved through the expression of cell surface adhesins that localise them at their preferred sites. Type 1 fimbriae are a particular adhesin that facilitate attachment of *Escherichia coli* to a number of mannose-containing cell-surface receptors within mammalian hosts. They are expressed by the majority (>80 %) of all *E. coli* isolates, commensal and pathogenic alike, and contribute to their colonisation and persistence at a number of different niches.

Expression of the type 1 fimbrial adhesin in *Escherichia coli* is phase variable. This reversible ON-OFF switching in expression is determined by the orientation of a 314 bp invertible element, which is catalysed by two recombinase proteins, FimB (switching in both directions) and FimE (switching predominantly from ON-to-OFF). The *fimB* recombinase gene is separated from the divergently transcribed *nanC* by a large (1.4 kb) intergenic region, and two regulatory proteins, NanR (Neu5Ac-responsive) and NagC (GlcNAc-responsive) bind to sequences far upstream (> 600 bp) of the *fimB* promoter to activate expression. Preliminary data had suggested that these regulatory proteins control *fimB* expression by an anti-repression mechanism that involves additional sequences proximal to *nanC*. In this study, I investigated the mechanism that determines distant activation of *fimB* by NanR and NagC. Contrary to previous findings, I provide evidence to suggest that *nanC*-proximal sequences do not inhibit *fimB*, and that a number of genetic elements are able to activate *fimB* expression at a distance. I show that NanR and NagC activate a single promoter for *fimB* from unusually large distances, and yet they do not appear to require additional, *cis*-active elements within the intervening sequence in order to function. The efficacy of activation by both these regulators is, however, sensitive to their precise location within the *nanC-fimB* intergenic region, and the significance of these data, with regards to a mechanism of activation is discussed.

Chapter 1:

Introduction

1.1 Background

There is remarkable diversity between different *Escherichia coli* isolates. *Escherichia coli* is the prototypic facultative anaerobe within the mammalian gastrointestinal tract, where the majority of isolates form part of the commensal microbial flora. However, there are a number of pathogenic strains that cause specific intestinal or extraintestinal diseases. In each instance, pathogenicity is associated with the acquisition of many virulence determinants by horizontal transfer from other microorganisms.

Attachment to the host cell surface is usually the first step in colonisation, and cell-surface structures that facilitate adhesion are often virulence factors of pathogenic bacteria. Type 1 fimbriae are cell surface adhesins that facilitate attachment of bacteria to a number of cell-surface receptors. They are interesting in that they are not restricted to one particular function or niche; whilst they are a virulence factor in the urinary tract, they are produced by the majority of *E. coli* isolates, pathogenic and commensal alike, and thought to play an important role in the colonisation of different niches.

1.2 Type 1 fimbriae

Many enteric bacteria produce peritrichously arranged proteinaceous appendages, otherwise known as fimbriae or pili, which extend from the cell surface and facilitate adhesion to specific target molecules (Duguid *et al.*, 1955). Type 1 fimbriae are one of the most common fimbrial types, with as many as 80 % of all *E. coli* strains being capable of their synthesis, and they are widely conserved throughout the Enterobacteriaciae family (Abraham *et al.*, 1988). They were originally defined by their capacity to agglutinate guinea pig erythrocytes in a mannose-sensitive fashion (Brinton, 1959; Duguid *et al.*, 1955), and have subsequently been shown to bind a number of mannose-containing molecules (Eden *et al.*, 1977; Duguid *et al.*, 1979; Hartley *et al.*, 1979; Ohman *et al.*, 1982; Knutton *et al.*, 1984; Fader *et al.*, 1988; Parkkinen *et al.*, 1988; Wold *et al.*, 1988; Parkkinen & Korhonen, 1989; Gbarah *et al.*, 1991).

Type 1 fimbriae are a virulence factor of uropathogenic *E. coli* (UPEC), as they promote bacterial persistence and enhance the inflammatory response to infection (Bahrani-Mougeot *et al.*, 2002; Connell *et al.*, 1996). They facilitate adhesion to the uroplakin 1a membrane glycoproteins, present on the luminal surface of the bladder epithelium (Hung *et al.*, 2002; Wu *et al.*, 1996; Zhou *et al.*, 2001). However, in addition to facilitating adhesion they also mediate invasion of these superficial epithelial cells,

triggering a host immune response that results in the infected cells being exfoliated and removed with the flow of urine (Martinez *et al.*, 2000; Mulvey *et al.*, 2001). To avoid clearance by exfoliation, intracellular UPEC re-emerge and eventually establish a persistent, quiescent bacterial reservoir within the bladder mucosa (Mulvey *et al.*, 2001). These intracellular bacteria mature into biofilms, creating pod-like bulges on the bladder surface (Anderson *et al.*, 2003). The bacteria within these pods are encased in a polysaccharide-rich matrix surrounded by a protective shell of uroplakin (Anderson *et al.*, 2003).

Type 1 fimbriae also facilitate binding to human brain microvascular endothelial cells by *E. coli* K1 (Teng *et al.*, 2005), which is a preliminary step for penetration into the central nervous system (CNS) and establishing neonatal meningitis (Kim, 2001). Furthermore, type 1 fimbrial expression within the CNS stimulates an inflammatory response that contributes to the brain dysfunction and neuronal injury that is associated with meningitis (Lee *et al.*, 2005).

The role of type 1 fimbriae in the gastrointestinal tract remains unclear. Type 1 fimbriae are, however, considered important for oropharyngeal colonisation (Bloch & Orndorff, 1990), they have an important role in the faecal-oral transmission of *E. coli* between mammalian hosts (Bloch *et al.*, 1992), and they are a potential virulence factor for Crohn's disease (Boudeau *et al.*, 2001). Another interesting role for type 1 fimbriae, is in the initial surface attachment of *E. coli* during biofilm formation (Genevaux *et al.*, 1999; Pratt & Kolter, 1998).

1.3 Structure and synthesis of type 1 fimbriae

The *fim* gene cluster, located at 98 minutes on the *E. coli* chromosome, is comprised of nine genes that are specifically required for the biogenesis, assembly and function of type 1 fimbriae (Orndorff & Falkow, 1984). They are organised into three discreet transcriptional units (Olsen & Klemm, 1994): seven structural genes, *fimAICDFGH*, that are transcribed as a polycistronic mRNA; and two regulatory genes, *fimB* and *fimE*. FimA is the major structural subunit of type 1 fimbriae, which comprises approximately 98 % of the entire structural protein (Klemm, 1984); whilst FimF, FimG and FimH are minor structural subunits that are present at a ratio less that 1:100 relative to FimA (Abraham *et al.*, 1987; Krogfelt & Klemm, 1988). FimC is a periplasmic chaperone protein, and FimD is an outer-membrane 'usher' that together facilitate the correct

assembly of structural subunits within a growing fimbrium. The precise role of the FimI protein remains unknown, but it shares extensive sequence homology with the major structural subunit, FimA, and is likely to have a similar structural role.

Structurally, type 1 fimbriae are composed of two distinct elements: a rigid, right-handed helical rod and a flexible fibrillar tip (Hahn *et al.*, 2002; Jones *et al.*, 1995). The rod component is a hollow tube formed from a helically coiled string of subunits, primarily composed of FimA (Choudhury *et al.*, 1999; Hahn *et al.*, 2002). The outer diameter of this tube is 6.9 nm, with estimates of the internal cavity diameter ranging between 2.1 and 2.5 nm and the approximate length is ~1.5-1.9 μm (Brinton, 1959; Hahn *et al.*, 2002). One helical repeat has 26 or 27 (FimA) monomers and is ~10-19 nm long (Hahn *et al.*, 2002). The fibrillar tip is 16 nm long, 3 nm wide (Jones *et al.*, 1995), and is composed of the three minor subunits (FimF, FimG & FimH) (Hahn *et al.*, 2002; Jones *et al.*, 1995; Saulino *et al.*, 2000). The sequential orientation of monomers in this fibrillar tip is FimF-FimG-FimH, such that FimH is the distal tip protein (Hahn *et al.*, 2002).

Translocation of each structural subunit across the cytoplasmic membrane occurs via the general secretory pathway (de Keyzer *et al.*, 2003), and assembly within the periplasm, into the growing fimbrium occurs via a highly conserved chaperone-usher (FimC-FimD) pathway (Soto & Hultgren, 1999). Each structural subunit contains a conserved C-terminal 'pilin' domain that has an immunoglobulin-like (Ig) β-barrel structure (Choudhury *et al.*, 1999; Sauer *et al.*, 1999). However, Ig-like folds are characterised by 7-9 anti-parallel β-strands, arranged into two sheets (Bork *et al.*, 1994), and the structural subunits only have 6 (Choudhury *et al.*, 1999; Sauer *et al.*, 1999). The chaperone proteins, such as FimC have two Ig-like domains (Hung *et al.*, 1996; Jones *et al.*, 1993; Kuehn *et al.*, 1993; Pellecchia *et al.*, 1998), and via a process known as 'strand complementation', one of the chaperone's β-strands serves as the structural subunit's seventh strand (Barnhart *et al.*, 2000; Choudhury *et al.*, 1999; Sauer *et al.*, 1999). These chaperones also serve as a catalyst, by binding to the non-native structural subunits and accelerating their folding by over 100-fold (Barnhart *et al.*, 2000; Vetsch *et al.*, 2004); they prevent structural subunits from prematurely forming subunit-subunit aggregates within the periplasm; and they protect structural subunits from proteolytic degradation within the periplasm (Jones *et al.*, 1995; Klemm, 1992).

The FimC-subunit complexes are targeted to the outer membrane-bound usher, FimD, which is able to discriminate between the different chaperone-subunit complexes and ensure that they are assembled into the growing fimbrium in the correct order (Dodson *et al.*, 1993; Saulino *et al.*, 1998). New subunits are incorporated into the growing fimbrium at its base (Lowe *et al.*, 1987), via a process known as donor strand exchange, in which the β-strand of the chaperone is replaced by an N-terminal extension of a neighbouring subunit (Barnhart *et al.*, 2000; Choudhury *et al.*, 1999; Sauer *et al.*, 1999). During donor strand exchange the new subunit undergoes a topological change that seals the N-terminal extension, of the neighbouring subunit within the Ig-like fold, forming a stable interaction between the two (Sauer *et al.*, 2002).

1.4 Phase variation

Phase variation is defined as the reversible switching in gene expression that results in a mixed or differentiated population. Typically there are no intermediate levels of expression of phase variable genes, with cells being differentiated into expressing (phase on) or non-expressing (phase off) fractions within any given population. Phase variation is essentially a stochastic event, although the frequency in switching between phases can be modulated in response to environmental factors, and is also heritable at the genetic or epigenetic level (van der Woude & Baumler, 2004).

Phase variable gene expression is thought to be a strategy that maximises growth and survival in unpredictable environments (Dybvig, 1993). Traditionally it has been considered as a mechanism for evading the host immune response, as the expression of a large number of immunogenic cell-surface structures, including type 1 fimbriae, is phase variable. Whilst there are several lines of evidence to suggest that this might be true for type 1 fimbriae (refer to chapters 1.8.6 & 1.8.7 for more details), it is now becoming apparent that a number of genes are phase variable in organisms that are not associated with a host, or do not confer antigenic properties (van der Woude, 2006). Irrespective of its purpose, phase variation of type 1 fimbriae appears to be important in each niche that they are utilized. Mutations that result in constitutive expression of type 1 fimbriae in commensal *E. coli* F18 inhibit its colonisation of the mouse large intestine (McCormick *et al.*, 1993), and they render a neonatal meningitis isolate, *E. coli* K1 strain RS218, incapable of inducing severe bacteremia (Xie *et al.*, 2006).

1.5 Genetic regulation of type 1 fimbriae

The phase variation of type 1 fimbriae is determined by the orientation of a 314 bp invertible element, *fimS*, that is located immediately upstream of the structural genes (*fimAICDFGH*) (figure 1.1) (Abraham *et al.*, 1985; Freitag *et al.*, 1985). The promoter for the structural genes is located within *fimS* (Olsen & Klemm, 1994), such that they are only transcribed when it is directed towards them (phase on, or fimbriate), and not when it is directed away (phase off, or afimbriate). Inversion of *fimS* is catalysed by two site-specific recombinase proteins encoded by *fimB* and *fimE* (Clegg *et al.*, 1985; Klemm *et al.*, 1985), that work independently of each other to determine orientation (Klemm, 1986; McClain *et al.*, 1991). Whilst FimB mediates inversion in both directions, FimE predominantly switches from on-to-off (Gally *et al.*, 1996; McClain *et al.*, 1991; Stentebjerg-Olesen *et al.*, 2000).

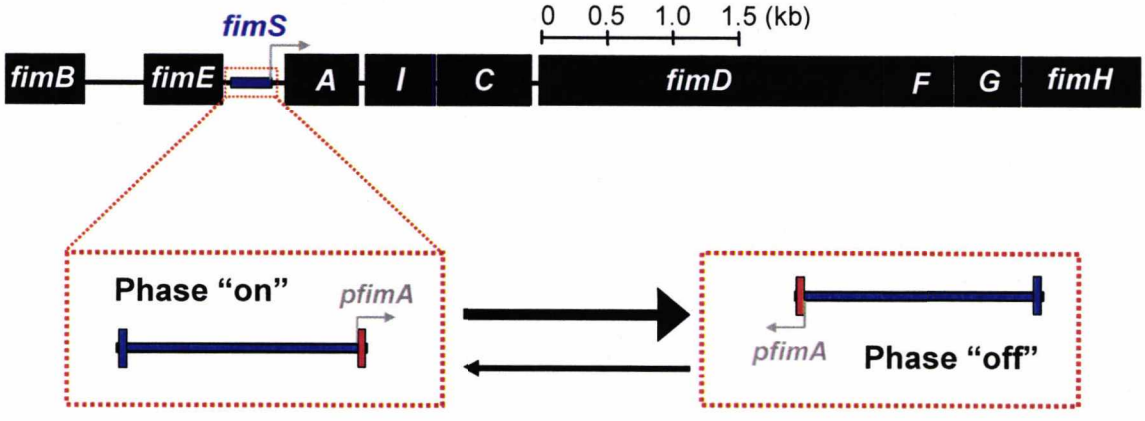


Figure 1.1: Schematic diagram of the *fim* gene cluster. The *fimB-fimH* fragment of the MG1655 genome defined (bp 4538525-4547279; Blattner *et al.*, 1997), as well as the promoter for the structural genes (*pfimA*) and the orientation of the invertible element (*fimS*) in on phase and off phase.

Orientation of *fimS* is necessary but not sufficient for type 1 fimbrial expression (McClain *et al.*, 1993); there appears to be additional levels of control that regulate expression when *fimS* is in the phase-on orientation to produce both fimbriate and afimbriate cells. The molecular mechanism that underpins this additional level of control, however, remains largely unknown.

1.6 The *fim* recombinases & recombination

The two *fim* recombinase proteins, FimB and FimE, belong to the λ-integrase family of site-specific recombinases (Argos *et al.*, 1986; Dorman & Higgins, 1987). Although there is considerable variation between members of this family at the primary amino acid level, their crystal structures, where determined, contain similar catalytic pockets in which a tetrad of residues is conserved (arginine-histidine-arginine-tyrosine) (Chen *et al.*, 2000; Esposito & Scocca, 1997; Guo *et al.*, 1997; Hickman *et al.*, 1997; Kwon *et al.*, 1997; Gopaul *et al.*, 1998; Nunes-Duby *et al.*, 1998; Subramanya *et al.*, 1997). In addition to this conserved tetrad, a fifth residue (a lysine) that lies outside of this catalytic pocket is loosely conserved within this family and, in addition to the conserved tetrad is required for catalytic properties in family members that possess it (Cao & Hayes, 1999). FimB and FimE are the smallest members of this family (both ~23 kDa), that share 48 % amino acid identity with each other and are thought to have evolved from the duplication of a common ancestral gene (Klemm, 1986). They are, however, noticeably different in that their conserved tetrads of amino acids are located at different positions: these are R47-H141-R144-Y176 in FimB (Burns *et al.*, 2000), and R41-H136-R139-Y171 in FimE (Smith & Dorman, 1999). Both FimB and FimE contain the conserved lysine residue, at the same position relative to the catalytic tetrad (K72 & K66, respectively), although the effect of these residues have yet to be determined. Moreover, specific residues that determine the orientational bias of one recombinase protein do not have the same effects on the other (Smith & Dorman, 1999)

The recombination event at the *fim* locus, catalysed by either *fim* recombinase protein is essentially the same as those catalysed by other λ-integrase family members at their respective target sites (Grainge & Jayaram, 1999; Grindley *et al.*, 2006). The *fim* invertible element, *fimS*, is a 296 bp sequence that is flanked by a 9 bp inverted repeat (5' TTGGGGCCA) at either end (Abraham *et al.*, 1985); in the on orientation these are known as the left (*fimA*-distal) and right (*fimA*-proximal) inverted repeats (IRL & IRR respectively). Inversion of *fimS* requires four molecules of either recombinase protein to bind four half-sites, two of which overlap each inverted repeat (Gally *et al.*, 1996). The recombinase proteins begin inversion by first introducing a single strand breakage at each site of strand exchange (IRL & IRR); the hydroxyl group of the conserved tyrosine residue nucleophilically attacks the scissile phosphodiester bond to form a transient 3' phosphotyrosine bond (Craig & Nash, 1983; Evans *et al.*, 1990; Gronostajski & Sadowski, 1985; Pargellis *et al.*, 1988), with the conserved arginine-histidine-arginine

residues facilitating this process (Grainge & Jayaram, 1999). This recombinase-DNA intermediate is resolved by nucleophilic attack of the 3' phosphotyrosine bond by the hydroxyl group of the free 5'end of the complementary strand, to form a four-way DNA junction otherwise known as a Holliday junction (Gopaul *et al.*, 1998; Hoess *et al.*, 1987; Meyer-Leon *et al.*, 1988l). Resolution of this Holliday intermediate is essentially the same as the events that lead to its formation. The hitherto inactive recombinase monomers form transient 3' phosphotyrosine bonds with the previously unaffected DNA strands, which are resolved after nucleophilic attack by the free 5' hydroxyl of opposing strands (Chen *et al.*, 2000; Grindley *et al.*, 2006; Guo *et al.*, 1997; Kho & Landy, 1994).

1.6.1 Orientational control

As is the case with a number of fimbriae, on-to-off switching of *fim* is much faster than off-to-on, such that there is a strong bias towards an afimbriate state (Blyn *et al.*, 1989; McClain *et al.*, 1993; van der Woude & Low, 1994). For type 1 fimbriae, this bias is largely governed by *fimE* expression and FimE-mediated recombination. In *E. coli* K12 strain MG1655, FimE-mediated on-to-off switching occurs at a much higher frequency than FimB-mediated switching in either direction (10^{-3} to 10^{-4} per cell per generation, versus 0.3 per cell per generation respectively) (Kulasekara & Blomfield, 1999; McClain *et al.*, 1993). The specificity of FimE switching in the on-to-off orientation is thought to arise for two reasons. Firstly, FimE shows a strong preference for the invertible element in the on orientation as a substrate for recombination *in vitro* (and *in vivo*) (Gally *et al.*, 1996); recombinase half sites flank IRL and IRR, such that the two half sites within *fimS* switch positions in each orientation, and FimE binds to IRL in the off orientation with a much lower affinity. Secondly, the orientation of the invertible element controls *fimE* expression by stabilising *fimE* mRNA in the on orientation, and mediating its degradation in the off orientation (Kulasekara & Blomfield, 1999; Sohanpal *et al.*, 2001). This orientation-specific control over *fimE* expression, known as orientational control (Kulasekara & Blomfield, 1999), is determined by Rho-dependent termination (Hinde *et al.*, 2005; Joyce & Dorman, 2002). Rho utilisation (rut) sites, required for Rho-dependent termination are incorporated into the *fimE* transcript when *fimS* is in the off orientation, but not in the on orientation.

1.7 Auxiliary factors required for *fim* recombination

The ability to bring strand exchange sites into close contact (synapsis) is a prerequisite for all recombination events, and in many instances this requires additional factors. In the case of *fim* inversion, three regulatory proteins significantly affect the rate of switching in both directions: Integration Host Factor, the Leucine-responsive Regulatory Protein and the Histone-like Nucleoid Structuring protein.

1.7.1 Integration Host Factor (IHF)

IHF was originally identified as a host-encoded protein required for site-specific integrative recombination of phage λ into the *E. coli* chromosome (Miller & Friedman, 1980). However, it has subsequently been classified as a global regulator that affects a number of recombination reactions, initiation of replication at oriC, condensation of the bacterial nucleoid and the transcription of many *E. coli* genes (Arfin *et al.*, 2000; Freundlich *et al.*, 1992; Grindley *et al.*, 2006; Hwang & Kornberg, 1992; Ryan *et al.*, 2002; Ryan *et al.*, 2004). The ability of IHF to bind and bend DNA is central to its function in each of these processes. It binds as a (~ 20 kDa) heterodimer (composed of IHF α & β subunits, encoded by *ihfA* & *ihfB*), making contact with the DNA exclusively via interactions with the phosphodiester backbone and minor groove (Rice *et al.*, 1996). Although IHF protects a sequence >25 bp, only 9 bp show any sequence conservation (Goodrich *et al.*, 1990); the conserved element has the consensus 5' WATCARXXXXTTR (where W is A or T; R is A or G; and X is not conserved, and can be any nucleotide). Upon binding to DNA, IHF introduces sharp bends (> 160°) in the local sequence (Kosturko *et al.*, 1989; Lorenz *et al.*, 1999; Rice *et al.*, 1996; Teter *et al.*, 2000; Thompson & Landy, 1988). The crystal structure of IHF complexed with a 35 bp sequence has been solved, and this determined the centre of the IHF-induced bend to between the fourth and fifth nucleotides upstream of the conserved 9 bp sequence (Rice *et al.*, 1996).

1.7.2 Leucine-responsive regulatory protein (Lrp)

Lrp is a global regulator in *E. coli* that affects the expression of a large number of genes (Brinkman *et al.*, 2003); DNA microarray analysis has shown that its regulon is at least 10 % of all genes in this organism (Tani *et al.*, 2002). In solution it exists as a homodimer (18.8 kDa each) at nanomolar concentrations (Chen *et al.*, 2001a; Willins *et al.*, 1991), but is known to bind DNA in a range of multimeric forms that include dimers, tetramers, octamers and hexadecamers at micromolar concentrations (Chen *et*

al., 2001a; Thaw *et al.*, 2006; Wang & Calvo, 1993a; Wang & Calvo, 1993b). Lrp dimers bind to a consensus sequence, 5' YAGHAWATTWTDCTR (where Y is C or T; H is not G; W is A or T; D is not C; and R is A or G) (Cui *et al.*, 1995), and in doing so introduce significant bends in the DNA (Wang & Calvo, 1993a); *in vitro* a single dimer introduces a ~52° bend in linear DNA fragments, with two dimers cooperatively increasing this to ~135°. It was originally characterised for its ability to regulate expression of the *ilvIH* operon (that encodes acetohydroxy acid synthase III, required for the biosynthesis of isoleucine, leucine and valine), which it activates in the absence of exogenous leucine and represses in its presence (Platko *et al.*, 1990; Ricca *et al.*, 1989). However, expression of genes at other loci can be repressed, activated or remain unaffected by exogenous leucine (Calvo & Matthews, 1994). Whilst leucine reduces the affinity of Lrp dimers for their consensus binding sites (Ernsting *et al.*, 1993), cooperative binding to multiple sites is increased such that certain multimeric states are favoured over others. In particular, there is evidence to suggest that exogenous leucine dissociates Lrp hexadecamers to form leucine-bound octameric complexes (Chen & Calvo, 2002; Chen *et al.*, 2005; Chen *et al.*, 2001b).

In the on orientation, IHF binds to a *fimE*-proximal site adjacent to *fimS* (site 1) (Eisenstein *et al.*, 1987), and an IRR-proximal site within it (site 2) (Blomfield *et al.*, 1997); whilst Lrp binds cooperatively to two high-affinity, IRL-proximal sites within *fimS* (figure 1.2) (Gally *et al.*, 1994). Mutations that disrupt the expression of *ihfA*, *ihfB* or *lrp* markedly decrease *fimS* switching in both directions, by either recombinase protein (Blomfield *et al.*, 1993; Eisenstein *et al.*, 1987), and mutations that reduce binding of IHF or Lrp to these sites within *fimS in vitro* lead to a significant decrease in recombination *in vivo* (Blomfield *et al.*, 1997; Gally *et al.*, 1994). The ability of both these proteins to introduce sharp bends in DNA upon binding has led to the hypothesis that they facilitate synapsis, bringing the two inverted repeats into juxtaposition and directly enhancing the recombination event itself (Blomfield *et al.*, 1993).

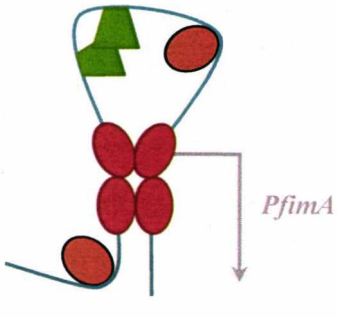


Figure 1.2: Schematic diagram of *fimS* during synapsis. IHF (orange spheres), Lrp (green blocks) and recombinase protein (red spheres) are defined.

1.7.3 H-NS

The Histone-like Nucleoid Structuring protein (H-NS) was originally named and characterised by its ability to bind and compact chromosomal DNA (Spassky *et al.*, 1984; Spurio *et al.*, 1992; Varshavsky *et al.*, 1977). It is a small (15 kDa), abundant protein that, in contrast to (basic) eukaryotic histones is globally neutral (pI 7.5). In addition to compacting chromosomal DNA, and constraining supercoils (Tupper *et al.*, 1994), it also serves as a transcriptional regulator, primarily as a repressor, for a large number (~ 200) of genes in *E. coli* and other gram-negative bacteria (Bertin *et al.*, 2001; Hommais *et al.*, 2001). Rather than recognising a specific sequence, H-NS shows high affinity for intrinsically curved A_n tracts (Lucht *et al.*, 1994; Rimsky *et al.*, 2001; Tanaka *et al.*, 1991; Yamada *et al.*, 1991; Zuber *et al.*, 1994). It forms a homodimer in solution, but can also form much larger oligomeric (>20-mer) complexes at specific DNA sites and at higher concentrations (Rimsky, 2004). The ability to form such oligomeric complexes is necessary for H-NS recognition of intrinsically curved DNA (Spurio *et al.*, 1997), but the mechanism of transcriptional regulation by these oligomeric structures remains unknown.

Loss of *hns* was originally characterised, in a *fimE* mutant background, by a rapid (100-fold greater) phase switching of type 1 fimbriae (Blomfield *et al.*, 1991b; Higgins *et al.*, 1988; Kawula & Orndorff, 1991; Spears *et al.*, 1986), and subsequent research has shown that it has both indirect and direct effects upon recombination. Indirectly it binds upstream of both *fimB* and *fimE* to alter their expression (Donato & Kawula, 1999; Donato *et al.*, 1997; Olsen & Klemm, 1994; Olsen *et al.*, 1998), and the subsequent de-repression of both recombinase genes (and concomitant increase in recombinase protein concentrations) is thought to largely account for the *hns* phenotype (this is defined further in the sections 1.8.2 & 1.10.1). H-NS is also thought to be a positive regulator of *lrp* expression (Levinthal *et al.*, 1994), and strongly influences *in vivo* DNA supercoiling (Higgins *et al.*, 1988). In addition, H-NS is thought to directly affect the expression of type 1 fimbriae by several methods. Gel mobility shift assays have shown that H-NS binds to sequences adjacent to and within the invertible element to stimulate expression of a *fimA-lacZ* reporter fusion, and fimbriation, in phase locked-on mutants (Schembri *et al.*, 1998). In addition to this effect on *fimAICDFGH* expression, *fimE* promoter activity, reading into the invertible element is able to inhibit FimB-mediated off-to-on inversion, and H-NS is required for this (O'Gara J & Dorman, 2000). Thus,

rather than simply stimulating or repressing expression of type 1 fimbriae, H-NS exerts a number of different effects at the level of recombinase expression, recombinase activity, and structural gene expression.

1.8 Environmental factors controlling *fim* recombination

Whilst expression of type 1 fimbriae is phase variable, there are a number of environmental factors that modulate the frequency of switching, or the expression of the structural genes: the branched-chain amino acids, temperature, growth rate, osmolarity and pH, DNA supercoiling, and the aminosugars *N*-acetylneuraminic acid and *N*-acetylglucosamine.

1.8.1 Branched-chain amino acids

Lrp is a global regulator that controls the expression of a large regulon; this includes genes involved in amino acid metabolism and fimbriae biosynthesis (Newman & Lin, 1995; Newman *et al.*, 1992). Inversion of *fimS* is modulated in response to the branched-chain amino acids and alanine (isoleucine, valine, leucine & alanine; IVLA) (Gally *et al.*, 1993), and this results from Lrp binding to a third site (site 3) within the invertible element (Roesch & Blomfield, 1998). In contrast to sites 1 & 2, occupancy of site 3 by Lrp inhibits recombination (Roesch & Blomfield, 1998). However, Lrp binds to site 3 with a much lower affinity than it does for sites 1 & 2; with binding to this third site being enhanced by occupancy of sites 1 & 2 by Lrp, and by additional, unknown factors (Lahooti *et al.*, 2005; Roesch & Blomfield, 1998). Exposure to IVLA selectively disrupts binding to this third site alone to stimulate recombination (Roesch & Blomfield, 1998).

When type 1 fimbriae are produced, a single cell will typically possess 200-500 fimbriae on their cell surface, which represents up to 8 % of the total cellular protein and a significant drain of resources. Other workers have suggested that exogenous leucine might be an indicator of nutritional status (Newman *et al.*, 1992), and regulation of type 1 fimbrial production in response to exogenous IVLA and Lrp might ensure that the cell only commits to producing them when there is sufficient resources available.

1.8.2 Temperature

Expression of type 1 fimbriae is regulated in response to temperature (Dorman & Bhriain, 1992; Kawula & Orndorff, 1991), such that off-to-on switching is 10-fold

higher at 37°C relative to 28°C (Gally *et al.*, 1993). This effect of temperature is thought to be indirect, as expression of the two *fim* recombinases is differentially affected by temperature; upon increasing temperature from 30°C to 37°C, repression of *fimE* increases 4-fold whilst that of *fimB* decreases two-fold (Olsen *et al.*, 1998). The expression of *hns* is stimulated (3- to 4-fold) during cold-shock (Brandi *et al.*, 1994; La Teana *et al.*, 1991), and as other workers have suggested for other genes, Olsen *et al.* (1998) have attributed these temperature effects on *fim* to the activity of H-NS which represses both *fimB* and *fimE* throughout this temperature range (Olsen *et al.*, 1998). However, this cold-shock induction of *hns* is associated with much larger temperature shifts (typically 37-10 °C), and there appears to be no change in the level of *hns* mRNA (and presumably H-NS concentration) between 30-37°C (Colonna *et al.*, 1995; Goransson *et al.*, 1990; Sonden & Uhlin, 1996). Whilst H-NS does repress a number of genes, in *E. coli* and other enteric bacteria, that are subjected to such thermoregulation, responsiveness to temperature (within this temperature range) has subsequently been attributed to other factors in several instances, and not H-NS (Tobe *et al.*, 1993; White-Ziegler & Low, 1992).

Despite the fact that the molecular mechanism that governs thermoregulation of *fim* remains uncharacterised, its significance is clear; expression of type 1 fimbriae is stimulated at temperatures it would encounter inside a mammalian host, and repressed at temperatures it would encounter outside of one.

1.8.3 Growth rate

FimB contains multiple copies of the rare UUG leucine codon (5 copies in *E. coli* 536, and 6 in K12), which is translated by the minor leucyl-tRNA ($tRNA_5^{leu}$) encoded by *leuX*. Several lines of evidence have shown that *leuX* indirectly affects expression of type 1 fimbriae by limiting the rate of translation of *fimB* mRNA (Morschhauser *et al.*, 1994; Newman *et al.*, 1994; Ritter *et al.*, 1997; Ritter *et al.*, 1995). This is thought to occur because it is the only leucyl-tRNA that does not show growth-dependent regulation, such that the concentration of $tRNA_5^{leu}$ decreases as growth rate increases (Rowley *et al.*, 1993).

Other workers have previously suggested that expression of *fimB* and *fimA* is repressed as cells enter stationary phase, and this phenotype is suppressed by a null mutation in *rpoS* (Dove *et al.*, 1997), which encodes an alternative sigma factor (known as σ^{38} or

σ^S) that is expressed during stationary phase (Jishage *et al.*, 1996). However, research carried out in our laboratory suggests the opposite (I. C. Blomfield, unpublished data), that *fimB* expression (and fimbriation, as a consequence) in fact increases during stationary phase. This observation has recently been confirmed and extended by Aberg *et al.* (2006), who showed that *fimB* expression (and fimbriation) is stimulated during stationary phase and could find no effect of an *rpoS* mutation. They discovered that type 1 fimbrial expression is stimulated by guanosine tetra-/ penta-phosphate ((p)ppGpp), the regulatory alarmone that induces a large number of genes in response to nutrient limitation (reviewed by Magnusson *et al.*, 2005). Thus, it appears that expression of type 1 fimbriae is rate-limited during rapid exponential growth, and stimulated during stationary phase and times of metabolic stress.

1.8.4 Osmolarity & pH

Schwan *et al.* (2002) have shown that a number of uropathogenic *E. coli* isolates down-regulate expression of type 1 fimbriae in response to high osmolarity and low pH during growth in LB; that is, a greater percentage of the population remains afimbriate (*fim*$^-$) under these conditions. A similar effect is observed during growth in LB supplemented with filter-sterilized human urine. These effects are primarily indirect, by modulating the expression of both recombinase genes (decreasing *fimB* expression, and increasing *fimE* expression). They provide evidence that the osmolarity-responsive EnvZ-OmpR system is required for this response (Forst *et al.*, 1990; Jo *et al.*, 1986; Mizuno & Mizushima, 1990), although the precise mechanism of control remains unclear.

1.8.5 DNA supercoiling

As is the case for many organisms, the *E. coli* genome is maintained in a negatively supercoiled state. This is principally determined by the opposing actions of DNA gyrase, Topoisomerase I, and to a lesser extent Toposiomerase IV; whilst DNA gyrase introduces negative supercoils, Topoisomerase I removes them and Topoisomerase IV removes positive supercoils (DiNardo *et al.*, 1982; Pruss & Drlica, 1989; Pruss *et al.*, 1982; Snoep *et al.*, 2002; Zechiedrich *et al.*, 2000). A number of external stimuli affect the global level of supercoiling, such as osmotic stress, oxygen tension, nutritional shifts and temperature change (Rui & Tse-Dinh, 2003). Furthermore, supercoiling induces torsional tension in DNA, and can therefore influence a number of cellular processes, such as initiation of replication, DNA looping and transcription (Funnell *et al.*, 1987; Lia *et al.*, 2003; Peter *et al.*, 2004).

Expression of type 1 fimbriae is known to be sensitive to supercoiling in at least two ways. Inversion of the *fim* invertible element is sensitive to localised supercoiling, with the inhibition of DNA gyrase introducing a bias towards off-to-on switching (Dove & Dorman, 1994). This bias in switching occurs in the absence of FimE and is attributed to a modulation of FimB specificity; recently, it was shown that this bias requires both relaxed DNA (as a consequence of DNA gyrase inhibition) and Lrp binding to sites 1 and 2 within *fimS* (Kelly *et al.*, 2006). Mutations that disrupt Lrp binding to either, or both of these sites reverse this bias towards on-to-off switching. In contrast, loss of Topoisomerase I or inhibition of DNA gyrase inhibit or stimulate *fimB* expression (respectively) (Dove & Dorman, 1994). These effects on *fimB* correlate well with the global level of supercoiling, and do not suggest that localised DNA supercoiling is a significant factor that controls expression of this gene.

1.8.6 Sialic acid

Sialic acid is the name given to the negatively charged, nine carbon monosaccharide *N*-acetylneuraminic acid (Neu5Ac) and its numerous (> 40) derivatives. Synthesis of these aminosugars is primarily restricted to mammals and a number of their pathogens, where they play a critical role in determining the outcome of interactions between the two. In higher eukaryotes, sialic acids are commonly found as the terminal sugar residue on the oligosaccharide chains of many cell-surface or serum glycoconjugates (Schauer, 2000; Vimr *et al.*, 2004). These terminally-sialylated-molecules facilitate a diverse range of cell-cell and small molecule recognition phenomena (reviewed by Schauer, 2000; Traving & Schauer, 1998; Varki, 1993). Sialic acid also serves as an anti-recognition molecule for mammalian hosts, such that sialylated surfaces are identified as 'self'. It is for this reason that a number of pathogenic microorganisms (including *E. coli* K1) incorporate sialic acid into their surface components, allowing them to masquerade as 'self' in order to evade the host immune system (Moran *et al.*, 1996; Vimr & Lichtensteiger, 2002). Sialic acids are also recognition determinants in the adhesion of many pathogenic viruses, bacteria and protozoa (Kelm & Schauer, 1997). The significance of sialic acid to commensal organisms is poorly understood beyond its use as a carbon, nitrogen and energy source (Chang *et al.*, 2004; Plumbridge & Vimr, 1999; Vimr *et al.*, 2004). Recently, our laboratory has shown that off-to-on phase variation of type 1 fimbriae in commensal *E. coli* K12 is significantly reduced in

the presence of sialic acid (El-Labany *et al.*, 2003). This effect is indirect upon recombination, as exogenous Neu5Ac inhibits expression of the *fimB* recombinase gene.

Coordination of type 1 fimbrial expression in response to sialic acid is interesting for a number of reasons. Despite the fact that sialic acid is relatively abundant in mammals, it is almost always present within the oligosaccharide chains of glycolipids or glycoproteins (Schauer, 2000; Vimr *et al.*, 2004); as a consequence, free sialic acid is typically scarce (Sillanaukee *et al.*, 1999). Furthermore, *E. coli* does not appear to possess a neuraminidase, required to cleave terminal sialic acid residues from these host oligosaccharide chains (Vimr, 1994), yet sialic acid appears to be a preferred carbon source for *E. coli* (Chang *et al.*, 2004; Snyder *et al.*, 2004; Vimr *et al.*, 2004). Thus, whilst other, neuraminidase$^+$ bacteria might generate free sialic acid for *E. coli* to utilise, free sialic acid levels rise during inflammation (Sillanaukee *et al.*, 1999), and our laboratory has previously proposed that *E. coli* interprets exogenous sialic acid as an indicator of immune activation (El-Labany *et al.*, 2003).

Upon activation of polymorphonuclear leukocytes (PMNs), sialidase is translocated from within an intracellular granule compartment to the plasma membrane (Cross & Wright, 1991), where most of the cell-associated sialic acid is thought to reside (DePierre *et al.*, 1980). This results in the desialylation of the cell surface, which is required for the migration of PMNs to the site of infection (Carlos & Harlan, 1994). However, this membrane-associated sialidase is also able to desialylate adjacent cells in the local environment (Cross & Wright, 1991), such that activation of PMNs is accompanied by an increase in the local free sialic acid levels. Type 1 fimbriae, as indeed all fimbriae, are highly immunogenic, therefore there is a clear rationale for decreasing the number of fimbriate (*fim*$^+$) bacteria in response to free sialic acid.

This is not just significant for commensal *E. coli*, within the gastrointestinal tract, but also for pathogenic *E. coli* at extraintestinal locations. Through their interaction with uroplakin 1a receptors on the bladder mucosa (Zhou *et al.*, 2001), UPEC isolates trigger an epithelial cell response, in an LPS-dependent manner (Fischer *et al.*, 2006; Hedlund *et al.*, 2001; Samuelsson *et al.*, 2004; Schilling *et al.*, 2001; Schilling *et al.*, 2003), that stimulates the production of three pro-inflammatory cytokines: interleukin-6 (IL-6), IL-8 and tumour necrosis factor α (TNFα) (Godaly *et al.*, 1998; Malaviya *et al.*, 1996; Samuelsson *et al.*, 2004). As IL-8 stimulates neutrophil migration (Baggiolini *et al.*,

1989; Godaly *et al.*, 1998), adherence of UPEC isolates to uroplakin 1a receptors results in an increase in localised free sialic acid levels (Cross *et al.*, 2003). There is a clear rationale for UPEC isolates recognising signals of immune activation, such as an increase in free sialic acid levels, because their interactions with the host actually trigger the immune response in this niche.

1.8.7 *N*-acetylglucosamine

In addition to sialic acid, off-to-on phase variation of type 1 fimbriae is also significantly reduced by exogenous N-acetylglucosamine (GlcNAc); and this effect is also indirect upon recombination, with exogenous GlcNAc inhibiting expression of the *fimB* recombinase gene (Sohanpal *et al.*, 2004). However, unlike sialic acid, which is a host-derived amino sugar and not essential for growth, GlcNAc is an essential component of the bacterial cell wall. In addition, GlcNAc is a constituent of the outer membrane lipopolysaccharide (LPS) and the polysaccharide capsules of gram-negative bacteria. Thus, as a phosphoenolpyruvate-dependent phosphotransferase system (PTS) sugar in *E. coli*, an increase in intracellular GlcNAc-6-P can be derived from a number of sources: *de novo* synthesis; environmental GlcNAc; as an intermediate in the catabolism of other aminosugars such as sialic acid (Plumbridge & Vimr, 1999); and cell wall recycling (Goodell, 1985; Park, 2001; Uehara *et al.*, 2005).

Like sialic acid, GlcNAc has also been proposed to be a signal of immune activation, and suppression of type 1 fimbrial synthesis in response to increased levels of GlcNAc might serve to limit inflammation (Sohanpal *et al.*, 2004). There are a number of observations that support this hypothesis. The first is that GlcNAc is an intermediate in the catabolism of sialic acid (Plumbridge & Vimr, 1999), such that increased levels of intracellular GlcNAc-6-P might be an indirect signal that the level of free sialic acid, within the host, is raised (refer to chapter 1.8.6 for more details). In addition, free GlcNAc levels rise during inflammation (Kriat *et al.*, 1991), and high levels of free *N*-acetyl-β-glucosaminidase are associated with upper urinary tract infections (Rodriguez-Cuartero *et al.*, 1998). It is also interesting to note that synthesis of curli, another adhesive cell-surface structure that is highly immunogenic (Bian *et al.*, 2000; Kikuchi *et al.*, 2005), is also inhibited by exogenous GlcNAc (Barnhart *et al.*, 2006).

1.9 Pathogenic Adaptations affecting *fim* in clinical isolates

Many workers have noted that there is considerable heterogeneity in the control of *fim* phase variation between clinical isolates. These differences were originally observed during growth on agar in the laboratory; whereas there is a strong selection for the afimbriate phase in commensal *E. coli* strains, some clinical isolates remain in the fimbriate phase and some continue to phase-vary (Hultgren *et al.*, 1986; Schwan *et al.*, 1992; Schwan *et al.*, 1994). Other workers have shown that this heterogeneity often results from differences at the level of recombination (Abraham *et al.*, 1986; Leathart & Gally, 1998; Lim *et al.*, 1998; Schwan *et al.*, 1992), and whilst the molecular mechanisms are not fully understood in each instance, a number of factors have been identified that might account for these differences.

There is evidence to suggest that recombinase specificity is modulated in clinical isolates by changes in the nucleotide sequence of the invertible element (Leathart & Gally, 1998). As a consequence, cloning invertible elements from certain clinical isolates into a commensal *E. coli* K12 background is sufficient for transfer of the switching behaviour between these strains (Leathart & Gally, 1998). However, exchange of these nucleotide sequences was unable to impart the same specificity from other clinical isolates (Leathart & Gally, 1998), which showed that there are additional factors that modulate recombination in some backgrounds. A number of additional recombinase proteins have been identified within the genomes of various pathogenic *E. coli* backgrounds: the neonatal meningitis isolate, *E. coli* K1 strain RS218, contains HbiF (Xie *et al.*, 2006); and a significant proportion of UPEC isolates, including strain CFT073, contain three *fim* recombinase homologues, IpuA, IpuB and IpbA (Bryan *et al.*, 2006). Furthermore, at least three of these have been shown to catalyse inversion of *fimS in vivo*; in contrast to FimB and FimE, HbiF and IpbA predominantly switch from off-to-on, whilst IpuA switches in both directions (Xie *et al.*, 2006). The precise role of these additional recombinase proteins, and their impact on the virulence of these two pathogens, remains unknown.

Many *E. coli* strains are capable of producing numerous types of fimbriae, each with their own target specificity, but it is very uncommon for more than one type to be produced at any given time (Nowicki *et al.*, 1984). This mutual exclusivity between alternative fimbriae is determined by cross-talk between fimbrial systems, such that synthesis of one type inhibits the synthesis of others. This has been characterised for

the two fimbrial adhesins that most uropathogenic *E. coli* isolates possess, type 1 fimbriae and P-fimbriae (Xia *et al.*, 2000). Cross-talk between these two fimbriae is determined by PapB (Holden *et al.*, 2001; Holden *et al.*, 2006; Xia *et al.*, 2000), the transcriptional regulator of the *pap* operon (Forsman *et al.*, 1989), which binds to *fimS* and inhibits FimB-mediated (off-to-on) recombination, whilst enhancing FimE-mediated (on-to-off) recombination (Holden *et al.*, 2001; Holden *et al.*, 2006; Xia *et al.*, 2000). In addition, cross-talk occurs between the related F1845-fimbrial (*daa*), S-fimbrial (*sfa*) and P-related fimbrial (*prf*) operons (Morschhauser *et al.*, 1994; van der Woude & Low, 1994), but only PapB and SfaB homologues are able to inhibit recombination of *fimS* (Holden *et al.*, 2001; Holden *et al.*, 2006; Xia *et al.*, 2000). There is also evidence that type 1 fimbrial expression might, in turn, be able to down-regulate the expression of P-fimbriae (Snyder *et al.*, 2005), but the molecular mechanism that that determines this effect remains unknown.

In addition to the differences in *fim* gene expression between *E. coli* isolates, there are also differences in adhesin specificity. Whilst the type 1 fimbriae of both commensal and pathogenic *E. coli* isolates bind with high affinity to trimannosyl (α1-3, α1-6 D-mannotriose) receptor residues, there is considerable variability (up to 15-fold) in their affinities for monomannosyl residues (Sokurenko *et al.*, 1997; Sokurenko *et al.*, 1994; Sokurenko *et al.*, 1995). These differences result from sequence variation within the *fimH* adhesin, with most (>80 %) commensal isolates binding to monomannosyl receptors with low affinity and most (>70 %) urinary tract isolates with high affinity (Sokurenko *et al.*, 1995). Only FimH variants that bind with high affinity to monomannosyl residues adhere to uroepithelial cells (Sokurenko *et al.*, 1997), and whilst this phenotype is advantageous for colonisation of (and virulence within) the urinary tract it is also thought to impair fitness for commensal ecology (Sokurenko *et al.*, 1998).

1.10 Regulation of *fimB* expression

The average intergenic region in *E. coli* K12 is 118 bp (Blattner *et al.*, 1997), and yet *fimB* is separated from the divergently transcribed *nanC* by 1.4 kb of non-coding sequence. Schwan *et al.* (1994) originally provided evidence that sequences far upstream of *fimB* might facilitate binding of a transcriptional activator, and subsequent analyses have defined both *fimB* proximal and *fimB*-distal regulatory elements that control expression of this recombinase gene.

1.10.1 *fimB*-proximal regulation

Arguably the single-most significant factor that determines precise levels of *fimB* expression is H-NS; gel mobility shift assays have shown that it binds avidly and cooperatively to multiple sequence elements upstream of *fimB in vitro* (Donato *et al.*, 1997), and *hns* null mutations significantly inhibit *fimB* expression (~5-fold) *in vitro* and *in vivo* as a consequence (Donato *et al.*, 1997; Olsen & Klemm, 1994; Olsen *et al.*, 1998). The *fimB* promoters have been mapped independently by several groups, and although there are discrepancies regarding precise transcriptional start sites there is evidence to suggest that there are two promoters, located ~150 bp and ~290 bp upstream of the *fimB* open reading frame (ORF) (Aberg *et al.*, 2006; El-Labany *et al.*, 2003; Olsen & Klemm, 1994; Schwan *et al.*, 1994). Whilst the mechanism of H-NS repression remains unknown, the sequences that it binds to *in vitro* are close to these promoters and its ability to inhibit *in vitro* transcription assays suggest that it directly affects promoter activity.

Recently, Aberg *et al.* (2006) showed that *fimB* expression is stimulated by the regulatory alarmone (p)ppGpp. Other workers have shown that (p)ppGpp can stimulate expression of genes in one of two ways: indirectly, by binding to and modulating the effect of regulatory proteins, or binding to RNAP and directly stimulating promoter activity. Aberg *et al.* (2006) presented evidence that suggested it might have a direct effect upon the *fimB* promoters.

1.10.2 *fimB*-distal regulation

The *nanC-fimB* intergenic region was first formally characterised by El-Labany *et al.* (2003), who constructed and analysed a series of deletion and replacement mutations in order to localise *cis*-active elements that control *fimB* expression (figure 1.3). This identified two short (< 30 bp) sequences (defined as 'region 1' and 'region 2'), located >600 bp upstream of the *fimB* promoters that stimulated expression. Two regulatory proteins were subsequently shown to bind to these sites (Sohanpal *et al.*, 2004): the *N*-acetylneuraminic acid (Neu5Ac) responsive regulator, NanR (region 1); and the *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P) responsive regulator, NagC (region 2/ NagC1). NagC also binds to a second site, within the Δ3 region (NagC2), and mutations that reduce the affinity of NagC for either of these operator sequences inhibit *fimB* expression (I. C. Blomfield, manuscript submitted for publication). As GlcNAc-6-

P is generated as an intermediate in Neu5Ac catabolism (Plumbridge & Vimr, 1999), exogenous Neu5Ac dissociates both NanR and NagC from their respective binding sites to inhibit *fimB* expression.

The NanR and NagC1 binding sites contain, or are immediately adjacent to a Dam-methylation site (5' GATC) (respectively), and these regulators are able to protect them from methylation both *in vitro* and *in vivo* (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004). Although DAM-methylation controls the phase variation of other genes in *E. coli* (Haagmans & van der Woude, 2000; Nou *et al.*, 1993), the significance of methylation protection at these two sites remains unknown, and does not appear to alter the binding affinities of either protein *in vitro* (I. C. Blomfield, personal communication). It is, however, diagnostic of NanR and NagC binding, and shows that they form stable nucleoprotein complexes which in order to become fully un-methylated must survive at least two rounds of DNA replication; this is because newly synthesised DNA is non-methylated, and in order for both DNA strands to remain un-methylated at any given site it must be protected from DAM-methylase activity through two rounds of synthesis. Analysis of the DAM-methylation status of sequences within the *yjhA-fim* intergenic region has also shown that the NanR and NagC1 operator sequences are not bound simultaneously, with methylation-protection occurring at either of these sites, but not both, at any given time. The mechanism that controls these alternative nucleoprotein complexes remains unknown.

The effects of NanR and NagC at these operator sequences are two-fold; they activate *fimB* expression at a distance (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004), and inhibit expression of the divergently transcribed *nanC* gene. Other workers have shown that *nanC* encodes a KdgM-family outer membrane porin that, in the absence of the non-specific porins OmpF and OmpC facilitates the uptake of Neu5Ac (Blot *et al.*, 2002; Condemine *et al.*, 2005). Expression of *nanC* is inhibited by NanR, whose binding site overlaps the -10 hexamer of the single *nanC* promoter; and NagC, at the nagC1 operator occludes CAP (Catabolite Activator Protein) from binding to an overlapping operator sequence.

Prior to this study, preliminary data had been obtained that suggested an unusual mechanism might account for activation of *fimB* expression by NanR and NagC (El-Labany *et al.*, 2003). When *fimB* was moved to an ectopic location at *lac,* the Δ2

region, which contains the NanR and NagC1 binding sites (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004) was only required in *cis* when the sequence extending into *nanC* were also present in *cis*. This led to the hypothesis that this sequence, upstream of the Δ2 region might contain *cis*-active inhibitory elements, the activity of which NanR and NagC serve to antagonise.

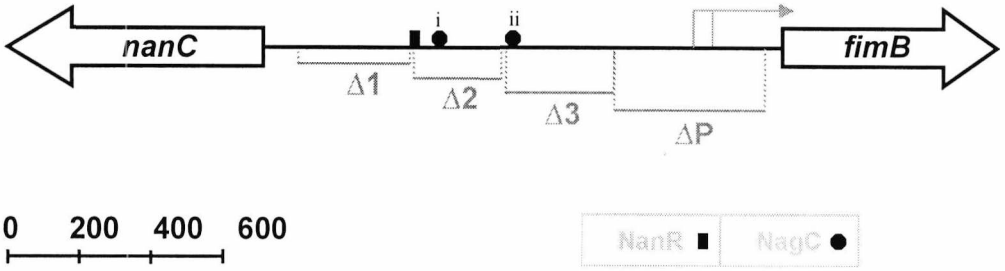


Figure 1.3: Schematic diagram of the *nanC-fimB* intergenic region. The NanR, NagC1 (i) and NagC2 (ii) binding sites are defined (Sohanpal *et al.*, 2004), as are the deletion mutations that were used to map *cis*-active elements within this region (El-Labany *et al.*, 2003).

1.10.3 Regulation of gene expression by GntR-family regulators

NanR belongs to the GntR family of regulators, which were originally defined by Haydon and Guest (1991) and named after the repressor of the gluconate operon in *Bacillus subtilis*. This family contains over 270 members, each defined by similarities within their N-terminal DNA binding domains, which contain a winged helix-turn-helix motif (Haydon & Guest, 1991). Although there is significant divergence in the C-terminal oligomerisation and effector-binding domains of GntR family members, they have subsequently been divided into five subfamilies which possess similarity (~ 55 %) throughout their entire sequences (Lee *et al.*, 2003; Rigali *et al.*, 2002). NanR belongs to the FadR-subfamily, which includes 40 % of all members of the GntR superfamily and is named after the regulator of fatty acid metabolism in *E. coli* (Rigali *et al.*, 2002).

In addition to the regulation of *fimB* and *nanC* expression (Condemine *et al.*, 2005; Sohanpal *et al.*, 2004), NanR is known to repress the *nanATEK-yhcH* operon (Kalivoda *et al.*, 2003), which is required for the uptake and catabolism of sialic acids (Plumbridge & Vimr, 1999). The NanR binding site is located within a 27 bp element that is

conserved at three positions within the *E. coli* K12 genome (Blattner *et al.*, 1997): the *fimB-nanC* intergenic region, upstream of *nanATEK-yhcH* and upstream of the *yjhBC* operon. In this way, the mechanism of *nanATEK-yhcH* repression is identical to that of *nanC*; it binds to a consensus site that overlaps a -10 promoter hexamer to occlude RNA polymerase (RNAP) from binding (Condemine *et al.*, 2005; Kalivoda *et al.*, 2003). The *yjhBC* operon remains uncharacterised, but the predicted product of *yjhB* has (>50 %) homology with the sialic acid transporter, NanT (Vimr & Troy, 1985), and the presence of this 27 bp conserved element suggests that it might also be under the transcriptional control of NanR.

There are currently nine members of the FadR-subfamily encoded within the *E. coli* genome (Rigali *et al.*, 2002), and whilst six of these are exclusively recognised as transcriptional repressors (Blanco *et al.*, 1986; Dong *et al.*, 1993; Quail & Guest, 1995; Rodionov *et al.*, 2000; Tong *et al.*, 1996), two (in addition to NanR) are known to function as transcriptional activators for certain genes: the (long-chain) fatty acid-responsive regulator, FadR, stimulates expression of the *fabA*, *fabB* and *iclR* genes (Campbell & Cronan, 2001; Gui *et al.*, 1996; Henry & Cronan, 1992); and the glycolate-responsive regulator, GlcC, stimulates expression of the *glcDEF* operon (Pellicer *et al.*, 1999). In each instance, activation by either of these regulators is thought to require direct contact with RNAP. Activation by FadR requires promoter proximal binding sites, between -40 to -53 bp relative to each transcriptional start site (Campbell & Cronan, 2001; Gui *et al.*, 1996; Henry & Cronan, 1992), a location where many transcriptional activators bind to mediate stimulatory protein-protein contacts with the RNAP holoenzyme ($\alpha_2\beta\beta'\sigma^{70}$) (Collado-Vides *et al.*, 1991; Hochschild & Dove, 1998). In contrast, GlcC binds to a consensus sequence -170 bp relative to the *glcD* promoter, and is entirely dependent upon IHF binding to a sequence at the mid-point between GlcC and the *glcD* promoter (Pellicer *et al.*, 1999). This strongly suggests that GlcC requires an IHF-induced bend, which would loop the intervening DNA and localise GlcC at the promoter region (Dworkin *et al.*, 1998).

1.10.4 Regulation of gene expression by ROK-family regulators

The ROK-family (Repressors, Orfs and Kinases) consists of both transcriptional regulators, primarily repressors, and sugar kinases (Titgemeyer *et al.*, 1994). They are divided into two classes, which are primarily differentiated by the presence or absence of an N-terminal helix-turn-helix DNA binding motif (Titgemeyer *et al.*, 1994). *E. coli*

possesses two ROK-family transcriptional regulators: NagC and Mlc. Whilst Mlc only functions as a transcriptional repressor (Kim *et al.*, 1999; Plumbridge, 1998a; Plumbridge, 1998b; Plumbridge, 1999; Tanaka *et al.*, 1999), NagC is known to function as a transcriptional activator for the *glmUS* operon, in addition to *fimB* (Plumbridge, 1995).

NagC binds as a dimer, and NagC sites typically function as pairs (Plumbridge, 2001). Often these pairs interact, via looping of the intervening DNA between these sites. With the exception of *fimB*, regulation by NagC requires at least one promoter-proximal binding site in order to serve as either a repressor or activator (Plumbridge, 2001). Repression occurs when one of its binding sites overlaps a promoter element (Plumbridge, 1999; Plumbridge, 2001; Plumbridge & Pellegrini, 2004). Activation of *glmUS* requires a binding site that is -47 bp relative to one promoter, where it appears to function as a classical transcriptional activator that directly contacts RNAP (Plumbridge, 1995).

1.11 Aims

Control of *fimB* expression by NanR and NagC is interesting for several reasons. They coordinate expression of *fimB*, hence off-to-on phase switching of type 1 fimbriae in response to Neu5Ac and GlcNAc, which are potentially key signals of a host inflammatory response (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004). Furthermore, they function at an uncharacteristically large distance from the *fimB* promoter region, where they were thought to activate *fimB* expression by antagonising more distant, *cis*-active elements. The aim of this study was to investigate the mechanism by which NanR and NagC activate *fimB* expression.

Chapter 2

Materials and methods

2.1 Bacterial strains, plasmids, bacteriophage and oligonucleotides

Descriptions and genotypes of bacterial strains used in this study, and the bacteriophage, plasmids and oligonucleotides used to construct them are listed in tables 2.1 – 2.4. All the bacterial strains constructed in this thesis are derivatives of *E. coli* K-12 strain MG1655, and nucleotide sequence fragments that are mentioned in this study are defined according to their position within the complete genome sequence (Blattner *et al.*, 1997). All plasmids used for allelic exchange are derivatives of the chloramphenicol-resistant, temperature sensitive vector pMAK705 (Hamilton *et al.*, 1989).

2.2 Media, growth and storage conditions

All reagents were obtained from Sigma-Aldrich UK unless otherwise stated, and were no less than standard laboratory grade. Water was purified through multiple de-ionising and filtration cartridges (Millipore Milli-Q®) prior to use, and is referred to here as de-ionised H_20 (dH_2O). Solutions were sterilised by autoclaving at 121 °C for 15 minutes when appropriate, or by filtration through sterile 0.2 μm filters (Nalgene). Solutions were stored by refrigeration at 4 °C in the dark unless otherwise stated. Where concentrations of buffer components are indicated by percentages, they refer to weight:volume ratios at room temperature (RT) for solids, and volume:volume ratios (also at RT) for liquids. Bacteria were frozen for long-term storage at −70 °C in lysogeny broth (LB) plus 15 % glycerol (Fisher Scientific UK). When required a small portion of the frozen stock was thawed and plated onto, or inoculated into, appropriate medium.

2.2.1 Liquid media

Lysogeny broth (also known as Luria-Bertani broth, or LB), Tryptose B1 broth (TB1) supplemented with maltose and magnesium (TBMM) and rich defined MOPS medium were used in this study. LB was prepared as described (Sambrook, 1989): 10 g of tryptone (BD; Becton-Dickinson & Co.), 5 g of sodium chloride (Fisher Scientific UK), and 5 g of yeast extract (Oxoid) per litre (autoclaved). TBMM was also prepared as described (Way *et al.*, 1984): Tryptose B1 (TB1) broth (10 g of tryptone, 5 g of sodium chloride per litre) was autoclaved and supplemented with filter-sterile 0.2 % maltose and 10 mM magnesium sulphate after cooling. Rich defined (RD) MOPS medium is minimal 3-(*N*-morpholino)-propanesulphonic acid (MOPS) media supplemented with nucleotide bases, vitamins and amino acids (Neidhardt *et al.*, 1974).

Table 2.1: Bacterial strains used in this study. The method of construction is defined for all strains that were made in this study.

Bacterial strain	Genotype	Reference, source or construction
AAEC189a	λ⁻ F⁻ *endA1 Thi-1 hsdR17 supE44 ΔlacU169 recA mcrA mcrB ΔfimB-H*	Laboratory collection (Blomfield *et al.*, 1991b)
C600	*λ⁻ F⁻ supE44 hsdR thi-1 thr-1 leuB6(Am) lacY1 tonA21 fhuA21 cyn101 glnV44(AS) rfbC1 glpR200*	*E. coli* genetic stock centre (CGSC3004)
J96	Genotype unknown (genome not sequenced)	(Hull *et al.*, 1981)
MG1655	λ⁻ F⁻ *rph-1*	*E. coli* genetic stock centre (Guyer, 1981)
BGEC043	MG1655 *ΔlacZYA fimB'-lacZYA* ΔEco0109-ClaI Ω sacB-kan*	Laboratory collection (Blomfield *et al.*, 1993)
BGEC696	MG1655 *ΔlacZYA fimB'-'lacZYA** Δ2 Ω sacB-kan*	Laboratory collection
BGEC711	MG1655 *ΔlacZYA fimB'-'lacZYA Δ2*	Laboratory collection
BGEC811	MG1655 *ΔlacZYA fimB'-'lacZYA ΔnanC-yjhT-yjhS Ω sacB-kan*	Laboratory collection
BGEC904	MG1655 *ΔlacZYA fimB'-'lacZYA P2* (mutation in -10 consensus)	Laboratory collection
BGEC905	MG1655 *ΔlacZYA fimB'-'lacZYA*	Laboratory collection
BGEC916	MG1655 *ΔlacZYA fimB'-'lacZYA P1* (mutation in -10 consensus)	Laboratory collection
KCEC023	MG1655 *ΔlacZYA fimB'-'lacZYA Δ819*	Laboratory collection
KCEC025	MG1655 *ΔlacZYA fimB'-'lacZYA Δ819 Δ2*	Laboratory collection
KCEC118	MG1655 *ΔlacZYA fimB'-'lacZYA Rm1*	Laboratory collection

KCEC171	MG1655 Δ*lacZYA fimB'-'lacZYA ΔP*	Laboratory collection
KCEC486	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5*	Laboraory collection
KCEC619	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S1)	mTn*10* transposon mutagenesis in Δ2 background
KCEC620	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S3)	mTn*10* mutagenesis in Δ2 background
KCEC621	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S1)	mTn*10* mutagenesis in Δ2 background
KCEC622	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S1)	mTn*10* mutagenesis in Δ2 background
KCEC623	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S5)	mTn*10* mutagenesis in Δ2 background
KCEC624	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S1)	mTn*10* mutagenesis in Δ2 background
KCEC625	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S1)	mTn*10* mutagenesis in Δ2 background
KCEC626	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S4)	mTn*10* mutagenesis in Δ2 background
KCEC627	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S2)	mTn*10* mutagenesis in Δ2 background
KCEC655	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#1) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC656	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 hns*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC657	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC658	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#2) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC659	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 hns*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC660	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#3) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC661	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#4) at undefined location	mTn*10* mutagenesis in Rm1 background

KCEC662	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC663	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#5) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC664	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC665	MG1655 Δ*lacZYA fimB'-'lacZYA* mTn*10* at (#6) undefined location	mTn*10* mutagenesis in Rm1 background
KCEC666	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 hns*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC667	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#7) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC668	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC669	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC670	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC671	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#8) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC672	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#9) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC673	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* mTn*10* (#10) at undefined location	mTn*10* mutagenesis in Rm1 background
KCEC674	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC675	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 yjhE-fecE*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC676	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 hns*::mTn*10*	mTn*10* mutagenesis in Rm1 background
KCEC677	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC678	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC679	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 lrp*::mTn*10*	mTn*10* mutagenesis in Rm5 background

KCEC680	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 hns*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC681	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC682	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 sdhB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC683	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC684	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC685	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC686	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 gatB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC687	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 crr*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC688	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC689	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC690	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC691	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5* mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC692	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC693	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC694	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC695	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC696	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 nanC-fimB*::mTn*10*	mTn*10* mutagenesis in Rm5 background
KCEC703	MG1655 Δ*lacZYA fimB'-'lacZYA yjhE-fecE*::mTn*10*	P1 transduction of mTn*10*, KCEC675 → BGEC905

KCEC708	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 yjhE-fecE*::mTn*10*	P1 transduction of mTn*10*, KCEC675 → KCEC486
KCEC755	MG1655 Δ*lacZYA fimB'-'lacZYA hns*::mTn*10*	P1 transduction of mTn*10*, KCEC676 → BGEC905
KCEC756	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 hns*::mTn*10*	P1 transduction of mTn*10*, KCEC676 → KCEC118
KCEC757	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 hns*::mTn*10*	P1 transduction of mTn*10*, KCEC676 → KCEC486
KCEC769	MG1655 Δ*lacZYA fimB'-'lacZYAΔ743 Ω 'cynX-TT$_{rrnB}$-[cyn-lac intergenic region]*	Allelic exchange, pSF001 → BGEC811
KCEC771	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 Δ743 Ω ('cynX-TT$_{rrnB}$-[cyn-lac intergenic region]*	Allelic exchange, pSF003 → BGEC811
KCEC773	MG1655 Δ*lacZYA fimB'-'lacZYA* Partial deletion of the Δ3 region (Δ3·1)	Allelic exchange, pSF005 → BGEC696
KCEC775	MG1655 Δ*lacZYA fimB'-'lacZYA* Partial deletion of the Δ3 region (Δ3·2)	Allelic exchange, pSF007 → BGEC696
KCEC777	MG1655 Δ*lacZYA fimB'-'lacZYA* Partial deletion of the Δ3 region (Δ3·3)	Allelic exchange, pSF009 → BGEC696
KCEC779	MG1655 Δ*lacZYA fimB'-'lacZYA* Partial deletion of the Δ3 region (Δ3·X)	Allelic exchange, pSF011 → BGEC696
KCEC781	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 Δ3·1*	Allelic exchange, pSF013 → BGEC696
KCEC783	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 Δ3·1*	Allelic exchange, pSF015 → BGEC696
KCEC785	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 Δ3·2*	Allelic exchange, pSF017 → BGEC696
KCEC787	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 Δ3·2*	Allelic exchange, pSF019 → BGEC696
KCEC789	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 Δ3·3*	Allelic exchange, pSF021 → BGEC696
KCEC791	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 Δ3·3*	Allelic exchange, pSF023 → BGEC696

KCEC793	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1 Δ3·X*	Allelic exchange, pSF025 → BGEC696
KCEC795	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5 Δ3·X*	Allelic exchange, pSF027 → BGEC696
KCEC797	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement (het. Repl.) within the Rm10-NagC2 region (Rm11)	Allelic exchange, pSF029 → BGEC696
KCEC799	MG1655 Δ*lacZYA fimB'-'lacZYA* Small het. Repl. within the Rm10-NagC2 region (Rm12)	Allelic exchange, pSF031 → BGEC696
KCEC801	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm13)	Allelic exchange, pSF033 → BGEC696
KCEC803	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm14)	Allelic exchange, pSF035 → BGEC696
KCEC805	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm15)	Allelic exchange, pSF037 → BGEC696
KCEC807	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm16)	Allelic exchange, pSF039 → BGEC696
KCEC809	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm17)	Allelic exchange, pSF041 → BGEC696
KCEC811	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm18)	Allelic exchange, pSF043 → BGEC696
KCEC813	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm19)	Allelic exchange, pSF045 → BGEC696
KCEC815	MG1655 Δ*lacZYA fimB'-'lacZYA* Small heterologous replacement within the Rm10-NagC2 region (Rm20)	Allelic exchange, pSF047 → BGEC696
KCEC822	MG1655 Δ*lacZYA fimB'-'lacZYA Δ2 nanC-fimB*::mTn*10* (S4) *ΔP Ω sacB-kan*	Allelic exchange, pSF075 → KCEC626

KCEC824	MG1655 Δ*lacZYA fimB'-'lacZYA* Δ2 Δ*P* Ω *sacB-kan* Δ*743* Ω *'cynX-TT*$_{rrnB}$*-[cyn-lac intergenic region]*	Allelic exchange, pSF075 → KCEC771
KCEC826	MG1655 Δ*lacZYA fimB'-'lacZYA P1+P2* mutations	Allelic exchange, pSF077 → BGEC043
KCEC828	MG1655 Δ*lacZYA fimB'-'lacZYA* Δ2 Δ*P nanC-fimB*::mTn*10* (S4)	Allelic exchange, pSF073 → KCEC822
KCEC830	MG1655 Δ*lacZYA fimB'-'lacZYA* Δ2 Δ*P* Δ*743* Ω *'cynX-TT*$_{rrnB}$*-[cyn-lac-intergenic region]*	Allelic exchange, pSF073 → KCEC824
KCEC832	MG1655 Δ*lacZYA fimB'-'lacZYA* Size-matched heterologous replacement of the Δ3 deletion (up to the 5' TTT that precedes NagC2) with *bla* sequence (Δ*3·0-HR*)	Allelic exchange, pSF049 → BGEC696
KCEC834	MG1655 Δ*lacZYA fimB'-'lacZYA* Size-matched heterologous replacement of the Δ3·1 deletion with *bla* sequence (Δ*3·1-HR*)	Allelic exchange, pSF051 → BGEC696
KCEC836	MG1655 Δ*lacZYA fimB'-'lacZYA* Size-matched heterologous replacement of the Δ3·2 deletion with *bla* sequence (Δ*3·2-HR*)	Allelic exchange, pSF053 → BGEC696
KCEC838	MG1655 Δ*lacZYA fimB'-'lacZYA* Size-matched heterologous replacement of the Δ3·3 deletion with *bla* sequence (Δ*3·3-HR*)	Allelic exchange, pSF055 → BGEC696
KCEC854	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* Δ*3·0-HR*	Allelic exchange, pSF057 → BGEC696
KCEC856	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5* Δ*3·0-HR*	Allelic exchange, pSF059 → BGEC696
KCEC858	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* Δ*3·1-HR*	Allelic exchange, pSF061 → BGEC696
KCEC860	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5* Δ*3·1-HR*	Allelic exchange, pSF063 → BGEC696
KCEC862	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* Δ*3·2-HR*	Allelic exchange, pSF065 → BGEC696
KCEC864	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5* Δ*3·2-HR*	Allelic exchange, pSF067 → BGEC696

KCEC866	MG1655 Δ*lacZYA fimB'-'lacZYA Rm1* Δ*3-3-HR*	Allelic exchange, pSF069 → BGEC696
KCEC868	MG1655 Δ*lacZYA fimB'-'lacZYA Rm5* Δ*3-3-HR*	Allelic exchange, pSF071 → BGEC696
KCEC870	MG1655 Δ*lacZYA fimB'-'lacZYA* P_{Klemm}	Allelic exchange, pSF079 → BGEC043

*/** Two *fimB* reporter fusion constructs are utilised in this study: a transcriptional fusion (derivatives of BGEC043 and defined here as *fimB'-lacZYA*); and a translational fusion (derivatives of BGEC696 and defined here as *fimB'-'lacZYA*).

Table 2.2: Bacteriophage used in this study.

Bacteriophage	Genotype	Reference/ source of construction
λNK1323	λ *b522 c1857 Pam80 nin5*	Nancy Kleckner (Way *et al.*, 1984)
$P1_{vir}$		(Miller, 1972)

Table 2.3: Plasmids used in this study. All plasmids with names beginning 'pSF' were constructed in this study.

Plasmids	Genotype	Reference, source of construction
pBR322	Amp^R (*bla* from Tn*3*) Tet^R (from pSC101) *rep* replicon (from pMB1)	Laboratory collection (Bolivar *et al.*, 1977)
pIB305	pMAK705 (temperature-sensitive pSC101 replicon) Cam^R	Laboratory collection (Blomfield *et al.*, 1991a)
pIB307	pIB305 (*lac* homology removed)	Laboratory collection (Blomfield *et al.*, 1991a)
pDG028	pIB307 Δ*lrp* Ω *sacB-kan* [*sacB-kan* fragment is bound by BamHI sites]	Laboratory collection
pIB346	pIB307 '*nanC - fimB-lacZYA* Δ*Eco0109- 5'-fimB* Ω *sacB-kan*	Laboratory collection
pIB347	*pIB307 'nanC- (5'-FimB-LacZYA)* [translational-fusion] Δ*P*	Laboratory collection
pIB413	pIB307 *5' nanC – 5' fimB* (HindIII-blunt-to-ClaI-blunt)	Laboratory collection

pIB462	pIB305 *'cynX -linker (HindIII-BamHI)-'lacI* (*lac* homology vector)	Laboratory collection (Kulasekara & Blomfield, 1999)
pIB622	pIB307 *5' yjhT- 5' fimB* (HindIII-blunt-to-ClaI-blunt)	Laboratory collection
pIB672	pIB347 *(P2 mutation: 5' TTTACT* converted to *5' GTGACT)*	Laboratory collection
pIB678	pIB347 *(P1 mutation: 5' TATAAT converted to 5' GCTAAT)*	Laboratory collection
pIB688	pIB462 Δ*(HindIII-BamHI)-linker Ω 'nanC-fimB-fimE* (HindIII-BamHI)	Laboratory collection (El-Labany *et al.*, 2003)
pIB690	pIB462 Δ*(HindIII-BamHI)-linker Ω 'nanC-Δ2-fimB-fimE* (HindIII-BamHI)	Laboratory collection (El-Labany *et al.*, 2003)
pSE001	pIB413 Rm1	Laboratory collection (El-Labany *et al.*, 2003)
pSE035	pIB347 Δ*P* [ΔP mutation denoted by Eco0109 site]	Laboratory collection (El-Labany *et al.*, 2003)
pBKS137	pIB413 *Rm5*	Laboratory collection (Sohanpal *et al.*, 2004)
pSF001	pIB622 Δ*743 Ω 'cynX-* TT_{rrnB} *- [cyn-lac inter-genic region]*	pIB688 (NcoI-blunt-SphI) fragment ligated into pIB622 (partial-HindIII-blunt, SphI) fragment

pSF003	pIB622 *Δ2 Δ743 Ω 'cynX-TT$_{rrnB}$ - [cyn-lac intergenic region]*	pIB690 small fragment ligated into pSF001 large fragment (both cut with SacII-SphI)
pSF005	pIB413 *Δ3·1*	PCR fragment ('ApaIIUPF-D31mutER', pIB413) ligated into pIB413 (both cut with ApaII-Eco0109)
pSF007	pIB413 *Δ3·2*	PCR fragment ('ApaIIUPF-D32Eco0R', pIB413) ligated into pIB413 (both cut with ApaII-Eco-0109)
pSF009	pIB413 *Δ3·3*	PCR fragment ('ApaIIUPF-D33EmR', pIB413) ligated into pIB413 (both cut with ApaII-Eco- 0109)
pSF011	pIB413 *Δ3·4*	PCR fragment ('ApaIIUPF-D34mutER', pIB413) ligated into pIB413 (both cut with ApaII-Eco-0109)
pSF013	pIB413 *Δ3·1 Rm1*	pSE001 small fragment ligated into pSF005 large fragment (both cut with HindIII-BlpI)
pSF015	pIB413 *Δ3·1 Rm5*	pBKS137 small fragment ligated into pSF005 large fragment (both cut with HindIII-BlpI)
pSF017	pIB413 *Δ3·2 Rm1*	pSE001 small fragment ligated into pSF007 large fragment (both cut with HindIII-BlpI)
pSF019	pIB413 *Δ3·2 Rm5*	pBKS137 small fragment ligated into pSF007 large fragment (both cut with HindIII-BlpI)
pSF021	pIB413 *Δ3·3 Rm1*	pSE001 small fragment ligated into pSF009 large fragment (both cut with HindIII-BlpI)
pSF023	pIB413 *Δ3·3 Rm5*	pBKS137 small fragment ligated into pSF009 large fragment (both cut with HindIII-BlpI)
pSF025	pIB413 *Δ3·4 Rm1*	pSE001 small fragment ligated into pSF011 large fragment (both cut with HindIII-BlpI)
pSF027	pIB413 *Δ3·4 Rm5*	pBKS137 small fragment ligated into pSF011 large fragment (both cut with HindIII-BlpI)
pSF029	pIB413 *Rm11*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm401Rev' & 'Rm401FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF031	pIB413 *Rm12*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm402Rev' & 'Rm402FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)

pSF033	pIB413 *Rm13*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm403Rev' & 'Rm403FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF035	pIB413 *Rm14*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm404Rev' & 'Rm404FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF037	pIB413 *Rm15*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm405Rev' & 'Rm405FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF039	pIB413 *Rm16*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm406Rev' & 'Rm406FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF041	pIB413 *Rm17*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm407Rev' & 'Rm407FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF043	pIB413 *Rm18*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm408Rev' & 'Rm408FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF045	pIB413 *Rm19*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm409Rev' & 'Rm409FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF047	pIB413 *Rm20*	PCR fragment ligated into pIB413 (both cut with ApaII-Eco0109). Mutagenesis primers ('Rm410Rev' & 'Rm410FWD') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.2 for details)
pSF049	pIB413 *Δ3·0-HR*	PCR fragment ('D3HetRepF-D30HRRev', pBR322) ligated into pIB413 (both cut with ApaII-Eco-0109)
pSF051	pIB413 *Δ3·1-HR*	PCR fragment ('D3HetRepF-D31HRRev', pBR322) ligated into pIB413 (both cut with ApaII-Eco-0109)

pSF053	pIB413 Δ*3·2-HR*	PCR fragment ('D3HetRepF-D32HRRev', pBR322) ligated into pIB413 (both cut with ApaII-Eco-0109)
pSF055	pIB413 Δ*3·3-HR*	PCR fragment ('D3HetRepF-D33HRRev', pBR322) ligated into pIB413 (both cut with ApaII-Eco-0109)
pSF057	pIB413 Δ*3·0-HR Rm1*	pSE001 small fragment ligated into pSF049 large fragment (both cut with HindIII-BlpI)
pSF059	pIB413 Δ*3·0-HR Rm5*	pBKS137 small fragment ligated into pSF049 large fragment (both cut with HindIII-BlpI)
pSF061	pIB413 Δ*3·1-HR Rm1*	pSE001 small fragment ligated into pSF051 large fragment (both cut with HindIII-BlpI)
pSF063	pIB413 Δ*3·1-HR Rm5*	pBKS137 small fragment ligated into pSF051 large fragment (both cut with HindIII-BlpI)
pSF065	pIB413 Δ*3·2-HR Rm1*	pSE001 small fragment ligated into pSF053 large fragment (both cut with HindIII-BlpI)
pSF067	pIB413 Δ*3·2-HR Rm5*	pBKS137 small fragment ligated into pSF053 large fragment (both cut with HindIII-BlpI)
pSF069	pIB413 Δ*3·3-HR Rm1*	pSE001 small fragment ligated into pSF055 large fragment (both cut with HindIII-BlpI)
pSF071	pIB413 Δ*3·3-HR Rm5*	pBKS137 small fragment ligated into pSF055 large fragment (both cut with HindIII-BlpI)
pSF073	pIB347 Δ*HindIII-BlpI* [HindIII site within *nanC*] Δ*P*	pSE035 cut with HindIII-BlpI, then blunted and re-ligated
pSF075	pSF073 Δ*P* Ω *sacB-kan*	pDG028 small fragment (BamHI-blunt) ligated into pSF075 large fragment (Eco0109-blunt)
pSF077	pIB347 *P1+P2*	pIB672 small fragment ligated into the pIB678 large fragment (both cut with HindIII-SphI)
pSF079	pIB347 P_{Klemm}	PCR fragment ligated into pIB347 (both cut with BlpI-Sph1). Mutagenesis primers ('P3mutR' & 'P3mutF'), 'standard' primers ('ApaIIUPF' & 'FimBUPR') and plasmid template (pIB413) were used to make this PCR product (see section 2.6.4.3 for details)

Table 2.4: Oligonucleotide primers used in this study. Non-annealing (red font) sequence and **REase sites** (**bold font**) are highlighted.

	Primer name	Nucleotide sequence (5' → 3')
*	ApaIIUPF	GTGATGTTTGCCATAGATTGC
*	D31mutER	GGGGAA**GGGACCT**GATGATTTTTTCATCTACATTCG
*	D32Eco0R	ACTACG**AGGTCCT**CCTTCCTCAATCACAGCCATTCGC
*	D33EmR	CATCA**AGGGCCC**TCTCCAGATCGACTTCAG
*	D34mutER	CCACA**AGGCCCT**AAACATATTTCATGATGAGAATTATGC
*	D3HetRepF	CATCAG**GCTGAGC**ATAATTCTCATCATGAAATATGTTTGAAGAACGTTTTCCAATGATGAGCAC
*	D30HRRev	CCACA**AGGCCCT**TGTGCAAAAAAGCGGTTAGCTCCTTCGG
*	D31HRRev	GGGGAA**GGGACCT**CAGCACTGCATAATTCTCTTACTGTCATGCC
*	D32HRRev	ACTACG**AGGTCCT**GTGAGTACTCAACCAAGTCATTCTGAG
*	D33HRRev	CATCAC**GGGCCCT**CACATAGCAGAACTTTAAAAGTGCTC
	DownP	GGCGAAGCTTCTGATGTTTAATTG
*	FimBUPR	GACGCGCACTAGCTGTGCTGG
*	mTn1x	CTCCCCTGAGCTTGAGGGG
*	mTn2x	GGCATCACTTCTTGGATAGGGATAAGG
*	mTn3x	GCGATCCCACCACCAGCC
*	mTn4x	GGCACCTTTGGTCACCAACGC
*	P3mutF	GAACATATCACATGAGGTATGGATAAGATTAGTTGCATTAATGATGAGG
*	P3mutR	GCAACTAATCTTATCCATACCTCATGTGATATGTTCTGTAACAAG
	Regmid	CCGCGGATCCTCTTAGAAAAAGATCG
*	Rm401FWD	CGGC**CCGCGG**TAAAGATCTTAGTTTAACTATTTGTTTTATAAATAATTTATTAAGAGTC
*	Rm401REV	CCGG**CCGCGG**CTTCTGGACCGCAAGTGTTCGTCTGGCTTCACATGGC
*	Rm402FWD	GCGG**CCGCGG**ATCATTAACTATTTGTTTTATAAATAATTTATTAAGAGTCTAAAC
*	Rm402REV	CCGC**CCGCGG**AATGTTTTTGAAGACCACCGCAAGTGTTCGTCTGGC
*	Rm403FWD	GCGG**CCGCGG**AACAATTATAAATAATTTATTAAGAGTCTAAACAAGGG
*	Rm403REV	CCGC**CCGCGG**TTACTAAGATCTTTAGTTTTTGAAGACCACCGC

*	Rm404FWD	GGCGCCGCGGTATTAAATTTAAGAGTCTAAACAAGGGGAGCTTTGCAAGC
*	Rm404REV	CCGCCCGCGGAACAAATAGTTAAACTAAGATCTTTAGTTTTTGAAGACC
*	Rm405FWD	GCGGCCGCGGTGAAGGGGAGCTTTGCAAGCTAACTCAGTGAGC
*	Rm405REV	CGCGCCGCGGAGAATTTAAATTATTTATAAAACAAATAGTTAAACTAAGATC
*	Rm406FWD	GCGGCCGCGGAACGAAGCTAACTCAGTGAGCTTGGTGAAAATCAGTG
*	Rm406REV	GCGGCCGCGGGGAAGTTTAGACTCTTAATAAATTATTTATAAAACAAATAGTTAAAC
*	Rm407FWD	GGCCCCGCGGCACAGCTTGGTGAAAATCAGTGTTTACCCGCCATCAGG
*	Rm407REV	GCGCCCGCGGTCGAAGCAAAGCTCCCCTTGTTTAGACTCTTAATAAATTATTTATAAAAC
*	Rm408FWD	CCGGCCGCGGACAGTGTTTACCCGCCATCAGGCTGAGC
*	Rm408REV	CCGGCCGCGGGGTTCGACACTGAGTTAGCTTGCAAAGCTCCCCTTGTTTAGAC
*	Rm409FWD	GGCCCCGCGGATGGGCGCATCAGGCTGAGCATAATTCTCATCATG
*	Rm409REV	GGCCCCGCGGCATTTTCACCAAGCTCACTGAGTTAGCTTGCAAAGCTCCCC
*	Rm410FWD	GGCGCCGCGGCCATAATTCTCATCATGAAATATGTTTCCTGGTTTGTGGC
*	Rm410REV	GCGCCCGCGGACTACGCGGGTAAACACTGATTTTCACCAAGCTCACTGAGTTAGC
*	YjhAF	CCGTCGCCCATTTCTTATGG

* Constructed in this study.

Minimal MOPS solution was made as a 10x concentrate containing the following reagents (per litre): MOPS (Melford Laboratories Ltd), 1.0 M (400 ml); *N*-[Tris-(hydroxymethyl)-methyl]-glycine (tricine), 1.0 M (40 ml); iron (II) sulphate heptahydrate, 0.01 M (10 ml); ammonium chloride, 1.9 M (50 ml); potassium sulfate, 0.276 M (10 ml); calcium chloride, 0.05 M (10 ml); magnesium chloride hexahydrate, 0.247 M (10 ml); sodium chloride, 5.0 M (100 ml); micronutrients[*5] (10 ml); and dH_2O (360 ml). Each reagent was dissolved individually, and added in the above order to prevent them from precipitating. After the addition of tricine, the pH of the solution was adjusted to 7.4 using potassium hydroxide (Fisher Scientific UK). Concentrated Minimal MOPS solution was filter sterilised, then stored in aliquots at −20 °C.

Minimal MOPS media, as described by Neidhardt *et al.* (1974) contains 0.4 % glucose and 1.32 mM dibasic potassium phosphate in addition to MOPS solution. However, in this study glycerol (Fisher Scientific UK) was substituted for glucose (also 0.4 %) because the latter would obscure the effects of *N*-acetylneuraminic acid and *N*-acetylglucosamine upon their addition to the medium. In order to acknowledge this alternative carbon source, rich defined MOPS medium is abbreviated to RD_{glyc} MOPS in this study. Stock 100x solutions of glycerol and dibasic potassium phosphate were autoclaved or filter sterilised (respectively), then refrigerated until required.

Nucleotide bases were made as two 10x solutions: the first contained adenine, cytosine and uracil (each 2 mM) dissolved in 0.015 M potassium hydroxide; and the second, containing only guanine (also 2 mM, and in 0.015 M potassium hydroxide). They were filter sterilised and stored in aliquots.

The following vitamins were first dissolved in 0.02 M potassium hydroxide: thiamine hydrochloride, calcium pantothenate, 4-aminobenzoic acid, 4-hydroxybenzoic acid and 2,3-dihydroxybenzoic acid (each 20 mM). They were then combined to make a single, 100x stock using dH_2O, with concentrations of vitamins being 1 mM and that of potassium hydroxide being 4 mM. The 100x stock solution was then filter sterilised and stored in aliquots.

The following L-amino acids were made as a 10x stock solution (per litre): alanine, 1.6 M (5 ml); arginine, 0.97 M (5 ml); asparagine, 114 mM (40 ml); aspartic acid, 1.03 M in 0.67 M potassium hydroxide (5 ml); glutamic acid, 1.2 M in 0.86 M potassium

hydroxide (5 ml); glutamine, 0.24 M (25 ml); glycine, 1.6 M (5 ml); histidine, 0.54 M (5 ml); isoleucine, 0.4 M (10 ml); leucine, 0.1 M (80 ml); lysine, 1.0 M (5 ml); phenylalanine, 0.1 M (40 ml); proline, 0.8 M (5 ml); serine, 0.8 M (125 ml); threonine, 0.8 M (5 ml); tryptophan, 0.1 M (10 ml); tyrosine, 20 mM in 0.1 M potassium hydroxide (100 ml); valine, 0.6 M (10 ml); dH_2O (515 ml). The solution was then filter sterilised and stored in aliquots. L-cysteine and L-methionine were freshly prepared when required, and added directly to the medium: 2.38 ml of 61 mM cysteine per litre, and 3.33 ml of 60 mM methionine.

The following micronutrient solution was added to the 10x MOPS concentrate (per 100 ml): ammonium molybdate tetrahydrate, 10 mM (30 µl); boric acid, 0.1 M (400 µl); cobalt (II) chloride, 10 mM (300 ml); copper (II) sulphate pentahydrate, 10 mM (100 µl); manganese (II) chloride, 0.01 M (800 µl); zinc sulphateheptahydrate, 10 mM (100 µl); dH_2O (98.27 ml). The micronutrient solution was filter sterilised, then refrigerated in aliquots until required.

Once all the ingredients were added to make 1x RD_{glyc} MOPS medium, the solution was filter sterilised prior to use. The media was refrigerated at 4 °C for a maximum of three days, after which it was discarded and a fresh batch made from concentrated stocks.

2.2.2 Solid media

L-agar was used for the enrichment (and selection) of particular strains, in numerous protocols used in this study; this medium is LB supplemented with 1.5 % agar (Difco). Tryptose B1 (TB1) agar was used to make bacteriophage λ plaques, and is TB1 broth supplemented with 10 g/ L agar; this was used in conjunction with TB1 top agar, which is TB1 broth supplemented with 7 g/ L agar. Lactose MacConkey agar (Difco) was used as an indicator medium to evaluate expression of *fimB-lacZ* reporter fusions during transposon mutagenesis. Sucrose agar was used to select for recombinant bacteria (Blomfield *et al.*, 1991b), and is L-agar with 6 % sucrose (Fisher Scientific UK) supplemented, and sodium chloride omitted. Bacteria were grown on solid media in static incubators at 28, 37 or 42 °C, depending on specific requirements of each protocol.

2.2.3 Addition of antimicrobial agents

Liquid and solid media was supplemented with antibiotics in order to select for particular strains and plasmids. Stock solutions of each antibiotic were made as follows: chloramphenicol (CAM), 34 mg/ ml in ethanol (Fisher Scientific UK); kanamycin A (Kan), 25 mg/ ml in dH_2O; and tetracycline hydrochloride (Tet), 10 mg/ ml in 50 % ethanol. Stock solutions were stored in aliquots at −20 °C; only those without ethanol (kanamycin) were filter sterilised prior to use. Working concentrations of each antibiotic were 25 µg/ ml (chloramphenicol) or 10 µg/ ml (kanamycin and tetracycline).

2.2.4 Additional supplements

Liquid and solid media were also supplemented with calcium chloride*, magnesium sulfate*, sodium citrate*, *N*-acetylneuraminic† acid and *N*-acetylglucosamine† when required. Each reagent was stored as a filter sterilised 1 M stock, either at RT (*) or −20 °C (†). Working concentrations are defined where they are used in this study.

2.3 Centrifugation

Two centrifuges were used in this study: for centrifugation of eppendorf tubes up to a maximum acceleration of 10, 000 x g, the Eppendorf MiniSpin® microcentrifuge was used; and for centrifugation of Falcon tubes, or eppendorf tubes at accelerations exceeding 10, 000 x g, the Sigma Laboratory Centrifuge 2K15 (Philip Harris Scientific). Three rotors were used with the Sigma 2K15 centrifuge: Sigma 256/ 97 Nr.12149, for centrifugation of 50 ml Falcon tubes; Sigma 147/ 98 Nr.12141-H, for 15 ml Falcon tubes; and Sigma 21/ 98 Nr.12148 for 1.5 & 2 ml eppendorf tubes.

2.4 Transformation of DNA

Transformation is the genetic alteration of a cell resulting from the introduction, uptake and expression of foreign genetic material. In order to facilitate these alterations in a bacterium such as *E. coli* K-12, transformation must be preceded by a method that increases their ability to uptake this foreign genetic material (their compentency). The protocols used to facilitate these two processes in this study are essentially as described Sambrook *et al.* (1989).

2.4.1 Competent cells protocol

The following method was used to increase the competency of *E. coli* cells: a culture was inoculated from −70 °C stocks into 10 ml of LB and incubated overnight at 37 °C in a shaking water bath (180 rpm). The following day, 8 µl of the overnight culture was sub-cultured into 28 ml of pre-warmed LB and incubated in a shaking water bath (180 rpm) at 37 °C until an OD_{600} of 0.5 was obtained. Five minutes before the culture reached this optical density (OD_{600} ~ 0.4375), 3.75 ml of sterile warm glycerol was slowly added to the culture. The culture was then incubated in ice water for 10 minutes, and centrifuged at 4 °C for 10 minutes at 1, 700 x g after this incubation. The supernatant was discarded and the pellet was gently suspended in an ice-cold magnesium chloride solution (0.1 M magnesium chloride, 15 % glycerol). The resuspended cells were then centrifuged at 4 °C for 8 minutes at 1, 530 x g. The supernatant was discarded and the pellet gently resuspended in 6.25 ml of ice-cold T-salts (75 mM calcium chloride, 6 mM magnesium chloride, 15 % glycerol). The cells were incubated on ice water for a minimum of 20 minutes, being gently agitated periodically; this was to increase exposure of cells to the calcium ions in solution, as they are responsible for increasing competency. They were then centrifuged at 4 °C for 6 minutes at 1380 x g. The supernatant was discarded and the pellet gently resuspended in 1.25 ml of ice-cold T-salts. Once resuspended the cells were transferred in 100 µl aliquots to pre-cooled eppendorf tubes, and stored at −70 °C. The (competent) cells remained at −70 °C for a minimum of 24 hours prior to use.

2.4.2 Assessing transformation efficiency of competent cells

The transformation efficiency of all new batches of competent cells was assessed prior to their use, in order to maximise the possibility that transformation would occur with desired DNA fragments. In order to do so, an aliquot of the competent cells was transformed (section 2.4.3) with 1 ng of super-coiled plasmid DNA (10 µl of a 0.1 ng/ µl stock). Both undiluted and a ten-fold dilution of cells were plated onto L-agar (containing a suitable antibiotic) and incubated at 37 °C overnight. The number of colonies present after incubation was used to assess the competency of cells, with 10-100 colonies from the ten-fold dilution (~ 10^6 to 10^7 colonies per µg DNA per ml cells) being desirable.

2.4.3 Transformation protocol

An aliquot of competent cells was thawed (either in hand or on ice), then placed in a sterile glass tube on ice. 5-10 µl of DNA was added to the competent cells, and incubated on ice for 30 minutes. Cells were then heat-shocked at 42 °C for 30 seconds in order to encourage uptake of the foreign DNA, then briefly incubated on ice. 1 ml of LB was added to the cells, and they were then incubated at 37 °C* for 1 hour with gentle shaking. 200µl of cells were plated onto 5 L-agar plates (containing a suitable antibiotic) (200 µl per plate). The plates were then incubated overnight at 37 °C*. The following day transformants were used to inoculate LB (containing a suitable antibiotic), and the inoculated media was incubated overnight at 37 °C* in a shaking water bath (180 rpm). An aliquot of this overnight culture was then stored in 15 % glycerol at −70 °C.

* When using temperature sensitive plasmids, the incubation temperature was altered to 28 °C.

2.5 Isolation and manipulation of DNA

Two commercial kits from QIAGEN Ltd were used in this study to extract and purify plasmid DNA: the QIAprep Spin Miniprep Kit, and QIAfilter Plasmid Midi Kit. DNA restriction and modification enzymes were purchased from New England Biolabs, Promega Corporation and Roche Applied Science. Amplification of DNA by PCR was carried out using a thermostable DNA polymerase from Roche Applied Science: TAQ DNA Polymerase for diagnostic and inverse-PCR; Pwo DNA Polymerase for plasmid DNA; Expand High Fidelity PCR System, and Expand Long Template PCR System for amplification of genomic DNA. PCR and REase-digestions were carried out according to the manufacturer's recommendations. Purification of DNA from amplification, restriction and modification reactions and from agarose gels was carried out using the Promega Wizard SV Gel & PCR Clean-up System. All the kits were used in accordance with the manufacturer's guidelines. Oligonucleotide primers were synthesised by MWG Biotech or QIAGEN Operon, and stored as 500 pmol/ µl stocks at −70 °C (long term) and 50 pmol/ µl stocks at −20 °C (for use in PCR). Sequencing of DNA was carried out by MWG Biotech or the Advanced Biotechnology Centre (ABC), Imperial College. Standard protocols were used for all other molecular genetic procedures (Ausubel *et al.*, 1987; Sambrook *et al.*, 1989).

2.5.1 Isolation of plasmid DNA

All plasmids used and constructed in this study were transformed into the bacterial strain AAEC189a for long-term storage at −70 °C. When required, the desired −70 °C stock was used to inoculate an aliquot of LB (containing a suitable antibiotic) and incubated overnight in a shaking water bath (180 rpm) at a suitable temperature; 28 °C for temperature-sensitive plasmids, or 37 °C for all others. A maximum of 5 ml of overnight culture was used in conjunction with the QIAprep Spin Miniprep kit, and 50-100 ml with the QIAfilter Plasmid Midi Kit. Overnight cultures were treated according to the manufacturer's guidelines, and the resulting plasmid stocks were stored at -20 °C.

2.5.2 Isolation of genomic DNA

Genomic DNA was isolated according to a standard protocol (Ausubel *et al.*, 1987), with minor modifications: 567 µl of an overnight culture was added directly to an eppendorf tube containing 30 µl of 10 % sodium dodecyl sulphate(SDS) (BDH Laboratory Supplies, UK) and 3 µl of Proteinase K (4 mg/ ml) (Fisher Scientific UK), and incubated at 65 °C for a minimum of 2 hours. 100 µl of sodium chloride (5 M) was added and the solution mixed gently. 80 µl of warm CTAB solution (10 % hexadecyltrimethylamonium bromide in 0.7 M sodium chloride) was added and mixed gently, and the solution was incubated at 65 °C for a further 10 minutes. 780 µl of chloroform:isoamyl alcohol (24:1) was added to the solution, mixed, and then centrifuged at 10, 000 x g for 5 minutes. The upper aqueous phase was transferred to a fresh eppendorf tube, and an equal volume of phenol:chloroform:isoamyl alcohol was added (25:24:1, saturated with a solution containing 10 mM tris-(hydroxymethyl)-aminomethane (TRIS), pH 8.0, and 1 mM ethylenediamine-tetraacetic acid (EDTA)). The solution was gently mixed, and then centrifuged at 10, 000 x g for 5 minutes. The upper aqueous phase was transferred to a fresh eppendorf tube and 0.6 volumes of isopropanol (Fisher Scientific UK) were added; this was gently mixed until a high molecular weight, DNA precipitate became visible. The solution was then centrifuged 10, 000 x g for 2 minutes, after which the supernatant was removed by aspiration. The DNA pellet was then washed with 1 ml of 70 % ethanol, and centrifuged at 10, 000 x g for 1 minute. The supernatant was again removed by aspiration, and the pellet briefly air-dried. The pellet was resuspended overnight in 10-50 µl of TE (10 mM TRIS hydrochloride (Fisher Scientific UK), pH 8.0, 1 mM EDTA (BDH Laboratory Supplies, UK)) to avoid shearing the genomic DNA.

2.5.3 Digestion of DNA with restriction endonucleases

Restriction endonucleases were used according to the manufacturer's guidelines; in the optimal volume of the supplied reaction buffer, for 1-2 hours at the optimal temperature. Partial digestions were carried out using (2-fold) serial dilutions of the restriction enzyme for one hour at the optimal reaction temperature. When required to digest DNA with more than one restriction enzyme, both restriction endonucleases were added to the same reaction and a compatible reaction buffer and reaction temperature was used (according to the manufacturer's guidelines). If more than two restriction endonucleases were required, or the restriction endonucleases were incompatible with each other, DNA was digested sequentially (with a maximum of two enzymes per reaction). After digestion, reaction mixtures were either purified with the Promega Wizard Gel and PCR Clean-up System, or gel-purified (see chapter 2.5.7). If the restriction endonucleases could not be heat-inactivated, they were subjected to a phenol:chloroform:isoamyl alcohol (25:24:1, saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) extraction prior to purification.

Restriction enzymes (restriction endonucleases or REase) are described in this study according to the revised nomenclature proposed by Roberts *et al.* (2003).

2.5.4 Modification of DNA using Klenow

Klenow (Roche Applied Science) was used to convert cohesive ends of digested DNA fragments to non-cohesive ends. This enzyme is the large fragment of DNA polymerase I from *E. coli* (lysogenic strain NM 964), and although it retains the 5'→3' polymerase and 3'→5' exonuclease activities of native DNA polymerase I, it lacks the 5'→3' exonuclease activity. The activities of Klenow were exploited in this study to both fill-in 3' recessed ends (in the presence of dNTPs), and cut-back 3'overhangs generated by digestion of DNA by restriction endonucleases. In this way, DNA fragments could be digested with restriction endonucleases that generate incompatible ends, and then covalently linked by ligation. These reactions were carried out according to the manufacturer's guidelines.

2.5.5 Ligation of DNA fragments with T4 DNA ligase

DNA fragments generated by REase digestion were covalently joined using T4 DNA ligase (Roche Applied Science). Reaction volume, time and temperature, and molar ratios of DNA fragments (vector:insert) were chosen in accordance to the

manufacturer's guidelines: these were typically 20 µl, overnight (>16 hours) at 4 °C (cohesive ends) or 16 °C (non-cohesive) with a molar ratio of 1:2. After incubation, 5-10 µl of the ligation mixture was used in the transformation of *E. coli* cells (strain AAEC189a, 100 µl).

2.5.6 Amplification of DNA by PCR

The polymerase chain reaction (PCR) facilitates *in vitro* amplification of a target DNA sequence. Amplification is achieved by enzymatic polymerisation of DNA primed from oligonucleotides annealed such that they flank the target sequence. Repetitive cycles of double-stranded DNA denaturation at high temperature, annealing of oligonucleotide primers at low temperature and elongation of synthesised DNA strands at intermediate temperature were performed in a thermal cycling heating-block (Progene-Techne thermocycler). All oligonucleotides used in this study were synthesised by MWG Biotech or Qiagen Operon.

PCR reactions were set up on ice, as follows: primer 1, 50 pmol/ µl (1 µl); primer 2, 50 pmol/ µl (1 µl); PCR buffer, 10x (10 µl); deoxynucleotides (2'-deoxyadenosine-5'-phosphate, 2'-deoxycytidine-5'-phosphate, 2'-deoxyguanosine-5'-phosphate, 2'-deoxythymidine-5'-phosphate) (Promega Corporation), 2 mM each (10 µl); DNA template, ~ 0.5 ng/ µl (2 µl); DNA Polymerase, (variable concentration, according to manufacturer's guidelines; typically 1-5 U/ µl) (0.75 µl); dH_2O, 75.25 µl. Only 0.5 µl of Pwo DNA Polymerase was used per reaction, and the volume of dH_2O was adjusted accordingly to take this into account.

Negative control PCR reactions were carried out in parallel, in which there was no DNA template present (and the volume of dH_2O adjusted accordingly). In order to maintain uniformity between PCR reactions and their controls, two 'master' mixes were made: one contained dNTPs, (2x) primers and dH_2O sufficient for both reactions; and the other contained PCR buffer, DNA polymerase and dH_2O (also sufficient for both reactions). Identical reaction volumes could then be made using (49 µl) aliquots of these two master mixes, combined with DNA template (2 µl dH_2O in the negative control). The volume of each master mix was at least 10 % greater than the total required, to take into account the volume lost during pipetting. Filter tips were used to aliquot all the reagents, to minimise the risk of contamination. Each reaction was then transferred to a thermocycler, which was set to cycle as follows: 95 °C for 30 seconds (this 'hot start'

dissociates the primers from the template, and prevents them from annealing to incorrect target sequences) (1 cycle); 95 °C for 30 seconds, 55 °C for 40 seconds, 72 °C for 45 seconds (10 cycles); the same temperatures and times as the previous 10 cycles, but with the elongation time at 72 °C increasing by 5 seconds every cycle (15 cycles); 72 °C for 7 minutes (to ensure that each strand synthesised was completed to the full length of the target sequence) (1 cycle); 10 °C (∞; until ready to process the reaction). If the target sequence exceeded 1 kb in length, the elongation time at 72 °C was increased by 1 minute for every addition kb of sequence; and the temperature of each elongation step was adjusted from 72 °C to 68 °C.

2.5.7 Agarose gel electrophoresis of DNA

Electrophoresis was used to separate DNA fragments according to size on a 0.8-1.5 % agarose gel (agarose LE, Roche Applied Science) in 1x TAE buffer (40 mM Tris base, 20 mM acetic acid and 1 mM EDTA, pH 8.5) (Sambrook, 1989 16). DNA samples were mixed with standard (1x) loading buffer that contained bromophenol blue, xylene cyanol FF and orange G, and loaded into the sample wells of the agarose gel. Samples were electrophoresed, typically at 5.0 V/ cm until they had migrated to a sufficient distance, and the gel was then stained by soaking in 0.5 μg/ ml ethidium bromide for 30-45 minutes. The gel was then de-stained in 1 mM magnesium sulphate (or dH_2O) for 20 minutes, to remove unbound ethidium bromide and facilitate identification of very small amounts of DNA. The DNA was visualised by exposing the gel to ultra-violet (UV) light and observing the fluorescence (Kodak Gel Logic 100) resulting from DNA-bound ethidium bromide. Molecular weight markers (λ DNA/ HindIII and ΦX174 DNA/ HaeIII, Promega Corporation) were loaded alongside the DNA samples prior to electrophoresis in order to estimate DNA fragment sizes.

Agarose gel electrophoresis was used in this study for two purposes. The first was as a diagnostic tool to check digestion and PCR products against each other, and/ or molecular weight markers; in this instance, the above protocol was used. The second was to separate DNA fragments within a digestion or PCR. For this purpose the electrophoresis, staining and de-staining of DNA samples was carried out according to the above protocol, but the visualisation step was not carried out (the Kodak Gel Logic 100 emits both long- and short-wavelength UV, and exposure to the latter would otherwise damage the DNA fragments). Instead, the desired DNA fragment (or band) was excised from the agarose gel using a clean razor blade. Identification and excision

of DNA bands was carried out in a dark-room using a UV trans-illuminator (minimum exposure, long-wavelength UV only) (UVGL-25 Mineralight lamp, Ultra-violet Products, California, USA). The excised DNA fragment could then be purified using a commercial kit (Promega Wizard Gel & PCR Clean-up System).

2.5.8 Nucleic acid quantification

The concentration of DNA in solution was determined spectrophotometrically by rearranging the Beer-Lambert Law:

[e2.1] $A = \varepsilon c l$ ('A' is absorbance; 'ε' is the molar extinction coefficient; 'c' is the concentration of the solute; and 'l' is the length of the light path in solution).

The absorbance was calculated at a wavelength of 260 nm (A_{260}), at which double stranded DNA has a molar extinction coefficient of 0.02 $(\mu g/ ml)^{-1} cm^{-1}$. The purity of DNA samples were determined by measuring A_{280}, and then calculating the A_{260}/ A_{280} ratio. For clean DNA preparations the ratio should be 1.8-2.0; a ratio greater than 2.0 indicates contamination with RNA, whilst lower than 2.0 indicates contamination of the preparation with protein or phenol.

2.5.9 Sequence determination of DNA

Modification of DNA can introduce unwanted sequence alterations, and in order to confirm that these had not been introduced, DNA fragments manipulated in this study were sequenced for errors. Sequence determination was also used in this study to identify sequences that were simply unknown. DNA was first amplified by PCR using a proof-reading DNA polymerase (such as Pwo or the Expand enzymes) and primers specific to the region of interest. The PCR product was checked on an agarose gel to confirm that the fragment amplified was approximately correct in size and that there were no unwanted products in either the PCR sample or the negative control. The PCR product was then gel purified (Promega Wizard Gel and PCR Clean-up System), and its concentration determined spectrophotometrically. The DNA was then sent to either MWG Biotech or ABC for sequencing. Upon receipt, the resulting pre-processed sequence was compared to the expected sequence; and I referred back to the raw sequence output file when differences between the two were encountered.

2.6 Genetic manipulation and strain construction

Transposon mutagenesis was used to mutagenize the *E. coli* K-12 genome and generalised transduction was used to move selectable genetic markers between strains (respectively); standard protocols were used (Kleckner *et al.*, 1991; Silhavy, 1984). Alternatively, non-selectable sequence alterations were cloned into a plasmid vector and transferred to a host strain by homologous recombination; the allelic exchange protocol described by Blomfield *et al.* (1991) was used to facilitate this transfer.

2.6.1 Transposons and transposon mutagenesis

Transposable elements (transposons) are discrete segments of DNA capable of moving from site to site within a genome, independent of any extensive sequence homology. Mobility is dependent upon the activity of a transposase enzyme that may (autonomously) or may not be (non-autonomously) encoded within the transposon. This transposase recognises a specific DNA sequence, located at either end of the transposon and facilitates excision and insertion at a different site. Transposons can be transferred into unrelated genomes after insertion into plasmid or viral vectors.

Tn*10* has been widely exploited by molecular geneticists to make insertions throughout bacterial genomes and disrupt open reading frames. It is a composite autonomous transposon consisting of a central tetracycline marker (T^C) bound by two IS*10* elements (left and right) in inverted orientation (Foster *et al.*, 1981). Transposition of Tn*10* is non-replicative, being excised from one site and inserted at another with preference for (but not restricted to) the symmetrical 5' XGCTXAGCX heptanucleotide (X can be any nucleotide) (Bender & Kleckner, 1986; Halling & Kleckner, 1982; Haniford *et al.*, 1991; Kleckner *et al.*, 1991). There is evidence that suggests the Tn*10* transposase makes contact with the sequences immediately adjacent to this heptanucleotide consensus, and these interactions also affect insertion (Halling & Kleckner, 1982; Pribil & Haniford, 2000).

A derivative of Tn*10* was used in this study to make genome-wide transposon insertions: mini-Tn*10* (mTn*10*). This derivative is a truncation of Tn*10* that only contains the T^C marker bound by the outermost 70 bp of IS*10* right (inverted) at both ends (figure 2.1). There are two main advantages to using mTn*10* in preference to Tn*10* as a mutagen. The first is that transposition frequency decreases by 40 % per kilobase of transposon length, and mTn*10* (2.9 kb) is much smaller that Tn*10* (10 kb). The

second is that mTn*10* (by definition of its name) is non-autonomous, and this increases the stability of insertions. In this study mTn*10* was transferred from the bacteriophage vector, λNK1323. The transposase is encoded in *cis* to mTn*10* within this vector; hence loss of the vector results in loss of the transposase (and the ability of mTn*10* to move to secondary sites). The transposase within λNK1323 bears two mutations, with adenine residues at bp 400 and 745 being altered to guanine residues (resulting in two amino acid substitution mutations, C134Y and C249Y, respectively). The two mutations (*ats1* and *ats2* respectively) alter the target site specificity of the Tn*10* transposase, and facilitate insertion of the transposon at sites that deviate significantly from the heptanucleotide consensus.

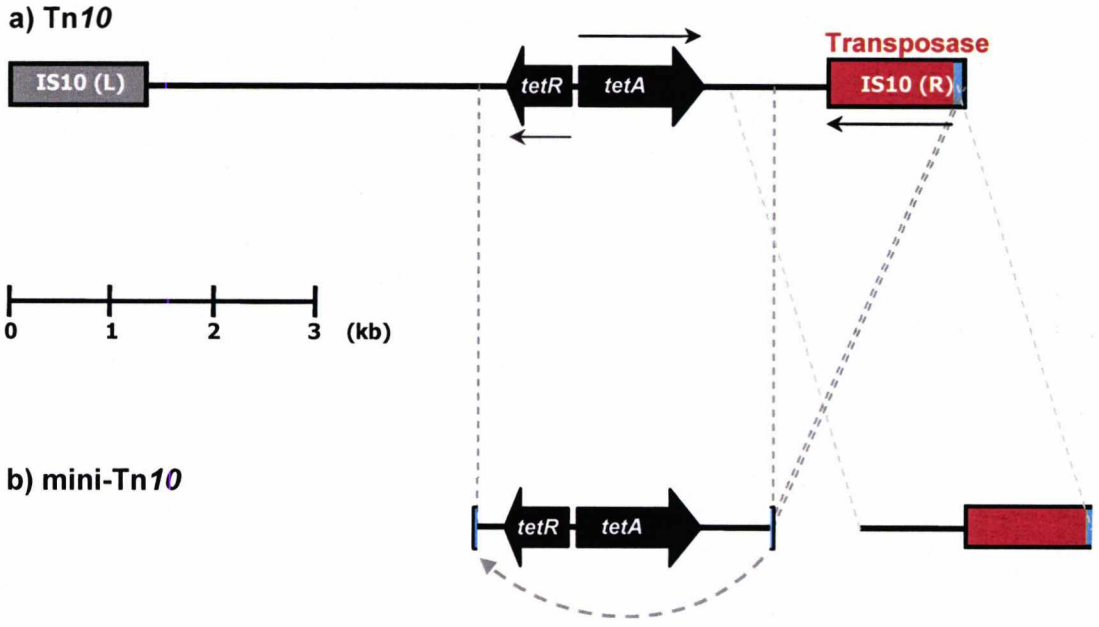


Figure 2.1: Schematic diagram defining functional elements of both Tn*10* and mTn*10*. The DNA fragments sub-cloned from Tn*10* to make mTn*10* are also defined. Bacteriophage λNK1323 served as a source of mTn*10* in this study, in which the transposase has two mutations conferring altered target site specificity (ATS) and is under the control of a P_{tac} inducible promoter (P_{tac} -ATS-Tpase). Arrows (⟶) define the direction of promoters for each gene.

Preparation of λNK1323 stocks and transposition of mTn*10* was carried out essentially as described by Way *et al.* (1984).

2.6.1.1 Preparation of bacteriophage λ lysate

Stocks of bacteriophage λNK1323 were prepared as a source of mTn*10* for transposon mutagenesis using the following protocol: an overnight culture of the permissive host, C600 was incubated overnight (> 14 hours) in TBMM at 37 °C in a shaking water bath (180 rpm). This host strain was used because it possesses an amber suppressor mutation within *leuB*, and λNK1323 has an amber codon within a gene required for DNA replication (*P80*). The following day a scraping of −70 °C λNK1323 stock was diluted in 100 µl LB, supplemented with 10 mM magnesium sulphate, and added to an equal volume of overnight culture and 2.5 ml of pre-cooled TB1 top agar (~ 45 °C). After mixing, this solution was immediately poured onto a plate containing 25 ml of TB1 agar, spread, and allowed to cool in a 37 °C incubator. Once set, the plate was inverted and incubated at 37 °C overnight. A fresh overnight (> 14 hours) culture of C600 in TBMM was also made (37 °C, 180 rpm). The following day 100 µl of overnight culture was sub-cultured into 10 ml of pre-warmed (37 °C) LB supplemented with 10 mM magnesium sulphate. A single λNK1323 plaque was carefully picked, using a Finn pipette tip, transferred into the flask of sub-cultured C600 cells and incubated at 37 °C in a shaking water bath until the former lysed the latter (~ 4-5 hours). Upon lysis 50 µl of chloroform was added, mixed well (vortexing) for 30 seconds and then allowed to sit on the bench for 10 minutes. The solution was then transferred to a Falcon tube and centrifuged at 4, 500 g for 10 minutes (4 °C). 5-7 ml of the supernatant was transferred to a fresh Falcon tube, and refrigerated until required. Fresh λNK1323 lysates were stored at 4 °C for a minimum of 24 hours prior to use.

2.6.1.2 Transposition from λ into the chromosome of *Escherichia coli*

Transposition was carried out according to the following protocol: 5 ml of TBMM was inoculated with the recipient strain (from -70 °C stock) and incubated overnight (> 14 hours) at 37 °C in a shaking water bath (180 rpm). The following day the culture was centrifuged for 10 minutes at 1, 467 x g (RT); the supernatant was discarded, the pellet resuspended in 1.5 ml of LB and then separated into 40 µl aliquots (x 24). After the addition of bacteriophage λ lysate* each solution was incubated for 15 minutes at room temperature (to allow the bacteriophage to enter the host cell), then a further 15 minutes at 37 °C (to encourage transposition to occur). After this second incubation 2 ml of LB, supplemented with 2.5 mM sodium pyrophosphate was added to each aliquot, mixed, and then centrifuged for 10 minutes at 2, 292 x g (RT). The supernatant was discarded, the pellets resuspended in 2 ml of fresh LB supplemented with 2.5 mM sodium

pyrophosphate and then incubated at 37 °C in a stationary water bath. After one hour, 100 µl of each aliquot was plated onto 5x Lactose MacConkey agar plates, supplemented with 10 µg/ ml tetracycline hydrochloride and incubated at 37 °C overnight (> 14 hours). The following day colonies with raised FimB-LacZ expression (lac^+) were picked with a sterile finn pipette tip, used to inoculate LB supplemented with 10 µg/ ml tetracycline, and then incubated at 37 °C overnight (> 14 hours) in a shaking water bath (180 rpm). A sample of each overnight culture was used to make −70 °C glycerol stocks.

* The actual volume of λNK1323 lysate used varied. New lysates were first tested by carrying out a small-scale transposition experiment, in which only a few aliquots of cells were incubated with 40 µl of bacteriophage (as described above). The transposition frequency (colonies per plate/ µl bacteriophage) was determined and used to calculate the volume of bacteriophage lysate required to obtain 300 colonies per plate.

2.6.1.3 Inverse PCR

Inverse PCR is a method defined by Ochman *et al.* (1988) that facilitates the rapid *in vitro* amplification of unknown DNA sequences that are flanked by known sequence. Unlike conventional PCR, in which primers flank the target region, inverse PCR uses primers that are orientated in opposite directions such that DNA polymerisation proceeds outwards from a known sequence. The template for these primers is a restriction fragment that has been ligated upon itself to form a circle. In this way unknown regions of DNA can be rapidly amplified without any prior knowledge of their sequence (figure 2.2). In this study inverse PCR was used to identify the region adjacent to mTn*10* insertions at unknown locations within the *E. coli* genome (after transposon mutagenesis).

Inverse PCR was carried out using the following protocol: genomic DNA was isolated from transposon insertion strains according to the protocol defined previously in this chapter (2.5.2); DNA pellets were resuspended in 10 µl of TE. DNA samples were analysed spectrophotometrically (chapter 2.5.9) and found to contain 0.1-0.3 µg DNA/ µl. The entire stock of genomic DNA was then digested overnight (> 16 hours) at 37 °C in a 100 µl volume with the restriction enzyme RsaI (20 U). After digestion the reaction was heat inactivated at 65 °C for 20 minutes. A ligation reaction was then carried out

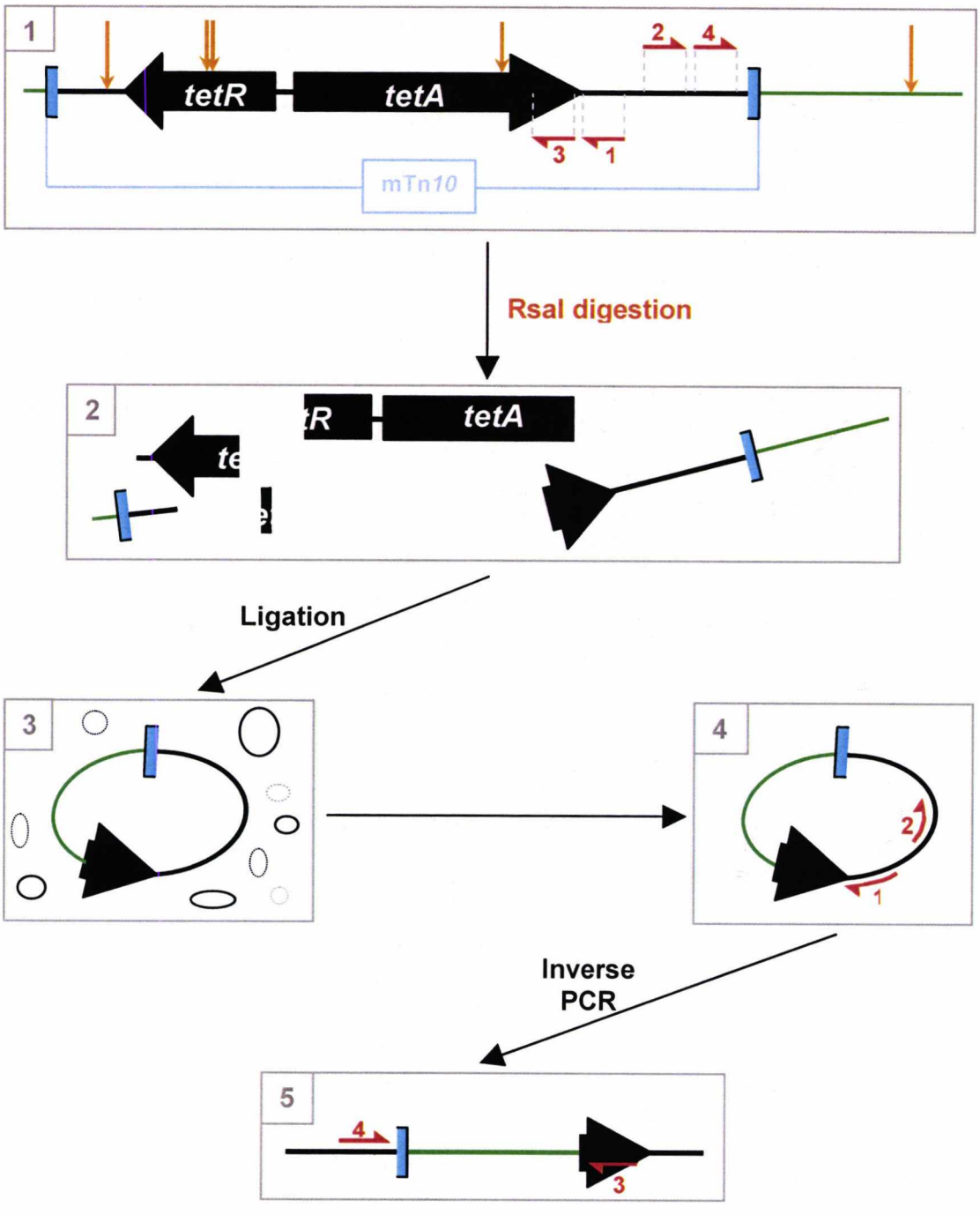


Figure 2.2: Schematic diagram defining the inverse PCR protocol used in this study. Genomic DNA [1] was initially digested with the restriction endonuclease, RsaI [2]. This REase recognises the tetranucleotide sequence, 5' GTAC, cleaving mTn*10* at four locations and the K-12 genome at 24, 074 locations. The reaction (REase) was heat-inactivated, diluted 1/10 and then ligated to form small, circularised DNA fragments [3]. Oligonucleotide primers, specific to the 1, 041 bp mTn*10* restriction fragment (*'tetA-IS10R*) were then used to carry out (inverse) PCR [4] (primers are labelled 1-4 here and correspond to 'mTn1x'-'mTn4x' respectively). In order to discriminate between erroneous and true inverse PCR products this PCR reaction [4] was used as template for a second PCR [5], with an alternative pair of primers (also specific to mTn*10*). This second PCR was gel-purified prior to sequencing.

overnight (> 16 hours) at 4 °C in a 50 µl reaction volume, with 5 units of T4 DNA ligase and 5 µl of the digested DNA. The following day inverse PCR was carried out using TAQ DNA polymerase, two inverse PCR primers ('mTn1x' & 'mTn2x') and 1 µl of the overnight ligation in a 50 µl reaction volume. At the same time, two control PCRs were carried out: one to show that there is sufficient DNA present for conventional PCR to work (using two primers, 'Regmid' & 'DownP' whose template remains undigested by RsaI); and the other to show that the DNA was digested to completion by RsaI (using two primers, 'DownP' & 'FimBUPR' whose template contains an RsaI restriction site). The inverse PCR was checked by running 10 % on an agarose gel and staining with ethidium bromide. In order to confirm that the product of this inverse PCR was correct, a second PCR was carried out using two primers ('mTn3x' & 'mTn4x') that anneal to it. This PCR was carried out in a 50 µl reaction volume with the inverse PCR product (1 µl of a 1/ 100 dilution) as template. After confirming the expected product sizes of these PCRs, they were gel-purified (refer to section 2.5.3), quantified (section 2.5.8) and then sent away for sequencing (section 2.5).

2.6.2 Generalised transduction using bacteriophage $P1_{vir}$

$P1_{vir}$ is a temperate bacteriophage that infects and lysogenizes a number of enteric bacteria, including *E. coli*. It has widely been exploited for its ability carry out generalised transduction and move genetic elements from one strain to another (Lennox, 1955), and in this study a virulent form of P1 ($P1_{vir}$) was used to transfer selectable genetic markers that were inserted at a number of locations within the *E. coli* genome. Generalised transduction was carried out using standard protocols (Silhavy, 1984).

2.6.2.1 Preparation of bacteriophage P1 lysate

5 ml of LB (supplemented with a suitable antibiotic) was inoculated with the donor strain (from -70 °C stock) and incubated overnight (> 14 hours) at 37 °C in a shaking water bath (180 rpm). The following day 200 µl of the donor strain was sub-cultured into 10 ml of fresh pre-warmed LB supplemented with 0.2 % glucose and 5 mM $CaCl_2$, and incubated at 37 °C in a shaking water bath (180 rpm). After 30 minutes, 200 µl of $P1_{vir}$ lysate was added to the culture and incubated until it lysed the donor strain cells; lysis by $P1_{vir}$ is far more efficient than λ, with the culture medium appearing as clear as fresh medium after approximately 1.5-2.5 hours. The lysed culture was transferred to a 50 ml Falcon tube, supplemented with 200 µl of chloroform, mixed well for 30 seconds

(vortexing) and then centrifuged for 10 minutes at 4, 500 x g at 4 °C. 5 ml of the supernatant was transferred to a sterile glass test-tube, taking care not to disturb the pellet. 100 µl of chloroform was added to the test-tube and vortexed well. The test-tube was then sealed with a test-tube cap and Parafilm® barrier film, and stored at 4 °C for a minimum of 24 hours prior to use.

2.6.2.2 Transfer of selectable genetic markers by generalised transduction

5 ml of LB was inoculated with the recipient strain (from -70 °C stock) and incubated overnight (> 14 hours) at 37 °C in a shaking water bath (180 rpm). The following day 200 µl of the recipient strain culture was added to an equal volume of a solution containing 20 mM magnesium sulphate and 10 mM calcium chloride, and then separated into three, 100 µl aliquots. 0, 10 and 50 µl of the desired $P1_{vir}$ lysate was added to each of these aliquots, whilst a fourth sample was made containing lysate alone (no recipient cells). Each aliquot was mixed gently but well (by pipetting), then incubated at 30 °C in a static water bath for 30 minutes. 1 ml of LB supplemented with 0.1 M sodium citrate was added to each sample, mixed (by inverting sample) and then incubated at 37 °C in a static water bath for a minimum of 2 hours. Each sample was plated onto L-agar supplemented with an appropriate antibiotic (100 µl per plate), and incubated at 37 °C overnight. The following day, transductants were re-streaked onto fresh L-agar plates (also containing an appropriate antibiotic) and incubated at 37 °C overnight; transductants were preferentially taken from plates that contained less (10 µl) bacteriophage. The following day transductants were used to inoculate 5 ml of LB (containing an appropriate antibiotic) and incubated overnight at 37 °C in a shaking water bath (180 rpm). A sample of each overnight culture was used to make −70 °C glycerol stocks.

2.6.3 Allelic exchange

Non-selectable mutations were cloned into plasmid vectors and then transferred to the *E. coli* chromosome by homologous recombination. Allelic exchange of plasmid DNA for chromosomal DNA was carried out according to the protocol described by Blomfield *et al.* (1991). This protocol consists of two separate steps (figure 2.3). The first involves the construction of an intermediate strain in which the region of interest is replaced by a *sacB-kan* cassette; for this purpose, the kanamycin resistance determinant (*kan*) is used to select for uptake of this cassette into the chromosome. The second step

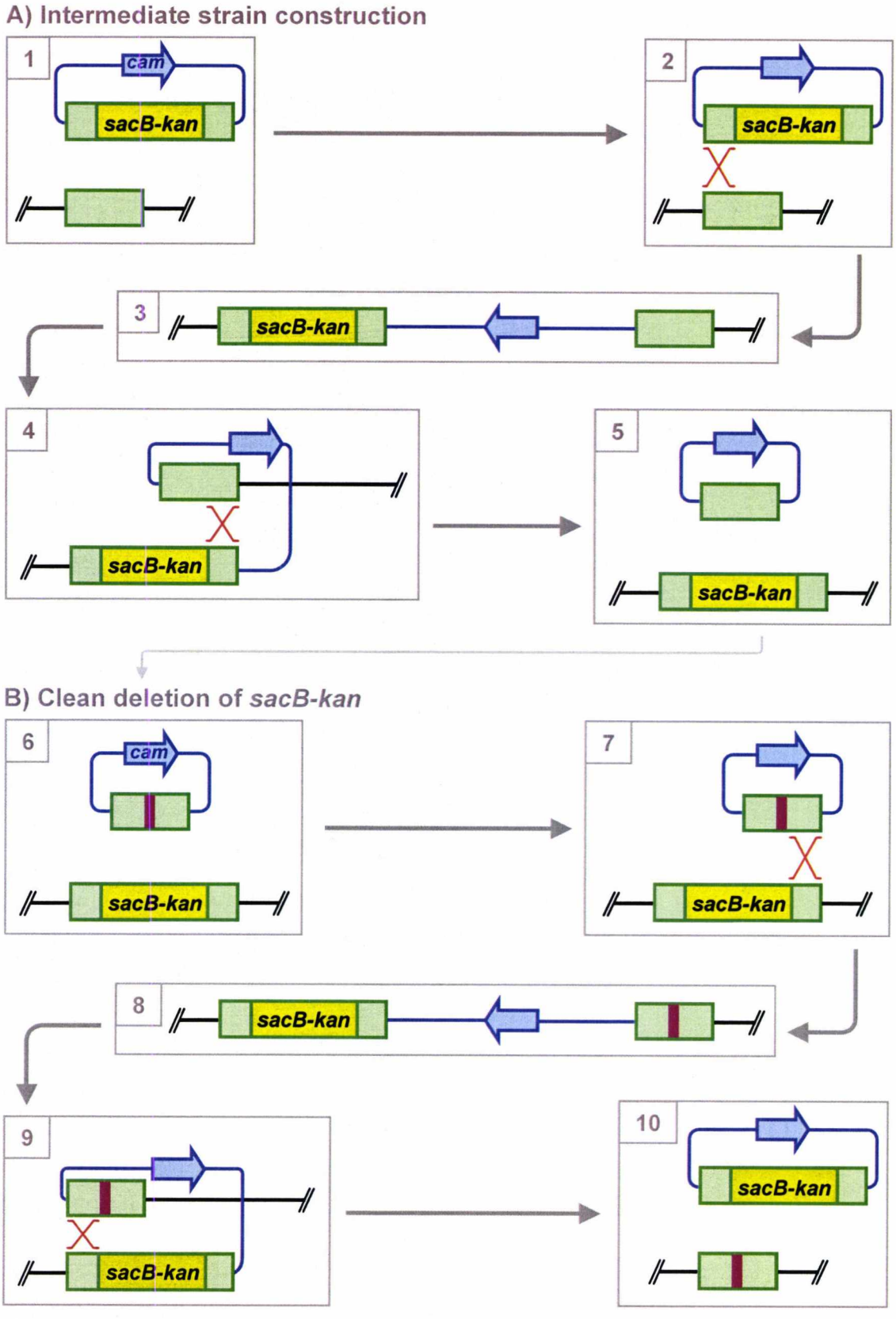


Figure 2.3: Schematic diagram defining the allelic exchange protocol used in this study. The intermediate strain was constructed by being transformed with a compatible *sacB-kan* plasmid (1), that was integrated into the recipient genome at the region of interest (2 & 3) and then cured, exchanging the region of interest for the *sacB-kan* cassette (4 & 5). The desired mutation was made in a similar manner, by exchange of this *sacB-kan* cassette for the plasmid-borne mutation (6 – 10).

involves the replacement of this cassette for the desired mutation (deletion or replacement); and to facilitate this step, the counter-selectable *sacB* marker is used.

The *sacB* gene encodes the levan sucrase enzyme from *Bacillus subtilis*, which polymerises sucrose into long polysaccharide chains that are lethal in gram negative bacteria. Both steps require a temperature sensitive plasmid vector, and derivatives of pMAK705 were used in this study.

2.6.3.1 Integration of the *sacB-kan* cassette

The recipient strain was transformed with a plasmid bearing the *sacB-kan* cassette using protocols described previously in this chapter (2.4.1 & 2.4.2), with just one exception: transformed cells were incubated overnight on pre-warmed L-agar plates (supplemented with 25 μg/ ml chloramphenicol) at 42 °C instead of 37 °C. pMAK705 has the pSC101 temperature sensitive replicon, for which 28 °C is the permissive temperature and 42 °C in non-permissive; incubation overnight at the non-permissive temperature (on solid media) is selective for integration of the plasmid into the genome. After this overnight incubation, up to four colonies were used to inoculate 5 ml of pre-warmed LB and incubated at 42 °C in a shaking water bath (180 rpm) for 4-5 hours. A Ten-fold dilution series (10^{-1} to 10^{-6}) was made of every culture, and 100 μl of each dilution was plated onto L-agar. Plates were then incubated at 28 °C overnight (without selection), in order to facilitate plasmid curing. The following day, single colonies were picked and re-streaked onto L-agar plates supplemented with either kanamycin A (10 μg/ ml), chloramphenicol (25 μg/ ml) or no antibiotic, and incubated at 37 °C overnight; colonies were plated according to a grid, so that the same colony could be identified on each plate. The following day, several Kan^+ Cam^- colonies were re-streaked onto fresh L-agar supplemented with kanamycin A (10 μg/ ml) to isolate single colonies, and incubated at 37 °C overnight. Single colonies were then used to inoculate 5 ml of LB supplemented with 10 μg/ ml kanamycin A, and incubated at 37 °C in a shaking water bath overnight. A sample of each overnight culture was used to make -70 °C glycerol stocks. $P1_{vir}$ lysates were made of these intermediate strains, and used to transduce the *sacB-kan* cassette into the recipient strain (according to the protocol described previously in this chapter). Strains constructed by generalised transduction, and not just allelic exchange were used as intermediates in the production of 'clean' (unmarked) mutations.

2.6.3.2 Exchange of the *sacB-kan* cassette for the desired mutation

The intermediate strain was transformed with the plasmid bearing the 'clean', non-selectable mutation, as described in section 2.6.3.1 (incubated overnight at 42 °C). Up to four colonies were used to inoculate 5 ml of pre-warmed LB and incubated at 42 °C in a shaking water bath (180 rpm) for 3 hours. Each culture was diluted ten-fold and a thousand-fold, and whilst 100 µl of both dilutions was plated onto sucrose agar plates, only the latter dilution was plated onto L-agar (several plates per dilution). Plates were then incubated at 28 °C overnight, in order to facilitate plasmid curing and, in the case of sucrose agar facilitate the exchange of the *sacB-kan* cassette for the clean deletion. If both allelic exchange and plasmid curing have been successful, the 10^{-1} dilution sucrose plates should yield a similar number of colonies to the 10^{-3} L-agar plates. The following day, single colonies were picked and re-streaked onto L-agar plates supplemented with either chloramphenicol (25 µg/ ml), kanamycin A (10 µg/ ml) or no antibiotic, and incubated at 37 °C overnight; colonies were again plated according to a grid, so that the same colony could be identified on each plate. The following day, several Cam^{-} Kan^{-} colonies were re-streaked onto fresh L-agar to isolate single colonies, and incubated at 37 °C overnight. Single colonies were then used to inoculate 5 ml of LB, and incubated at 37 °C in a shaking water bath overnight. A sample of each overnight culture was used to make -70 °C glycerol stocks.

2.6.4 Construction of deletion and replacement mutations

Non-selectable sequence alterations were constructed and cloned into a plasmid vector (pMAK705 derivative) using REase digestion (and modification) and PCR mutagenesis. The PCR mutagenesis protocols used in this study are defined in the following sub-sections.

2.6.4.1 PCR-generated deletion mutations

Deletions were made either by incorporating REase sites within 3' tails of primers, and cloning digested PCR fragments into a plasmid digested with the same enzymes. In order to do so, a number of REase sites present within *nanC*, and the *nanC-fimB* intergenic region were utilised (figure 2.4).

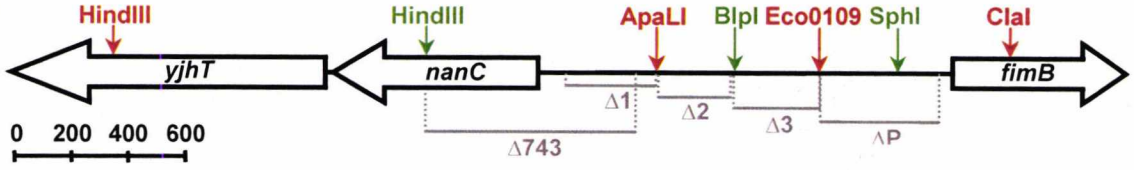


Figure 2.4: Schematic diagram defining restriction endonuclease sites, within the *yjhT-fimB* region of the *E. coli* K-12 genome, that were used in this study.

Standard protocols were used to make PCR deletion mutations. The mutagenesis PCR was carried out using pIB413 as a template and specific oligonucleotide mutagenesis primers (refer to chapter 2.5.6 for this protocol). A small percentage of this PCR (~ 5-10 %) was checked on a suitable agarose gel; if no erroneous products were identified, the remainder was purified (using the Promega Wizard SV Gel & PCR Clean-up System) and then subjected to the desired REase reaction(s) (chapter 2.5). This digestion reaction was then purified (using the same commercial kit as before). This digested PCR fragment, containing the desired deletion could then be cloned into a suitable plasmid vector fragment (that has been digested with the same restriction enzymes).

2.6.4.2 PCR-generated replacement mutations

Replacement mutations were made using two different methods in this study. Replacement mutations between Rm10 and NagC2 (Rm11-20) were split between two non-annealing primer tails, and incorporated within two primers that anneal upstream (reverse strand) and downstream (forward strand) of the region of interest (figure 2.5). They were then used to carry out two separate PCR reactions, which could be subsequently digested and ligated to form a single product, containing the complete replacement mutation. In order to do so, each replacement mutation contained a unique SacII REase site (5' CCGCGG). The ligated PCR products were then used as a template for a third PCR, using the two outside primers. This third PCR was then be cloned into the *nanC-fimB* sequence, within a plasmid, by REase digestion and ligation.

Standard protocols were used to construct these replacement mutations. The two mutagenesis PCRs (1 and 2 in figure 2.5) were carried out as described previously, using pIB413 as a template, one standard oligonucleotide primer and one mutagenesis primer (refer to chapter 2.5.6 for this protocol). A small percentage of each PCR (~ 5-10 %) was checked on a suitable agarose gel; if no erroneous products were identified,

the remainder was gel-purified and then digested with the REase SacII (chapters 2.5.7 and 2.5.3 respectively). The digestion reactions were purified using the Promega Wizard SV Gel & PCR system, and then ligated together overnight (chapter 2.5.5). This ligation reaction was amplified by PCR using the same two standard primers that were used in the initial two mutagenesis PCRs and 2 µl of the ligation reaction (as template). This PCR was then purified, digested with suitable restriction enzymes and then purified once again. This digested PCR fragment, containing the desired replacement mutation could then be cloned into a suitable plasmid vector fragment (that has been digested with the same restriction enzymes).

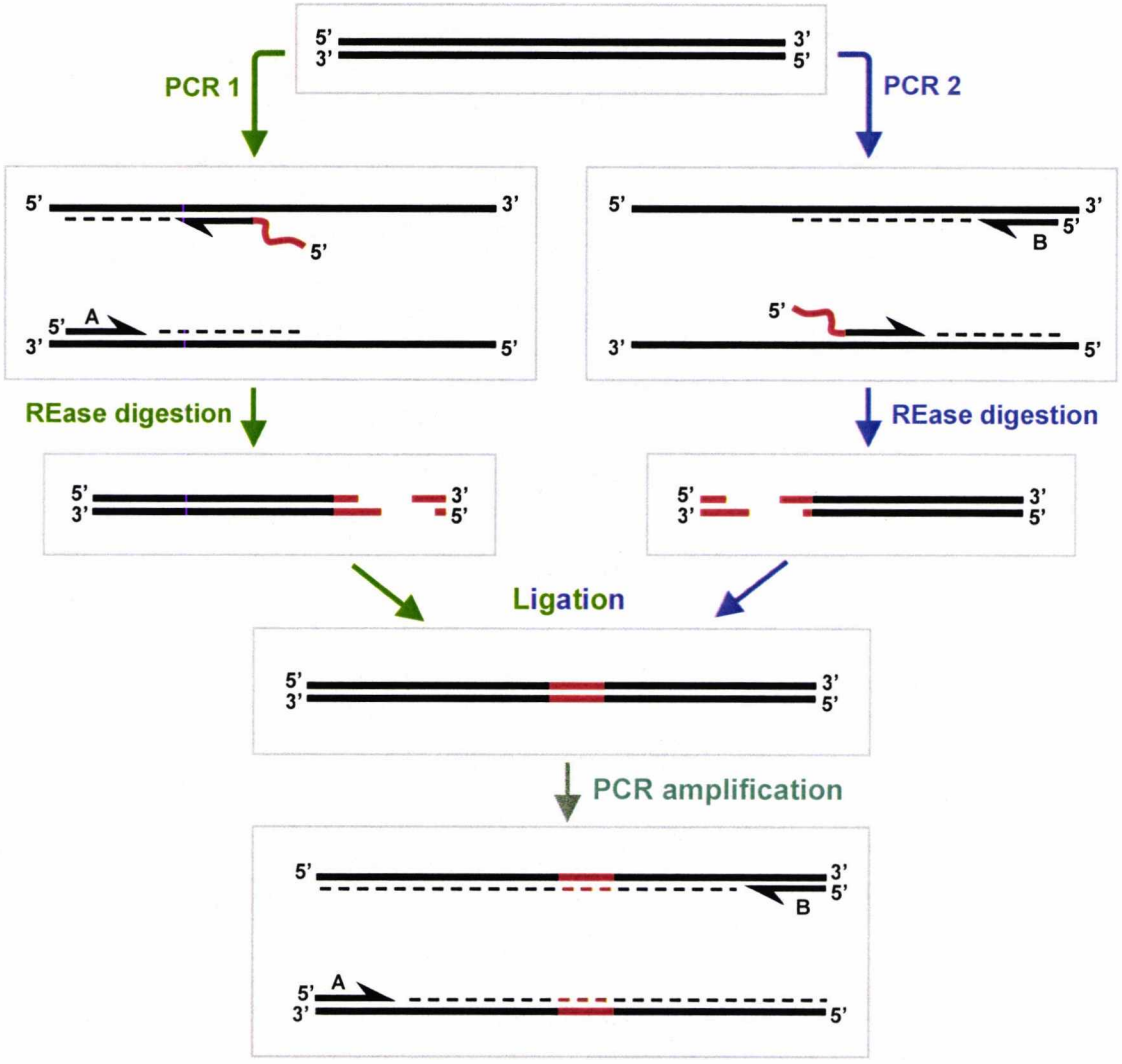


Figure 2.5: Schematic diagram defining the PCR mutagenesis protocol used to construct replacement mutations between Rm10 and NagC1. Two PCRs were initially carried out (PCR 1 & 2) using the same template (pIB413), a 'standard' primer (A, 'ApaIIUPF' or B, 'FimBUPR') and a mutagenesis primer ('RM4xxREV' or 'RM4xxFWD' respectively). The PCR products were digested with the REase SacII, ligated together and then amplified by PCR. This final PCR product, containing the desired replacement mutation was then cloned into a suitable vector.

Heterologous replacement of the Δ3 region was achieved by using primers that anneal to the β-lactamase gene, *bla*, within pRS415. The resulting PCR product was cloned into a plasmid vector by incorporating REase sites within each primer (BlpI and Eco0109). The primer that annealed to the forward strand ('D3HetRepF') contained a 38 bp non-annealing tail in order to incorporate NagC2 site and BlpI site that are present in the Δ3 deletion background (figure 2.6).

A) D3HetRepF nucleotide sequence (5' → 3'):

CATCAGGCTGAGCATAATTCTCATCATGAAATATGTTTGAAGAACGTTTTCCAATGATGAGCAC

B) *bla* nucleotide sequence (5' → 3'):

```
1-49      GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC
          ('D3HetRepF')          ('D33HRRev')

50-99     GGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATAC

100-149   ACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCAT
                ('D32HRRev')

150-199   CTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCAT
                    ('D31HRRev')

200-249   GAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGA

250-274   AGGAGCTAACCGCTTTTTTGCACA
          ('D30HRRev')
```

Figure 2.6: The nucleotide sequence of the primers used to construct size-matched fragments of *bla* sequence within the Δ3 region. The nucleotide sequence of primer 'D3HetRepF' is given (A), defining the region that anneals to the *bla* target sequence (in green), and the non-annealing tail. This tail contains contiguous sequence extending 6 bp beyond the BlpI site (underlined), upstream of the Δ3 region, to 3 bp beyond the NagC2 site, within this region (MG1655 genomic sequence, bp 4537750-4537788). The fragment of the *bla* nucleotide sequence used as a template for PCR is also given (B), in which primer sites are defined. Note that only the forward strand is given above, such the reverse compliment is highlighted for each primer that anneals to the reverse strand ('D30HRRev' – 'D33HRRev').

The Δ3 heterologous replacement mutations were constructed using the same method as the deletion mutations defined in chapter 2.6.4.1; the only differences being the oligonucleotide primers and template used for PCR.

2.6.4.3 PCR-generated substitution mutations

Overlap extension mutagenesis (Ho *et al.*, 1989) was used to make small substitution mutations that disrupt the RNA-polymerase, σ_{70} -10 consensus sequence defined by Klemm *et al.* (1994) (figure 2.7). In this method, two DNA fragments are synthesised by PCR that have overlapping ends. These fragments can then be combined in a fusion reaction, in which the overlapping ends anneal. DNA polymerase can synthesise the complimentary sequence from the free 3' hydroxyl groups of this fusion product to form a complete, double-stranded DNA sequence that extends both upstream and downstream of this region of overlap. Standard primers can then be used to amplify this sequence. If the desired mutation is incorporated within this region of overlap, the resulting PCR product can be cloned into a region of interest (using natural REase sites) without producing any other sequence alterations.

The following method was used in this study to mutate the −10 promoter consensus of Klemm's *fimB* promoter (Olsen & Klemm, 1994): The two mutagenesis PCRs (1 and 2 in figure 2.7) were carried out using pIB413 as a template, one standard oligonucleotide primer ('ApalIUPF' or 'FimBUPR') and one mutagenesis primer ('P3 MutR' or 'P3MutF' respectively) (Standard PCR conditions were used, as defined in chapter 2.5.6). A small percentage of each PCR (~ 5-10 %) was checked on a suitable agarose gel; if no erroneous products were identified, the remainder was gel-purified (chapter 2.5.7). A third PCR (3 in figure 2.7) was carried out using the products of the first two PCRs as template, and the two standard primers used previously ('ApalIUPF' and 'FimBUPR'). This PCR was carried out using the following cycles: 95 °C for 30 seconds (1 cycle); 95 °C for 30 seconds, 48 °C for 1 minute, 72 °C for 1 minute (10 cycles); 95 °C for 30 seconds, 55 °C for 45 seconds, 72 °C for 45 seconds (25 cycles); 72 °C for 7 minutes (1 cycle); 10 °C (until ready to process the reaction). A small sample of the reaction was checked on an agarose gel, and then purified using the Promega Wizard SV Gel & PCR system. This PCR product was then digested with BlpI and SphI overnight and then purified. This PCR fragment could then be cloned into pIB347 digested with the same restriction enzymes.

All other plasmids constructed in this study were made by sub-cloning DNA fragments between plasmids.

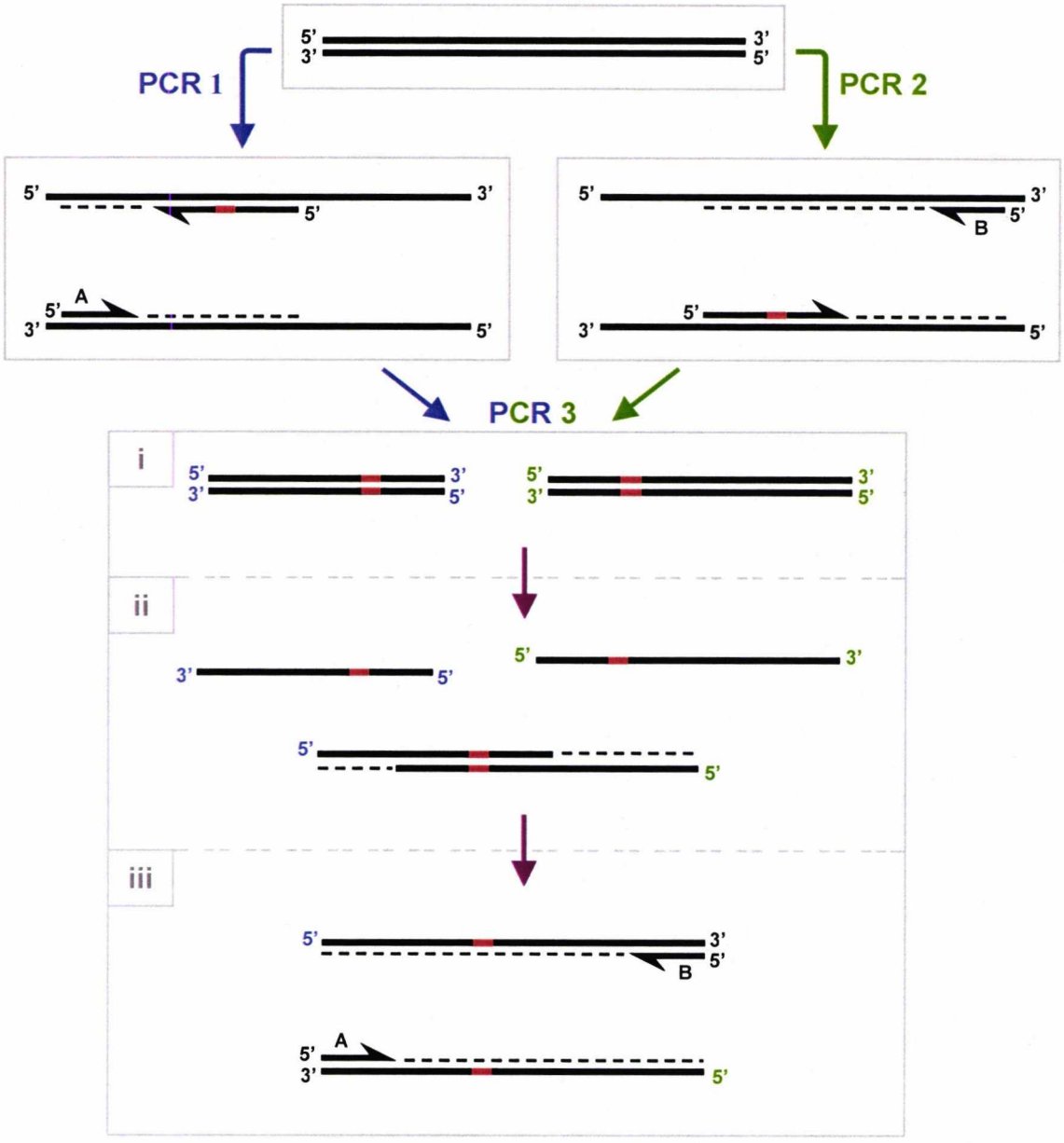


Figure 2.7: Schematic diagram defining the overlap extension mutagenesis protocol. Initially two PCRs were carried out (PCR1 & PCR2); each had the same template (pIB413), one standard oligonucleotide primer (A, 'ApaIIUPF' or B, 'FimBUPR') and one mutagenesis primer ('P3MutR' or 'P3MutF', respectively). These two PCR products were then combined in a third PCR (PCR 3). During the initial cycles of this PCR (i) the two products anneal to each other about their region of overlap, and DNA polymerase synthesises full-length DNA sequence from the free 3' hydroxyl groups (ii). These full-length sequences can then serve as templates for two standard primers (A & B) (iii).

2.7 β-galactosidase assays

Translational *lacZYA* (LacZ) reporter fusions were constructed to report the expression of *fimB* in this study. The β-galactosidase assay was used to quantify this expression and was carried out essentially as described by Miller *et al.* (1972).

Bacterial strains were grown in 5 ml of RD_{glyc} MOPS medium (both with and without any supplements that were specific to the assay) overnight at 37 °C in a shaking water bath (180 rpm) prior to assay; media was inoculated from -70 °C glycerol stocks. The following day, each strain was sub-cultured into a minimum of 15 ml of fresh RD_{glyc} MOPS medium (± supplements) and incubated as before (37 °C, 180 rpm) until they reached an optical density of 0.2 at a wavelength of 600 nm (OD_{600}). The inoculum used to sub-culture each strain was determined spectrophotometrically such that they would all grow for the same number of generations to reach the desired OD. It has previously been suggested that strains should be grown for a minimum of 3.3 generations during exponential phase prior to assaying for β-galactosidase activity (Neidhardt, 1990), and in this study they were grown for a minimum of 7 generations.

Once they reached the desired optical density, strains were incubated in ice-water for a minimum of 20 minutes. The OD_{600} was recorded after this incubation, and then 250 µl of cells was added to a test-tube containing 250 µl of Z-buffer (60 mM dibasic sodium phosphate heptahydrate; 40 mM monobasic sodium phosphate; 10 mM potassium chloride (BDH Laboratory Supplies, UK); 1 M magnesium sulphateheptahydrate; and 50 mM 2-mercaptoethanol; adjusted to pH 8.0 and filter-sterilised). 5 µl of 0.1 % SDS and 10 µl of chloroform were added, and the test tube was then vortexed for 10 seconds. The solution was pre-incubated for 5 minutes at 28 °C, and then the β-galactosidase assay was started by adding 100 µl of 4 mg/ ml 2-nitrophenyl-β-D-galactopyranoside (ONPG). The assay was stopped by adding 250 µl of 1 M sodium carbonate, after which the OD of the reaction was recorded at 420 nm and 550 nm. β-galactosidase activity was calculated in Miller units according to the following equation (Miller, 1972):

[e2.2] $$[1000\ (OD_{420} - 1.75 \times OD_{550})]/\ [t \times OD_{600}\ (V/\ 2)]$$

('t' is the time or duration of the assay; V is the volume of cells present in the reaction)

The above equation (e2.2) differs from that proposed by Miller *et al.* (1972) in that the volume is halved. This is because the quantities of every reagent used in this protocol are half of those proposed by in the original protocol.

2.7.3 Statistical analysis (I): 95 % confidence limits

β-galactosidase data given in this study are mean values calculated from a minimum of 4 assays. 95 % confidence limits are also shown for all β-galactosidase assays presented in this study, in order to define the error associated with each of these mean values. 95 % confidence limits were calculated in this study using Minitab™ (v13.1) statistical software, but are given by the following equation:

[e2.3] $$\mu \pm t_{(\alpha/2,\, n-1)}.\sigma/\sqrt{n}$$

('μ' is the arithmetic mean; 'n' is the sample size; 'σ' is the sample standard deviation; 'α' is the desired significance level; and '$t_{(\alpha/2,\, n-1)}$' is the upper critical value of the t-distribution with '*n*-1' degrees of freedom).

All β-galactosidase assays presented in this study have 95 % confidence limits that lie within 10 % of the mean.

2.7.4 Statistical analysis (II): 95 % confidence limits for a ratio

The relative effect of one β-galactosidase assay value to another is also calculated in this study; that is, the amount of expression lost as a consequence of combining two mutations. Relative effect was calculated according to the following equation:

[e2.4] $$100 - (m1/\, m2) \times 100$$

(m1 is the first mean β-galactosidase assay value; and m2 is the second)

The 95 % confidence limits were calculated for this percentage ratio according to the following equation (Kendall, 1977):

[e2.5] $$\text{Relative effect (\%)} \pm 100 - (m1/\, m2) \times 100 + (1.96 \times 100 \times SE)$$

Where SE is the standard error of the (m1/ m2) ratio, and is given by the following equation:

[e2.6] $\sqrt{[s^2/(4 \times m2^2) + m1 \wedge 2s^2/(4 \times m2^4)]}$

(Here $a \wedge n$ indicates a raised to the power of n)

Where s^2 is the pooled error variance, which is given by the following equation:

[e2.7] $s^2 = (95\ \%\ \text{c.l. of m1})^2 + (95\ \%\ \text{c.l. of m2})^2/2$

(95 % c.l. is the 95 % confidence limit for each mean value)

Chapter 3

Searching for the distant, *trans*-active repressor of *fimB*

3.1 Summary

The *fimB* recombinase gene is separated from the divergently transcribed *nanC* gene by a large, 1.4 kb intergenic region, and sequences far (> 700 bp) upstream are known to regulate expression (El-Labany *et al.*, 2003). Furthermore, preliminary data suggested that these distant activating elements bind regulatory proteins that might stimulate expression of *fimB* by antagonising inhibitory elements that lay even further upstream (El-Labany *et al.*, 2003). The data presented in this chapter attempts to identify these distant, *trans*-active inhibitory elements for *fimB*. Whilst several elements were identified in this study, which encode regulatory proteins or whose functions remain unknown, none had the phenotype of a distant repressor of *fimB* that could be antagonised by *fimB* activators.

3.2 Introduction

Two regulatory proteins, NanR and NagC control expression of the *fimB* (Sohanpal *et al.*, 2004). They occupy *cis*-active sites, greater than 700 bp upstream of the *fimB* orf and stimulate expression (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004). When *fimB* is moved to an ectopic location at *lac*, a deletion that removes both NanR and NagC binding sites (Δ2) only affects FimB recombination when sequences extending into *nanC* accompany it at *lac* (El-Labany *et al.*, 2003). This defined a 743 bp region, without which NanR or NagC have no effects upon *fimB* expression. This also led to the hypothesis that NanR and NagC serve to antagonise a repressor of *fimB*, and that this repressor has a binding site located within the 743 bp region.

Transposon mutagenesis was used in this chapter in order to identify genome-wide insertion mutations that affected the activity of the distant repressor of *fimB*. This chapter characterises a number of transposon insertions that were capable of increasing expression of *fimB* in the absence of NanR- or NagC-activation, or both. The location of these transposon insertions was determined using genetic linkage experiements and inverse PCR, and their responsiveness to loss of activating elements (NanR and nagC) was determined. Whilst a number of transposon insertions are characterised in this study, *fimB* expression remained responsive to NanR and NagC in each instance.

3.3 Transposon mutagenesis in the Δ2 background identifies nine independent insertions that affect *fimB* expression

In order to identify the distant repressor of *fimB*, transposon mutagenesis was employed using the mTn*10* transposon, and BGEC711 (*fimB'-'lacZYA Δ2*) as the recipient strain. This strain was used because it has the Δ2 mutation, which removes both NanR and NagC binding sites; and if these regulators serve to antagonise a distant repressor, then pseudorevertants disrupting expression or activity of the repressor should suppress the reduction in *fimB* expression that results from the Δ2 mutation. A translational reporter fusion was used in this instance in order to identify mutations that increase transcription or translation of *fimB*, as the precise mechanism of distant *fimB* regulation remained unknown. After screening in excess of 100,000 colonies, only 9 independently isolated insertions were identified that increased *fimB* expression (strains KCEC619 – KCEC627). Phenotypically, each of these strains was indistinguishable from one another on MacConkey lactose agar (figure 3.1). In the presence of 10 μg/ ml tetracycline, expression of the *fimB-lacZ* reporter fusion was elevated, relative to the wild-type; whilst they were comparable to the Δ2 control in the absence of tetracycline. As tetracycline resistance is achieved in each of these isolates by induction of the divergently transcribed genes, *tetA* and *tetR*, within mTn*10*, this showed that induction of these genes is required for the effect upon *fimB*, and not insertion of mTn*10* alone.

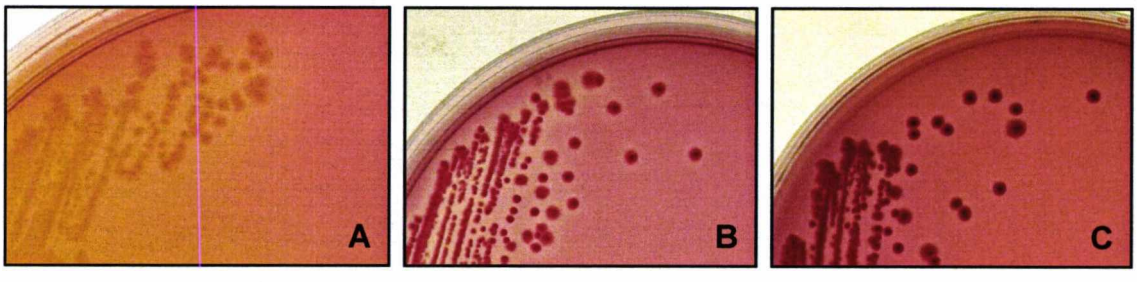


Figure 3.1 Phenotypic analysis of transposon insertion mutants. Cells were streak-plated onto MacConkey lactose agar, with (image C) or without (A & B) 10 μg/ ml tetracycline, and then incubated at 37 °C for 18 hours. The images above are of the following strains: A, BGEC711 (Δ2); B, BGEC905 (Wt); and C, KCEC619 (> Wt).

3.4 Genetic linkage of transposon insertions to the *fim* locus (i)

Whilst the near-randomly inserted transposons might have disrupted factors that regulate *fimB* expression, the possibility still remained that they might have inserted within the *nanC-fimB* intergenic region, and somehow circumvented distant repression

of *fimB*. In order to discriminate between local insertions, within the *nanC-fimB* intergenic region, or elsewhere within the genome and presumably a *trans*-active factor, genetic linkage experiments were carried out. The kanamycin resistance marker from BGEC696 (*fimB'-'lacZYA Δ2ΩsacB-kan*) was transduced (by P1 transduction) into each of the transposon insertion strains. In this way, the kanamycin resistance marker would displace the tetracycline resistance marker of the mTn*10* transposon, if the latter was located near-to or within the *fim* locus. Genetic linkage was defined as the percentage of transductants that had become tetracycline sensitive upon uptake of the kanamycin resistance determinant.

Using 30-transductants per transposon insertion, all nine insertions were found to be tetracycline sensitive, and therefore 100 % linked to the *fim* locus (data not shown). Subsequent PCR analysis confirmed that they were all located within the *nanC-fimB* intergenic region (data not shown). Coupled with the fact that these strains increased expression of *fimB-lacZ* in a tetracycline dependant fashion, this evidence suggested that they served as promoter replacements for *fimB*. It was for this reason that these strains were discarded in the search for *trans*-active repressors of *fimB*.

3.5 Transposon mutagenesis, in the region 1 mutant background

It was unclear why transposon insertions in the Δ2 background and unlinked to the *fim* locus could be not be identified. However, it was possible that NanR and NagC, being two very different regulatory proteins might overcome different repressors that bind far-upstream of *fimB*. It was for this reason that transposon mutagenesis was next carried out in the Rm1 mutant strain (KCEC118; *fimB'-'lacZYA Rm1*), as this heterologous replacement mutation removes the NanR binding site at region 1 (El-Labany *et al.*, 2003). At least 80,000 colonies were screened in this background, and this identified 20 independently isolated insertions that increased *fimB* expression in the presence of the Rm1 mutation (summarised in figure 3.2). The phenotypes of each strain, on MacConkey lactose-agar are given in table 3.1.

3.6 Genetic linkage of transposon insertions to the *fim* locus (ii)

Genetic linkage experiments were carried out on the new transposon insertion strains as previously described in this chapter. Of the 20 isolates in the Rm1 background, 6 were 100 % linked to the *fim* locus (table 3.2). Like the transposon insertion isolates that

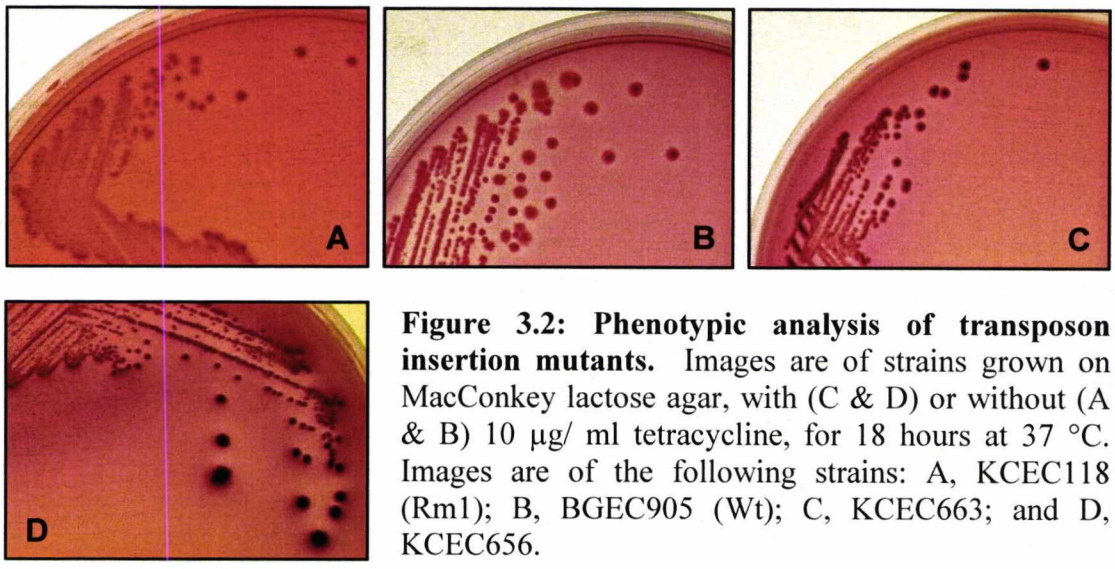


Figure 3.2: Phenotypic analysis of transposon insertion mutants. Images are of strains grown on MacConkey lactose agar, with (C & D) or without (A & B) 10 µg/ ml tetracycline, for 18 hours at 37 °C. Images are of the following strains: A, KCEC118 (Rm1); B, BGEC905 (Wt); C, KCEC663; and D, KCEC656.

Colony phenotypes:		Strains
10 µg/ml tetracycline	No tetracycline	
≈ Wt	≈ Rm1	KCEC657, KCEC658, KCEC662, KCEC664, KCEC666, KCEC667, KCEC668, KCEC669, KCEC670, KCEC673, KCEC675.
≈ Wt	≈ Wt	KCEC655, KCEC660, KCEC671.
> Wt	> Wt	KCEC663, KCEC672, KCEC674.
>> Wt	>> Wt	KCEC656, KCEC659, KCEC676.

Table 3.1: Phenotypic analysis of transposon insertion strains. Transposon insertion isolates, categorised according to *fimB-lacZ* expression. Isolates were characterised on MacConkey lactose agar, in both the presence and absence of 10 µg/ ml tetracycline, after incubation at 37 °C for 18 hours.

linkage to *fim* (%)	Strains
100	KCEC657, KCEC662, KCEC664, KCEC668, KCEC669, KCEC670.
97	KCEC674.
37	KCEC675.
23	KCEC671.
0	KCEC655, KCEC656, KCEC658, KCEC659, KCEC660, KCEC663, KCEC666, KCEC667, KCEC672, KCEC673, KCEC676.

Table 3.2: Genetic linkage of transposon insertions to the *fim* locus (ii). The kanamycin resistance determinant (*kan*) from BGEC696 was transduced into each of the strains given above, using the lytic phage $P1_{vir}$. Transductants were then screened for tetracycline sensitivity on L-agar, after incubation at 37 °C for 18 hours. Genetic linkage is given as the percentage of transductants that became sensitive to tetracycline upon acquisition of the KanR determinant. Strains are colour-coded here for direct comparison with their phenotypes on MacConkey lactose agar (table 3.4).

were identified in the Δ2 background, these 6 isolates were also tetracycline-dependent; and discarded in the search for the distant repressor of *fimB*. One insertion was 97 % linked to the *fim* locus; although the relatively low sample number (30) in these linkage experiments meant that this isolate might be no further away from *fimB* than those that were 100 % linked. Two isolates had transposon insertions that were 37 % and 23 % linked to the *fim* locus, which suggested that they were relatively distant (>26 kb and >36 kb respectively). The remaining 11 insertions were identified as unlinked (>93 kb) to the *fim* locus. The fact that isolates in this last group had a number of different phenotypes suggested that they might occupy a number of different sites within the *E. coli* genome.

3.7 Quantification of *fim*-unlinked, transposon insertion phenotypes

β-galactosidase assays were used to quantify the effects of all transposon insertions that were not 100 % linked to the *fim* locus (figure 3.3). Eight of the sixteen strains tested had either no effect (KCEC663), or were in fact inhibited in rich defined media (KCEC strains 655, 658, 660, 667, 771, 772 & 773). These transposon insertions suggested that there are mutations that only effect *fimB* expression on MacConkey lactose agar, and not in defined-rich MOPS medium. It is for this reason that they were discarded, as they could not be affecting the distant repressor of *fimB*, which does have a phenotype

in rich media. The remaining nine transposon insertions appeared to fall into one of three categories: those that result in modest (~2 x Wt) increases in *fimB* expression (KCEC674 & 675); those that significantly (~4 x Wt) increased expression of *fimB* (KCEC666); and those that stimulate *fimB* expression to a level far greater (~6 x) than the wild-type (KCEC656, 659 & 676).

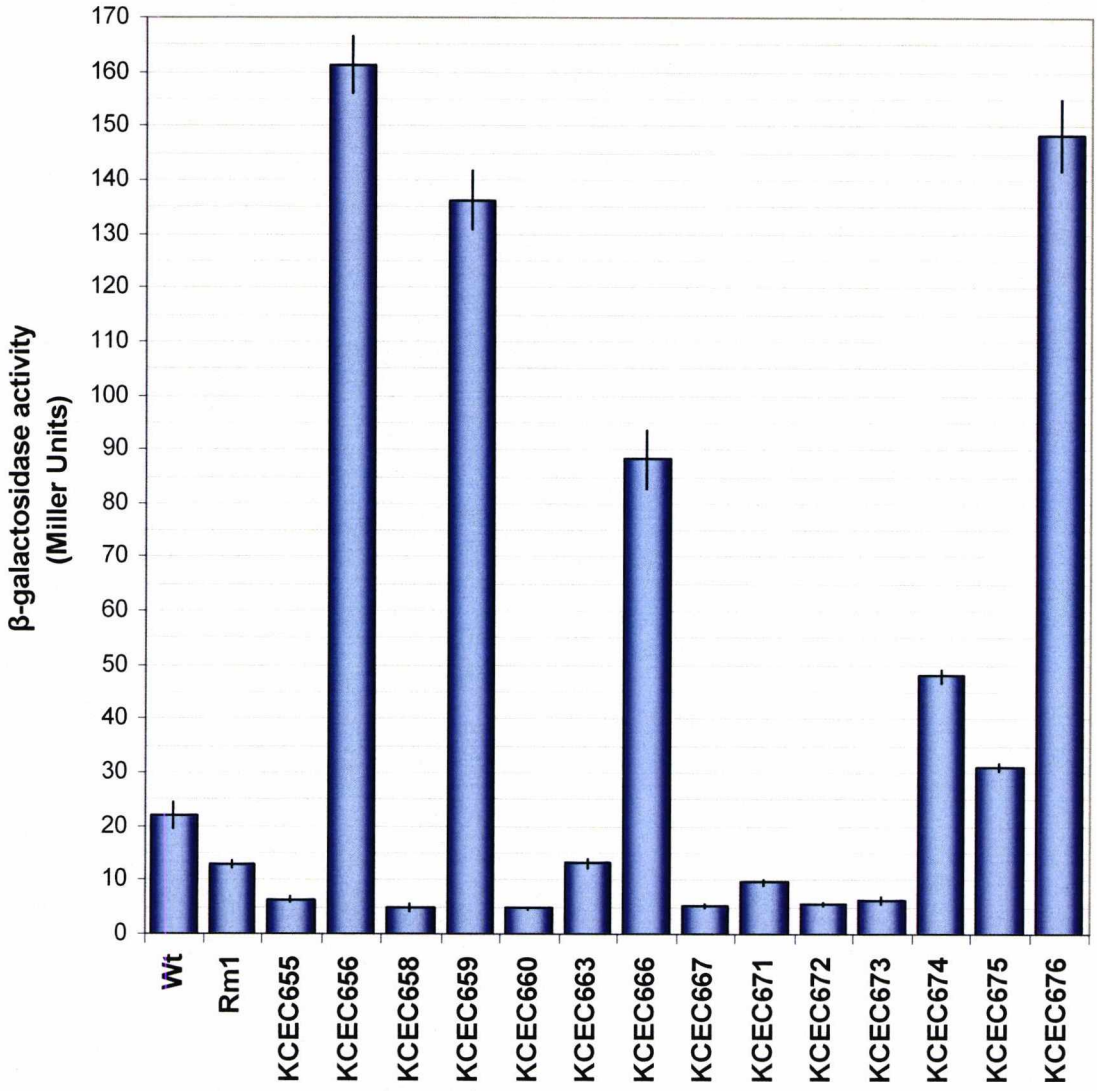


Figure 3.3: Quantification of *fim*-unlinked, transposon insertion phenotypes. β-galactosidase activity of *fim*-unlinked, or *fim*-distal transposon insertion strains, in the Rm1 background. Cells were grown in rich defined MOPS media, with glycerol as the sole carbon source, plus 10 μg/ ml tetracycline (with exception of the two control strains that were grown in the absence of tetracycline). Control strains are BGEC905 (Wt) and KCEC118 (Rm1).

3.8 Identification of transposon insertion sites

Inverse-PCR was used to determine the location of each transposon insertion that had a phenotype in rich defined MOPS media (table 3.3). Although eight transposon insertion strains were investigated, only two sites within the *E. coli* genome were identified: within or near to *hns*, and within the *yjhV-fecABCDE* intergenic region (figures 3.4 and 3.5). Whilst HNS had previously been shown to inhibit *fimB* expression (Donato *et al.*, 1997), and all of transposon insertions at that locus would clearly disrupt expression of the gene that encodes it, the insertion within the *yjhV-fecABCDE* intergenic was not so easy to interpret. The genes closest to the transposon insertion site, one of unknown function (*yjhV*) and the other belonging to an operon that facilitates the uptake of iron complexes (*fecABCDE*) are both convergent upon the intergenic region in which the transposon is inserted; and it was unclear what function it would serve at this location. The transposon of strain KCEC674 was located within the *nanC-fimB* intergenic region, and this was consistent with the 97 % genetic linkage to that locus, that was presented previously in this chapter.

Strain	mTn*10*-flanking sequence (5' → 3'):	Genomic position (bp)	Genomic strand	Location of mTn*10*
KCEC656	CATTCTCTTGCCTGC...	1,292,084	+	Within *hns*
KCEC659	CGCTGAGCACGTTTA...	1,291,868	+	Within *hns*
KCEC666	TATTGGGGTGGTTTG...	1,292,165	+	Upstream of *hns*
KCEC674	GATTAAATACCTGAG...	4,537,260	+	Within *nanC-fimB* intergenic region
KCEC675	CGCCCAAGCCTGTAC	4,508,075	+	Within the *yjhV-fecABCDE* intergenic region
KCEC676	AGCGAAGCACTTAAA...	1,292,142	-	Within *hns*

Table 3.3: Identification of transposon insertion sites. Inverse-PCR was used to identify the sequence adjacent to each transposon. This sequence was then located within the *E. coli* K-12 MG1655 genome (Blattner *et al.*, 1997) using the Colibri database (http://genolist.pasteur.fr/Colibri/).

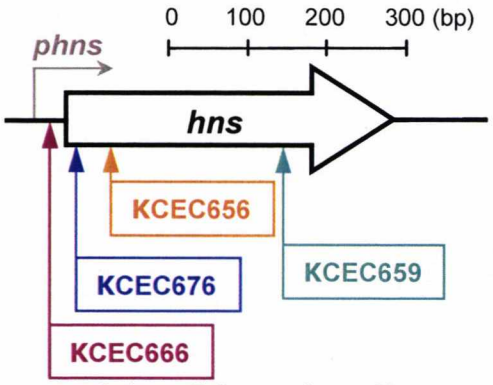


Figure 3.4: Schematic diagram of transposon insertion sites at the *hns* locus.

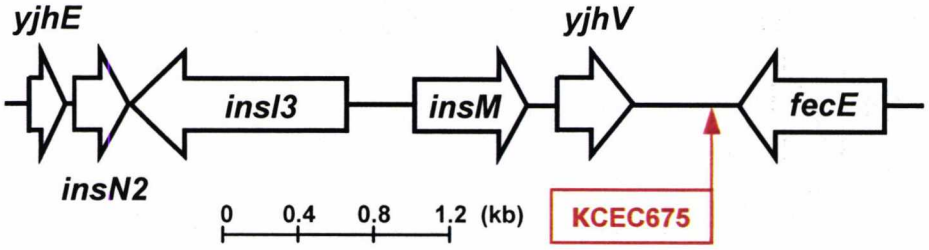


Figure 3.5: Schematic diagram of the transposon insertion site within the *yjhE-fecE* intergenic region. Since the *E. coli* K-12 (MG1655) genome was sequenced (Blattner *et al.*, 1997), the *yjhE-fecE* intergenic region has been defined in greater detail (Gerdes *et al.*, 2003; Serres *et al.*, 2001). Three genes have been assigned names and functions: *insN2*, which encodes transposase InsN; *insI3*, which encodes transposase InsI; and *insM*, which encodes transposase InsM. A fourth gene, that lies immediately upstream of mTn*10* in strain KCEC675, *yjhV*, is of unknown function.

3.9 Responsiveness of transposon insertions to the Rm1 and Rm5 mutations

Although the transposon insertions at the *hns* locus and within the *yjhE-fecE* intergenic region identify factors that inhibit *fimB* expression, it was unclear if either affected the distant repressor of *fimB*. In order to test this, transposon insertions from KCEC675 (*yjhE-fecE* mTn*10* insertion) and KCEC676 (insertion within *hns*, between codon 1 and 2) were moved into the wild-type and Rm5 backgrounds (the Rm5 replacement mutation removes the NagC1 binding site, otherwise known as region 2; El-Labany *et al.* 2003; Sohanpal *et al.*, 2004), and their relative effects upon the *fimB-lacZ* reporter fusion were compared (figure 3.6). Even though both transposon insertions significantly increased *fimB* expression, they still remained responsive to activators bound within the Δ2 region. These data therefore suggested that neither transposon insertion had any effect upon the distant repressor of *fimB*.

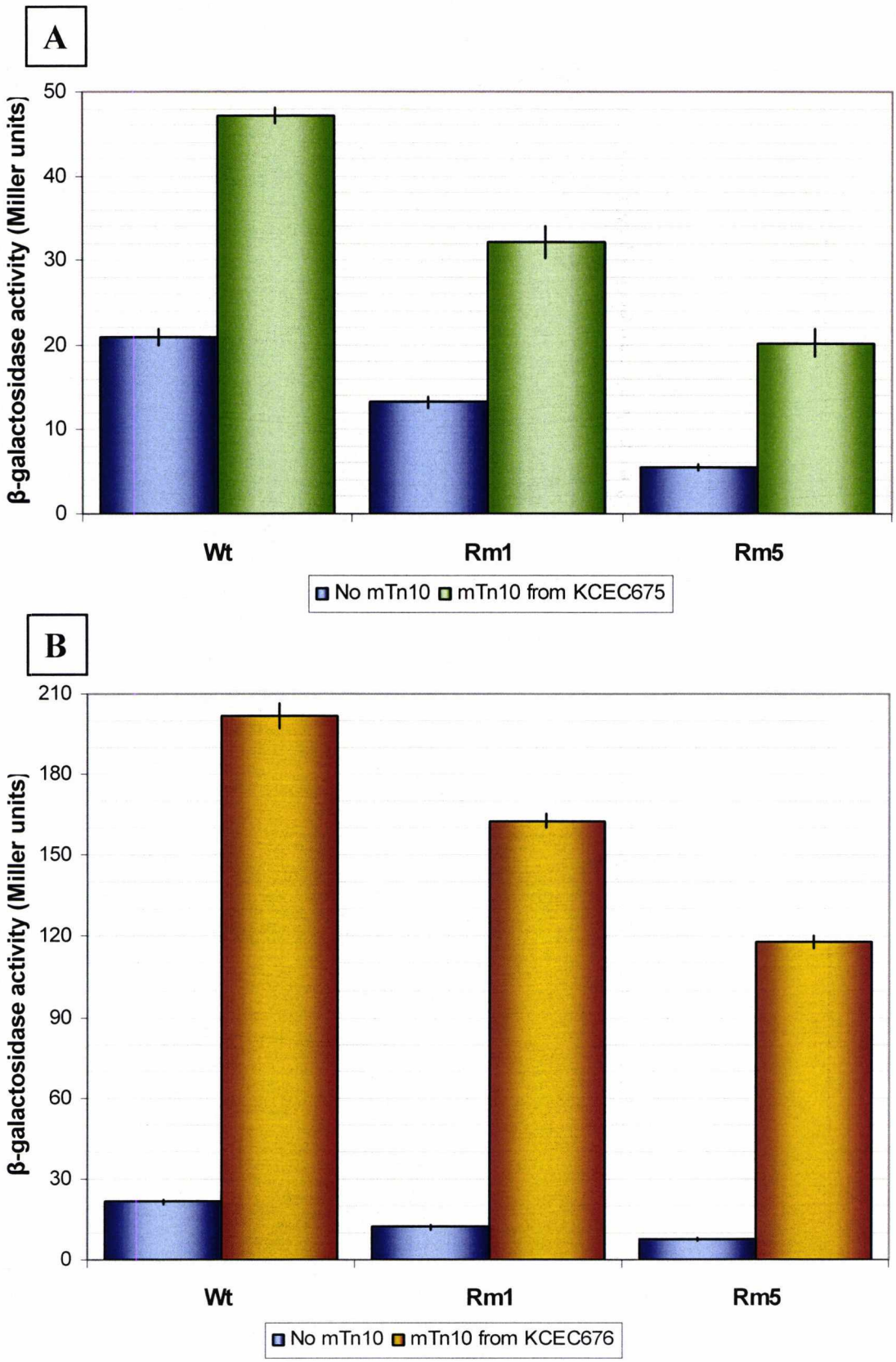


Figure 3.6: Responsiveness of transposon insertions to the Rm1 and Rm5 mutations. β-galactosidase activity of transposon insertions, from KCEC675 (A) and KCEC676 (B), in the wild-type, Rm1 and Rm5 backgrounds. All the strains were grown in RD_{glyc} MOPS media; only KCEC675 and its derivatives were supplemented with 10 μg/ ml tetracycline (*). Strains used are as follows (left-to-right): (above) BGEC905, KCEC703*, KCEC118, KCEC675*, KCEC486 and KCEC708*; (below) BGEC905, KCEC755, KCEC118, KCEC756, KCEC486 and KCEC757.

3.10 Transposon mutagenesis in the region 2 background

Notwithstanding the fact that transposon mutagenesis in the Rm1 background had failed to identify the distant repressor of *fimB*, it still seemed plausible that searching in the region 2 background might be successful. It was for this reason that transposon mutagenesis was employed once again, this time using KCEC486 (*fimB'-'lacZYA Rm5*) as the recipient; the Rm5 mutation disrupts the NagC1 binding site, also known as region 2 (Sohanpal *et al.*, 2004). After screening in excess of 50, 000 colonies, twenty isolates were identified in this background that raised expression of *fimB-lacZ* on MacConkey lactose agar. Analysis of *fimB-lacZ* expression and genetic linkage to the *fim* locus, for each of these isolates, is given in table 3.4. Like the transposon mutants in the region 1 background, these isolates could be separated into a number of different groups according to their expression patterns of *fimB-lacZ*, in the presence or absence of tetracycline. However, genetic linkage experiments showed that each isolate conformed to just one of two different classes: all isolates that were tetracycline-dependant were also 100 % linked to the *fim* locus; whilst all tetracycline-independent phenotypes were unlinked to *fim*. Of the twenty isolates identified, only six belonged to the latter of these two groups.

3.11 Further analysis of transposon isolates in the Rm5 background

The six transposon insertions unlinked to *fim*, in the Rm5 background, were then subjected to inverse-PCR; and this identified six different insertion sites throughout the *E. coli* K12 genome (table 3.5). Two of these insertions disrupted genes that encode the regulatory proteins, *hns* and *lrp*. Whilst the former has already been disproved as the distant repressor of *fimB*, earlier in this chapter, so too had the latter by other workers in the Blomfield laboratory; Blomfield *et al.* (1993) have previously shown that an *lrp* mutation stimulates *fimB* expression during growth on MacConkey agar, but subsequent analysis showed that it has no discernable effect in RD_{glyc} MOPS (I. C. Blomfield, personal communication). Three of the remaining insertions were within genes hitherto unknown to affect expression of *fimB*: within *sdhB*, a gene that encodes a one subunit (SdhB) of the tetrameric succinate dehydrogenase enzyme of *E. coli* (Spencer & Guest, 1982); within *gatB*, a gene that encodes a subunit (II^B) of the galacticol-specific transporter (II^{Gat}) of the phosphoenolpyruvate:carbohydrate PTS system (Nobelmann & Lengeler, 1996); and within *crr*, that encodes one component of

Colony phenotypes:		Strains	Genetic linkage (%)
10 µg/ ml tetracycline	No tetracycline		
≈ Wt	≈ Rm5	KCEC677, KCEC678, KCEC681, KCEC683, KCEC684, KCEC685, KCEC688, KCEC693, KCEC694, KCEC696.	100
≈ Wt	≈ Wt	KCEC679, KCEC686.	0
> Wt	> Wt	KCEC680, KCEC682, KCEC687.	0
>> Wt	≈ Rm5	KCEC689, KCEC690, KCEC692, KCEC695.	100
>> Wt	>> Wt	KCEC691.	0

Table 3.4: Analysis of transposon mutants in the Rm5 background
Transposon insertion isolates were categorised according to *fimB-lacZ* expression on MacConkey lactose agar in both the presence and absence of tetracycline, after incubation at 37 °C for 18 hours. The kanamycin resistance determinant from BGEC696 was then transduced into each of the strains, and transductants were screened for tetracycline sensitivity on L-agar, after incubation at 37 °C for 18 hours. Genetic linkage is given as the percentage of transductants that became sensitive to tetracycline upon acquisition of the Kan^R determinant.

Strain	mTn*10*-flanking sequence (5' → 3'):	Genomic position (bp)	Strand	Location
KCEC679	GGTTAAGCAGCGCCG...	932,016	-	Within *lrp*
KCEC680	CGCGCTTTCTTCTTC...	1,292,008	+	Within *hns*
KCEC682	TATTGCGCATAGAAT...	757,252	-	Within *sdhB*
KCEC686	AACCGCGCCTCCGCA...	2,172,548	+	Within *gatB*
KCEC687	GGTGAAACCCCGGTT...	2,534,331	+	Within *crr*
KCEC691	TAC...	?	?	Inconclusive

Table 3.5: Identification of transposon insertion sites. Inverse-PCR was used to identify the sequence adjacent to each transposon. This sequence was subsequently located within the *E. coli* K12 (MG1655) genome using either the Colibri database (http://genolist.pasteur.fr/Colibri/) or the NCBI BLAST database (http://www.ncbi.nlm.nih.gov/BLAST/).

the glucose-specific enzyme II of the PTS system (IIA^{Glc}) (Meadow *et al.*, 1982; reviewed by Siebold *et al.*, 2001). The sixth transposon insertion, of KCEC691 could not be located using this protocol. This was because the end of the transposon (5' CATCAG), and the genomic sequence that was adjacent to it (5' TAC) formed an RsaI restriction endonuclease site (5' GTAC). As the inverse-PCR protocol required digestion with RsaI, these three nucleotides (5' TAC) were all that could be obtained of the adjacent sequence, and was insufficient to locate the transposon insertion site.

Strains KCEC682, KCEC686, KCEC687 and KCEC691 were subsequently assayed for β-galactosidase activity, in rich defined MOPS media with glycerol as the sole carbon source. However, none of these transposon insertions had any effect upon *fimB* expression in this media (data not shown).

3.12 Discussion

Expression of the FimB recombinase gene is stimulated by two regulatory proteins, NanR and NagC. Preliminary data suggested that these proteins serve to antagonise a repressor that lies even further upstream, and here, transposon mutagenesis was used to identify it. The intention was to identify insertions that were able to compensate for the loss of elements that would otherwise antagonise this repressor.

Conservative estimates suggest that using a transposon such as mTn*10* should hit any gene of interest after screening approximately 10, 000 colonies (Kleckner *et al.*, 1991). However, after screening over ten times this number of colonies, only nine insertions were identified that were able to compensate for the Δ2 mutation. All nine insertions were located, by genetic linkage experiments and PCR analysis, within the *nanC-fimB* intergenic region.

NanR and NagC are two very different regulators, and the Δ2 mutation disrupts stimulation of *fimB* expression by both of them. The inability to identify insertions that could compensate for the Δ2 mutation, led to a new hypothesis being tested: that these two regulators might serve to antagonise different repressors of *fimB*. Transposon mutagenesis identified a total of 40 independently isolated insertions that could compensate for either the Rm1 or Rm5 mutations, which disrupt NanR or NagC binding respectively. However, genetic linkage experiments (and inverse

PCR) determined that 21 of these insertions were linked to the *fim* locus; that is, 35 % of the insertions isolated in the Rm1 background, and 70 % of those isolated in the Rm5 background.

Transposon insertion sites of genetically linked, but distant insertions, and insertions unlinked to the *fim* locus could be further characterised by their effects upon a *fimB-lacZ* reporter fusion in liquid media. Of these 20 insertions, 12 had no effect in rich defined MOPS media with glycerol as the sole carbon source. The locations of the remaining 8 insertions, that did have an effect in this media, were identified by inverse-PCR. These insertions were within three loci: *hns*, within the *yjhV-fecE* intergenic region and *lrp*. Whilst all three insertions increase expression of *fimB*, NanR and NagC still stimulate expression in these backgrounds. This showed that whilst disrupting inhibitory factors, insertions at any of these loci do no affect the distant repressor of *fimB*, and to this end, transposon mutagenesis was unsuccessful.

Chapter 4:

Analysis of the *nanC-fimB* intergenic region

4.1 Summary

Two regulatory proteins, NanR and NagC, bind to sequences far (>700 bp) upstream of the *fimB* recombinase gene in order to stimulate expression, and preliminary data suggested that they serve to antagonise more distant inhibitory elements for *fimB* (El-Labany *et al.*, 2003). A number of artificial constructs within the *nanC-fimB* intergenic region were characterised in this chapter in order to understand better the mechanism of distant regulation for *fimB*. These initially brought into question the repressor/ anti-repressor model that was based upon the preliminary data, led to experiments that subsequently suggested it was incorrect, and then accounted for the effects observed in the preliminary data.

4.2 Introduction

It had previously been shown that an ectopic copy of *fimB* requires sequences upstream of the Δ2 region in *cis* in order to remain responsive to regulatory elements contained within the Δ2 region (El-Labany *et al.*, 2003). This observation led to the hypothesis that the sequences upstream of the Δ2 region bind one or more repressors of *fimB*, and that regulatory elements within it serve to antagonise them. Previous data in this thesis outlined a random, genome-wide search for this distant repressor of *fimB*; and whilst it identified a number of factors that affect *fimB* expression, it remained unsuccessful for this purpose.

This chapter focuses on sequences located within the *nanC-fimB* intergenic region. Transposon *cis*-insertions, previously isolated in this study, were investigated in more detail; their precise location within the genome was determined, as was their effects upon *fimB* expression under a number of different conditions. The repressor-binding region was subsequently deleted, and together these experiments gave results that were inconsistent with the repressor/ anti-repressor model of *fimB* regulation. In addition, control experiments were carried out that account for the results obtained when *fimB* was moved to this ectopic location, and suggest that the distant repressor/ anti-repressor model is incorrect.

4.3 Identification of transposon insertion sites within the *nanC-fimB* intergenic region

Transposon mutagenesis had previously identified nine independently isolated insertions in the Δ2 background, all of which were located near-to or within the *fim*

locus (refer to section 3.2-3.3 for more information). Here, two PCR reactions were carried out per insertion strain in order to identify their precise locations. By using one primer that was specific to the *nanC-fimB* intergenic region ('FimBUPR'), and one of two primers that were specific to mTn*10* ('mTn1x' & 'mTn2x'), the orientation of each transposon could also be determined. This confirmed that all nine transposon insertions were orientated such that the *tetA*-proximal end was closest to *fimB*. The PCR products were then sequenced, to determine the precise location of each transposon (table 4.1).

Strain	Sequence adjacent to transposon (5' → 3')	Insertion site
KCEC619	TACAAAGCCCGCATAAACAA	S1
KCEC620	TGTTTGCCATAGATTGCGAG	S3
KCEC621	(See KCEC619)	S1
KCEC622	(See KCEC619)	S1
KCEC623	CATTTTGTGGCAAATGGAGT	S5
KCEC624	(See KCEC619)	S1
KCEC625	(See KCEC619)	S1
KCEC626	TGCAC....(Δ2 deletion junction)	S4
KCEC627	TACTCAATCCATTTTACCCA	S2

Table 4.1: Identification of transposon insertion sites. Summary of transposon insertion sites for strains KCEC619 to KCEC627. In each instance, a 30 bp sequence adjacent to each transposon was identified, with no mismatches to a single site within the MG1655 genome. These insertion sites are summarised in figure 4.1

Of the nine transposon insertion strains sequenced, five had transposons inserted at the S^1 location, 1,170 bp upstream of the *fimB* orf; whilst a single strain had a transposon insertion at each of the remaining locations, ranging between 944 bp and 402 bp upstream of the *fimB* orf (figure 4.1). This in itself was intriguing because three of the insertion sites are downstream of the 743 bp repressor binding region, and the remaining six insertions are within it, yet all are tetracycline-dependant for compensation of the Δ2 mutation. This suggested that either the distant repressor retained activity over a large range of separation distances (± 2.93 kb), or the distant repressor, anti-repressor hypothesis was incorrect.

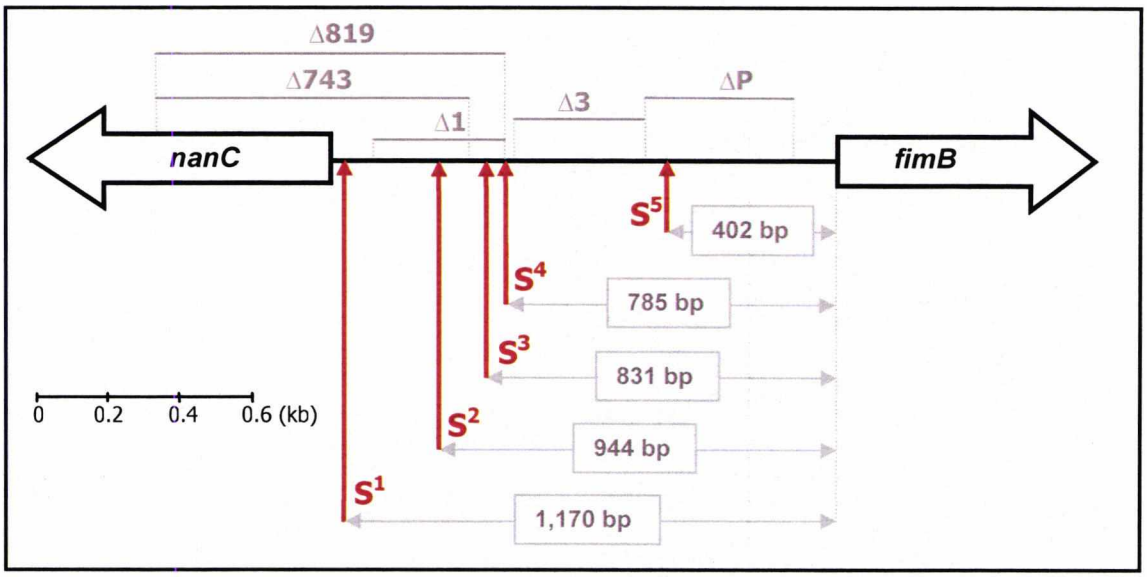


Figure 4.1: Identification of transposon insertion sites. Schematic diagram of the *nanC-fimB* intergenic region of BGEC711. Transposon insertion sites (**S**) are depicted, along with their distances from the *fimB* orf (located at bp 4538525 within the MG1655 genome).

4.4 Quantification of transposon *cis*-insertion phenotypes

Having determined the insertion sites of each transposon insertion strain, their effects upon the expression of a *fimB-lacZ* reporter fusion were determined (figure 4.2). This showed that compensation for the Δ2 mutation is largely tetracycline dependant, yet in the absence of tetracycline there is still partial compensation. It was unclear why more distant insertions, 800-1,000 bp upstream of the *fimB* orf were better able to compensate than insertions that were closer.

It is important to note the expression of each transposon insertion strain was relatively variable in the absence of tetracycline. As an example, the expression of strain KCEC626 (S^4 insertion) varied between 20 – 80 % of the wild-type between data sets, with the data presented in figure 4.2 being reproduced most frequently. Although the actual levels of expression varied greatly, the expression of each insertion mutant remained consistent relative to one another; such that the trends shown in figure 4.2 were reproduced in each instance. The source of variation of these strains was never identified.

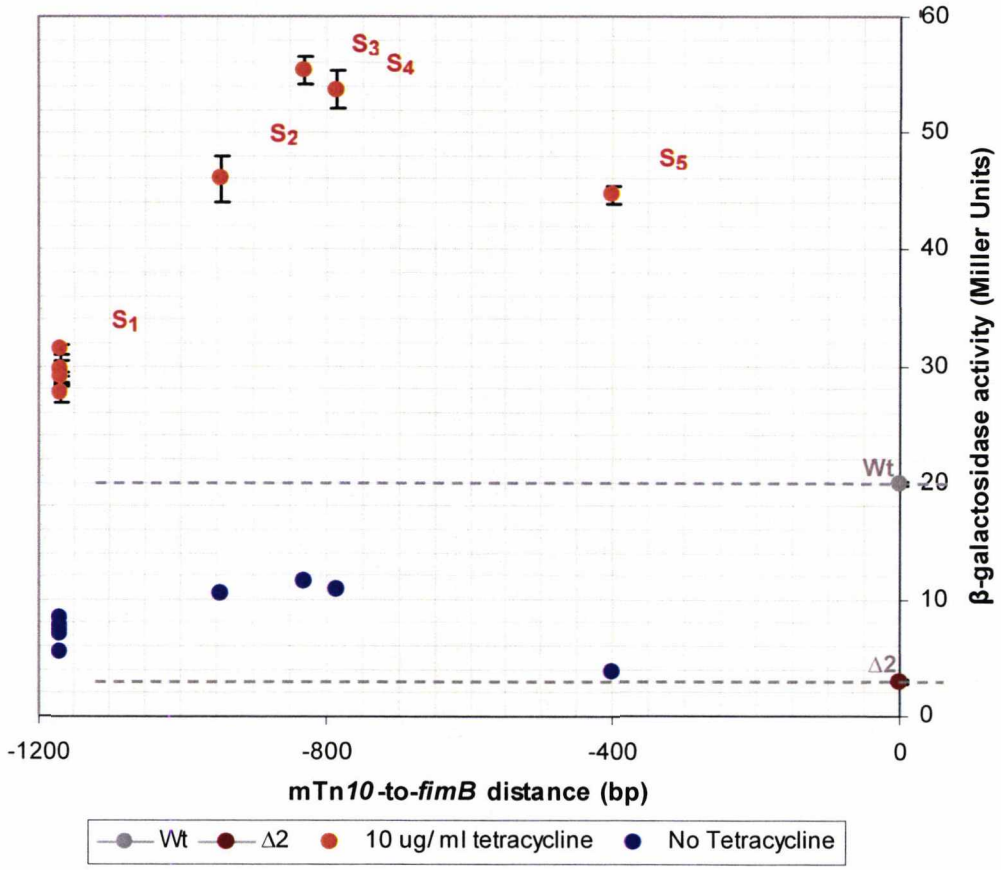


Figure 4.2: Quantification of transposon *cis*-insertion sites. β-galactosidase produced by strains containing transposons inserted within the *nanC-fimB* intergenic region, plus control strains. Data is plotted against the distance of each transposon from the *fimB* orf (in bp). Strains used are as follows: Wt, BGEC905; Δ2, BGEC711; transposon insertion strains, KCEC619 to KCEC627.

4.5 Transposon *cis*-insertions re-activate the *fimB* promoters

Other workers have shown that the promoters within mTn*10* can read beyond the ends of the transposon and transcribe genes that are situated downstream (Takiff *et al.*, 1992). Although there appeared to be a curious pattern between transposon insertion site and its effect on *fimB* expression, the tetracycline-dependant effect of mTn*10* insertions might be explained by transcriptional read-through. In order to test this, the promoter region of *fimB* was removed (ΔP) from KCEC626 (mTn*10* at the S4 location) so that transcriptional read-through from the *tetA* promoter could be quantified (figure 4.3).

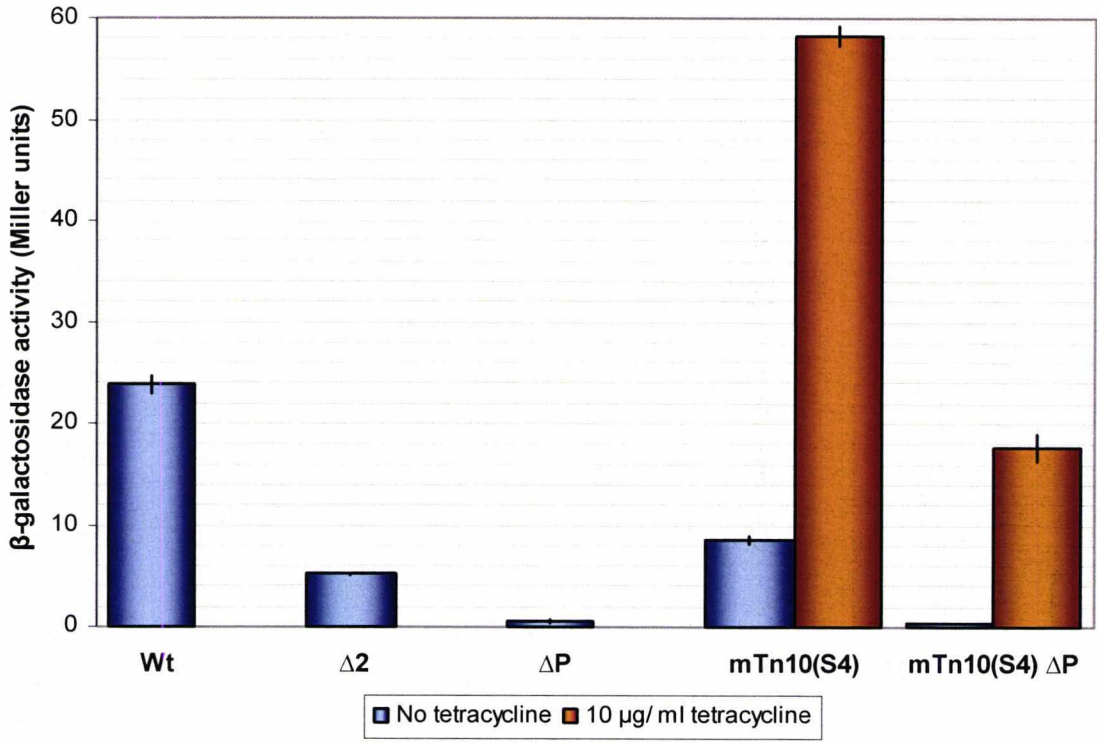


Figure 4.3: Transposon *cis*-insertions re-activate the *fimB* promoters. β-galactosidase produced by strains containing transposons inserted at the S^4 location, ± the ΔP mutation. Strains used are as follows: Wt, BGEC905; Δ2, BGEC711; ΔP, KCEC171; mTn*10*(S^4), KCEC626; mTn*10*(S^4) ΔP, KCEC828.

In the absence of tetracycline, the transposon insertion had no effect when the native *fimB* promoters were removed. This suggested that the transposon can partially compensate for loss of the Δ2 mutation, in a promoter-independent fashion, by re-activating the native *fimB* promoters. The effect of mTn*10* in the presence of tetracycline was also greatly diminished when the *fimB* promoter region was deleted, which suggested that the native *fimB* promoters can be re-activated by promoter activity that originates upstream. However, it is also possible that the ΔP region might contain elements that facilitate anti-termination of transcripts that originate upstream. As both hypotheses would account for the observations made from these data, the effect of *tetA* promoter activity on *fimB* expression remained unclear.

4.6 Deletion of the distant repressor region does not compensate for the loss of activators

All nine of the transposon *cis*-insertions were located either within the 743 bp repressor binding region, or separating it from the native *fimB* promoters.

However, displacement of the repressor binding region only resulted in partial compensation of the Δ2 mutation, and this led us to question the distant repressor, anti-repressor model. As a consequence an 819 bp sequence was deleted, which contains the 743 bp repressor binding region and (for simplicity) the 76 bp that separates it from the Δ2 region (figure 4.1) (MG1655 genome sequence, bp 4536665-4537484). This deletion, both with and without the Δ2 mutation was constructed, and the effects on *fimB* expression quantified (figure 4.4).

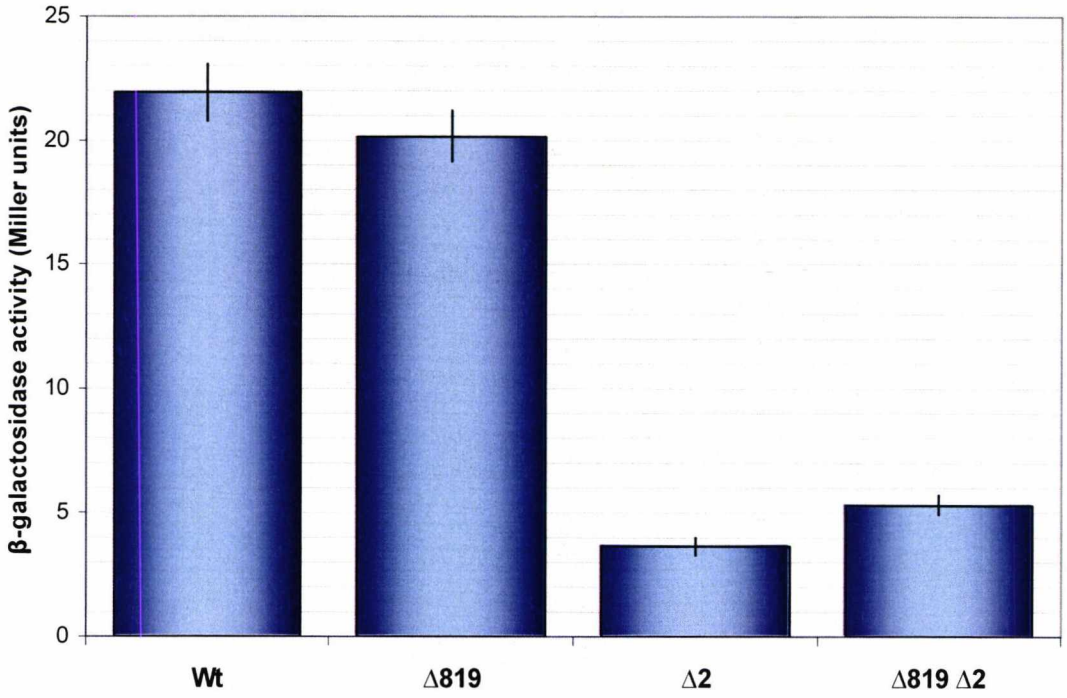


Figure 4.4: Deletion of the distant repressor region does not compensate for the loss of activators. β-galactosidase produced by a FimB-LacZ translational fusion containing the Δ819 mutation, ± the Δ2 mutation. Strains used are as follows: Wt, BGEC905; Δ819, KCEC023; Δ2, BGEC711; and Δ819 Δ2, KCEC025.

If the model was correct, deletion of the repressor-binding region (Δ819) might have no effect, or might increase expression; and the Δ2 mutation would have no effect when combined with it. However the Δ819 mutation slightly decreased expression relative to the wild-type, and only partially compensated for loss of the Δ2. These data suggested that activators within the Δ2 region do not serve to antagonise inhibitory elements located within this region.

4.7 Post-*lac* sequence is able to compensate for the loss of *fimB* activators

Although the distant repressor, anti-repressor model was a logical hypothesis to explain the merodiploid data, analysis of the transposon *cis*-insertion strains and the Δ819 constructs suggested that it was incorrect. Alternatively the merodiploid data could be explained if expression of *fimB* in these merodiploid constructs was affected by the foreign sequence that is brought into close proximity, rather than additional sequence from the *nanC-fimB* intergenic region, if this foreign sequence was somehow able to compensate for loss of native *fimB* activators (the Δ2 mutation). In order to test this hypothesis, the sequence present in the short merodiploid construct, immediately upstream of the ectopic copy of *fimB*, was cloned into the *nanC-fimB* intergenic region (figure 4.5) (this sequence, located downstream of the *lac* locus is referred to as post-*lac* sequence in this study). This merodiploid sequence was used to replace the Δ819 mutation, and contains 1,015 bp of the *cynX* orf (MG1655 genome sequence, bp 359355-360370); 63 bp of the *cynX-lacA* intergenic region (MG1655 genome sequence, bp 360371-360434); and four tandem repeats of the *E. coli rrnB* transcriptional terminator (724 bp) from the cloning vector pRS415 (Simons *et al.*, 1987) (the 165 bp *rrnB* terminator region can also be located within the MG1655 genome, 4169336-4169501). β-galactosidase assays were used to quantify the effects of these constructs on the expression of a *fimB-lacZ* reporter fusion (figure 4.6 & figure 4.7).

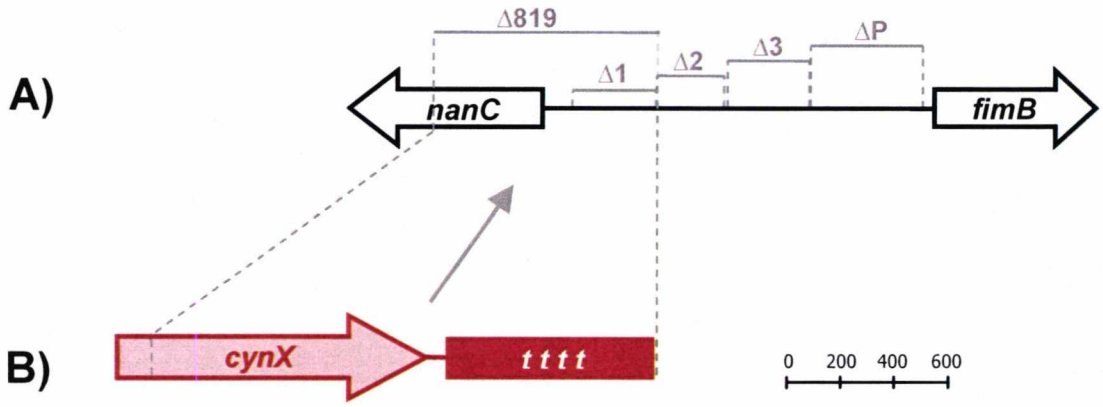


Figure 4.5: Schematic diagram defining the wild-type *nanC-fimB* intergenic region (A), and the post-*lac* sequence that was inserted to make strain KCEC679 (B) ('*tttt*' defines the four tandem copies of the *rrnB* terminator region from pRS415).

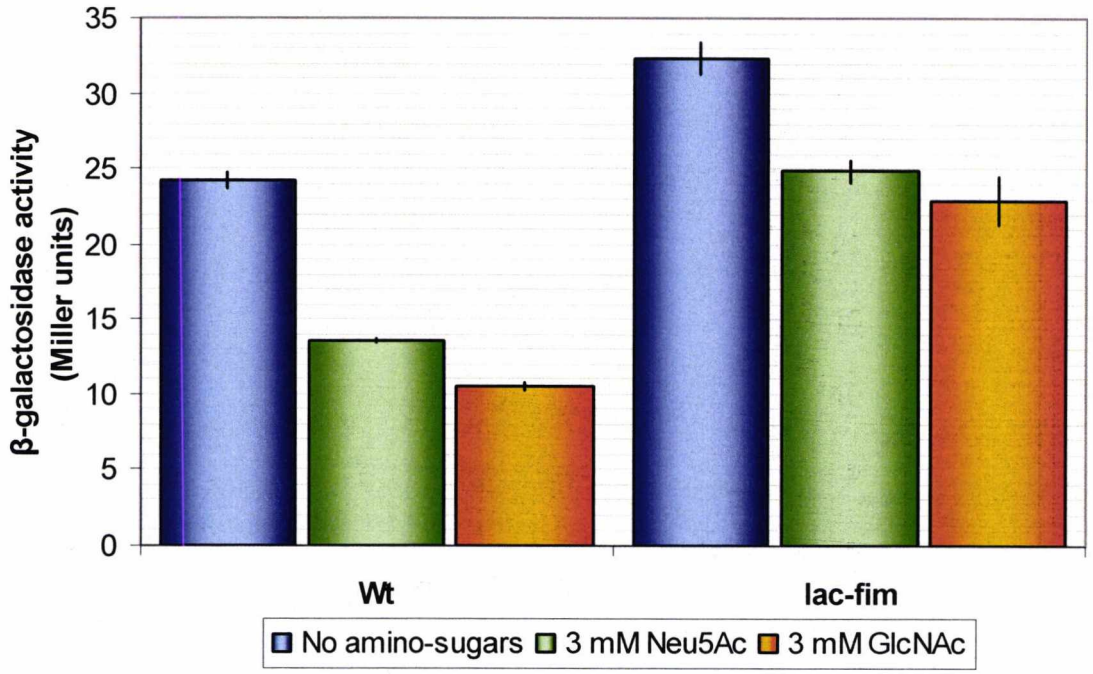


Figure 4.6: Post-*lac* sequence is able to compensate for loss of *fimB* activators (i). β-galactosidase produced by a *fimB-lacZ* translational fusion in the presence of the post-*lac* hybrid construct. Cells were grown in RD MOPs media supplemented with no aminosugars; 3 mM Neu5Ac; or 3 mM GlcNAc. Strains used are as follows: Wt, BGEC905; *lac-fim* hybrid, KCEC769.

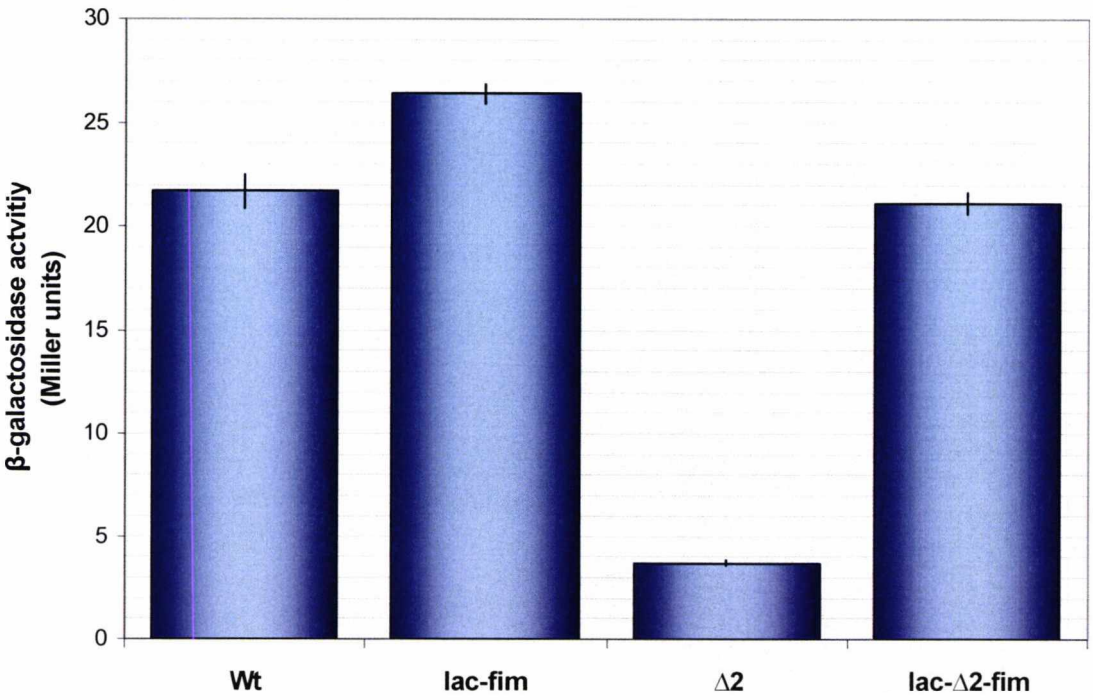


Figure 4.7: Post-*lac* sequence is able to compensate for loss of *fimB* activators (ii). β-galactosidase produced by a *fimB-lacZ* translational fusion, in the presence of the post-*lac* hybrid construct, ± the Δ2 mutation. Strains used are as follows: Wt, BGEC905; *lac-fim*, KCEC769; *lac*-Δ2-*fim*, KCEC771.

When the Δ2 region was present, post-*lac* sequence resulted in a slight increase in expression of a *fimB-lacZ* reporter fusion. Although this construct remained responsive to both Neu5Ac and GlcNAc, it maintained wild-type levels of expression in the presence of either aminosugar. In addition, the post-*lac* sequence was also able to maintain wild-type levels of *fimB* expression when combined with the Δ2 mutation. These data suggested that the post-*lac* sequence and activators within the Δ2 region have an additive effect when combined. However, in the presence of the post-*lac* sequence these activators appear to function less efficiently, as wild-type levels of expression are maintained in their absence, yet both combined do not increase expression to double the wild-type level. Most importantly, though, this showed that the post-*lac* sequence is able to compensate for loss of the *fimB* activators, and strongly suggested that the distant repressor, anti-repressor model was incorrect.

4.8 Post-*lac* sequence is able to re-activate the *fimB* promoters

Transposon insertions upstream of *fimB* were able to partially compensate for the Δ2 mutation, and they might do so by re-activating the native *fimB* promoters from a distance. Post-*lac* sequence was also able to compensate for the Δ2 mutation, at a similar distance, and this suggested that it might also re-activate the *fimB* promoters. In order to test this, post-*lac* sequence was combined with both the Δ2 and ΔP mutations, and expression of a *fimB-lacZ* reporter fusion was used to measure the effect (figure 4.8).

This showed that the post-*lac* sequence has no phenotype when combined with the ΔP mutation. It suggested that, like the transposon *cis*-insertions, post-*lac* sequence is able to re-activate the native *fimB* promoters. However unlike the transposon *cis*-insertions, it suggested that this activity of post-*lac* sequence was entirely independent of transcriptional read-through. Whilst much of the phenotype associated with the transposon *cis*-insertions was dependant upon transcriptional read-through, post-*lac* sequence contains no known promoters and is separated from the *fim* sequence by four tandem repeats of the *E. coli rrnB* transcriptional terminator.

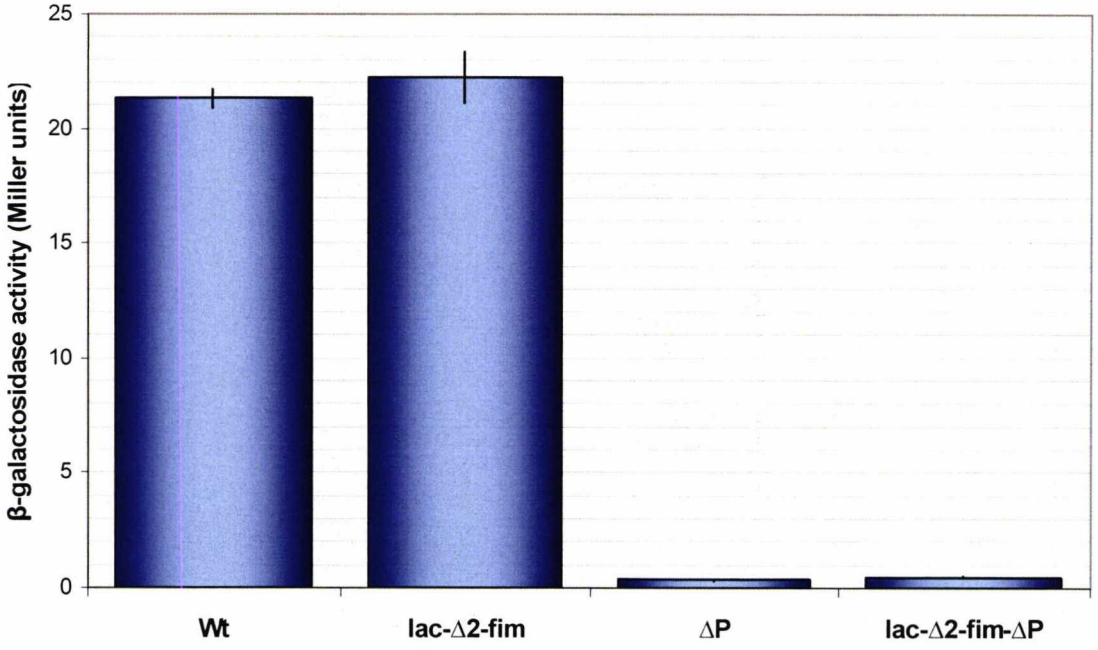


Figure 4.8: Post-*lac* sequence is also able to re-activate the *fimB* promoters. β-galactosidase produced by a *fimB-lacZ* translational fusion in the presence of the post-*lac* hybrid and Δ2 mutations, +/- the ΔP mutation. Strains used are as follows: Wt, BGEC905; *lac*-Δ2-*fim*, KCEC771; ΔP, KCEC171; *lac*-Δ2-*fim*-ΔP, KCEC828.

4.9 Sequence comparison of the *nanC-fimB* intergenic regions from *E. coli* K-12 strains MG1655 and J96.

Without the distant repressor, ant-repressor model, it was unclear how NanR and NagC stimulate expression of *fimB* at a distance. A potential answer was found in the literature regarding *fimB*, as other workers had shown that the uropathogenic *E. coli* strain J96 has a weak promoter within the Δ2 region (~ 900 bp upstream of *fimB*)(Schwan *et al.*, 1994). Analysis of the transposon insertion strains previously in this chapter had suggested that transcripts originating upstream, from mTn*10* might stimulate the native *fimB* promoters within the ΔP region; and other workers had shown that the ΔP mutation accounts for *fimB* promoter activity in *E. coli* K-12 (MG1655)(El-Labany *et al.*, 2003). Taken together, these data led to the hypothesis that this third promoter, far upstream, might serve a regulatory function and stimulate promoters located downstream in the ΔP region.

An assumption that was central to this hypothesis was that regulation of *fimB* expression in the commensal *E. coli* K-12 strain, MG1655, and the uropathogenic strain, J96, was the same. Initially the intergenic regions of these two strains were

compared by sequence analysis, to determine whether there were any alterations that might affect *fimB* expression (Figure 4.9). The only published sequence for this region of the J96 genome (Schwan *et al.*, 1994) differed strikingly from that of MG1655 throughout the Δ2 region, with the NanR, NagC1 and IHF binding sites each being altered. The sequences downstream of the Δ2 region, extending to the *fimB* orf were essentially the same, with only a few single-base substitutions and deletions (data not supplied). In order to confirm that this J96 sequence was correct, genomic DNA from this strain was amplified by PCR, and two replicate samples were sequenced. These two samples were first compared to each other, and then aligned to the published MG1655 and J96 sequences (Blattner *et al.*, 1997; Schwan *et al.*, 1994). This showed that the sequence published by Schwan *et al.*, (1994) is incorrect, and the Δ2 region in J96 differs from MG1655 by just two substitution mutations within the Δ4 region.

In order to account for the differences between the published *E. coli* J96 sequence, and the one defined in this study, the published sequence was further analysed using BLAST homology searches (Altschul *et al.*, 1990). This identified two discreet elements within this sequence: a 1, 087 bp element that, in agreement with the data presented here, bears homology to the region upstream of *fimB* in numerous *E. coli* genomes; and a 272 bp upstream element (the J96A sequence, as defined in figure 4.9) that bears homology to a number of cloning vectors, predominantly those derived from pBR322/ pUC series plasmids (homology data not presented here). As Schwan *et al.* (1994) cloned J96 genomic sequence into plasmid derivatives of pUC18/19 prior to sequencing, this suggests that they might have erroneously published a partial vector fragment in their sequence.

4.10 Replacement mutagenesis of the Δ4 – NagC2 region

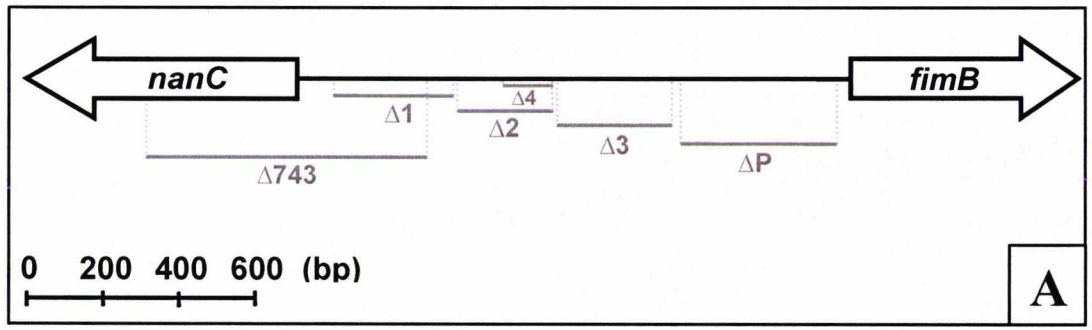
Schwan *et al.* (2004) had mapped this third promoter to approximately 900 bp upstream of the *fimB* orf, which placed it within the Δ2 region (Schwan *et al.*, 1994). The fact that there were few changes to the J96 sequence suggested that MG1655 might also possess this third promoter of *fimB*. Other workers had analysed a 133 bp fragment of the Δ2 region with a series of replacement mutations, and this had only identified the NanR and NagC1 binding sites (El-Labany *et al.*, 2003). The remaining 119 bp of the Δ2 region (the Δ4 region; 792 – 911 bp upstream of the *fimB* orf) remained uncharacterised. Here, small (12-14 bp)

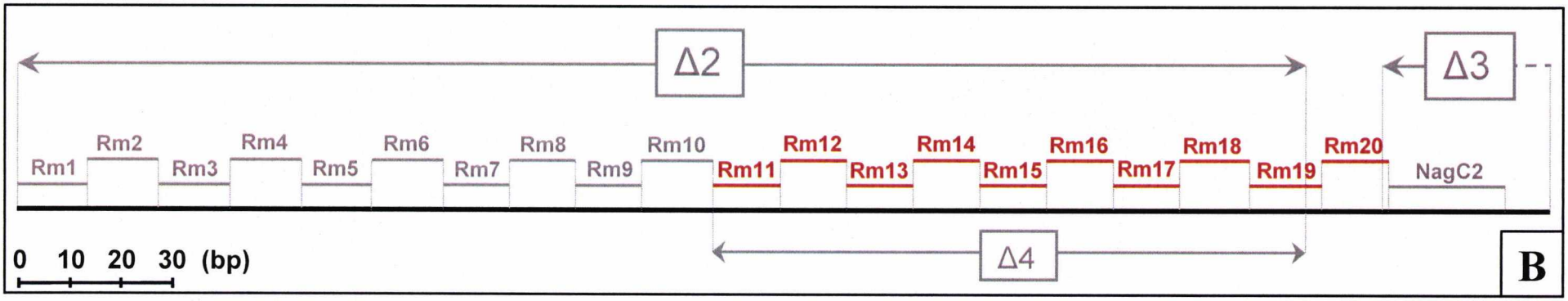
heterologous replacement mutations were constructed in order to analyse the Δ4 region, the Δ3 region up to NagC2, and the 15 bp that separates the two (Figure 4.10). β-galactosidase assays were used to quantify the effects of these replacement mutations upon a *fimB-lacZ* reporter fusion (4.11).

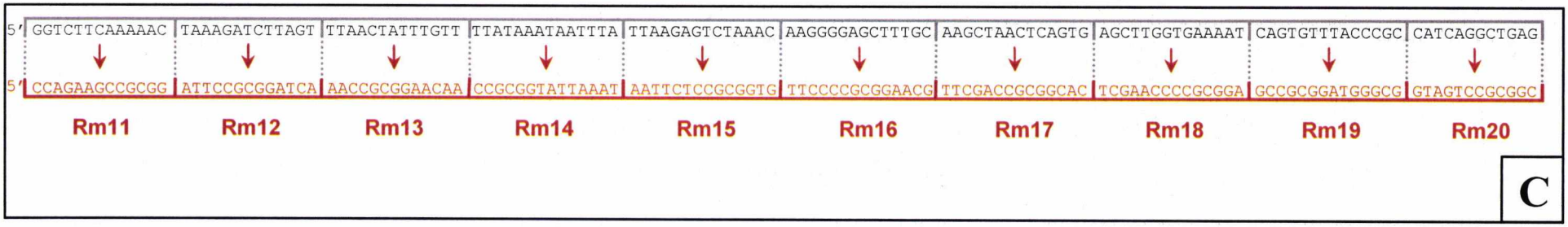
Figure 4.9: Sequence comparison of the *nanC-fimB* intergenic regions from *E. coli* strains K-12 (MG1655) and J96. A 376 bp partial sequence, from the *nanC-fimB* intergenic region of MG1655 (bp 4537370-4537746 of the MG1655 genome sequence) was aligned with two sequences for strain J96: 'J96A' was published by Schwan *et al.* (1994), and 'J96B' is from this study. This sequence contains a partial fragment of the Δ1 region (bp 1-122) and the entire Δ2 region (bp 124-376); the NanR, NagC1 and IHF binding sites are **highlighted** (bp 119-141, 182-203 and 285-296 respectively). Sequences were aligned and edited using ClustalW and Jalview (Clamp *et al.*, 2004; Thompson *et al.*, 1994). The region downstream of this sequence was also determined in this study, and was in good agreement to the published sequences for MG1655 & J96 (data not supplied).

```
J96A/                GACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGG
K-12/ 1-49           GTTGATCTCATTTTCATATTCCTTTCAGACTGAAGT-GCGTTTTCATCAC
J96B/                GTTGATCTCATTTTCATATTCCTTTCAGACTGAAGT-GCGTTTTCATCAC

J96A/                TTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGAT
K-12/ 50-99          CCGATGAGAAAGAACTGAAGTGATGTTTGCCATAGATTGCGAGATGCCCT
J96B/                CCGATGAGAAAGAACTGAAGTGATGTTTGCCATAGATTGCGAGATGCCCT

J96A/                GGCGCCAACAGTCCCCCGGGC-CACGGGCCTGCCC-ACCATACCCACGCCGA
K-12/ 100-149        GTCTCATTGACTATCTTGGTGCACCTTTATACCTGTTATACCAGATCAAA
J96B/                GTCTCATTGACTATCTTGGTGCACCTTTATACCTGTTATACCAGATCAAA

J96A/                AA-CAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGT
K-12/ 150-199        AATCACGCAATCCATACAACAAAACCAGATTTGCAATTCGTGTCACAAAA
J96B/                AATCACGCAATCCATACAACAAAACCAGATTTGCAATTCGTGTCACAAAA

J96A/                GATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGT--GAT
K-12/ 200-248        TATGTCGATCTTTTTCTAAGAGGAAGATGC-CATGTGAAGCCAGACGAAC
J96B/                TATGTCGATCTTTTTCTAAGAGGAAGATGC-CATGTGAAGCCAGACGAAC

J96A/                GCCGGCCACGATGCGTCCGCGTAGAGGATCTTAGTTTAACTATT-GTTTT
K-12/ 249-298        ACTTGCGGTGGTCTTCAAAAACTAAAGATCTTAGTTTAACTATTTGTTTT
J96B/                ACTTGCGGTGGTCTTCAAAAACTAAAGATCTTAGTTTAACTATTTGTTTT

J96A/                ATAAATAATTTGTT-AGAGTCTAAACAAGGGGAGCTTTGCAAGCTAACTC
K-12/ 299-348        ATAAATAATTTATTAAGAGTCTAAACAAGGGGAGCTTTGCAAGCTAACTC
J96B/                ATAAATAATTTGTTTAGAGTCTAAACAAGGGGAGCTTTGCAAGCTAACTC

J96A/                AGTGAGCTTGGTGAAAATCAGTGTTTAC
K-12/ 349-376        AGTGAGCTTGGTGAAAATCAGTGTTTAC
J96B/                AGTGAGCTTGGTGAAAATCAGTGTTTAC
```

Figure 4.10: Schematic diagram defining replacement mutations (Rm) 11-20. For reference, deletion mutations within the *nanC-fimB* intergenic region are defined (A), as are the original replacement mutations (Rm) within the Δ2 region (B) (El-Labany *et al.*, 2003), alongside mutations constructed in this study (Rm 11-20). The nucleotide sequences of both the wild-type and replacement mutations are also defined (C) (MG1655 genome sequence, bp 4537627-4537762).







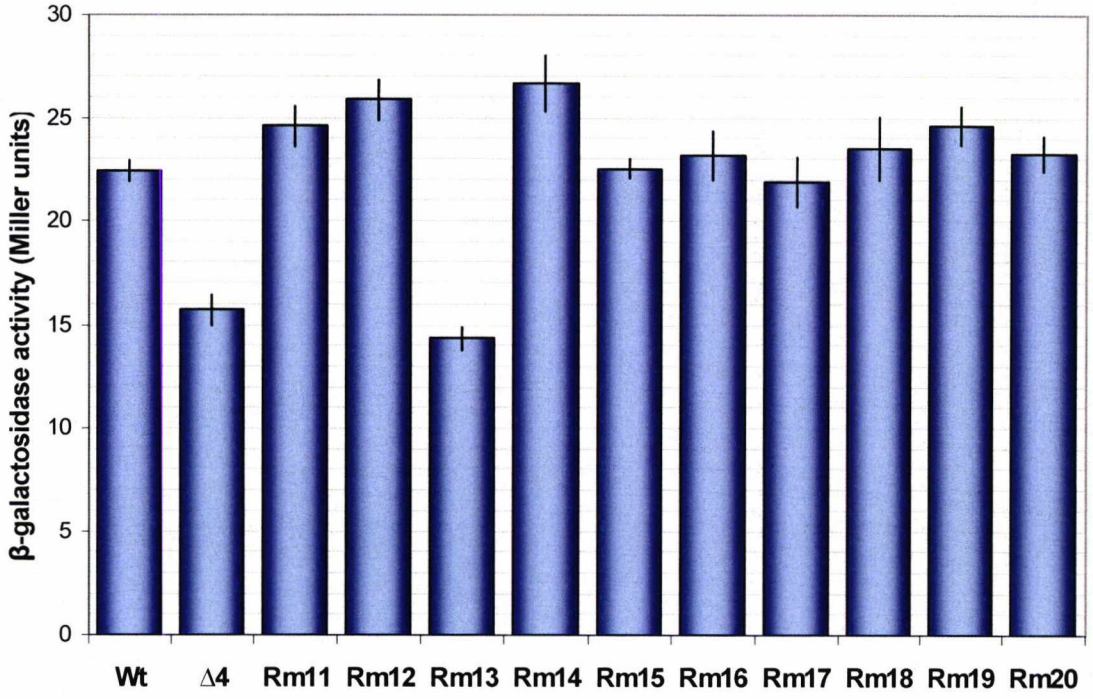


Figure 4.11: Replacement mutagenesis of the Δ4-NagC2 region.
β-galactosidase produced by a *fimB-lacZ* translational fusion containing small heterologous replacement mutations, Rm11-20. Strains used are as follows: Wt, BGEC905; Δ4, BGEC881; Rm11 KCEC797; Rm12, KCEC799; Rm13, KCEC801; Rm14, KCEC803; Rm15, KCEC805; Rm16, KCEC807; Rm17, KCEC809; Rm18, KCEC811; Rm19, KCEC813; and Rm20, KCEC815.

Only one replacement mutation, Rm13, had an effect on *fimB* expression, which was comparable to the effect of the Δ4 mutation. This mutation (Rm13) removes a sequence that is in good agreement with the consensus binding site of IHF (Goodrich *et al.*, 1990), and other workers within the laboratory have subsequently shown that IHF binds to this sequence *in vitro* to facilitate cooperative binding of NagC to the NagC1 and NagC2 sites (I. C. Blomfield, manuscript submitted for publication). Although MG1655 might still possess a third promoter, upstream of *fimB*, replacement mutagenesis showed that either it is not located in this region or it has no effect upon *fimB* expression.

4.11 Discussion

Other workers had shown that activators within the Δ2 region are not required for expression of an ectopic copy of *fimB* (El-Labany *et al.*, 2003), and this led to the hypothesis that these activators serve to antagonise one or more inhibitory elements that are located in an upstream, 743 bp region. Transposon mutagenesis was employed previously in this study, in order to identify this distant repressor of *fimB*;

and whilst it identified a number of insertions that affect *fimB* expression, *fimB* expression remained dependant upon activation by these activators, NanR and NagC.

Transposon mutagenesis had, however, isolated nine independent mTn*10* insertions within or close to the *fim* locus, in the Δ2 background. All nine insertions were located within the *nanC-fimB* intergenic region, at five different positions ranging from 402 bp to 1,170 bp upstream of the *fimB* orf. Each mTn*10* insertion was orientated such that its tetracycline-inducible (*tetA*) promoter was directed towards *fimB*; and whilst *fimB* expression was significantly increased upon induction of this promoter, each insertion exerted a smaller, yet discernible induction of *fimB* in the absence of tetracycline.

Transposon insertions within the *nanC-fimB* intergenic region were interesting for two reasons. Firstly, they highlighted an unusual relationship between insertion site and stimulation of *fimB* expression; with those in excess of 800 bp upstream of the *fimB* orf having a greater effect than one 400 bp upstream, within the ΔP region. Removal of the native *fimB* promoter region in one such transposon insertion strain showed that the ΔP region is required for mTn*10* to stimulate expression of *fimB*. This suggested that promoter activity originating far upstream can activate the native *fimB* promoters, or that the *fimB* promoter region facilitates anti-termination of transcripts originating upstream. Secondly, each of these transposon insertions had very little effect upon *fimB* expression in the absence of tetracycline, even though they displaced the repressor binding region by 2.9 kb. As no known (σ^{70}) prokaryotic transcriptional regulators retain activity over such large separation distances, this led us to question to the validity of this distant repressor, anti-repressor model.

Deletion of an 819 bp sequence, which contains the repressor binding region, had almost no effect upon *fimB* expression, and was unable to compensate for loss of the *fimB* activators, NanR and NagC. This suggested that the distant repressor, anti-repressor model was incorrect, but it remained unclear how the *fimB* expression pattern at an ectopic location could be accounted for. In order to do so, an alternative hypothesis was proposed in which sequences at this ectopic location might be able to supplement the role of *fimB* activators when located immediately upstream. In order to test this, the sequence immediately upstream of *fimB* in the short merodiploid (defined as 'post-*lac*') was cloned into the *fim* locus, and as proposed by this

hypothesis, it was able to compensate for the Δ2 mutation. Furthermore, much like the transposon insertions this sequence was entirely dependant upon the *fimB* promoter region for its function. However, unlike the transposon insertions it did not achieve this through promoter activity.

Without the distant repressor, anti-repressor model, it became unclear how NanR and NagC now served to activate *fimB* expression at such a distance. A potential explanation was found in the literature, as other workers had identified a third promoter for *fimB* promoter far (~ 900 bp) upstream in a uropathogenic *E. coli* strain, J96 (Schwan *et al.*, 1994). Intriguingly this third promoter would be situated within the Δ2 region, at a similar distance from the *fimB* orf as the transposon insertion that had the greatest effect on expression of this gene (831 bp).

Comparison of the sequences in this region, for both strains (J96 and MG1655) showed that there were many differences between the two. However, this fragment of the J96 genome was amplified in this study and re-sequenced; this showed that the published J96 sequence is incorrect, and that the two genomes differ in this region by only two substitution mutations. A series of short, heterologous replacement mutations (Rm) were constructed in order to identify this promoter within the remaining portion of the Δ2 region that was hitherto uncharacterised. These mutations analysed the 149 bp of sequence that separates the Rm10 mutation (El-Labany *et al.*, 2003) from the NagC2 site (Sohanpal *et al.*, 2004). Whilst they failed to identify a promoter, one defined a potential IHF site, within this region that fully accounts for the effect of the Δ4 mutation. Other workers have subsequently shown that IHF does indeed bind to this sequence, and in doing so it facilitates the cooperative binding of NagC sites (NagC1 & NagC2) *in vitro* (I. C. Blomfield, personal communication).

Chapter 5:

Further analysis of the *nanC-fimB* intergenic region

5.1 Summary

Two regulatory proteins, NanR and NagC, bind far upstream of *fimB* in order to activate expression. Preliminary data suggested that these might serve to antagonise inhibitory elements that lay further upstream of *fimB*; however, data presented in the previous chapter suggested that this is incorrect. In light of these data, it therefore remained unclear how NanR and NagC might stimulate *fimB* expression at a distance. In this chapter, deletion and replacement mutations were used to analyse the sequence that separates these distant activators from the *fimB* gene. This identified a single promoter, which is entirely responsible for *fimB* expression in *E. coli* K12. Mutational analysis also identified additional *cis*-active elements downstream of NanR and NagC that affect *fimB* expression independently. In addition, further mutational analysis also determined that activation of *fimB* expression by NanR and NagC is sensitive to their separation distance from the *fimB* promoter.

5.2 Introduction

Other workers have previously shown that two regulatory proteins, NanR and NagC, bind to sequences far upstream (≥ 740 bp) of *fimB* and stimulate its expression (Sohanpal *et al.*, 2004); a third regulatory protein, IHF, binds to a sequence 859 bp upstream of *fimB* (between the NagC1 and NagC2 binding sites) and is required for activation by NagC (I. C. Blomfield, manuscript submitted for publication) (figure 5.1). Contrary to previous findings (El-Labany *et al.*, 2003), data presented previously in this study suggested that these regulatory proteins do not serve to antagonise a distant repressor and that deletion of the 819 bp region upstream of their binding sites has little or no effect upon their activity. At the same time they do not appear to require any additional localised elements, within the Δ2 region in order to regulate expression of the *fimB* promoters from such a distance. It therefore remained unclear how these regulators control *fimB* expression over such large distances.

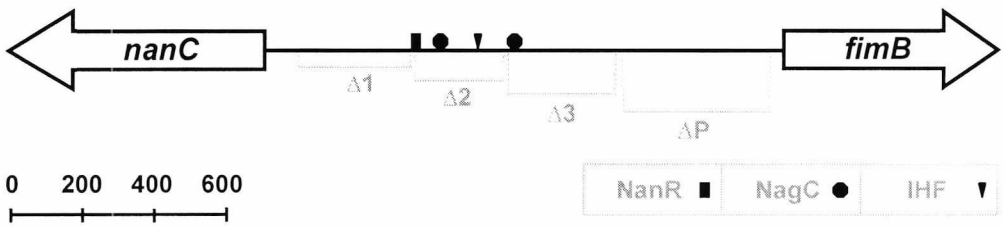


Figure 5.1: Schematic diagram of the *nanC-fimB* intergenic region. Deletion mutations within this region are also defined, as are the NanR, NagC and IHF binding sites.

This chapter investigates the role of sequences downstream of these activators, in order to understand better the mechanism of *fimB* activation. Small (2-3 bp) replacement mutations were used to identify *fimB* promoters for *Escherichia coli* K12, as defined previously by other workers in other *E. coli* strains (Olsen *et al.*, 1994; Schwan *et* al., 1994), and this identified a single promoter under the conditions tested. In addition, a series of deletion and heterologous replacement mutations were used to analyse the Δ3 region, up to but not including the NagC2 binding site, for *cis*-active elements that regulate *fimB* expression. These were combined with mutations that disrupt the NanR or NagC binding sites (Rm1 or Rm5, respectively) (El-Labany *et al.*, 2003), in order to determine the effects of each mutation on the efficacy of activation by each regulatory protein.

5.3 Analysis of the ΔP region (I): identification of Schwan's *fimB* promoters

In the absence of the distant repressor, anti-repressor model, NanR and NagC appeared to stimulate *fimB* expression at a distance, either directly or indirectly. Other workers had provided evidence that the ΔP region contains two native *fimB* promoters (El-Labany *et al.*, 2003; Olsen & Klemm, 1994; Schwan *et al.*, 1994), and this raised the possibility that activation by NanR or NagC might be specific for one, or both, of the promoters. Two independent research groups had provided evidence for these two promoters; Schwan *et al.* (1992 & 1994) used northern blot analysis and primer extension, and Donato *et al.* (1997) used *in vitro* transcription assays to identify them in RNA or DNA (respectively) derived from the uropathogenic *E. coli* strain J96 (Hull *et al.*, 1981). In order to confirm that these two promoters are present in *E. coli* K-12 strain MG1655, the predicted -10 hexamers were mutated (to 5' GCTAAT for P1, and 5' GTGACT for P2, respectively) (figure 5.2A & 5.2B), and their effects upon β-galactosidase activity of a *fimB-lacZ* reporter fusion were quantified (figure 5.3).

The P1 and P2 mutations, individually, had relatively little effect upon expression of *fimB*; and the combination of both mutations did not even halve expression. When compared to the dramatic reduction in *fimB* expression that results from the ΔP mutation, these data suggested that these (P1 and P2) mutations were only decreasing *fimB* promoter activity, and not abolishing it. More specifically, it suggested that these mutations were not disrupting the -10 hexamers of either *fimB* promoter.

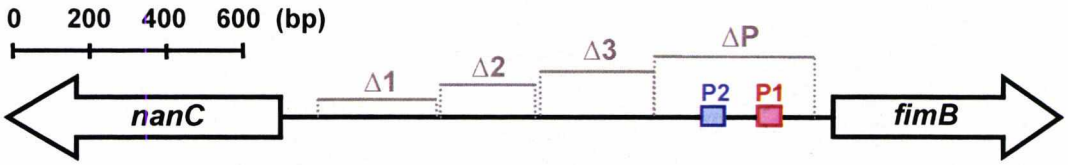


Figure 5.2A: Schematic diagram depicting the *fimB* promoters (P1 & P2) defined by Schwan *et al.* (1994).

Figure 5.2B: The nucleotide sequence of the promoters defined by Schwan *et al.* (1994). The predicted consensus sites (-35 & -10) of each promoter (P1 & P2), and the transcriptional start sites (+1) are also defined. The MG1655 genomic positions, for the sequences encompassing the P1 & P2 promoters defined here, are bp 4538326-4538386 & 4538185-4538245 respectively.

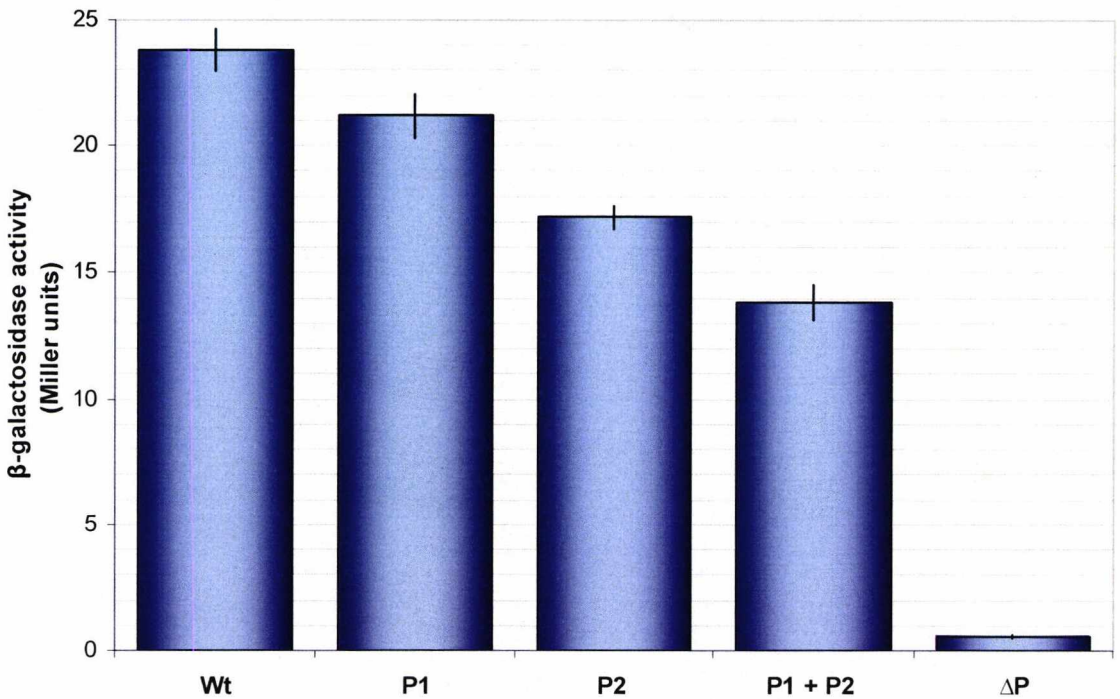


Figure 5.3: Analysis of the ΔP region (I): identification of Schwan's *fimB* promoters. The predicted -10 consensus sites, of both promoters defined by Schwan *et al.* (1994) were mutated and their effects quantified by measuring β-galactosidase activity of a *fimB-lacZ* reporter fusion. Strains used are as follows: Wt, BGEC905; P1 (mutation), BGEC916; P2, BGEC904; (P1 + P2), KCEC826; ΔP, KCEC171.

5.4 Analysis of the ΔP region (II): identification of Klemm's *fimB* promoter

In contrast to Schwan *et al.* (1994), other workers had defined only a single, alternative *fimB* promoter in an *E. coli* K-12 strain VL584 (*Δfim*) bearing the *fim* fragment of K12 strain CSH50 (Miller, 1972). In order to confirm the presence of this promoter in *E. coli* K-12 strain MG1655, the predicted -10 consensus for this promoter (P_{Klemm}) was also mutated (from 5' TATATA to 5' GGTATG) (figure 5.4A & 5.4B). Its effect was again quantified by measuring β-galactosidase activity of a *fimB-lacZ* reporter fusion (figure 5.5).

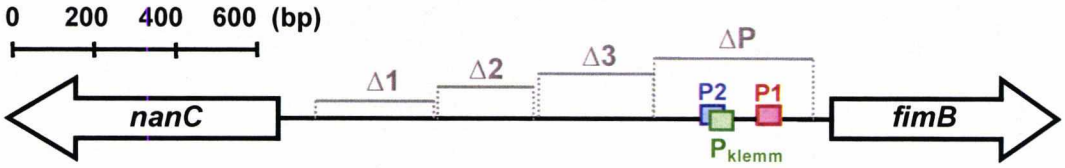


Figure 5.4A: Schematic diagram depicting the *fimB* promoter defined by Olsen *et al.* (1994).

P_{Klemm}

5' TGTTGATTTTACTTGTTACAGAACATATCACATGATATATAGATAAGATTAGTTGCATTA

-35 -10 +1

Figure 5.4B: The nucleotide sequence of the promoter defined by Olsen *et al.* (1994). The predicted consensus sites of this promoter (-35 & -10), and the transcriptional start site (+1) are also defined. The MG1655 genomic position of the sequence defined here is bp 4538207-4538267.

The effect of this mutation (P_{Klemm}) was comparable to that of the ΔP. As it entirely removed all promoter activity within this region, it suggested that the *fimB* recombinase gene of *E. coli* K-12 strain MG1655 has a single promoter, or only one is active under these growth conditions, and that this mutation had disrupted it. Other workers, within Dr. I.C. Blomfield's laboratory have subsequently shown that mutagenesis of the predicted -35 consensus site of this promoter has a comparable effect upon *fimB* expression as the mutation defined here (I. C. Blomfield, personal communication). These data suggest that NanR and NagC serve to activate a single promoter for *fimB*.

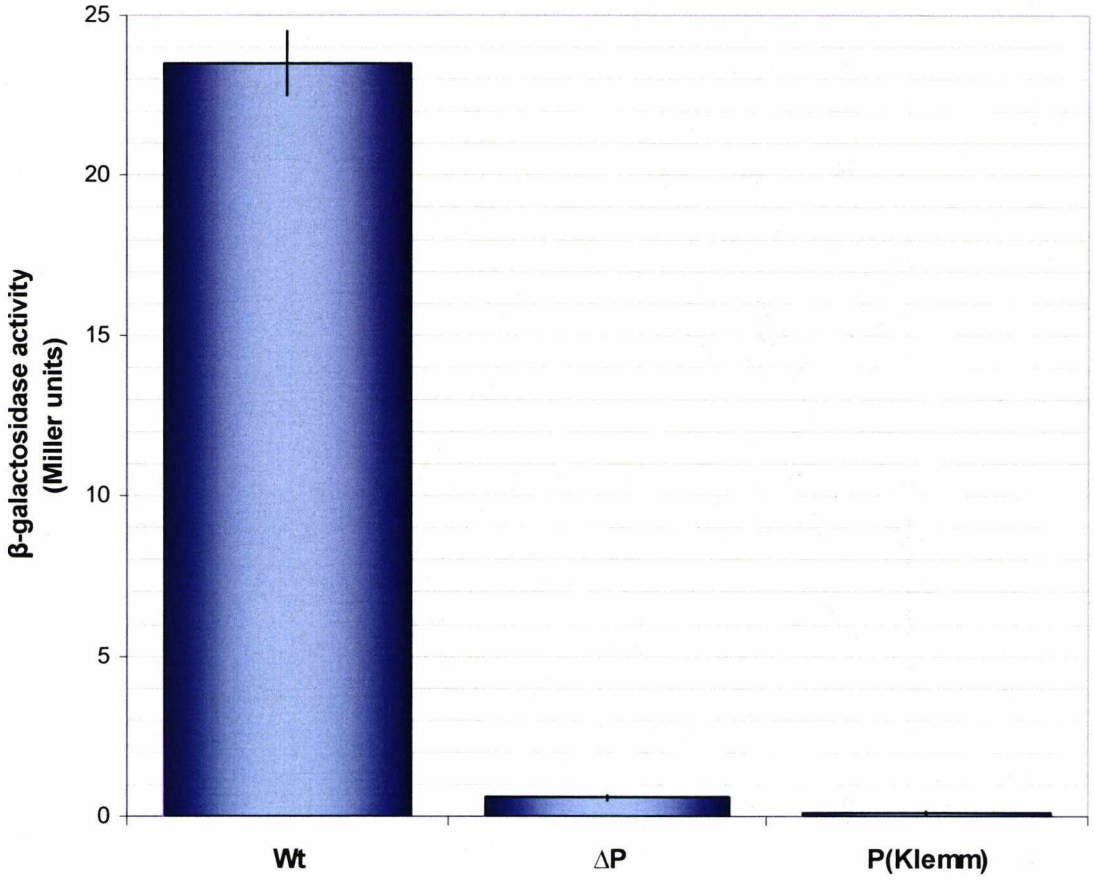


Figure 5.5: Analysis of the ΔP region (II): identification of Klemm's *fimB* promoter. The predicted -10 consensus site, of the promoter defined by Olsen & Klemm (1994) was mutated (P_{Klemm}) and its effect quantified by measuring β-galactosidase activity of a *fimB-lacZ* reporter fusion. Strains used are as follows: Wt, BGEC905; ΔP, KCEC171; P_{Klemm} (mutation), KCEC870.

5.5 Heterologous replacement of the sequence adjacent to NagC2

NanR and NagC are able activate *fimB* expression despite being separated from the *fimB* promoter by an unusually large intervening sequence (≥ 470 bp), and it seemed reasonable to suppose that additional *cis*-active elements that lie within the intervening sequence might be required for this to occur. In order to begin searching for these *cis*-active elements, the region adjacent to the NagC2 binding site (Δ3·X; figure 5.6) was replaced with heterologous sequence from within the β-lactamase (*bla*) gene of pBR322 (Δ3·X-heterologous replacement; Δ3·X-HR) (Bolivar *et al.*, 1977). In order to minimise the possibility that this mutation would reduce the affinity of NagC for the NagC2 site, the three thymine residues that are immediately downstream of it were also retained in this construct. The expression of a *fimB-lacZYA* reporter fusion was used to determine the effect of this mutation on *fimB* expression (figure 5.7).

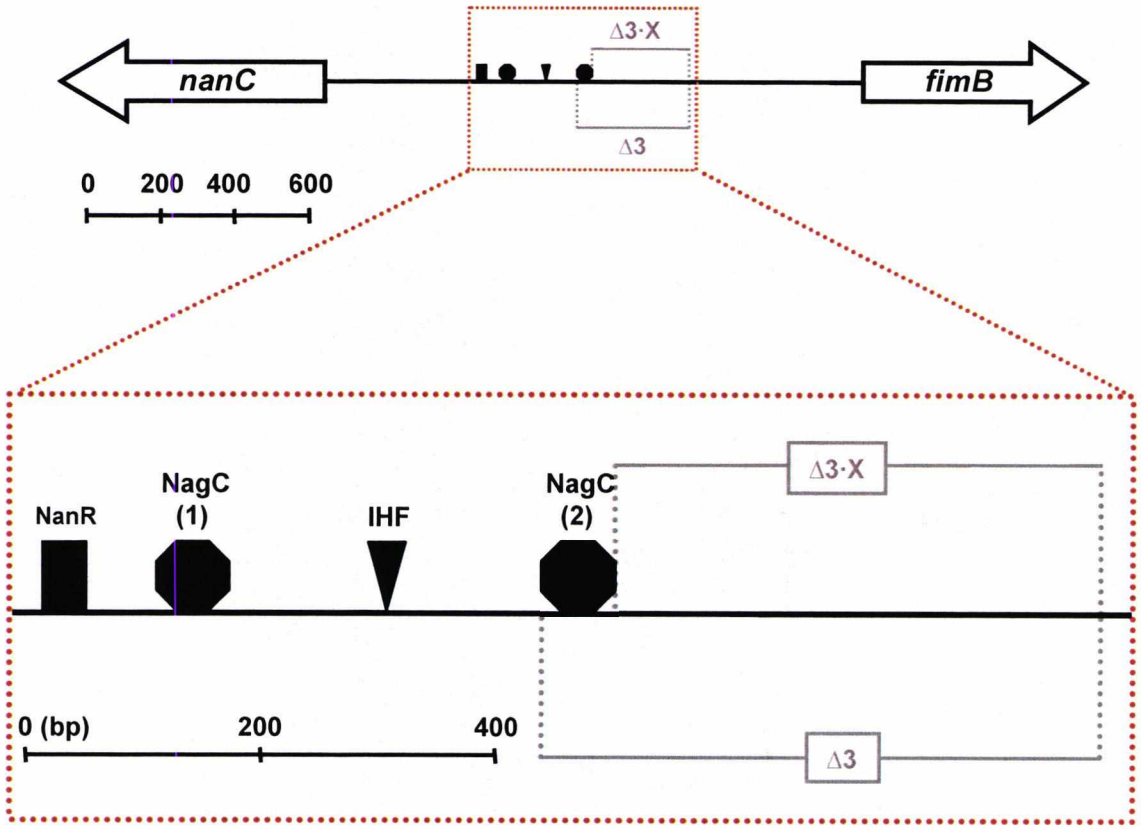


Figure 5.6: Schematic diagram of the *nanC-fimB* intergenic region, defining the Δ3·X mutation. For reference, the Δ3 mutation is also defined. The MG1655 genomic sequence, deleted by the Δ3·X mutation, is located at bp 4537788-4538063.

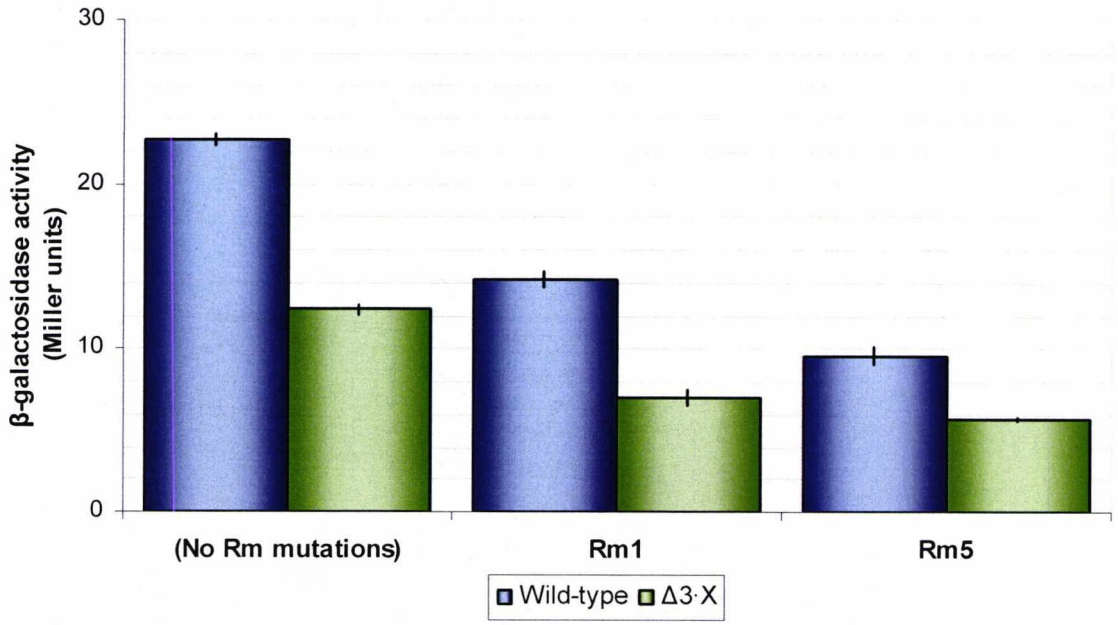


Figure 5.7: Heterologous replacement of the sequence adjacent to NagC2. β-galactosidase produced by a *fimB-lacZ* translational fusion in the presence of the Δ3·X-HR mutation (heterologous replacement of the Δ3·X region), ± the Rm1 or Rm5 mutations. Strains used are as follows (left-to-right): Wt, BGEC905; Δ3·X-HR, KCEC832; Rm1, KCEC118; Δ3·X-HR Rm1, KCEC854; Rm5, KCEC486; Δ3·X Rm5, KCEC856.

Heterologous replacement of the Δ3·X region decreased *fimB* expression 1.9-fold relative to the wild-type, which suggested that this region does in fact contain *cis*-active stimulatory sequences for *fimB*. However, when combined with replacement mutations (Rm) that remove the NanR or NagC binding sites (Rm1 or Rm5 respectively; El-Labany *et al.*, 2006), *fimB* expression remained responsive to the loss of either activator. Thus, whilst this region does appear to contain stimulatory *cis*-active elements, they do not appear to be required for NanR and NagC to function.

5.6 Deletion analysis of the Δ3·X region

In order to localise elements within the Δ3·X region that stimulate *fimB* expression, four partial deletion mutations were constructed. Each deletion begins at the Eco0109 restriction endonuclease site that denotes the *fimB*-proximal end of the Δ3·X (and Δ3) region, but terminates at different positions at the *fimB*-distal end (figure 5.8). Expression of a *fimB-lacZ* reporter fusion was used to quantify the effects of these partial deletions (figure 5.9).

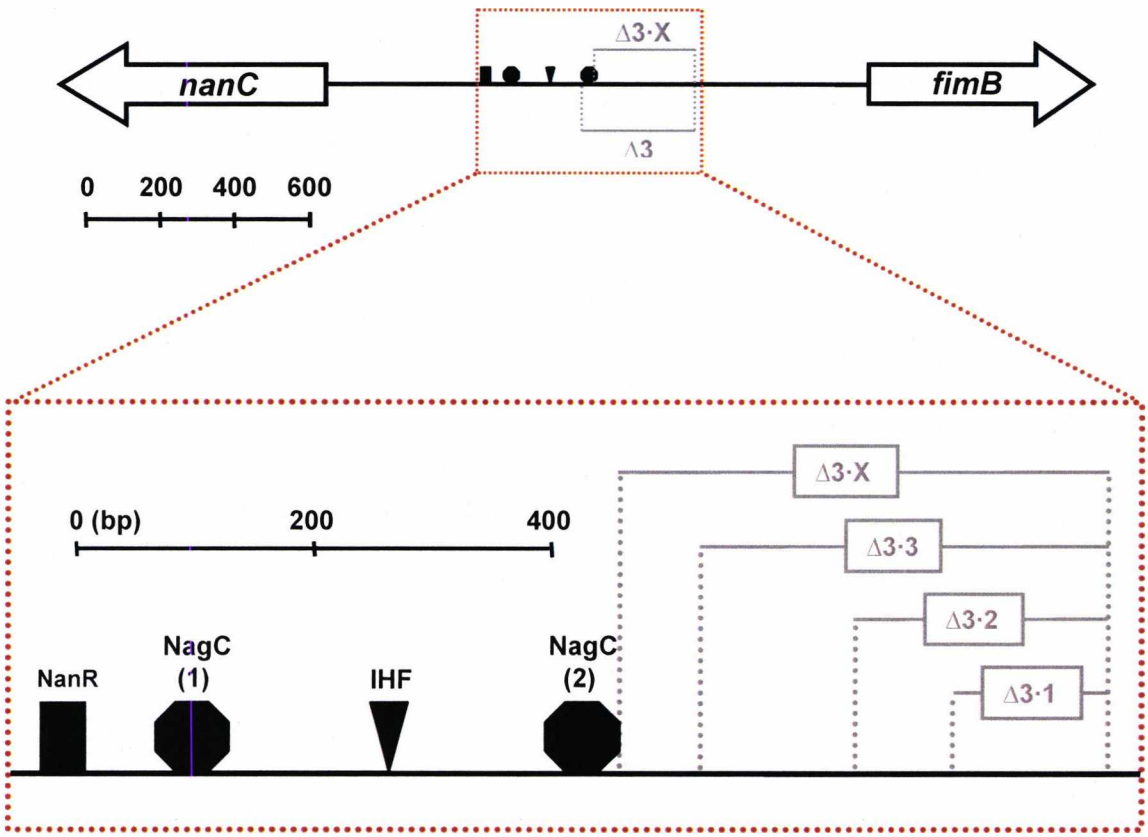


Figure 5.8: Schematic diagram of the *nanC-fimB* intergenic region, defining deletion mutations within the Δ3·X region. The MG1655 genomic positions of each deletion are as follows: Δ3·X, bp 4537788-4538063; Δ3·3, 4537834-4538063; Δ3·2, 4537921-4538063; and Δ3·1, 4537978-4538063.

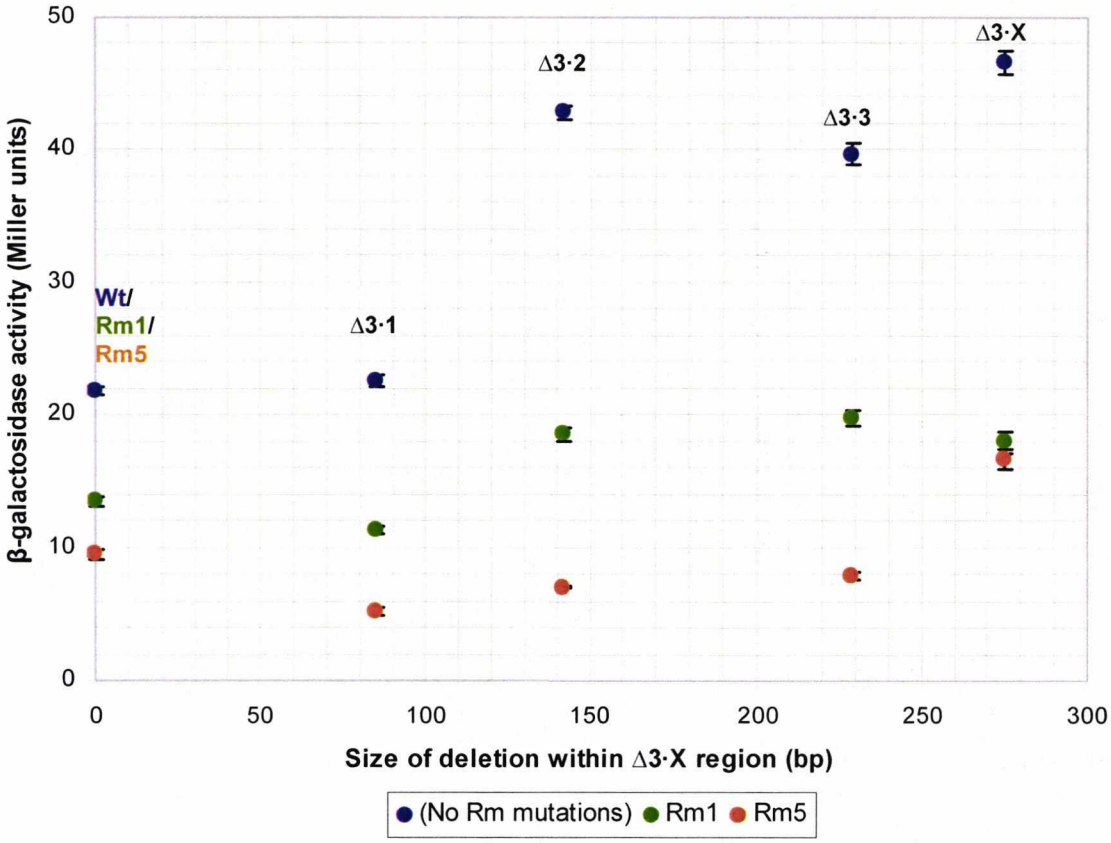


Figure 5.9: Deletion analysis of the Δ3·X region. Partial deletions, within the Δ3·X region, were combined with the Rm1 or Rm5 mutation, and their effects were quantified by measuring the β-galactosidase activity of a *fimB-lacZ* reporter fusion. Strains used are as follows: Wt, BGEC905; Rm1, KCEC118; Rm5, KCEC486; Δ3·1, KCEC773; Δ3·1 Rm1, KCEC781; Δ3·1 Rm5, KCEC783; Δ3·2, KCEC775; Δ3·2 Rm1, KCEC785; Δ3·2 Rm5, KCEC787; Δ3·3, KCEC777; Δ3·3 Rm1, KCEC789; Δ3·3 Rm5, KCEC791; Δ3·X, KCEC779; Δ3·X Rm1, KCEC793; Δ3·X Rm5, KCEC795.

Surprisingly, rather than decreasing expression (by removing stimulatory elements) each partial deletion of the Δ3·X region increased *fimB* expression relative to the wild-type; the only exception was the Δ3.1 deletion, which varied between data sets from having almost no effect (± 1-2 Miller units) to slightly inhibiting (-5 Miller units) *fimB* expression. Whilst this small (and variable) inhibition by the Δ3·1 deletion might represent the loss of *cis*-active stimulatory sequences, the effects of the other partial deletions were difficult to interpret. These data showed that efficacy of activation by NanR or NagC, as defined by *fimB* expression in the presence of the Rm1 and Rm5 mutations (respectively; El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004), is clearly affected by each deletion mutation within the Δ3·X region.

In order to simplify these data, and quantify the efficacy of NanR and NagC activation in each of these deletion backgrounds, the 'relative effect' of the Rm1 and Rm5 mutations was calculated; that is, the percentage expression lost as a result of the Rm1 or Rm5 mutations in each deletion background (Wt, Δ3·1, Δ3·2, Δ3·3 & Δ3·X; figure 5.10)

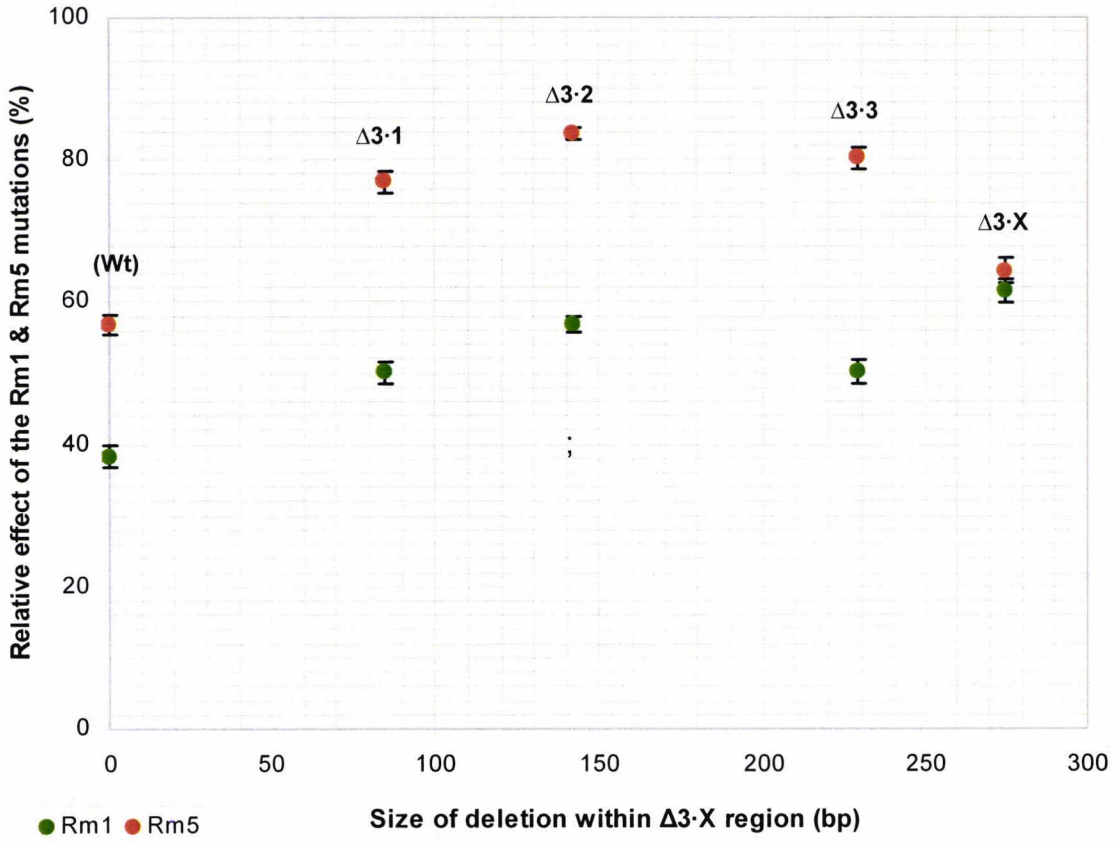


Figure 5.10: Responsiveness of Δ3·X partial deletion mutants to the Rm1 and Rm5 mutations (II). The data presented in figure 5.9 was used to calculate the relative effect of Rm1 and Rm5 mutations in each deletion background. Relative effect is given by the following equation in this instance: [100 – (mean1/ mean2) × 100] (refer to chapter 2.7.2 for details of the 95 % confidence limits of this ratio).

This suggested that the efficacy of NanR and NagC activation might increase as a function of deletion size; the simplest interpretation is that these deletions reduce the separation distance between these activators and the *fimB* promoter. this relationship is not linear, though, with the efficacy in the Δ3·3 background being lower than in the Δ3·2 and Δ3·X backgrounds. In contrast, the efficacy of NagC activation progressively increases in the Δ3·1 and Δ3·2 backgrounds, but progressively decreases

in the Δ3·3 and Δ3·X backgrounds. With exception of the Δ3·1 background, the effect of each deletion mutation appears to correlate with the modulation in efficacy of both NanR and NagC, such that increased efficacies of one or both of these activators results in a compound effect on *fimB* expression. Whereas the efficacy of NanR and NagC increases in the Δ3·1 background, relative to the wild type, the actual level of expression decreases, or does not change relative to the wild-type. This suggested that the cis-active stimulatory sequences, that function independently of NanR and NagC, might be located within the Δ3·1 region.

5.7 Partial heterologous replacements of the Δ3·X region

The relative effect of NanR and NagC suggested that deletions within the Δ3·X region largely affect *fimB* expression by altering the efficacies of these two activators as a function of distance. If this is true, we reasoned that the same relationship would also be found in heterologous replacement constructs that correspond in size to each of these partial deletions. In order to test this hypothesis, truncated fragments of the *bla* sequence were inserted into the Δ3·X deletion, which correspond to the size of deletions in the Δ3·1, Δ3·2 and Δ3·3 backgrounds. The effect of these heterologous replacement mutations (Δ3·1-HR-to-Δ3·3-HR), relative to their size-matched deletion equivalents (Δ3·1-to-Δ3·3) was measured using a *fimB-lacZ* reporter fusion (figure 5.11).

The level of expression in the two Δ3·1 deletion backgrounds (wild-type and heterologous replacement) were very similar, and notwithstanding the fact that this deletion in the wild-type sequence has a variable effect between data sets (*cf* figures 5.9 & 5.11), this also suggested that it might contain stimulatory elements for *fimB*. However, if the Δ3·X region only contains stimulatory elements, that are located within the Δ3·1 region, then the level of expression in each larger (wild-type) deletion background should also be very similar to its size-matched, heterologous-replacement equivalent. In contrast, the levels of expression in the Δ3·2 and Δ3·3 wild-type backgrounds are significantly higher than their size-matched heterologous replacement equivalents, which suggested that there might be inhibitory elements for *fimB* within the Δ3·2 region. Thus, whilst the heterologous replacement data shows that *fimB* expression is sensitive to separation distance, between distant activators and the *fimB* promoter region, deletions within the wild-type sequence are harder to interpret because they alter the separation distance and remove regulatory elements.

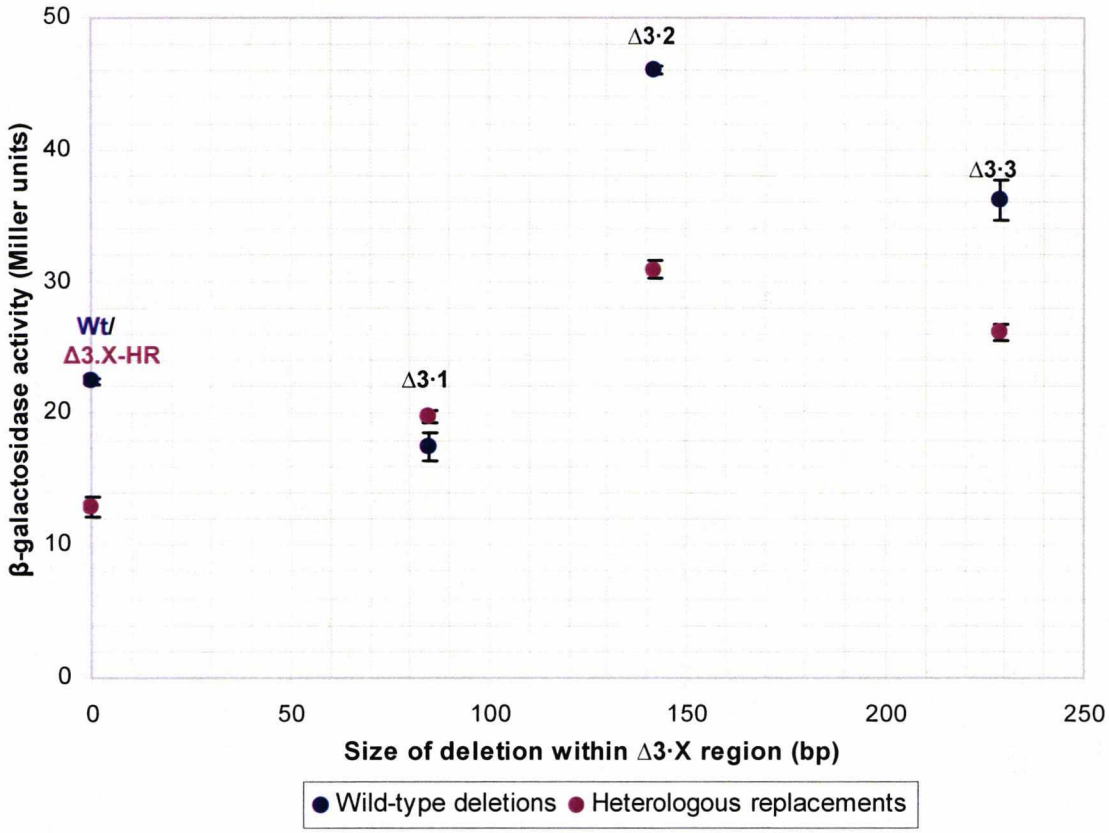


Figure 5.11: Heterologous replacement of the Δ3·X region. Size-matched equivalents of each partial deletion within the Δ3·X region were constructed using heterologous sequence from within the β-lactamase gene of pBR322 (Bolivar *et al.*, 1977). The effect of each deletion and heterologous replacement, within the Δ3·X region was calculated by measuring the expression of a *fimB-lacZ* reporter fusion. Strains used were as follows: Wt, BGEC905; Δ3·0-Heterologous Replacement (HR), KCEC832; Δ3·1, KCEC773; Δ3·1-HR, KCEC834; Δ3·2, KCEC775; Δ3·2-HR, KCEC836; Δ3·3, KCEC77; Δ3·3-HR, KCEC838.

Whilst deletions within the wild-type Δ3·X region alter the efficacies of NanR and NagC activation, it remained unclear whether this resulted from the alteration in separation distance, or from loss of additional regulatory elements. In order to discriminate between the two, heterologous replacement constructs were combined with the Rm1 or Rm5 mutations and their relative effects were calculated (figure 5.12). In this way we could determine the effect of distance upon efficacy of NanR and NagC, independently of these additional regulatory elements.

The relative effects of NanR and NagC, in each wild-type deletion construct and its size-matched, heterologous replacement equivalent were very similar, with only small differences between the two in each instance. This suggested that the efficacy of NanR and NagC in each of these partial deletions is largely determined by distance,

and whilst the differences between the two data sets (wild-type and heterologous replacements) might result from the presence or absence of these additional regulatory elements, their effects upon efficacy of NanR and NagC are very minor in comparison.

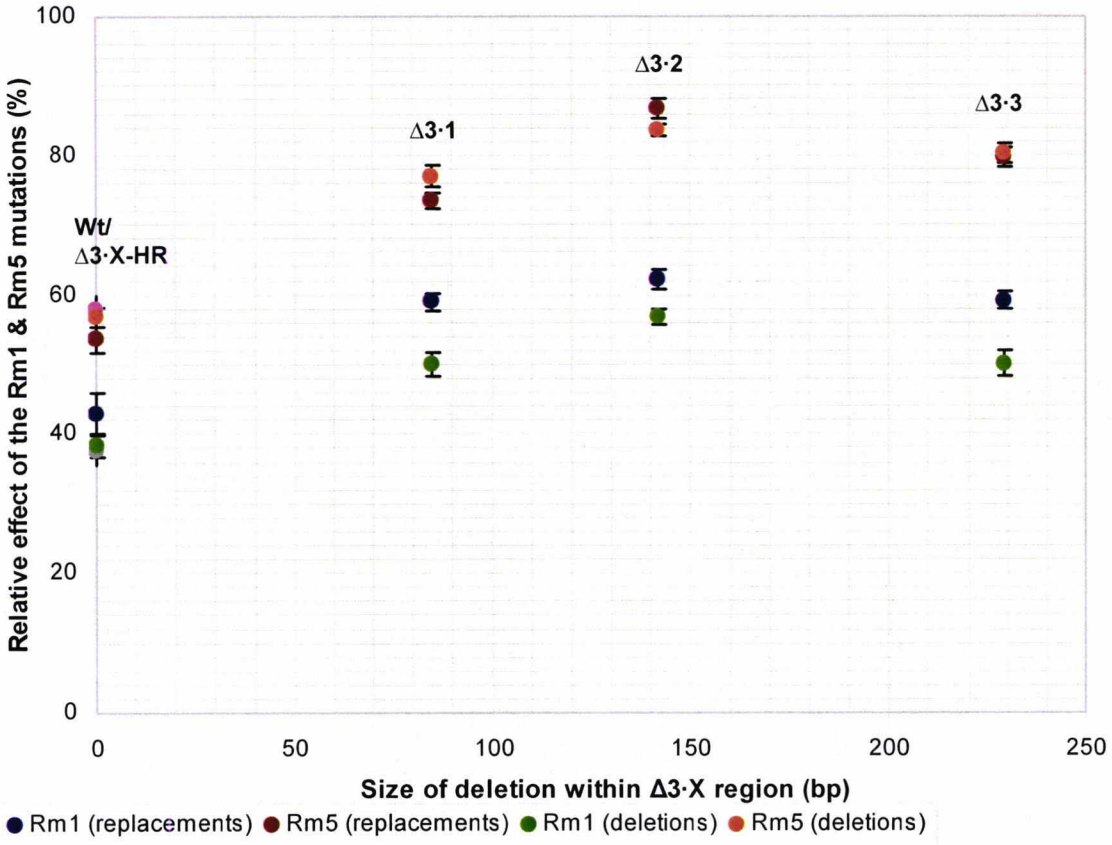


Figure 5.12: Responsiveness of the heterologous Δ3·X replacement mutants to the Rm1 and Rm5 mutations (II). The relative effects of the Rm1 and Rm5 mutations in each heterologous replacement (within the Δ3·X) background were calculated, and these were plotted against their relative effects in the partial deletion background (refer to figure 5.7 for details; raw data for these calculations is given in appendix A1). Although the relative effects of Rm1 and Rm5 in each (deletion versus heterologous replacement) background are calculated from two discrete data sets, the relative effects in the wild-type are reproduced in both; those produced alongside the heterologous replacement background are identified by grey (**Rm1**) and pink (**Rm5**) spherical markers, and are in good agreement with their values in the deletion background.

5.8 Summary

Two regulatory proteins, NanR and NagC, serve to stimulate expression of the *fimB* recombinase (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004), and a third, IHF, facilitates the latter to this end (I.C. Blomfield, personal communication). Previously in this study, the sequence upstream of the binding sites for these regulatory proteins was analysed, and no additional *cis*-active elements were identified that are required

for distant activation of *fimB* by NanR or NagC. Furthermore, the sequence that separates each regulatory protein's binding site from the next appears to serve no active role in their function. Yet these regulatory proteins, binding to sequences no less than 740 bp upstream of the *fimB* recombinase gene serve to activate its expression, and the mechanism with which they do so remained unclear.

Here, the sequence downstream of these regulatory proteins was analysed in order to better understand the mechanism of distant activation by these regulators. In order to define promoters within the ΔP region of *E. coli* K-12 strain MG1655, evidence from different *E. coli* strains was tested in order to define promoters within this region. Data obtained in this study suggested that, like the K-12 strain CSH50 (Olsen & Klemm, 1994), MG1655 possesses a single promoter within this region, its transcriptional start site being located 270 bp upstream of the *fimB* gene. This suggests that both NanR and NagC serve to activate a single promoter, despite binding to sequences that are ≥ 470 bp upstream.

The majority of the Δ3 region (El-Labany *et al.*, 2003), up to but not including the NagC2 binding site (defined here as the Δ3·X region), was also analysed for its effect upon *fimB* expression. Through the analysis of a number of partial deletion constructs, and their size-matched heterologous replacement constructs within this region, two main observations can be made. The first is that the Δ3·X region is likely to contain additional *cis*-active elements which are either stimulatory or inhibitory for *fimB* expression. Although their precise location or function was not determined in this study, there is evidence to suggest that stimulatory elements lie within an 84 bp, *fimB*-proximal fragment of the Δ3·X region (the Δ3·1 region); and inhibitory elements are within a 58 bp element of the Δ3·2 region (that is not present in the Δ3·1). There is also evidence that suggests these additional elements function independently of NanR and NagC to stimulate *fimB* expression.

The second observation is that efficacy of NanR- and NagC-mediated activation is determined by their precise location within the *nanC-fimB* intergenic region. Deletion mutations, within the Δ3·X region, have different effects upon activation by NanR and NagC. Whilst the efficacy of NanR activation increases with progressively larger deletions, NagC appears to have an optimum distance (~140 bp) after which larger deletions reduce its efficacy.

Chapter 6:

Discussion

6.1 Background

Type 1 fimbriae serve a number of functions for *E. coli*. They are a virulence factor of UPEC isolates (Bahrani-Mougeot *et al.*, 2002; Connell *et al.*, 1996; Langermann *et al.*, 1997), enhance infection of the central nervous system by *E. coli* K1 (Teng *et al.*, 2005), contribute to biofilm formation (Genevaux *et al.*, 1999; Pratt & Kolter, 1998) and are a potential virulence factor in Crohn's disease (Boudeau *et al.*, 2001). However, despite the fact that many commensal *E. coli* isolates produce type 1 fimbriae, their role remains largely unknown.

Phase variation of type 1 fimbriae is determined by a 314 bp invertible element that contains the promoter for the structural genes (Abraham *et al.*, 1985; Freitag *et al.*, 1985; Olsen & Klemm, 1994). The orientation of this invertible element is determined by two site specific recombinase proteins, FimB and FimE (Clegg *et al.*, 1985; Klemm *et al.*, 1985). FimB is the only recombinase that efficiently switches from off-to-on (afimbriate-to-fimbriate) (Gally *et al.*, 1996; McClain *et al.*, 1991; Stentebjerg-Olesen *et al.*, 2000), such that factors controlling expression of this recombinase determine, at least in part, the conditions under which type 1 fimbrial expression occurs.

The principle aim of this study was to investigate the mechanism that determines distant activation of the *fim* recombinase gene, *fimB*, by two regulatory proteins, NanR and NagC. Previous findings had shown that when *fimB* was placed at an ectopic location, within the *cynTSX-lacZYA* intergenic region, a deletion that removes NanR and NagC operator sites only inhibits FimB recombination if sequences further upstream were also present at *lac* (El-Labany *et al.*, 2003). This led to a model in which NanR and NagC might serve to antagonise one or more *cis*–active inhibitory elements that repress *fimB* expression from a large (> 1.1 kb) distance.

6.2 Summary of findings

Transposon mutagenesis was used in this study as a means to identify this distant repressor, or repressors of *fimB*. Whilst this was unsuccessful to this end, a number of transposon insertions within the *nanC-fimB* intergenic region were isolated that appeared to question the validity of this distant repressor, anti-repressor hypothesis. In order to test this hypothesis more directly, the putative repressor-binding region was deleted and its effect upon *fimB* expression determined. This showed that, in contrast to the ectopic copy of *fimB* at *lac*, expression of *fimB* at *fim* is almost unaffected by the

presence or absence of this region. Furthermore, I showed that sequences immediately upstream of *fimB* in these ectopic *lac* constructs are able to compensate for distant *fimB* activators when they are present at *fim*. This therefore strongly suggested that the distant repressor, anti-repressor hypothesis, despite being a logical interpretation of previous findings, was nonetheless incorrect.

Neither NanR nor NagC are known to activate any other *E. coli* promoters at such a distance, and we therefore considered the possibility that they might require additional *cis*-active elements closer to *fimB* for function. Previously, the 103 bp sequence downstream of the NanR binding site had been analysed using site-directed mutagenesis (El-Labany *et al.*, 2003), and in this study I continued this analysis down to the NagC2 binding site. This identified a single (13 bp) element, located between the NagC1 and NagC2 operator sequences that decreased *fimB* expression. This element contains a sequence that is in good agreement to the consensus binding site of IHF (Goodrich *et al.*, 1990), and other workers within our laboratory have subsequently shown that IHF binds here *in vitro*, where it facilitates cooperative binding of NagC to the NagC1 and NagC2 binding sites (I. C. Blomfield, personal communication).

A number of observations were also made in this study regarding the sequence that is downstream of the NagC2 binding site. Using site-directed mutagenesis, I defined a 3 bp substitution mutation that abolishes *fimB* expression and is in good agreement with a predicted *fimB* promoter element (Olsen & Klemm, 1994). In contrast to previous findings, in which two *fimB* transcriptional start sites have been mapped (Schwan *et al.*, 1994), this suggests that only a single *fimB* promoter accounts for expression of this gene under the conditions tested. In addition, the 275 bp sequence immediately downstream of the NagC2 consensus was analysed using a series of deletion and heterologous replacement constructs. Whilst this analysis suggests that there are both stimulatory and inhibitory *cis*-active elements within this region, neither NanR nor NagC require elements within this region in order to activate *fimB* expression at a distance. This analysis does, however, show that the efficacy of both NanR and NagC activation is determined by their precise locations within the *nanC-fimB* intergenic region.

6.2.1 Genome wide transposon insertions that affect *fimB* expression

Transposon mutagenesis was carried out in this study with the intention of identifying a distant repressor of *fimB* expression that two regulatory proteins, NanR and NagC, were thought to antagonise. In order to test this hypothesis, the mutagenesis protocol was carried out in various mutant backgrounds where NanR- or NagC-mediated antagonism (or both) was removed. This analysis was unsuccessful to this end, and subsequent data (also presented in this study) suggested that the model of distant repression and anti-repression was incorrect. However, transposon mutagenesis did nonetheless identify a number of insertions that affect *fimB* expression, and several observations can be made as a consequence.

One of the most striking observations that can be made is the low frequency with which insertions affecting *fimB* expression were isolated. Kleckner (1991) proposed that 10,000 colonies might need to be screened in order to isolate mTn*10* insertions within any particular gene, and the fact that so few *fim*-unlinked insertions were identified in this study might suggest that there are few factors that directly affect *fimB* expression. Indeed, subsequent analysis presented in this study suggests that there are relatively few *cis*-active elements within the *nanC-fimB* intergenic region that regulate expression of *fimB*. Alternatively, disruption of regulatory elements that control *fimB* expression might be non-permissive for growth on MacConkey agar, or otherwise essential for growth. Other workers have identified essential genes with mTn*10* insertions that are located between their promoters and the orf, thereby disrupting transcription originating from the native promoters and supplementing it with a tetracycline-inducible promoter that originates from within its boundaries (Takiff *et al.*, 1992). However, the detection of such a mutation, in this study, would depend on there being a discernable difference between native promoter activity and tetracycline-inducible (mTn*10*) promoter activity, which might not be the case.

Blomfield *et al.* (1993) used a similar method to identify mTn*10-kan* insertions that affect expression of a *fimB-lacZYA* transcriptional reporter fusion on MacConkey agar, and were able isolate insertions of interest at a much higher (approximately five-fold) frequency than encountered in this study. In order to determine whether the reporter fusion could account for this difference, I also carried out transposon mutagenesis in an isogenic Δ2 mutant (to the strain used by Blomfield *et al.*, 1993); however, this had no discernable effect in this study (data not supplied). Whilst this suggests that there

might not be many insertions that can suppress the loss of NanR- or NagC-mediated activation of *fimB* expression, it remains unclear why so few insertions were isolated in this study.

Nevertheless, transposon mutagenesis did identify two loci that do affect *fimB* expression during growth in RD_{glyc} MOPS media: *hns*, and a mTn*10* insertion between the divergently transcribed *yjhV* gene and the *fecABCDE* operon. H-NS is a well documented repressor of *fimB* expression, and other workers have shown that it binds avidly and independently to multiple elements downstream of the NanR and NagC binding sites (Donato *et al.*, 1997; Olsen & Klemm, 1994). Despite the (~ 10-fold) increase in *fimB* expression that results from mTn*10* insertions within *hns*, responsiveness to mutations that disrupt NanR or NagC activation is only slightly less than the isogenic wild-type. Thus whilst these activators might serve to antagonise H-NS repression to some extent in the wild-type (hns^+), the effect of these activators appears to be largely independent of H-NS repression. Many genes that are repressed by H-NS require activators to antagonise this repression (Atlung & Ingmer, 1997; Schroder & Wagner, 2002), and although much less common, there are a few known examples in which activators are still required in the absence of H-NS: stimulation of *virB*, which encodes a transcriptional regulator of the invasion genes of *Shigella flexneri*, by VirF (Tobe *et al.*, 1993); stimulation of *rscA*, which encodes the transcriptional activator of the capsular (colanic acid) polysaccharide synthesis genes, by RscB (Ebel & Trempy, 1999); and stimulation of the *flhDC* operon, which encodes the master regulators of the flagellar regulon, by CAP (Soutourina *et al.*, 1999). Mechanistically it remains unclear why activation of these genes is H-NS independent.

The way in which the mTn*10* insertion of strain KCEC675 affects *fimB* expression remains unknown. This mTn*10* insertion was located within the intergenic region that separates the *yjhV* gene from the *fecABCDE* operon, and is 374 bp downstream of the former and 183 bp downstream of the latter. The phenotype of this insertion is tetracycline-dependent, and transcription originating from within mTn*10* might alter the expression of adjacent genes by producing anti-sense RNA to the mRNA of either, or both, or to one of the many regulatory proteins that serve to inhibit *fimB* expression; without further analysis the activity of this mTn*10* insertion remains unclear. The *yjhV* gene encodes a 109-residue protein of unknown function. BLAST analysis (Altschul *et al.*, 1990) shows that it has homology to a putative oxidoreductase family

of unknown function, that is loosely conserved across a number of bacterial genera. The *fecABCDE* operon encodes a number of membrane proteins that facilitate uptake of exogenous ferric citrate; an outer-membrane transporter (FecA), a periplasmic, ferric-citrate binding protein (FecB), and three ABC (ATP binding cassette) superfamily proteins (FecC, FecD & FecE) (Visca *et al.*, 2002). Whilst the possibility of regulating *fimB* expression (and type 1 fimbriation) in response to a previously unknown signal is interesting, without further research it remains unclear which genes this mTn*10* insertion affects the expression of. In any case, data presented in this study shows that *fimB* expression remains equally responsive to loss of NanR and NagC in the presence of this insertion. Thus, despite potentially identifying an additional factor that affects *fimB* expression, it appears to be unrelated to distant activation by NanR or NagC.

Other workers have also previously shown that *fimB* expression is slightly elevated in *lrp* null mutants during growth on MacConkey agar (Blomfield *et al.*, 1993), consistent with the effect observed in this study (~ Wt levels of *fimB* expression in the Rm5 background). However, this mutation has no discernable effect upon *fimB* expression during growth in LB or RD_{glyc} MOPS media (I. C. Blomfield, personal communication), and as a consequence was not investigated further in this study. A number of other insertions were identified that had MacConkey agar-specific phenotypes, and for the same reason they were not analysed further in this study.

6.2.2 Artificial genetic constructs that can functionally replace NanR and NagC for activation of *fimB* expression

A number of transposon (mTn*10*) insertions were isolated within the *nanC-fimB* intergenic region that varied in their ability to suppress the Δ2 mutation both in the presence and absence of tetracycline (presented in chapter 3). Other workers have shown that transcription originating from within mTn*10* can transcribe adjacent genes (Takiff *et al.*, 1992), and this is arguably the simplest explanation that might account for the effect of these insertions upstream of *fimB*. One such insertion was located 402 bp upstream of the *fimB* orf (presented in chapter 3; strain KCEC623) and in this way might supplement the role of native *fimB* promoters. In addition, all mTn*10* insertions located within the *nanC-fimB* intergenic region in this study were orientated such that the tetracycline-inducible *tetA* promoter was directed towards *fimB*. Whilst expression of both *tetR* and *tetA* is tetracycline-inducible, transcription originating from the *tetA*

promoter is significantly greater (7- to 11-fold) than that of *tetR* (Daniels & Bertrand, 1985; Hillen *et al.*, 1984).

If transcription originating from within mTn*10* serves as a 'promoter replacement', to transcribe the *fimB* gene (FimB-LacZYA reporter fusion), then insertions that are located significantly further upstream might be expected to have a reduced capacity to this end due to the effects of polarity. Indeed, *nanC*-proximal mTn*10* insertions were significantly reduced in their ability to stimulate *fimB* expression, and whilst the link remains unproven, this correlates with an increase in untranslated mRNA (2.07 kb versus 1.07 kb).

The fact that mTn*10* insertions closer to the mid-point of the *nanC-fimB* intergenic region stimulate expression to a greater extent than a *fimB*-proximal insertion is, however, inconsistent with the effects of polarity. Furthermore, deletion of the native *fimB* promoter region (ΔP) dramatically reduces expression which is also inconsistent with the effects of polarity, as this reduces the amount of untranslated mRNA (and potential transcription termination sites) by placing mTn*10* closer to the *fimB* orf. This suggests that this region might contain anti-termination factors that are required for promoter replacement by my distal insertions, or that transcription from a location upstream and directed towards the native *fimB* promoters somehow stimulates them. Whilst other workers had previously identified a weak transcript that originated near to the mid-point of the *nanC-fimB* intergenic region (in the UPEC isolate J96) (Schwan *et al.*, 1994), I could find no evidence for such a promoter in *E. coli* K12 (MG1655). Whilst the tetracycline-dependent effects of these mTn*10* insertions were intriguing, they were not analysed any further in this study.

In addition to the tetracycline-dependent effects of transposon insertions within the *nanC-fimB* intergenic region, all but one could partially compensate for loss of *fimB* activators in a tetracycline-independent manner. Whilst it is possible that low-level transcription from the *tetA* promoter, in the absence of tetracycline, might account for this effect (Daniels & Bertrand, 1985; Hillen *et al.*, 1984), it is interesting to note that the *fimB* proximal insertion (at the S5 location) is the only one that has no phenotype in the absence of tetracycline (Chapter 4.3; figure 4.1). Not only is this inconsistent with the effects of polarity on transposon insertions at each location, it is also interesting because it lies between two fragments that H-NS has been shown to bind to

in vitro (Donato *et al.*, 1997). Other workers have shown that H-NS binds avidly to the 464 bp *fimB* promoter region, and have postulated that these interactions alone are sufficient for repression of *fimB*; the inability of this insertion to antagonise H-NS repression also suggests that this might be correct. Consistent with the tetracycline-dependent effect of a transposon insertion located at the mid-point of the *nanC-fimB* intergenic region; in the absence of tetracycline this same insertion was entirely dependent upon the native *fimB* promoter region for function.

Previous data had suggested that activating elements within the Δ2 region were only required for *fimB* expression when sequences closer to *nanC* were present in *cis* (El-Labany *et al.*, 2003). These data was obtained using merodiploid constructs, in which fragments of the *nanC-fimB* intergenic region (and the *fimB* gene) had been moved to the *lac* locus. Several lines of evidence, presented previously in this study, questioned the validity of this hypothesis: the inability to identify such a repressor (Chapter 3); the inability of transposon insertions, between the putative repressor-binding region and the *fimB* promoter region, to suppress this repression (Chapter 4.3); and that deletion of the putative *cis*-active sequences, required for distant repression, had little effect on *fimB* expression or dependency on distant activators (NanR & NagC) (Chapter 4.5). In order to explain the results obtained in the merodiploid constructs, I showed that post-*lac* sequence and *fimB* activators can synergistically activate *fimB* expression when it is located immediately upstream, within the *nanC-fimB* intergenic region, and that it can substitute for them in their absence (Δ2 mutation) (Chapter 4.6). In contrast to transposon insertions within the *nanC-fimB* intergenic region, the post-*lac* sequence contains no known promoters and was separated from the *fim* sequence by four tandem repeats of the *rrnB* transcriptional terminator region (El-Labany *et al.*, 2003). Like these transposon insertions, the effect of post-*lac* sequence was also entirely dependent on the native *fimB* promoter region.

The effect of either mTn*10* insertions or post-*lac* sequence on *fimB* expression remains unknown. Tetracycline resistance of mTn*10* results from the expression of two divergent genes, which other workers have shown can serve as promoter replacements for adjacent genes (Takiff *et al.*, 1992); in addition, IHF binds to the IS*10* fragment that flanks mTn*10* at both ends (Huisman *et al.*, 1989), and TetR regulatory protein binds to operator sites that overlap the *tetR* and *tetA* promoters (Hillen *et al.*, 1984; Meier *et al.*, 1988). For these reasons mTn*10* insertions could potentially affect *fimB*

expression in two ways; at least partially as a promoter replacement, and to alter the topology of the surrounding DNA. In contrast, the post-*lac* sequence contains four tandem repeats of the *rrnB* transcriptional terminator region (El-Labany *et al.*, 2003), the *cyn-lac* intergenic region and a truncated fragment of the *cynX* gene; none of which are known to bind regulatory proteins, or contain a promoter. Whilst the mechanism by which both transposon insertions and post-*lac* sequence activate *fimB* expression remains unclear, they do show that expression of this gene is sensitive to a number of different elements when they are situated within the *nanC-fimB* intergenic region, at a distance from the *fimB* promoter region.

Irrespective of its mechanism of activation, the effect of post-*lac* sequence on *fimB* expression was an important control experiment, which showed that our original hypothesis, that NanR and NagC antagonise distant inhibitory elements (El-Labany *et al.*, 2003), was almost certainly incorrect. With a series of merodiploid constructs, we had previously shown that an ectopic copy of *fimB* only requires activating elements, within the Δ2 region (NanR, NagC1 & IHF), if sequences closer to *nanC* were present in *cis* (El-Labany *et al.*, 2003). These data suggest that activating elements are no longer required because sequences brought into juxtaposition, the post-*lac* sequence, can compensate for them in their absence. Rather than showing that activating elements are only required when sequences closer to *nanC* are present in *cis*, these data suggests that the post-*lac* sequence can no longer activate *fimB* expression when located 819 bp upstream of these activating elements.

6.2.3 Replacement mutagenesis of the Δ4-NagC2 region

At other loci, both NanR and NagC require promoter-proximal binding sites in order to regulate gene expression, and yet their respective positions within the *nanC-fimB* intergenic region are far (≥ 470 bp) upstream of the *fimB* promoter. It therefore seemed logical that they might require *cis*-active elements, within the intervening sequence in order to stimulate *fimB* expression. Previously, a partial fragment of the Δ2 region (Δ5) had been analysed for *cis*-active elements by constructing a series of small (12-14 bp) heterologous replacement mutations (El-Labany *et al.*, 2003); this identified the NanR (Rm1 & Rm2) and NagC1 (Rm5 & Rm6) binding sites. In this study, these heterologous replacement mutations were extended to encompass the remainder of the Δ2 region, the Δ3 region up to the NagC2 site, and the (15 bp) sequence that separates the two. This identified a single replacement mutation (Rm13)

that reduces *fimB* expression 1.6-fold (Chapter 4.9), accounts for the phenotype of the Δ4 mutation and contains a sequence that is in good agreement with the IHF binding site consensus (Goodrich *et al.*, 1990).

Other workers within the laboratory have subsequently analysed the role of IHF at this site (I. C. Blomfield, manuscript submitted for publication). Mutations that prevent IHF and NanR from binding to their consensus sites, within the *nanC-fimB* intergenic region have a compound effect upon *fimB* expression in the wild type (I. C. Blomfield, manuscript submitted for publication). This effect of IHF in the wild-type is consistent with the observation that it is required for the cooperative binding of NagC to the NagC1 and NagC2 sites *in vitro* (I. C. Blomfield, manuscript submitted for publication), and suggests that it serves little or no role in NanR-mediated activation. However, in the Δ3 background NanR is entirely dependent upon IHF in order to activate *fimB*, and mutations that prevent either from binding almost entirely abolish expression.

IHF binding is known to introduce a sharp (>160 °) bend in DNA (Lorenz *et al.*, 1999; Rice *et al.*, 1996; Teter *et al.*, 2000), and DNA looping facilitates interactions between regulators and RNAP when they are bound at distant DNA sites (Gralla, 1991; Rippe *et al.*, 1995; Schleif, 1992; Vilar & Leibler, 2003; Vilar & Saiz, 2005). This dependency upon IHF suggests that NanR might contact the RNAP in order to stimulate *fimB* expression, which is consistent with the mechanisms of activation by many other regulators, including at least one other FadR-subfamily member (GlcC) (Campbell & Cronan, 2001; Collado-Vides *et al.*, 1991; Gui *et al.*, 1996; Henry & Cronan, 1992; Hochschild & Dove, 1998; Pellicer *et al.*, 1999). It suggests that, in the presence of the Δ3 deletion the intervening DNA is sufficiently inflexible that IHF is required for loop formation in order to localise NanR near to the promoter region. In contrast, the intrinsic flexibility of DNA over greater distances, such as the wild type *nanC-fimB* intergenic region, facilitates loop formation independently of such DNA-bending proteins (Figure 6.1) (Bellomy & Record, 1990).

It is interesting to note that NanR becomes dependent upon IHF in the Δ3 background, as the intervening sequence is still relatively large (452 bp); the activity of DNA-bending proteins typically becomes essential for looping of much shorter sequences (<150 bp), where the DNA is relatively inflexible both torsionally and laterally (Perez-

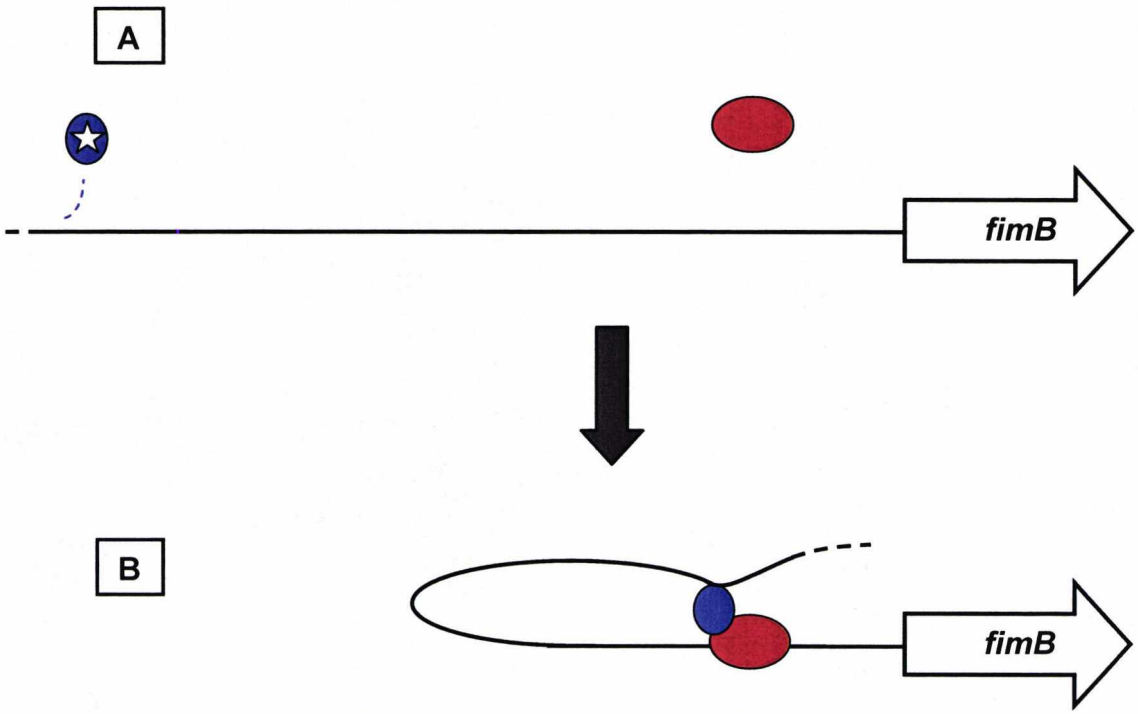


Figure 6.1 : Schematic diagram defining the model of NanR activation. In the presence of Neu5Ac, NanR (blue sphere) is unable to bind DNA, and as consequence RNAP (red sphere) alone cannot form closed or open complex formation (A). In the absence of Neu5Ac, NanR binds upstream of *fimB* and, via DNA looping that is faciliated by the large intervening sequence, contacts RNAP to stimulate closed or open complex formation (B).

Martin *et al.*, 1994; Saiz *et al.*, 2005; Vilar & Saiz, 2005; Wang & Giaever, 1988). Furthermore, the predicted centre of the IHF-induced bend does not, in this instance, localise NanR at the *fimB* promoter region (Rice *et al.*, 1996); instead it would localise approximately -150 bp (relative to the transcriptional start site defined by Aberg *et al,* 2006; Figure 6.XB). This suggests that there might be additional factors that determine IHF dependency in the Δ3 background. Other workers have shown that DNA supercoiling and DNA curvature can affect interactions between proteins bound to distant DNA sites (Blumberg *et al.*, 2005; Goodman & Nash, 1989; Liu *et al.*, 2001; Merlitz *et al.*, 1998; Rippe *et al.*, 1995). Whilst *fimB* expression does not appear to be significantly affected by DNA supercoiling (Dove & Dorman, 1994), the *nanC-fimB* intergenic region is relatively AT-rich (64 %) and contains numerous A_nT_m tracts ($n+m > 4$) that can affect the conventional B-DNA structure of DNA, and increase rigidity (Beveridge *et al.*, 2004; Carrera & Azorin, 1994; Dlakic & Harrington, 1996; Hagerman, 1990; Hagerman, 1992; Koo *et al.*, 1986; McConnell & Beveridge, 2001). Furthermore, H-NS binds to AT-rich and/or intrinsically curved sequences (Lucht *et*

al., 1994; Rimsky *et al.*, 2001; Tanaka *et al.*, 1991; Yamada *et al.*, 1991), and exhibits high affinity for multiple *fimB*-proximal fragments of the *nanC-fimB* intergenic region (Donato *et al.*, 1997). These suggest that the DNA sequence within this region might be both curved and inflexible, and for this reason might require a DNA-bending protein, such as IHF, to constrain the local DNA and facilitate NanR-RNAP contacts in the Δ3 background.

6.2.4 Analysis of *fimB* promoters

In chapter 5, I showed that a 3 bp substitution mutation within the ΔP region completely abolishes *fimB* expression during growth in RD_{glyc} MOPS media; this substitution mutation (5' TATATA to 5' GGTATG) alters the putative -10 hexamer of the promoter that had previously been identified by Olsen and Klemm (1994). Other workers within the our laboratory have subsequently shown that alteration of the putative -35 hexamer of this promoter (5' TTGTTA to 5' GGATTA) also abolishes *fimB* expression (I. C. Blomfield, unpublished data), and have recently mapped the transcriptional start site to 2 bp upstream of the site identified by Olsen & Klemm (1994). Taken together, these data show that *fimB* expression is determined by a single promoter during growth in RD_{glyc} MOPS media. Thus both NanR and NagC stimulate transcription from the same promoter, despite binding to sequences far upstream (-744, -692 & -470 bp for NanR, NagC1 & NagC2 sites respectively).

The sequence of this *fimB* promoter is consistent with the observation of other workers, who have shown that *fimB* expression *in vitro* requires the house-keeping sigma factor, σ^{70} (Donato *et al.*, 1997; Schwan *et al.*, 1994) (figure 6.2). The 17 bp spacer between the -10 and -35 hexamers is characteristic of promoters recognised by RNA polymerase containing the σ^{70} subunit (Harley & Reynolds, 1987; Hawley & McClure, 1983; Kirkegaard *et al.*, 1983; Mulligan *et al.*, 1985; Siebenlist *et al.*, 1980; Stefano & Gralla, 1982; Warne & deHaseth, 1993), and both the -10 and -35 hexamers deviate from the *E. coli* consensus at only two positions (Harley & Reynolds, 1987; Hawley & McClure, 1983; Lisser & Margalit, 1993). These suboptimal promoter hexamers are consistent with the observations made both here and in other studies that suggest *fimB* expression results from moderate strength promoter activity, even in the absence of H-NS (Donato *et al.*, 1997; El-Labany *et al.*, 2003; Schwan *et al.*, 1994; Sohanpal *et al.*, 2004).

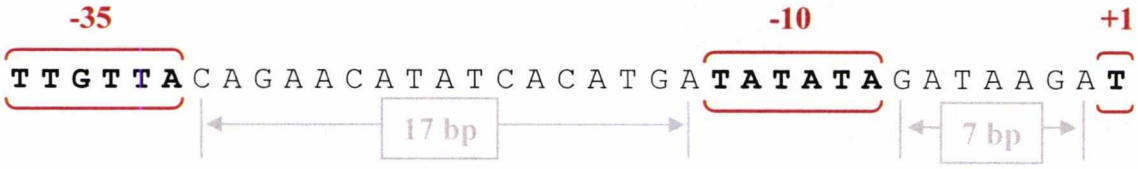


Figure 6.2: The (P2) promoter region of *fimB*. The -35 and -10 hexamers, as defined in this study are highlighted, as is the transcriptional start site ('+1', as defined by Aberg *et al.*, 2006).

In each instance, other workers have used primer extension protocols in order to identify the transcriptional start sites for *fimB* (Aberg *et al.*, 2006; Olsen & Klemm, 1994; Schwan *et al.*, 1994), and this method can erroneously predict transcriptional start sites as a result of premature termination of cDNA synthesis (Boorstein & Craig, 1989; Sambrook, 1989). This might therefore explain the subtle (2 bp) differences in start sites that were predicted by Olsen and Klemm (1994) and Aberg *et al.* (2006) for P2, and for those predicted by Schwan *et al.* (1994) and Aberg *et al.* (2006) for P1 (3 bp). However, it remains unclear why the transcriptional start site identified by Schwan *et al.* (1994), for P2, is 20 bp further upstream of the site identified by Aberg *et al.* (2006); this cannot be accounted for by differences in DNA sequence of the *E. coli* strains that were analysed, and is inconsistent with the data presented in this study. Incidentally, mutations that correspond to the -10 hexamers of the two promoters identified by Schwan *et al.* (1994) were made in this study, and they did inhibit *fimB* expression to some extent (Chapter 5). One of these mutations is -39 bp, relative to the promoter that was identified in this study and might affect the activity of this promoter. It remains unclear, however, what effect the other mutation might be having, as it is +103 bp (relative to the promoter identified in this study) and the P1 promoter appears to play no role in *fimB* expression under the conditions tested in this study. Nevertheless there is supporting evidence to suggest that the promoter identified in this study is correct for two commensal *E. coli* K12 strains (MG1655 & CSH50), and that promoter activity at this site predominates during growth in rich media (LB or RD_{glyc} MOPS) (Aberg *et al.*, 2006; Olsen & Klemm, 1994).

6.2.5 Analysis of the Δ3·X region

In chapter 5 the Δ3 region, up to but not including the NagC2 consensus binding site (defined as the Δ3·X region in this study) was analysed in order to determine its role in NanR- and NagC-mediated activation of fimB expression. This suggested the Δ3·X

region might contain multiple cis-active elements, that either stimulate or inhibit *fimB* expression, but these are not required for activation by NanR or NagC; the location of these additional elements, and their function was not further characterised in this study. However, the combined analysis of partial deletions, and their size-matched heterologous replacement equivalents within this region showed that the efficacy of activation by both NanR and NagC is determined by their precise locations within the *nanC-fimB* intergenic region (for details of heterologous replacement mutations within this region, refer to chapters 2.6.4.2 & 5.6). With the exception of the Δ3·3 background, progressively larger deletions within this region increase the level *fimB* expression, and in each background the precise level of expression appears to be determined by the combined efficacies of NanR and NagC.

Research carried out by other workers within our laboratory, and presented previously in this chapter suggests that NanR might activate *fimB* by contacting the promoter region, in a DNA looping-dependent fashion (chapter 6.2.3); whereas NanR does not require IHF in order to activate *fimB* expression in the wild-type, it becomes entirely dependent upon it in the presence of the Δ3 mutation (a 304 bp deletion between the IHF binding site and the *fimB* promoter; El-Labany *et al.*, 2003). It is interesting to note that the efficacy of activation by NanR, as a result of partial deletions within the same region exhibits a periodicity, whereby it progressively increases in the Δ3·1, Δ3·2 and Δ3·X backgrounds (85 bp, 142 bp & 275 bp deletions, respectively), but is slightly reduced in the Δ3·3 background (229 bp deletion). It remains unclear why activation by NanR follows this pattern, but two hypotheses might account for it: alterations to the correct alignment of NanR and IHF relative to promoter region; or the transition from an IHF-independent to IHF-dependent phenotype. DNA-bending proteins, such as IHF are known to form loops with characteristic stereospecificity (Perez-Martin *et al.*, 1994), and small alterations in the precise location of many distant regulators often results in a periodicity in their activation or repression that reflects the number of helical turns of the intervening DNA (Dunn *et al.*, 1984; Lee & Schleif, 1989; Newlands *et al.*, 1992; Stauffer & Stauffer, 1998; Ushida & Aiba, 1990; Zacharias *et al.*, 1992); that is, the regulator has to be present on the same helical 'face' of the DNA as RNAP in order for contact between the two to occur, with sequence deletions or additions that are integral multiples of the DNA helical turn permitting this interaction and half-integral multiples preventing it. If you assume that the helical repeat of this particular sequence is ~10.4 bp per helical turn (Wang, 1979),

then the Δ3·1, Δ3·2, Δ3·3 and Δ3.X mutations remove 8.2, 13.7, 22.0 and 26.4 helical turns respectively. If you assume that NanR is largely IHF-independent in the Δ3·1 background, and is IHF-dependent in the Δ3·2, Δ3·3 and Δ3·X backgrounds, then these data tentatively suggest that half-integral, rather than integral reductions in sequence are required for NanR to activate efficiently in these backgrounds. In order to confirm whether this might account for the effect of the Δ3·3 mutation on *fimB* expression, however, further work would be required; determining the dependency of NanR on IHF in each of these backgrounds, and multiple half- and quarter-integral reductions and additions in sequence that would test this hypothesis further. Alternatively, we can assume that the transition from an IHF-independent to -dependent phenotype occurs at some point within this series of deletions, and this might account for the reduced efficacy of NanR in the Δ3·3 background. Essentially, in this background the intervening sequence between NanR and the promoter region might be sufficiently inflexible for IHF-independent interactions to occur, yet at the same time IHF is unable to localise NanR close enough to the promoter region. An IHF-induced bend would localise NanR at a position -225 bp, versus -150 bp in the Δ3·3 and Δ3 backgrounds (respectively). Again, without further analysis this hypothesis also cannot be proven.

The mechanism of NagC-mediated activation remains unclear, as the interaction of dimers bound at two sites (NagC1 & NagC2), via IHF-dependent looping of the intervening DNA would seemingly tether both dimers at a promoter distal location (I. C. Blomfield, manuscript submitted for publication); -470 bp, relative to the transcriptional start site identified by Aberg *et al.* (2006). However, activation of *glmUS* by NagC requires promoter-proximal binding (-47 bp to the P1 promoter of *glmUS*) that is suggestive of direct RNAP contact (Plumbridge, 1995). Furthermore, it is interesting to note that in contrast to activation by NanR, whose efficacy mostly increases with each reduction in separation distance, NagC appears to have an optimum distance (defined by the Δ3·2 mutation) after which further reductions inhibit activation. Whilst NanR presumably becomes dependent upon IHF for activation in one of these partial deletion backgrounds, NagC has no DNA bending protein bound within the intervening sequence to facilitate loop formation and promoter contact. This therefore suggests that NagC might also require the inherent flexibility that the wild type separation distance affords for activation, and is unable to contact the *fimB* promoter region when it is reduced.

This mechanism of NagC activity remains highly speculative without further research being conducted, especially as other data presented in this study shows that *fimB* expression is sensitive to a number of genetic elements when situated at similar distances, and which do not clearly facilitate stimulatory protein-protein contacts with RNAP (Chapters 4.3-7, & 6.2.2). In contrast, other workers have shown that IHF activates expression of the *ilvGMEDA* operon of *E. coli* from a promoter distal site (centred at -92 bp), without making direct promoter contact (Pagel *et al.*, 1992; Parekh & Hatfield, 1996). Activation in this instance occurs as a result of an AT-rich upstream element in which the DNA duplex is destabilized by negative superhelical energy (Sheridan *et al.*, 1998); IHF serves to stabilise this element, transferring the negative superhelical energy downstream to the promoter region to enhance open complex formation. A similar DNA structural transmission mechanism also regulates expression of the *leuV* operon of *E. coli*, which has an upstream, stress induced duplex destabilized (SIDD) element that is stabilized by another transcriptional regulator, FIS (Opel *et al.*, 2004). SIDD elements are also found at genetic loci that require frequent duplex destabilization for function, such as promoter regions in *E. coli* (Benham & Bi, 2004; Wang *et al.*, 2004), and bacterial, viral and fungal origins of replication (Ak & Benham, 2005; Kowalski & Eddy, 1989; Polonskaya *et al.*, 2004).

Intriguingly, the WebSIDD algorithm predicts a SIDD element within the *nanC-fimB* intergenic region, that would require very little free energy to destabilize at physiologically relevant levels of DNA supercoiling (σ = -0.015) (Figure 6.3; Bi & Benham, 2004). More specifically, this predicted SIDD element is immediately downstream of the IHF binding site, such that IHF might transfer negative superhelical energy towards *fimB* (figure 6.4) (Opel *et al.*, 2004; Sheridan *et al.*, 1998). Furthermore, WebSIDD also predicts a SIDD element within the mTn*10* sequence and to a lesser extent the post-*lac* sequence.

Whilst the possibility of there being an SIDD element upstream of *fimB* is interesting, further work is again required in order to confirm its presence and activity. Structural transmission mechanisms have been shown to act over comparable distances in plasmid constructs (Spirito & Bossi, 1996), but it remains unclear whether it would effect *fimB* expression at such a distance within chromosomal DNA. Furthermore, as IHF or FIS alone are able to stabilise SIDD complexes at other loci (Opel *et al.*, 2004; Sheridan *et al.*, 1998), it remains unclear why NagC would be required for

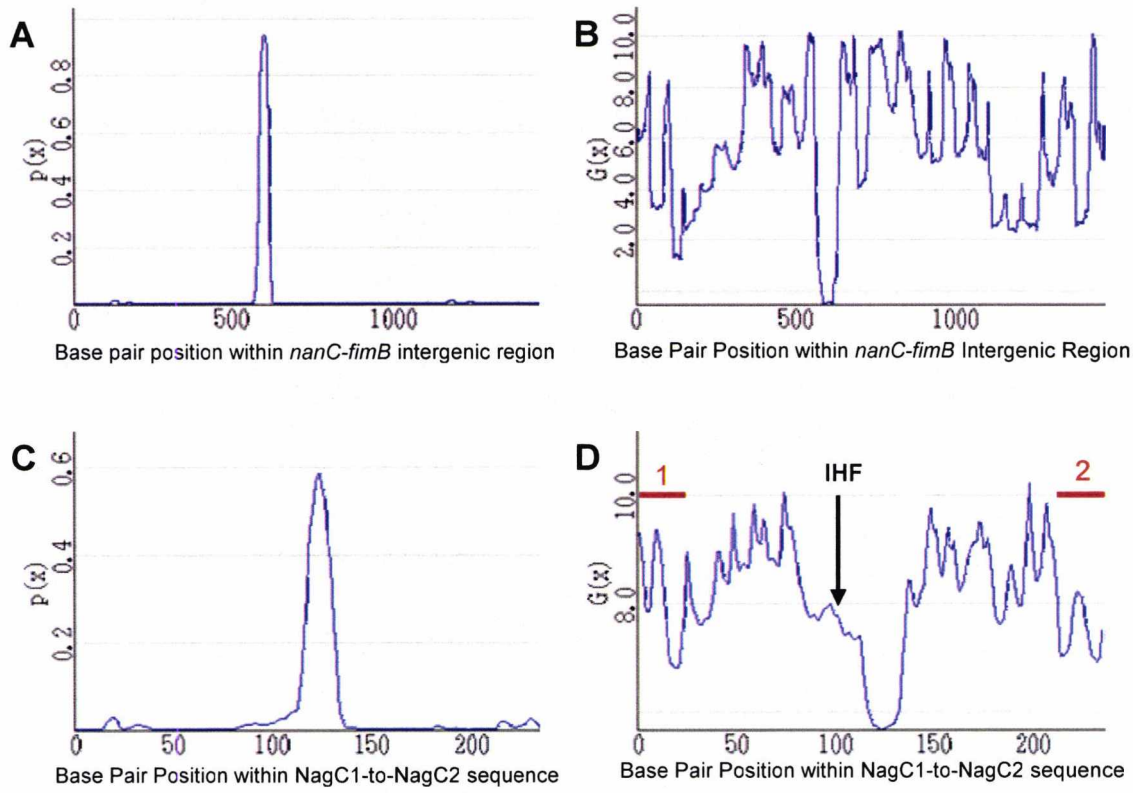
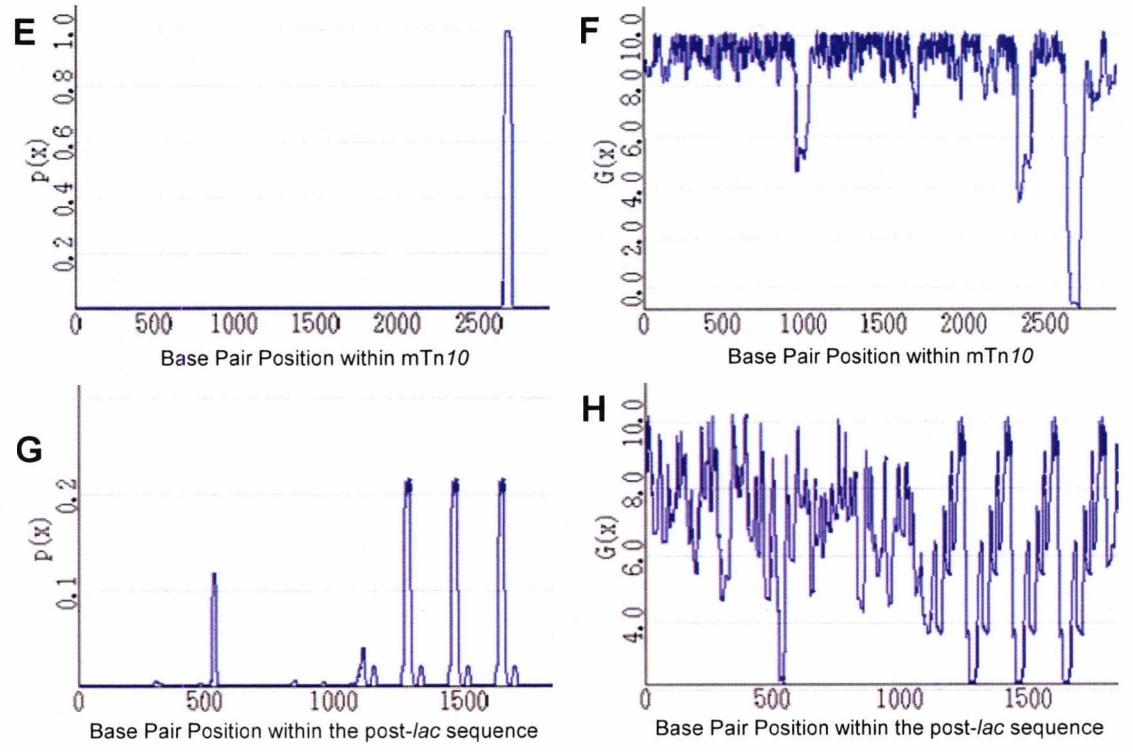


Figure 6.3: SIDD profiles of the *nanC-fimB* intergenic region. The probability of denaturation, p(x), for the entire *nanC-fimB* intergenic region (A), the partial sequence between the two NagC sites (NagC1 & NagC2) (C), mTn*10* (E), and the post-*lac* sequence (G), are defined; and the predicted free energy required for DNA duplex destabilization, ΔG(x), for each of these sequences (B, D, F & H, respectively) are defined. The NagC1 (1) and NagC2 (2) binding sites are defined, as is the predicted centre of the IHF bend, in chart D. SIDD calculations were made using the WebSIDD server (Bi & Benham, 2004), with the following variables: linear DNA type; 310.0 Kelvin; 0.01 M NaCl; σ = -0.055; 12.0 kcal threshold energy.



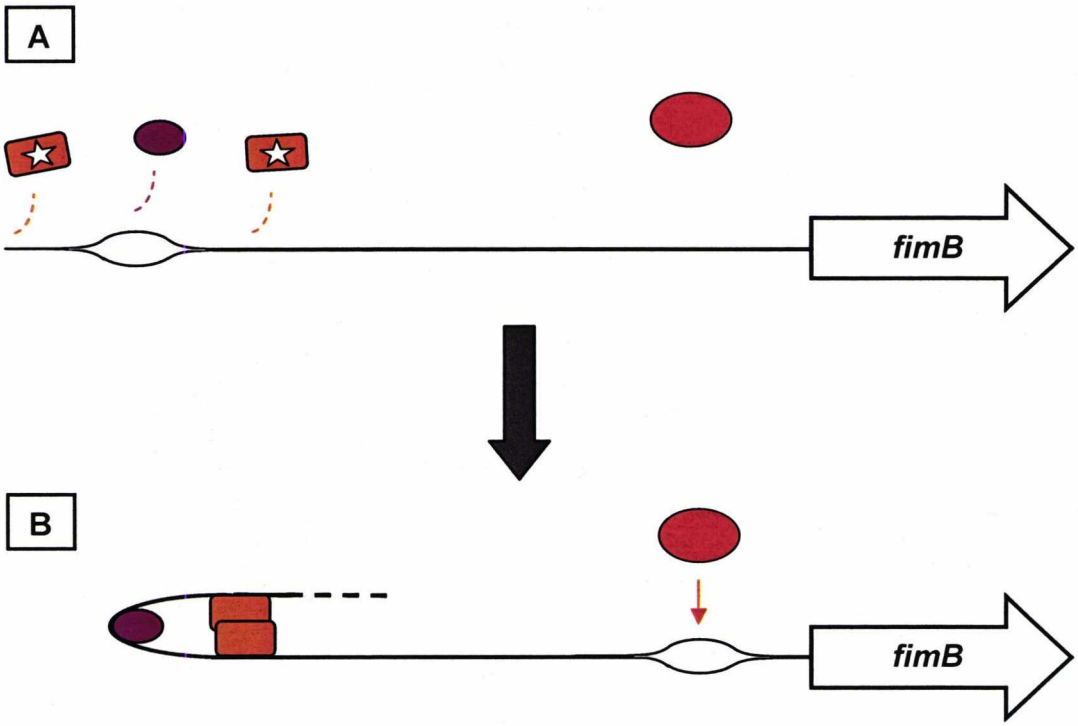


Figure 6.4:Schematic diagram defining the structural transmission model of NagC activation. In the presence of GlcNAc, NagC (orange rectangle) is unable to bind DNA and presumably IHF (purple sphere) alone is insufficient to prevent the upstream SIDD element from destabilising; this imposes torsional constraints upon the downstream promoter, increasing the energy required for open complex formation and inhibiting initiation of transcription (RNAP defined by red sphere) (A). In the absence of GlcNAc, IHF facilitates loop formation for the two NagC sites to interact, thereby stabilising the upstream SIDD element, transferring negative superhelical energy towards the promoter and lowering the energy required for open complex formation to occur (B).

stabilisation at this locus, or whether this effect might in fact be independent of NagC. In addition, small, heterologous replacement mutations, made within this study (Rm11-Rm16) lie within this predicted SIDD element, yet only loss of the IHF binding site (Rm13) decreased expression (refer to chapter 4.9 for more details). Whilst these replacement mutations would alter the sequence of this predicted SIDD element, each replacement mutation was very small (12-14 bp); with exception of the CG-rich SacII site within each of these mutations, the remainder of each mutation consisted of mostly A-to-T or T-to-A transitions which would not be expected to decrease the SIDD potential of this region. This might, therefore, explain why these mutations did not affect expression even though they are within this potential SIDD element. Replacement of this entire SIDD element, as other workers have done in the identification of SIDD elements at other loci (Opel *et al.*, 2004; Sheridan *et al.*, 1998), would therefore be a logical experiment to test this hypothesis.

6.3 Models for NanR- and NagC-mediated activation of *fimB* expression

I propose that NanR activates *fimB* expression at a distance by making direct promoter contact. The dependency of NanR upon IHF in the Δ3 background suggests that DNA looping is required for it to function, and I have shown that altering the distance that separates it from the *fimB* promoter modulates its efficacy. I propose that NanR is able to make direct promoter contact in the wild-type, independently of IHF, because the larger intervening sequence in this background is more flexible. Whilst NanR is not recognised as a transcriptional activator for any other *E. coli* genes, other members of the FadR-subfamily are. In each instance, members of this family appear to activate transcription via direct promoter contact, and the sequence homology between members of this family suggests that NanR might also retain this capacity.

I propose that NagC activates *fimB* transcription by direct promoter contact. In contrast to NanR, whose efficacy increased when brought closer to the *fimB* promoter, NagC appeared to depend upon a more promoter-distal location to efficiently activate *fimB*. I propose that NagC requires the inherent flexibility of the wild-type intervening sequence to facilitate direct promoter contact, and its efficacy decreases when this intervening sequence is reduced because this flexibility, to some extent, is lost. NagC is able to activate the *glmUS* operon (Plumbridge, 1995), but occupies a site immediately upstream of RNAP to do so, at a location that is suggestive of direct RNAP contact. This model remains highly speculative, however, especially as there might be a SIDD element between the two NagC sites. Whilst there is a possibility that NagC might activate *fimB* via a structural transmission mechanism, the requirement of NagC for a promoter-distal location appears inconsistent with this mechanism.

6.4 Future directions

There are a number of experiments that would build upon observations made in this study. In chapter 5 I provided evidence for additional *cis*-active elements that reside within the Δ3·X region, although their precise locations and functions remain unknown. Construction of short (<14 bp) heterologous replacement mutations in order to scan this region would serve to define these *cis*-active elements, at which point homology searches could be used in order to potentially determine the function of these smaller sequence fragments. Although the *fimB* promoter elements were defined in chapter 5, the remainder of the ΔP region (406 bp) remains uncharacterised and

scanning this region (also with short heterologous replacement mutations) would help to define any further *cis*-active elements located within it. According to the promoter defined by Olsen *et al.* (1994) and in this study, *fimB* has a large (269 bp) untranslated region and it would be interesting to determine whether this has a role in the transcription or translation of this gene.

Previously in this chapter (sections 6.2 & 6.3), I have suggested potential mechanisms with which NanR and NagC might serve to activate *fimB* expression: direct RNAP contact, or by stabilising a SIDD element. Although I proposed that the former might be correct for both of these regulatory proteins, it would be interesting to replace this SIDD element with sequence that is not predicted to be destabilised by superhelical energy, to see if it does affect *fimB* expression (Opel *et al.*, 2004; Sheridan *et al.*, 1998). In order to provide additional evidence that the former is correct, and that NagC (as well as NanR) might activate *fimB* expression via direct RNAP contact, it would be interesting to make additional deletion/ replacement mutations within the Δ3·X region, of differing sizes, in order to confirm the relationship between separation distance and efficacy. Combining IHF mutations (in *ihfA* or *ihfB*) with each of these partial deletion/ heterologous replacement mutation backgrounds, with and without the Rm1 mutation (which removes the NanR binding site upstream of *fimB*; El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004) would also determine the point at which NanR becomes dependent upon IHF for activation. In addition, reversing the NanR or NagC binding sites, or replacing them with binding sites for other regulators would help to determine whether the effects of these activators are specific and require direct RNAP contact. Ultimately, though, in order to prove that either NanR or NagC activate *fimB* expression via direct contact with RNAP, a lot of further work would be required. In order to do so, the crystal structures of NanR- and NagC-RNAP complexes could be solved, to identify the domains required for interaction and to introduce mutations into NanR and NagC that remove them.

In chapter 4 I compared the genomic sequence of the *nanC-fimB* intergenic region from the UPEC strain J96 to that of the commensal K12 strain MG1655, and this showed that the entire region is remarkably well conserved between the two. In addition to MG1655 and J96, there are a number of other *E. coli* genomes, either fully or partially sequenced, in which the *nanC-fimB* intergenic region is known; and several observations can be made when they are compared to each other (homology

data not supplied). In general, this entire intergenic region is remarkably well conserved (>99 %) within all known *E. coli* genomes and, notwithstanding the fact that spacing of *cis*-active elements appears to be important for precise levels of *fimB* expression (defined in more detail in chapters 5.5-7 & 6.2.5), it remains unclear why this sequence should be so well conserved when a large proportion of it appears to serve no purpose. Furthermore, the most prevalent verotoxigenic *E. coli* serotype in the UK and USA, 0157, cannot produce type 1 fimbriae because of a 16 bp deletion within *fimS* that renders them phase-locked off (Li *et al.*, 1997; Roe *et al.*, 2001); and yet the *nanC-fimB* intergenic regions of two 0157 strains that have been sequenced are remarkably similar to that of MG1655 (Hayashi *et al.*, 2001; Perna *et al.*, 2001), and repairing *fimS* restores wild-type (MG1655) switching in type 1 fimbrial expression (Roe *et al.*, 2001). The role of FimB in 0157 isolates, or the effect of sequences throughout this intergenic region upon expression of *nanC* or other genes remains unknown, and would be an interesting area to expand upon in the future.

Comparison of the sequences within this region of three UPEC strains (536, CFT073 and UT189) identifies a 12 bp insertion, within the Δ3·1 region, that is not present in the genome of MG1655 (Chen *et al.*, 2006; Hochhut *et al.*, 2006; Welch *et al.*, 2002). This 12 bp insertion results in a frameshift that produces a putative orf, of unknown function and directed towards *nanC*, and the effect of this insertion (via spacing of NanR and NagC, and potentially via transcription of this orf) remains unclear. Likewise, it would be interesting to investigate the role of this insertion, both in terms of its effect upon *fimB* expression but also the presence and significance of this putative orf.

6.5 Final conclusion

This study attempts to define the mechanism of distant activation of *fimB* by NanR and NagC, and ultimately it is unsuccessful to this end. However, I have provided evidence that strongly suggests that they do not serve to antagonise more distant inhibitory elements, despite the preliminary data on which this study is based (El-Labany *et al.*, 2003). Furthermore, I have provided evidence that is suggestive of particular mechanisms that could be readily tested in the future. Distant regulation of *fimB* remains a unique and interesting example of gene regulation, and further work to this end is certainly warranted.

References

Aberg, A., Shingler, V. & Balsalobre, C. (2006). (p)ppGpp regulates type 1 fimbriation of Escherichia coli by modulating the expression of the site-specific recombinase FimB. *Mol Microbiol* **60**, 1520-1533.

Abraham, J. M., Freitag, C. S., Clements, J. R. & Eisenstein, B. I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of Escherichia coli. *Proc Natl Acad Sci U S A* **82**, 5724-5727.

Abraham, J. M., Freitag, C. S., Gander, R. M., Clements, J. R., Thomas, V. L. & Eisenstein, B. I. (1986). Fimbrial phase variation and DNA rearrangements in uropathogenic isolates of Escherichia coli. *Mol Biol Med* **3**, 495-508.

Abraham, S. N., Sun, D., Dale, J. B. & Beachey, E. H. (1988). Conservation of the D-mannose-adhesion protein among type 1 fimbriated members of the family Enterobacteriaceae. *Nature* **336**, 682-684.

Abraham, S. N., Goguen, J. D., Sun, D., Klemm, P. & Beachey, E. H. (1987). Identification of two ancillary subunits of Escherichia coli type 1 fimbriae by using antibodies against synthetic oligopeptides of fim gene products. *J Bacteriol* **169**, 5530-5536.

Ak, P. & Benham, C. J. (2005). Susceptibility to superhelically driven DNA duplex destabilization: a highly conserved property of yeast replication origins. *PLoS Comput Biol* **1**, e7.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Anderson, G. G., Palermo, J. J., Schilling, J. D., Roth, R., Heuser, J. & Hultgren, S. J. (2003). Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* **301**, 105-107.

Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehle, M. M., Paegle, E. S. & Hatfield, G. W. (2000). Global gene expression profiling in Escherichia coli K12. The effects of integration host factor. *J Biol Chem* **275**, 29672-29684.

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V., Pierson, L. S., 3rd & et al. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* **5**, 433-440.

Atlung, T. & Ingmer, H. (1997). H-NS: a modulator of environmentally regulated gene expression. *Mol Microbiol* **24**, 7-17.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G. & Struhl, K. (1987). *Current protocols in molecular biology.* New York: John Wiley and Sons.

Baggiolini, M., Walz, A. & Kunkel, S. L. (1989). Neutrophil-activating peptide-1/interleukin 8, a novel cytokine that activates neutrophils. *J Clin Invest* **84**, 1045-1049.

Bahrani-Mougeot, F. K., Buckles, E. L., Lockatell, C. V., Hebel, J. R., Johnson, D. E., Tang, C. M. & Donnenberg, M. S. (2002). Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic Escherichia coli virulence determinants in the murine urinary tract. *Mol Microbiol* **45**, 1079-1093.

Barnhart, M. M., Lynem, J. & Chapman, M. R. (2006). GlcNAc-6P levels modulate the expression of Curli fibers by Escherichia coli. *J Bacteriol* **188**, 5212-5219.

Barnhart, M. M., Pinkner, J. S., Soto, G. E., Sauer, F. G., Langermann, S., Waksman, G., Frieden, C. & Hultgren, S. J. (2000). PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc Natl Acad Sci U S A* **97**, 7709-7714.

Bellomy, G. R. & Record, M. T., Jr. (1990). Stable DNA loops in vivo and in vitro: roles in gene regulation at a distance and in biophysical characterization of DNA. *Prog Nucleic Acid Res Mol Biol* **39**, 81-128.

Bender, J. & Kleckner, N. (1986). Genetic evidence that Tn10 transposes by a nonreplicative mechanism. *Cell* **45**, 801-815.

Benham, C. J. & Bi, C. (2004). The analysis of stress-induced duplex destabilization in long genomic DNA sequences. *J Comput Biol* **11**, 519-543.

Bertin, P., Hommais, F., Krin, E., Soutourina, O., Tendeng, C., Derzelle, S. & Danchin, A. (2001). H-NS and H-NS-like proteins in Gram-negative bacteria and their multiple role in the regulation of bacterial metabolism. *Biochimie* **83**, 235-241.

Beveridge, D. L., Dixit, S. B., Barreiro, G. & Thayer, K. M. (2004). Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers* **73**, 380-403.

Bi, C. & Benham, C. J. (2004). WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA. *Bioinformatics* **20**, 1477-1479.

Bian, Z., Brauner, A., Li, Y. & Normark, S. (2000). Expression of and cytokine activation by Escherichia coli curli fibers in human sepsis. *J Infect Dis* **181**, 602-612.

Blanco, C., Ritzenthaler, P. & Kolb, A. (1986). The regulatory region of the uxuAB operon in Escherichia coli K12. *Mol Gen Genet* **202**, 112-119.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1474.

Bloch, C. A. & Orndorff, P. E. (1990). Impaired colonization by and full invasiveness of Escherichia coli K1 bearing a site-directed mutation in the type 1 pilin gene. *Infect Immun* **58**, 275-278.

Bloch, C. A., Stocker, B. A. & Orndorff, P. E. (1992). A key role for type 1 pili in enterobacterial communicability. *Mol Microbiol* **6**, 697-701.

Blomfield, I. C., Kulasekara, D. H. & Eisenstein, B. I. (1997). Integration host factor stimulates both FimB- and FimE-mediated site-specific DNA inversion that controls phase variation of type 1 fimbriae expression in Escherichia coli. *Mol Microbiol* **23**, 705-717.

Blomfield, I. C., Vaughn, V., Rest, R. F. & Eisenstein, B. I. (1991a). Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol* **5**, 1447-1457.

Blomfield, I. C., McClain, M. S., Princ, J. A., Calie, P. J. & Eisenstein, B. I. (1991b). Type 1 fimbriation and fimE mutants of Escherichia coli K-12. *J Bacteriol* **173**, 5298-5307.

Blomfield, I. C., Calie, P. J., Eberhardt, K. J., McClain, M. S. & Eisenstein, B. I. (1993). Lrp stimulates phase variation of type 1 fimbriation in Escherichia coli K-12. *J Bacteriol* **175**, 27-36.

Blot, N., Berrier, C., Hugouvieux-Cotte-Pattat, N., Ghazi, A. & Condemine, G. (2002). The oligogalacturonate-specific porin KdgM of Erwinia chrysanthemi belongs to a new porin family. *J Biol Chem* **277**, 7936-7944.

Blumberg, S., Tkachenko, A. V. & Meiners, J. C. (2005). Disruption of protein-mediated DNA looping by tension in the substrate DNA. *Biophys J* **88**, 1692-1701.

Blumenstock, E. & Jann, K. (1982). Adhesion of piliated Escherichia coli strains to phagocytes: differences between bacteria with mannose-sensitive pili and those with mannose-resistant pili. *Infect Immun* **35**, 264-269.

Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H. & Low, D. A. (1989). Phase-variation of pyelonephritis-associated pili in Escherichia coli: evidence for transcriptional regulation. *Embo J* **8**, 613-620.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. & Boyer, H. W. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**, 95-113.

Boorstein, W. R. & Craig, E. A. (1989). Primer extension analysis of RNA. *Methods Enzymol* **180**, 347-369.

Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* **242**, 309-320.

Boudeau, J., Barnich, N. & Darfeuille-Michaud, A. (2001). Type 1 pili-mediated adherence of Escherichia coli strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. *Mol Microbiol* **39**, 1272-1284.

Brandi, A., Pon, C. L. & Gualerzi, C. O. (1994). Interaction of the main cold shock protein CS7.4 (CspA) of Escherichia coli with the promoter region of hns. *Biochimie* **76**, 1090-1098.

Brinkman, A. B., Ettema, T. J., de Vos, W. M. & van der Oost, J. (2003). The Lrp family of transcriptional regulators. *Mol Microbiol* **48**, 287-294.

Brinton, C. C., Jr. (1959). Non-flagellar appendages of bacteria. *Nature* **183**, 782-786.

Bryan, A., Roesch, P., Davis, L., Moritz, R., Pellett, S. & Welch, R. A. (2006). Regulation of type 1 fimbriae by unlinked FimB- and FimE-like recombinases in uropathogenic Escherichia coli strain CFT073. *Infect Immun* **74**, 1072-1083.

Burns, L. S., Smith, S. G. & Dorman, C. J. (2000). Interaction of the FimB integrase with the fimS invertible DNA element in Escherichia coli in vivo and in vitro. *J Bacteriol* **182**, 2953-2959.

Calvo, J. M. & Matthews, R. G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in Escherichia coli. *Microbiol Rev* **58**, 466-490.

Campbell, J. W. & Cronan, J. E., Jr. (2001). Escherichia coli FadR positively regulates transcription of the fabB fatty acid biosynthetic gene. *J Bacteriol* **183**, 5982-5990.

Cao, Y. & Hayes, F. (1999). A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type IB topoisomerases. *J Mol Biol* **289**, 517-527.

Carlos, T. M. & Harlan, J. M. (1994). Leukocyte-endothelial adhesion molecules. *Blood* **84**, 2068-2101.

Carrera, P. & Azorin, F. (1994). Structural characterization of intrinsically curved AT-rich DNA sequences. *Nucleic Acids Res* **22**, 3671-3680.

Chang, D. E., Smalley, D. J., Tucker, D. L., Leatham, M. P., Norris, W. E., Stevenson, S. J., Anderson, A. B., Grissom, J. E., Laux, D. C., Cohen, P. S. & Conway, T. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. *Proc Natl Acad Sci U S A* **101**, 7427-7432.

Chen, S. & Calvo, J. M. (2002). Leucine-induced dissociation of Escherichia coli Lrp hexadecamers to octamers. *J Mol Biol* **318**, 1031-1042.

Chen, S., Rosner, M. H. & Calvo, J. M. (2001a). Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from Escherichia coli. *J Mol Biol* **312**, 625-635.

Chen, S., Iannolo, M. & Calvo, J. M. (2005). Cooperative binding of the leucine-responsive regulatory protein (Lrp) to DNA. *J Mol Biol* **345**, 251-264.

Chen, S., Hao, Z., Bieniek, E. & Calvo, J. M. (2001b). Modulation of Lrp action in Escherichia coli by leucine: effects on non-specific binding of Lrp to DNA. *J Mol Biol* **314**, 1067-1075.

Chen, S. L., Hung, C. S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R. R., Ozersky, P., Armstrong, J. R., Fulton, R. S., Latreille, J. P., Spieth, J., Hooton, T. M., Mardis, E. R., Hultgren, S. J. & Gordon, J. I. (2006). Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proc Natl Acad Sci U S A* **103**, 5977-5982.

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* **6**, 885-897.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, S. D. (1999). X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. *Science* **285**, 1061-1066.

Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics* **20**, 426-427.

Clegg, S., Pruckler, J. & Purcell, B. K. (1985). Complementation analyses of recombinant plasmids encoding type 1 fimbriae of members of the family Enterobacteriaceae. *Infect Immun* **50**, 338-340.

Collado-Vides, J., Magasanik, B. & Gralla, J. D. (1991). Control site location and transcriptional regulation in Escherichia coli. *Microbiol Rev* **55**, 371-394.

Colonna, B., Casalino, M., Fradiani, P. A., Zagaglia, C., Naitza, S., Leoni, L., Prosseda, G., Coppo, A., Ghelardini, P. & Nicoletti, M. (1995). H-NS regulation of virulence gene expression in enteroinvasive Escherichia coli harboring the virulence plasmid integrated into the host chromosome. *J Bacteriol* **177**, 4703-4712.

Condemine, G., Berrier, C., Plumbridge, J. & Ghazi, A. (2005). Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in Escherichia coli. *J Bacteriol* **187**, 1959-1965.

Connell, I., Agace, W., Klemm, P., Schembri, M., Marild, S. & Svanborg, C. (1996). Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. *Proc Natl Acad Sci U S A* **93**, 9827-9832.

Craig, N. L. & Nash, H. A. (1983). The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell* **35**, 795-803.

Cross, A. S. & Wright, D. G. (1991). Mobilization of sialidase from intracellular stores to the surface of human neutrophils and its role in stimulated adhesion responses of these cells. *J Clin Invest* **88**, 2067-2076.

Cross, A. S., Sakarya, S., Rifat, S., Held, T. K., Drysdale, B. E., Grange, P. A., Cassels, F. J., Wang, L. X., Stamatos, N., Farese, A., Casey, D., Powell, J., Bhattacharjee, A. K., Kleinberg, M. & Goldblum, S. E. (2003). Recruitment of murine neutrophils in vivo through endogenous sialidase activity. *J Biol Chem* **278**, 4112-4120.

Cui, Y., Wang, Q., Stormo, G. D. & Calvo, J. M. (1995). A consensus sequence for binding of Lrp to DNA. *J Bacteriol* **177**, 4872-4880.

Daniels, D. W. & Bertrand, K. P. (1985). Promoter mutations affecting divergent transcription in the Tn10 tetracycline resistance determinant. *J Mol Biol* **184**, 599-610.

de Keyzer, J., van der Does, C. & Driessen, A. J. (2003). The bacterial translocase: a dynamic protein channel complex. *Cell Mol Life Sci* **60**, 2034-2052.

DePierre, J. W., Lazdins, J. & Karnovsky, M. L. (1980). The determination and localization of sialic acid in guinea-pig granulocytes. *Biochem J* **192**, 543-550.

DiNardo, S., Voelkel, K. A., Sternglanz, R., Reynolds, A. E. & Wright, A. (1982). Escherichia coli DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. *Cell* **31**, 43-51.

Dlakic, M. & Harrington, R. E. (1996). The effects of sequence context on DNA curvature. *Proc Natl Acad Sci U S A* **93**, 3847-3852.

Dodson, K. W., Jacob-Dubuisson, F., Striker, R. T. & Hultgren, S. J. (1993). Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. *Proc Natl Acad Sci U S A* **90**, 3670-3674.

Donato, G. M. & Kawula, T. H. (1999). Phenotypic analysis of random hns mutations differentiate DNA-binding activity from properties of fimA promoter inversion modulation and bacterial motility. *J Bacteriol* **181**, 941-948.

Donato, G. M., Lelivelt, M. J. & Kawula, T. H. (1997). Promoter-specific repression of fimB expression by the Escherichia coli nucleoid-associated protein H-NS. *J Bacteriol* **179**, 6618-6625.

Dong, J. M., Taylor, J. S., Latour, D. J., Iuchi, S. & Lin, E. C. (1993). Three overlapping lct genes involved in L-lactate utilization by Escherichia coli. *J Bacteriol* **175**, 6671-6678.

Dorman, C. J. & Higgins, C. F. (1987). Fimbrial phase variation in Escherichia coli: dependence on integration host factor and homologies with other site-specific recombinases. *J Bacteriol* **169**, 3840-3843.

Dorman, C. J. & Bhriain, N. N. (1992). Thermal regulation of fimA, the Escherichia coli gene coding for the type 1 fimbrial subunit protein. *FEMS Microbiol Lett* **78**, 125-130.

Dove, S. L. & Dorman, C. J. (1994). The site-specific recombination system regulating expression of the type 1 fimbrial subunit gene of Escherichia coli is sensitive to changes in DNA supercoiling. *Mol Microbiol* **14**, 975-988.

Dove, S. L., Smith, S. G. & Dorman, C. J. (1997). Control of Escherichia coli type 1 fimbrial gene expression in stationary phase: a negative role for RpoS. *Mol Gen Genet* **254**, 13-20.

Duguid, J. P., Clegg, S. & Wilson, M. I. (1979). The fimbrial and non-fimbrial haemagglutinins of Escherichia coli. *J Med Microbiol* **12**, 213-227.

Duguid, J. P., Smith, I. W., Dempster, G. & Edmunds, P. N. (1955). Non-flagellar filamentous appendages (fimbriae) and haemagglutinating activity in Bacterium coli. *J Pathol Bacteriol* **70**, 335-348.

Dunn, T. M., Hahn, S., Ogden, S. & Schleif, R. F. (1984). An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci U S A* **81**, 5017-5020.

Dworkin, J., Ninfa, A. J. & Model, P. (1998). A protein-induced DNA bend increases the specificity of a prokaryotic enhancer-binding protein. *Genes Dev* **12**, 894-900.

Dybvig, K. (1993). DNA rearrangements and phenotypic switching in prokaryotes. *Mol Microbiol* **10**, 465-471.

Ebel, W. & Trempy, J. E. (1999). Escherichia coli RcsA, a positive activator of colanic acid capsular polysaccharide synthesis, functions To activate its own expression. *J Bacteriol* **181**, 577-584.

Eden, C. S., Eriksson, B. & Hanson, L. A. (1977). Adhesion of Escherichia coli to human uroepithelial cells in vitro. *Infect Immun* **18**, 767-774.

Eisenstein, B. I., Sweet, D. S., Vaughn, V. & Friedman, D. I. (1987). Integration host factor is required for the DNA inversion that controls phase variation in Escherichia coli. *Proc Natl Acad Sci U S A* **84**, 6506-6510.

El-Labany, S., Sohanpal, B. K., Lahooti, M., Akerman, R. & Blomfield, I. C. (2003). Distant cis-active sequences and sialic acid control the expression of fimB in Escherichia coli K-12. *Mol Microbiol* **49**, 1109-1118.

Ernsting, B. R., Denninger, J. W., Blumenthal, R. M. & Matthews, R. G. (1993). Regulation of the gltBDF operon of Escherichia coli: how is a leucine-insensitive

operon regulated by the leucine-responsive regulatory protein? *J Bacteriol* **175**, 7160-7169.

Esposito, D. & Scocca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* **25**, 3605-3614.

Evans, B. R., Chen, J. W., Parsons, R. L., Bauer, T. K., Teplow, D. B. & Jayaram, M. (1990). Identification of the active site tyrosine of Flp recombinase. Possible relevance of its location to the mechanism of recombination. *J Biol Chem* **265**, 18504-18510.

Fader, R. C., Gondesen, K., Tolley, B., Ritchie, D. G. & Moller, P. (1988). Evidence that in vitro adherence of Klebsiella pneumoniae to ciliated hamster tracheal cells is mediated by type 1 fimbriae. *Infect Immun* **56**, 3011-3013.

Fischer, H., Yamamoto, M., Akira, S., Beutler, B. & Svanborg, C. (2006). Mechanism of pathogen-specific TLR4 activation in the mucosa: fimbriae, recognition receptors and adaptor protein selection. *Eur J Immunol* **36**, 267-277.

Forsman, K., Goransson, M. & Uhlin, B. E. (1989). Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in E. coli pili biogenesis. *Embo J* **8**, 1271-1277.

Forst, S., Delgado, J., Rampersaud, A. & Inouye, M. (1990). In vivo phosphorylation of OmpR, the transcription activator of the ompF and ompC genes in Escherichia coli. *J Bacteriol* **172**, 3473-3477.

Foster, T. J., Davis, M. A., Roberts, D. E., Takeshita, K. & Kleckner, N. (1981). Genetic organization of transposon Tn10. *Cell* **23**, 201-213.

Freitag, C. S., Abraham, J. M., Clements, J. R. & Eisenstein, B. I. (1985). Genetic analysis of the phase variation control of expression of type 1 fimbriae in Escherichia coli. *J Bacteriol* **162**, 668-675.

Freundlich, M., Ramani, N., Mathew, E., Sirko, A. & Tsui, P. (1992). The role of integration host factor in gene expression in Escherichia coli. *Mol Microbiol* **6**, 2557-2563.

Funnell, B. E., Baker, T. A. & Kornberg, A. (1987). In vitro assembly of a prepriming complex at the origin of the Escherichia coli chromosome. *J Biol Chem* **262**, 10327-10334.

Gally, D. L., Rucker, T. J. & Blomfield, I. C. (1994). The leucine-responsive regulatory protein binds to the fim switch to control phase variation of type 1 fimbrial expression in Escherichia coli K-12. *J Bacteriol* **176**, 5665-5672.

Gally, D. L., Leathart, J. & Blomfield, I. C. (1996). Interaction of FimB and FimE with the fim switch that controls the phase variation of type 1 fimbriae in Escherichia coli K-12. *Mol Microbiol* **21**, 725-738.

Gally, D. L., Bogan, J. A., Eisenstein, B. I. & Blomfield, I. C. (1993). Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in Escherichia coli K-12: effects of temperature and media. *J Bacteriol* **175**, 6186-6193.

Gbarah, A., Gahmberg, C. G., Ofek, I., Jacobi, U. & Sharon, N. (1991). Identification of the leukocyte adhesion molecules CD11 and CD18 as receptors for type 1-fimbriated (mannose-specific) Escherichia coli. *Infect Immun* **59**, 4524-4530.

Genevaux, P., Bauda, P., DuBow, M. S. & Oudega, B. (1999). Identification of Tn10 insertions in the rfaG, rfaP, and galU genes involved in lipopolysaccharide core biosynthesis that affect Escherichia coli adhesion. *Arch Microbiol* **172**, 1-8.

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabasi, A. L., Oltvai, Z. N. & Osterman, A. L. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **185**, 5673-5684.

Godaly, G., Frendeus, B., Proudfoot, A., Svensson, M., Klemm, P. & Svanborg, C. (1998). Role of fimbriae-mediated adherence for neutrophil migration across Escherichia coli-infected epithelial cell layers. *Mol Microbiol* **30**, 725-735.

Goodell, E. W. (1985). Recycling of murein by Escherichia coli. *J Bacteriol* **163**, 305-310.

Goodman, S. D. & Nash, H. A. (1989). Functional replacement of a protein-induced bend in a DNA recombination site. *Nature* **341**, 251-254.

Goodrich, J. A., Schwartz, M. L. & McClure, W. R. (1990). Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). *Nucleic Acids Res* **18**, 4993-5000.

Gopaul, D. N., Guo, F. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *Embo J* **17**, 4175-4187.

Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K. & Uhlin, B. E. (1990). Transcriptional silencing and thermoregulation of gene expression in Escherichia coli. *Nature* **344**, 682-685.

Grainge, I. & Jayaram, M. (1999). The integrase family of recombinase: organization and function of the active site. *Mol Microbiol* **33**, 449-456.

Gralla, J. D. (1991). Transcriptional control--lessons from an E. coli promoter data base. *Cell* **66**, 415-418.

Grindley, N. D., Whiteson, K. L. & Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**, 567-605.

Gronostajski, R. M. & Sadowski, P. D. (1985). The FLP recombinase of the Saccharomyces cerevisiae 2 microns plasmid attaches covalently to DNA via a phosphotyrosyl linkage. *Mol Cell Biol* **5**, 3274-3279.

Gui, L., Sunnarborg, A. & LaPorte, D. C. (1996). Regulated expression of a repressor protein: FadR activates iclR. *J Bacteriol* **178**, 4704-4709.

Guo, F., Gopaul, D. N. & van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-46.

Guyer, M. S., Reed, R. R., Steitz, J. A., Low, K. B. (1981). *Identification of a sex-factor-affinity site in E. coli as gamma delta. Cold Spring Harb. Symp. Quant. Biol.* **45** (part 1): 135-140.

Haagmans, W. & van der Woude, M. (2000). Phase variation of Ag43 in Escherichia coli: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. *Mol Microbiol* **35**, 877-887.

Hagerman, P. J. (1990). Sequence-directed curvature of DNA. *Annu Rev Biochem* **59**, 755-781.

Hagerman, P. J. (1992). Straightening out the bends in curved DNA. *Biochim Biophys Acta* **1131**, 125-132.

Hahn, E., Wild, P., Hermanns, U., Sebbel, P., Glockshuber, R., Haner, M., Taschner, N., Burkhard, P., Aebi, U. & Muller, S. A. (2002). Exploring the 3D molecular architecture of Escherichia coli type 1 pili. *J Mol Biol* **323**, 845-857.

Halling, S. M. & Kleckner, N. (1982). A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. *Cell* **28**, 155-163.

Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P. & Kushner, S. R. (1989). New method for generating deletions and gene replacements in Escherichia coli. *J Bacteriol* **171**, 4617-4622.

Haniford, D. B., Benjamin, H. W. & Kleckner, N. (1991). Kinetic and structural analysis of a cleaved donor intermediate and a strand transfer intermediate in Tn10 transposition. *Cell* **64**, 171-179.

Harley, C. B. & Reynolds, R. P. (1987). Analysis of E. coli promoter sequences. *Nucleic Acids Res* **15**, 2343-2361.

Hartley, C. L., Neumann, C. S. & Richmond, M. H. (1979). Adhesion of commensal bacteria to the large intestine wall in humans. *Infect Immun* **23**, 128-132.

Hawley, D. K. & McClure, W. R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res* **11**, 2237-2255.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C. G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. & Shinagawa, H. (2001). Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11-22.

Haydon, D. J. & Guest, J. R. (1991). A new family of bacterial regulatory proteins. *FEMS Microbiol Lett* **63**, 291-295.

Hedlund, M., Frendeus, B., Wachtler, C., Hang, L., Fischer, H. & Svanborg, C. (2001). Type 1 fimbriae deliver an LPS- and TLR4-dependent activation signal to CD14-negative cells. *Mol Microbiol* **39**, 542-552.

Henry, M. F. & Cronan, J. E., Jr. (1992). A new mechanism of transcriptional regulation: release of an activator triggered by small molecule binding. *Cell* **70**, 671-679.

Hickman, A. B., Waninger, S., Scocca, J. J. & Dyda, F. (1997). Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell* **89**, 227-237.

Higgins, C. F., Dorman, C. J., Stirling, D. A., Waddell, L., Booth, I. R., May, G. & Bremer, E. (1988). A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. *Cell* **52**, 569-584.

Hillen, W., Schollmeier, K. & Gatz, C. (1984). Control of expression of the Tn10-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the tet operon regulatory region. *J Mol Biol* **172**, 185-201.

Hinde, P., Deighan, P. & Dorman, C. J. (2005). Characterization of the detachable Rho-dependent transcription terminator of the fimE gene in Escherichia coli K-12. *J Bacteriol* **187**, 8256-8266.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.

Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., Gottschalk, G., Carniel, E. & Hacker, J. (2006). Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic Escherichia coli strain 536. *Mol Microbiol* **61**, 584-595.

Hochschild, A. & Dove, S. L. (1998). Protein-protein contacts that activate and repress prokaryotic transcription. *Cell* **92**, 597-600.

Hoess, R., Wierzbicki, A. & Abremski, K. (1987). Isolation and characterization of intermediates in site-specific recombination. *Proc Natl Acad Sci U S A* **84**, 6840-6844.

Holden, N. J., Uhlin, B. E. & Gally, D. L. (2001). PapB paralogues and their effect on the phase variation of type 1 fimbriae in Escherichia coli. *Mol Microbiol* **42**, 319-330.

Holden, N. J., Totsika, M., Mahler, E., Roe, A. J., Catherwood, K., Lindner, K., Dobrindt, U. & Gally, D. L. (2006). Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic Escherichia coli. *Microbiology* **152**, 1143-1153.

Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J. P., Danchin, A. & Bertin, P. (2001). Large-scale monitoring of pleiotropic

regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol Microbiol* **40**, 20-36.

Huisman, O., Errada, P. R., Signon, L. & Kleckner, N. (1989). Mutational analysis of IS10's outside end. *Embo J* **8**, 2101-2109.

Hull, R. A., Gill, R. E., Hsu, P., Minshew, B. H. & Falkow, S. (1981). Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection Escherichia coli isolate. *Infect Immun* **33**, 933-938.

Hultgren, S. J., Schwan, W. R., Schaeffer, A. J. & Duncan, J. L. (1986). Regulation of production of type 1 pili among urinary tract isolates of Escherichia coli. *Infect Immun* **54**, 613-620.

Hung, C. S., Bouckaert, J., Hung, D., Pinkner, J., Widberg, C., DeFusco, A., Auguste, C. G., Strouse, R., Langermann, S., Waksman, G. & Hultgren, S. J. (2002). Structural basis of tropism of Escherichia coli to the bladder during urinary tract infection. *Mol Microbiol* **44**, 903-915.

Hung, D. L., Knight, S. D., Woods, R. M., Pinkner, J. S. & Hultgren, S. J. (1996). Molecular basis of two subfamilies of immunoglobulin-like chaperones. *Embo J* **15**, 3792-3805.

Hwang, D. S. & Kornberg, A. (1992). Opening of the replication origin of Escherichia coli by DnaA protein with protein HU or IHF. *J Biol Chem* **267**, 23083-23086.

Jishage, M., Iwata, A., Ueda, S. & Ishihama, A. (1996). Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol* **178**, 5447-5451.

Jo, Y. L., Nara, F., Ichihara, S., Mizuno, T. & Mizushima, S. (1986). Purification and characterization of the OmpR protein, a positive regulator involved in osmoregulatory expression of the ompF and ompC genes in Escherichia coli. *J Biol Chem* **261**, 15252-15256.

Jones, C. H., Pinkner, J. S., Nicholes, A. V., Slonim, L. N., Abraham, S. N. & Hultgren, S. J. (1993). FimC is a periplasmic PapD-like chaperone that directs assembly of type 1 pili in bacteria. *Proc Natl Acad Sci U S A* **90**, 8397-8401.

Jones, C. H., Pinkner, J. S., Roth, R., Heuser, J., Nicholes, A. V., Abraham, S. N. & Hultgren, S. J. (1995). FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc Natl Acad Sci U S A* **92**, 2081-2085.

Joyce, S. A. & Dorman, C. J. (2002). A Rho-dependent phase-variable transcription terminator controls expression of the FimE recombinase in Escherichia coli. *Mol Microbiol* **45**, 1107-1117.

Kalivoda, K. A., Steenbergen, S. M., Vimr, E. R. & Plumbridge, J. (2003). Regulation of sialic acid catabolism by the DNA binding protein NanR in Escherichia coli. *J Bacteriol* **185**, 4806-4815.

Kawula, T. H. & Orndorff, P. E. (1991). Rapid site-specific DNA inversion in Escherichia coli mutants lacking the histonelike protein H-NS. *J Bacteriol* **173**, 4116-4123.

Kelly, A., Conway, C., T, O. C., Smith, S. G. & Dorman, C. J. (2006). DNA supercoiling and the Lrp protein determine the directionality of fim switch DNA inversion in Escherichia coli K-12. *J Bacteriol* **188**, 5356-5363.

Kelm, S. & Schauer, R. (1997). Sialic acids in molecular and cellular interactions. *Int Rev Cytol* **175**, 137-240.

Kendall, M., Stuart, A., (1977). *The Advanced Theory of Statistics*, 4th ed (vol.1). pp. 247. Mcmillan Publishing, New York.

Kho, S. H. & Landy, A. (1994). Dissecting the resolution reaction of lambda integrase using suicide Holliday junction substrates. *Embo J* **13**, 2714-2724.

Kikuchi, T., Mizunoe, Y., Takade, A., Naito, S. & Yoshida, S. (2005). Curli fibers are required for development of biofilm architecture in Escherichia coli K-12 and enhance bacterial adherence to human uroepithelial cells. *Microbiol Immunol* **49**, 875-884.

Kim, K. S. (2001). Escherichia coli translocation at the blood-brain barrier. *Infect Immun* **69**, 5217-5222.

Kim, S. Y., Nam, T. W., Shin, D., Koo, B. M., Seok, Y. J. & Ryu, S. (1999). Purification of Mlc and analysis of its effects on the pts expression in Escherichia coli. *J Biol Chem* **274**, 25398-25402.

Kirkegaard, K., Buc, H., Spassky, A. & Wang, J. C. (1983). Mapping of single-stranded regions in duplex DNA at the sequence level: single-strand-specific cytosine methylation in RNA polymerase-promoter complexes. *Proc Natl Acad Sci U S A* **80**, 2544-2548.

Kleckner, N., Bender, J. & Gottesman, S. (1991). Uses of transposons with emphasis on Tn10. *Methods Enzymol* **204**, 139-180.

Klemm, P. (1984). The fimA gene encoding the type-1 fimbrial subunit of Escherichia coli. Nucleotide sequence and primary structure of the protein. *Eur J Biochem* **143**, 395-399.

Klemm, P. (1986). Two regulatory fim genes, fimB and fimE, control the phase variation of type 1 fimbriae in Escherichia coli. *Embo J* **5**, 1389-1393.

Klemm, P. (1992). FimC, a chaperone-like periplasmic protein of Escherichia coli involved in biogenesis of type 1 fimbriae. *Res Microbiol* **143**, 831-838.

Klemm, P., Jorgensen, B. J., van Die, I., de Ree, H. & Bergmans, H. (1985). The fim genes responsible for synthesis of type 1 fimbriae in Escherichia coli, cloning and genetic organization. *Mol Gen Genet* **199**, 410-414.

Knutton, S., Lloyd, D. R., Candy, D. C. & McNeish, A. S. (1984). In vitro adhesion of enterotoxigenic Escherichia coli to human intestinal epithelial cells from mucosal biopsies. *Infect Immun* **44**, 514-518.

Koo, H. S., Wu, H. M. & Crothers, D. M. (1986). DNA bending at adenine . thymine tracts. *Nature* **320**, 501-506.

Kosturko, L. D., Daub, E. & Murialdo, H. (1989). The interaction of E. coli integration host factor and lambda cos DNA: multiple complex formation and protein-induced bending. *Nucleic Acids Res* **17**, 317-334.

Kowalski, D. & Eddy, M. J. (1989). The DNA unwinding element: a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. *Embo J* **8**, 4335-4344.

Kriat, M., Vion-Dury, J., Fayre, R., Maraninchi, D., Harle, J. R., Confort-Gouny, S., Sciaky, M., Fontanarava, E., Viout, P. & Cozzone, P. J. (1991). Variations of plasma sialic acid and N-acetylglucosamine levels in cancer, inflammatory diseases and bone marrow transplantation: a proton NMR spectroscopy study. *Biochimie* **73**, 99-104.

Krogfelt, K. A. & Klemm, P. (1988). Investigation of minor components of Escherichia coli type 1 fimbriae: protein chemical and immunological aspects. *Microb Pathog* **4**, 231-238.

Kuehn, M. J., Ogg, D. J., Kihlberg, J., Slonim, L. N., Flemmer, K., Bergfors, T. & Hultgren, S. J. (1993). Structural basis of pilus subunit recognition by the PapD chaperone. *Science* **262**, 1234-1241.

Kulasekara, H. D. & Blomfield, I. C. (1999). The molecular basis for the specificity of fimE in the phase variation of type 1 fimbriae of Escherichia coli K-12. *Mol Microbiol* **31**, 1171-1181.

Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**, 126-131.

La Teana, A., Brandi, A., Falconi, M., Spurio, R., Pon, C. L. & Gualerzi, C. O. (1991). Identification of a cold shock transcriptional enhancer of the Escherichia coli gene encoding nucleoid protein H-NS. *Proc Natl Acad Sci U S A* **88**, 10907-10911.

Lahooti, M., Roesch, P. L. & Blomfield, I. C. (2005). Modulation of the sensitivity of FimB recombination to branched-chain amino acids and alanine in Escherichia coli K-12. *J Bacteriol* **187**, 6273-6280.

Langermann, S., Palaszynski, S., Barnhart, M., Auguste, G., Pinkner, J. S., Burlein, J., Barren, P., Koenig, S., Leath, S., Jones, C. H. & Hultgren, S. J. (1997). Prevention of mucosal Escherichia coli infection by FimH-adhesin-based systemic vaccination. *Science* **276**, 607-611.

Leathart, J. B. & Gally, D. L. (1998). Regulation of type 1 fimbrial expression in uropathogenic Escherichia coli: heterogeneity of expression through sequence changes in the fim switch region. *Mol Microbiol* **28**, 371-381.

Lee, D. H. & Schleif, R. F. (1989). In vivo DNA loops in araCBAD: size limits and helical repeat. *Proc Natl Acad Sci U S A* **86**, 476-480.

Lee, J., Shin, S., Teng, C. H., Hong, S. J. & Kim, K. S. (2005). FimH adhesin of Escherichia coli K1 type 1 fimbriae activates BV-2 microglia. *Biochem Biophys Res Commun* **334**, 917-923.

Lee, M. H., Scherer, M., Rigali, S. & Golden, J. W. (2003). PlmA, a new member of the GntR family, has plasmid maintenance functions in Anabaena sp. strain PCC 7120. *J Bacteriol* **185**, 4315-4325.

Lennox, E. S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1**, 190-206.

Levinthal, M., Lejeune, P. & Danchin, A. (1994). The H-NS protein modulates the activation of the ilvIH operon of Escherichia coli K12 by Lrp, the leucine regulatory protein. *Mol Gen Genet* **242**, 736-743.

Li, B., Koch, W. H. & Cebula, T. A. (1997). Detection and characterization of the fimA gene of Escherichia coli O157:H7. *Mol Cell Probes* **11**, 397-406.

Lia, G., Bensimon, D., Croquette, V., Allemand, J. F., Dunlap, D., Lewis, D. E., Adhya, S. & Finzi, L. (2003). Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc Natl Acad Sci U S A* **100**, 11373-11377.

Lim, J. K., Gunther, N. W. t., Zhao, H., Johnson, D. E., Keay, S. K. & Mobley, H. L. (1998). In vivo phase variation of Escherichia coli type 1 fimbrial genes in women with urinary tract infection. *Infect Immun* **66**, 3303-3310.

Lisser, S. & Margalit, H. (1993). Compilation of E. coli mRNA promoter sequences. *Nucleic Acids Res* **21**, 1507-1516.

Liu, Y., Bondarenko, V., Ninfa, A. & Studitsky, V. M. (2001). DNA supercoiling allows enhancer action over a large distance. *Proc Natl Acad Sci U S A* **98**, 14883-14888.

Lorenz, M., Hillisch, A., Goodman, S. D. & Diekmann, S. (1999). Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence resonance energy transfer. *Nucleic Acids Res* **27**, 4619-4625.

Lowe, M. A., Holt, S. C. & Eisenstein, B. I. (1987). Immunoelectron microscopic analysis of elongation of type 1 fimbriae in Escherichia coli. *J Bacteriol* **169**, 157-163.

Lucht, J. M., Dersch, P., Kempf, B. & Bremer, E. (1994). Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled proU operon of Escherichia coli. *J Biol Chem* **269**, 6578-6578.

Malaviya, R., Ikeda, T., Ross, E. & Abraham, S. N. (1996). Mast cell modulation of neutrophil influx and bacterial clearance at sites of infection through TNF-alpha. *Nature* **381**, 77-80.

Martinez, J. J., Mulvey, M. A., Schilling, J. D., Pinkner, J. S. & Hultgren, S. J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *Embo J* **19**, 2803-2812.

McClain, M. S., Blomfield, I. C. & Eisenstein, B. I. (1991). Roles of fimB and fimE in site-specific DNA inversion associated with phase variation of type 1 fimbriae in Escherichia coli. *J Bacteriol* **173**, 5308-5314.

McClain, M. S., Blomfield, I. C., Eberhardt, K. J. & Eisenstein, B. I. (1993). Inversion-independent phase variation of type 1 fimbriae in Escherichia coli. *J Bacteriol* **175**, 4335-4344.

McConnell, K. J. & Beveridge, D. L. (2001). Molecular dynamics simulations of B '-DNA: sequence effects on A-tract-induced bending and flexibility. *J Mol Biol* **314**, 23-40.

McCormick, B. A., Klemm, P., Krogfelt, K. A., Burghoff, R. L., Pallesen, L., Laux, D. C. & Cohen, P. S. (1993). Escherichia coli F-18 phase locked 'on' for expression of type 1 fimbriae is a poor colonizer of the streptomycin-treated mouse large intestine. *Microb Pathog* **14**, 33-43.

Meadow, N. D., Saffen, D. W., Dottin, R. P. & Roseman, S. (1982). Molecular cloning of the crr gene and evidence that it is the structural gene for IIIGlc, a phosphocarrier protein of the bacterial phosphotransferase system. *Proc Natl Acad Sci U S A* **79**, 2528-2532.

Meier, I., Wray, L. V. & Hillen, W. (1988). Differential regulation of the Tn10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators O1 and O2. *Embo J* **7**, 567-572.

Merlitz, H., Rippe, K., Klenin, K. V. & Langowski, J. (1998). Looping dynamics of linear DNA molecules and the effect of DNA curvature: a study by Brownian dynamics simulation. *Biophys J* **74**, 773-779.

Meyer-Leon, L., Huang, L. C., Umlauf, S. W., Cox, M. M. & Inman, R. B. (1988). Holliday intermediates and reaction by-products in FLP protein-promoted site-specific recombination. *Mol Cell Biol* **8**, 3784-3796.

Miller, H. I. & Friedman, D. I. (1980). An E. coli gene product required for lambda site-specific recombination. *Cell* **20**, 711-719.

Miller, J. H. (1972). *Experiments in molecular genetics.* New York: Cold Spring Harbor Laboratory Press.

Mizuno, T. & Mizushima, S. (1990). Signal transduction and gene regulation through the phosphorylation of two regulatory components: the molecular basis for the osmotic regulation of the porin genes. *Mol Microbiol* **4**, 1077-1082.

Moran, A. P., Prendergast, M. M. & Appelmelk, B. J. (1996). Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. *FEMS Immunol Med Microbiol* **16**, 105-115.

Morschhauser, J., Vetter, V., Emody, L. & Hacker, J. (1994). Adhesin regulatory genes within large, unstable DNA regions of pathogenic Escherichia coli: cross-talk between different adhesin gene clusters. *Mol Microbiol* **11**, 555-566.

Mulligan, M. E., Brosius, J. & 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**, 3529-3538.

Mulvey, M. A., Schilling, J. D. & Hultgren, S. J. (2001). Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. *Infect Immun* **69**, 4572-4579.

Neidhardt, F. C., Bloch, P. L. & Smith, D. F. (1974). Culture medium for enterobacteria. *J Bacteriol* **119**, 736-747.

Neidhardt, F. C., Ingraham, J. L., Schaechter, M. (1990). *Physiology of the bacterial cell, a molecular approach.* Sunderland Massachusetts: Sinauer Associates Inc.

Newlands, J. T., Josaitis, C. A., Ross, W. & Gourse, R. L. (1992). Both fis-dependent and factor-independent upstream activation of the rrnB P1 promoter are face of the helix dependent. *Nucleic Acids Res* **20**, 719-726.

Newman, E. B. & Lin, R. (1995). Leucine-responsive regulatory protein: a global regulator of gene expression in E. coli. *Annu Rev Microbiol* **49**, 747-775.

Newman, E. B., D'Ari, R. & Lin, R. T. (1992). The leucine-Lrp regulon in E. coli: a global response in search of a raison d'etre. *Cell* **68**, 617-619.

Newman, J. V., Burghoff, R. L., Pallesen, L., Krogfelt, K. A., Kristensen, C. S., Laux, D. C. & Cohen, P. S. (1994). Stimulation of Escherichia coli F-18Col- type-1 fimbriae synthesis by leuX. *FEMS Microbiol Lett* **122**, 281-287.

Nobelmann, B. & Lengeler, J. W. (1996). Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism. *J Bacteriol* **178**, 6790-6795.

Nou, X., Skinner, B., Braaten, B., Blyn, L., Hirsch, D. & Low, D. (1993). Regulation of pyelonephritis-associated pili phase-variation in Escherichia coli: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation. *Mol Microbiol* **7**, 545-553.

Nowicki, B., Rhen, M., Vaisanen-Rhen, V., Pere, A. & Korhonen, T. K. (1984). Immunofluorescence study of fimbrial phase variation in Escherichia coli KS71. *J Bacteriol* **160**, 691-695.

Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* **26**, 391-406.

O'Gara J, P. & Dorman, C. J. (2000). Effects of local transcription and H-NS on inversion of the fim switch of Escherichia coli. *Mol Microbiol* **36**, 457-466.

Ochman, H., Gerber, A. S. & Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Ohman, L., Hed, J. & Stendahl, O. (1982). Interaction between human polymorphonuclear leukocytes and two different strains of type 1 fimbriae-bearing Escherichia coli. *J Infect Dis* **146**, 751-757.

Olsen, P. B. & Klemm, P. (1994). Localization of promoters in the fim gene cluster and the effect of H-NS on the transcription of fimB and fimE. *FEMS Microbiol Lett* **116**, 95-100.

Olsen, P. B., Schembri, M. A., Gally, D. L. & Klemm, P. (1998). Differential temperature modulation by H-NS of the fimB and fimE recombinase genes which control the orientation of the type 1 fimbrial phase switch. *FEMS Microbiol Lett* **162**, 17-23.

Opel, M. L., Aeling, K. A., Holmes, W. M., Johnson, R. C., Benham, C. J. & Hatfield, G. W. (2004). Activation of transcription initiation from a stable RNA promoter by a Fis protein-mediated DNA structural transmission mechanism. *Mol Microbiol* **53**, 665-674.

Orndorff, P. E. & Falkow, S. (1984). Organization and expression of genes responsible for type 1 piliation in Escherichia coli. *J Bacteriol* **159**, 736-744.

Pagel, J. M., Winkelman, J. W., Adams, C. W. & Hatfield, G. W. (1992). DNA topology-mediated regulation of transcription initiation from the tandem promoters of the ilvGMEDA operon of Escherichia coli. *J Mol Biol* **224**, 919-935.

Parekh, B. S. & Hatfield, G. W. (1996). Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. *Proc Natl Acad Sci U S A* **93**, 1173-1177.

Pargellis, C. A., Nunes-Duby, S. E., de Vargas, L. M. & Landy, A. (1988). Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J Biol Chem* **263**, 7678-7685.

Park, J. T. (2001). Identification of a dedicated recycling pathway for anhydro-N-acetylmuramic acid and N-acetylglucosamine derived from Escherichia coli cell wall murein. *J Bacteriol* **183**, 3842-3847.

Parkkinen, J. & Korhonen, T. K. (1989). Binding of plasminogen to Escherichia coli adhesion proteins. *FEBS Lett* **250**, 437-440.

Parkkinen, J., Virkola, R. & Korhonen, T. K. (1988). Identification of factors in human urine that inhibit the binding of Escherichia coli adhesins. *Infect Immun* **56**, 2623-2630.

Pellecchia, M., Guntert, P., Glockshuber, R. & Wuthrich, K. (1998). NMR solution structure of the periplasmic chaperone FimC. *Nat Struct Biol* **5**, 885-890.

Pellicer, M. T., Fernandez, C., Badia, J., Aguilar, J., Lin, E. C. & Baldom, L. (1999). Cross-induction of glc and ace operons of Escherichia coli attributable to pathway intersection. Characterization of the glc promoter. *J Biol Chem* **274**, 1745-1752.

Perez-Martin, J., Rojo, F. & de Lorenzo, V. (1994). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol Rev* **58**, 268-290.

Perna, N. T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin,

J., Yen, G., Schwartz, D. C., Welch, R. A. & Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature* **409**, 529-533.

Peter, B. J., Arsuaga, J., Breier, A. M., Khodursky, A. B., Brown, P. O. & Cozzarelli, N. R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biol* **5**, R87.

Platko, J. V., Willins, D. A. & Calvo, J. M. (1990). The ilvIH operon of Escherichia coli is positively regulated. *J Bacteriol* **172**, 4563-4570.

Plumbridge, J. (1995). Co-ordinated regulation of amino sugar biosynthesis and degradation: the NagC repressor acts as both an activator and a repressor for the transcription of the glmUS operon and requires two separated NagC binding sites. *Embo J* **14**, 3958-3965.

Plumbridge, J. (1998a). Control of the expression of the manXYZ operon in Escherichia coli: Mlc is a negative regulator of the mannose PTS. *Mol Microbiol* **27**, 369-380.

Plumbridge, J. (1998b). Expression of ptsG, the gene for the major glucose PTS transporter in Escherichia coli, is repressed by Mlc and induced by growth on glucose. *Mol Microbiol* **29**, 1053-1063.

Plumbridge, J. (1999). Expression of the phosphotransferase system both mediates and is mediated by Mlc regulation in Escherichia coli. *Mol Microbiol* **33**, 260-273.

Plumbridge, J. (2001). DNA binding sites for the Mlc and NagC proteins: regulation of nagE, encoding the N-acetylglucosamine-specific transporter in Escherichia coli. *Nucleic Acids Res* **29**, 506-514.

Plumbridge, J. & Vimr, E. (1999). Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli. *J Bacteriol* **181**, 47-54.

Plumbridge, J. & Pellegrini, O. (2004). Expression of the chitobiose operon of Escherichia coli is regulated by three transcription factors: NagC, ChbR and CAP. *Mol Microbiol* **52**, 437-449.

Polonskaya, Z., Benham, C. J. & Hearing, J. (2004). Role for a region of helically unstable DNA within the Epstein-Barr virus latent cycle origin of DNA replication oriP in origin function. *Virology* **328**, 282-291.

Pratt, L. A. & Kolter, R. (1998). Genetic analysis of Escherichia coli biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* **30**, 285-293.

Pribil, P. A. & Haniford, D. B. (2000). Substrate recognition and induced DNA deformation by transposase at the target-capture stage of Tn10 transposition. *J Mol Biol* **303**, 145-159.

Pruss, G. J. & Drlica, K. (1989). DNA supercoiling and prokaryotic transcription. *Cell* **56**, 521-523.

Pruss, G. J., Manes, S. H. & Drlica, K. (1982). Escherichia coli DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell* **31**, 35-42.

Quail, M. A. & Guest, J. R. (1995). Purification, characterization and mode of action of PdhR, the transcriptional repressor of the pdhR-aceEF-lpd operon of Escherichia coli. *Mol Microbiol* **15**, 519-529.

Ricca, E., Aker, D. A. & Calvo, J. M. (1989). A protein that binds to the regulatory region of the Escherichia coli ilvIH operon. *J Bacteriol* **171**, 1658-1664.

Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-1306.

Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem* **277**, 12507-12515.

Rimsky, S. (2004). Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Curr Opin Microbiol* **7**, 109-114.

Rimsky, S., Zuber, F., Buckle, M. & Buc, H. (2001). A molecular mechanism for the repression of transcription by the H-NS protein. *Mol Microbiol* **42**, 1311-1323.

Rippe, K., von Hippel, P. H. & Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem Sci* **20**, 500-506.

Ritter, A., Gally, D. L., Olsen, P. B., Dobrindt, U., Friedrich, A., Klemm, P. & Hacker, J. (1997). The Pai-associated leuX specific tRNA5(Leu) affects type 1 fimbriation in pathogenic Escherichia coli by control of FimB recombinase expression. *Mol Microbiol* **25**, 871-882.

Ritter, A., Blum, G., Emody, L., Kerenyi, M., Bock, A., Neuhierl, B., Rabsch, W., Scheutz, F. & Hacker, J. (1995). tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic Escherichia coli. *Mol Microbiol* **17**, 109-121.

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Kruger, D. H., Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P. C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M., Wilson, G. G. & Xu, S. Y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**, 1805-1812.

Rodionov, D. A., Mironov, A. A., Rakhmaninova, A. B. & Gelfand, M. S. (2000). Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. *Mol Microbiol* **38**, 673-683.

Rodriguez-Cuartero, A., Lopez-Fernandez, A. & Perez-Blanco, F. (1998). Urinary N-acetyl-beta-N-glucosaminidase in patients with urinary tract infection. *Eur Urol* **33**, 348-350.

Roe, A. J., Currie, C., Smith, D. G. & Gally, D. L. (2001). Analysis of type 1 fimbriae expression in verotoxigenic Escherichia coli: a comparison between serotypes O157 and O26. *Microbiology* **147**, 145-152.

Roesch, P. L. & Blomfield, I. C. (1998). Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the fim switch to stimulate site-specific recombination in Escherichia coli. *Mol Microbiol* **27**, 751-761.

Rowley, K. B., Elford, R. M., Roberts, I. & Holmes, W. M. (1993). In vivo regulatory responses of four Escherichia coli operons which encode leucyl-tRNAs. *J Bacteriol* **175**, 1309-1315.

Rui, S. & Tse-Dinh, Y. C. (2003). Topoisomerase function during bacterial responses to environmental challenge. *Front Biosci* **8**, d256-263.

Ryan, V. T., Grimwade, J. E., Nievera, C. J. & Leonard, A. C. (2002). IHF and HU stimulate assembly of pre-replication complexes at Escherichia coli oriC by two different mechanisms. *Mol Microbiol* **46**, 113-124.

Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E. & Leonard, A. C. (2004). Escherichia coli prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol Microbiol* **51**, 1347-1359.

Saiz, L., Rubi, J. M. & Vilar, J. M. (2005). Inferring the in vivo looping properties of DNA. *Proc Natl Acad Sci U S A* **102**, 17642-17645.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Samuelsson, P., Hang, L., Wullt, B., Irjala, H. & Svanborg, C. (2004). Toll-like receptor 4 expression and cytokine responses in the human urinary tract mucosa. *Infect Immun* **72**, 3179-3186.

Sauer, F. G., Pinkner, J. S., Waksman, G. & Hultgren, S. J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543-551.

Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. & Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058-1061.

Saulino, E. T., Bullitt, E. & Hultgren, S. J. (2000). Snapshots of usher-mediated protein secretion and ordered pilus assembly. *Proc Natl Acad Sci U S A* **97**, 9240-9245.

Saulino, E. T., Thanassi, D. G., Pinkner, J. S. & Hultgren, S. J. (1998). Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *Embo J* **17**, 2177-2185.

Schauer, R. (2000). Achievements and challenges of sialic acid research. *Glycoconj J* **17**, 485-499.

Schembri, M. A., Olsen, P. B. & Klemm, P. (1998). Orientation-dependent enhancement by H-NS of the activity of the type 1 fimbrial phase switch promoter in Escherichia coli. *Mol Gen Genet* **259**, 336-344.

Schilling, J. D., Mulvey, M. A., Vincent, C. D., Lorenz, R. G. & Hultgren, S. J. (2001). Bacterial invasion augments epithelial cytokine responses to Escherichia coli through a lipopolysaccharide-dependent mechanism. *J Immunol* **166**, 1148-1155.

Schilling, J. D., Martin, S. M., Hunstad, D. A., Patel, K. P., Mulvey, M. A., Justice, S. S., Lorenz, R. G. & Hultgren, S. J. (2003). CD14- and Toll-like receptor-dependent activation of bladder epithelial cells by lipopolysaccharide and type 1 piliated Escherichia coli. *Infect Immun* **71**, 1470-1480.

Schleif, R. (1992). DNA looping. *Annu Rev Biochem* **61**, 199-223.

Schroder, O. & Wagner, R. (2002). The bacterial regulatory protein H-NS--a versatile modulator of nucleic acid structures. *Biol Chem* **383**, 945-960.

Schwan, W. R., Seifert, H. S. & Duncan, J. L. (1992). Growth conditions mediate differential transcription of fim genes involved in phase variation of type 1 pili. *J Bacteriol* **174**, 2367-2375.

Schwan, W. R., Seifert, H. S. & Duncan, J. L. (1994). Analysis of the fimB promoter region involved in type 1 pilus phase variation in Escherichia coli. *Mol Gen Genet* **242**, 623-630.

Serres, M. H., Gopal, S., Nahum, L. A., Liang, P., Gaasterland, T. & Riley, M. (2001). A functional update of the Escherichia coli K-12 genome. *Genome Biol* **2**, RESEARCH0035.

Sheridan, S. D., Benham, C. J. & Hatfield, G. W. (1998). Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. *J Biol Chem* **273**, 21298-21308.

Siebenlist, U., Simpson, R. B. & Gilbert, W. (1980). E. coli RNA polymerase interacts homologously with two different promoters. *Cell* **20**, 269-281.

Siebold, C., Flukiger, K., Beutler, R. & Erni, B. (2001). Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). *FEBS Lett* **504**, 104-111.

Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984). *Experiments with gene fusions.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sillanaukee, P., Ponnio, M. & Jaaskelainen, I. P. (1999). Occurrence of sialic acids in healthy humans and different disorders. *Eur J Clin Invest* **29**, 413-425.

Smith, S. G. & Dorman, C. J. (1999). Functional analysis of the FimE integrase of Escherichia coli K-12: isolation of mutant derivatives with altered DNA inversion preferences. *Mol Microbiol* **34**, 965-979.

Snoep, J. L., van der Weijden, C. C., Andersen, H. W., Westerhoff, H. V. & Jensen, P. R. (2002). DNA supercoiling in Escherichia coli is under tight and subtle homeostatic control, involving gene-expression and metabolic regulation of both topoisomerase I and DNA gyrase. *Eur J Biochem* **269**, 1662-1669.

Snyder, J. A., Haugen, B. J., Buckles, E. L., Lockatell, C. V., Johnson, D. E., Donnenberg, M. S., Welch, R. A. & Mobley, H. L. (2004). Transcriptome of uropathogenic Escherichia coli during urinary tract infection. *Infect Immun* **72**, 6373-6381.

Snyder, J. A., Haugen, B. J., Lockatell, C. V., Maroncle, N., Hagan, E. C., Johnson, D. E., Welch, R. A. & Mobley, H. L. (2005). Coordinate expression of fimbriae in uropathogenic Escherichia coli. *Infect Immun* **73**, 7588-7596.

Sohanpal, B. K., Kulasekara, H. D., Bonnen, A. & Blomfield, I. C. (2001). Orientational control of fimE expression in Escherichia coli. *Mol Microbiol* **42**, 483-494.

Sohanpal, B. K., El-Labany, S., Lahooti, M., Plumbridge, J. A. & Blomfield, I. C. (2004). Integrated regulatory responses of fimB to N-acetylneuraminic (sialic) acid and GlcNAc in Escherichia coli K-12. *Proc Natl Acad Sci U S A* **101**, 16322-16327.

Sokurenko, E. V., Chesnokova, V., Doyle, R. J. & Hasty, D. L. (1997). Diversity of the Escherichia coli type 1 fimbrial lectin. Differential binding to mannosides and uroepithelial cells. *J Biol Chem* **272**, 17880-17886.

Sokurenko, E. V., Courtney, H. S., Ohman, D. E., Klemm, P. & Hasty, D. L. (1994). FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among fimH genes. *J Bacteriol* **176**, 748-755.

Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A. & Hasty, D. L. (1995). Quantitative differences in adhesiveness of type 1 fimbriated Escherichia coli due to structural differences in fimH genes. *J Bacteriol* **177**, 3680-3686.

Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, C., Schembri, M. A. & Hasty, D. L. (1998). Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A* **95**, 8922-8926.

Sonden, B. & Uhlin, B. E. (1996). Coordinated and differential expression of histone-like proteins in Escherichia coli: regulation and function of the H-NS analog StpA. *Embo J* **15**, 4970-4980.

Soto, G. E. & Hultgren, S. J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly. *J Bacteriol* **181**, 1059-1071.

Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A. & Bertin, P. (1999). Multiple control of flagellum biosynthesis in Escherichia coli: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the flhDC master operon. *J Bacteriol* **181**, 7500-7508.

Spassky, A., Rimsky, S., Garreau, H. & Buc, H. (1984). H1a, an E. coli DNA-binding protein which accumulates in stationary phase, strongly compacts DNA in vitro. *Nucleic Acids Res* **12**, 5321-5340.

Spears, P. A., Schauer, D. & Orndorff, P. E. (1986). Metastable regulation of type 1 piliation in Escherichia coli and isolation and characterization of a phenotypically stable mutant. *J Bacteriol* **168**, 179-185.

Spencer, M. E. & Guest, J. R. (1982). Molecular cloning of four tricarboxylic acid cyclic genes of Escherichia coli. *J Bacteriol* **151**, 542-552.

Spirito, F. & Bossi, L. (1996). Long-distance effect of downstream transcription on activity of the supercoiling-sensitive leu-500 promoter in a topA mutant of Salmonella typhimurium. *J Bacteriol* **178**, 7129-7137.

Spurio, R., Falconi, M., Brandi, A., Pon, C. L. & Gualerzi, C. O. (1997). The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *Embo J* **16**, 1795-1805.

Spurio, R., Durrenberger, M., Falconi, M., La Teana, A., Pon, C. L. & Gualerzi, C. O. (1992). Lethal overproduction of the Escherichia coli nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Mol Gen Genet* **231**, 201-211.

Stauffer, L. T. & Stauffer, G. V. (1998). Spacing and orientation requirements of GcvA-binding sites 3 and 2 and the Lrp-binding region for gcvT::lacZ expression in Escherichia coli. *Microbiology* **144 (Pt 5)**, 1417-1422.

Stefano, J. E. & Gralla, J. D. (1982). Spacer mutations in the lac ps promoter. *Proc Natl Acad Sci U S A* **79**, 1069-1072.

Stentebjerg-Olesen, B., Chakraborty, T. & Klemm, P. (2000). FimE-catalyzed off-to-on inversion of the type 1 fimbrial phase switch and insertion sequence recruitment in an Escherichia coli K-12 fimB strain. *FEMS Microbiol Lett* **182**, 319-325.

Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J. & Wigley, D. B. (1997). Crystal structure of the site-specific recombinase, XerD. *Embo J* **16**, 5178-5187.

Takiff, H. E., Baker, T., Copeland, T., Chen, S. M. & Court, D. L. (1992). Locating essential Escherichia coli genes by using mini-Tn10 transposons: the pdxJ operon. *J Bacteriol* **174**, 1544-1553.

Tanaka, K., Muramatsu, S., Yamada, H. & Mizuno, T. (1991). Systematic characterization of curved DNA segments randomly cloned from Escherichia coli and their functional significance. *Mol Gen Genet* **226**, 367-376.

Tanaka, Y., Kimata, K., Inada, T., Tagami, H. & Aiba, H. (1999). Negative regulation of the pts operon by Mlc: mechanism underlying glucose induction in Escherichia coli. *Genes Cells* **4**, 391-399.

Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O. & Matthews, R. G. (2002). Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci U S A* **99**, 13471-13476.

Teng, C. H., Cai, M., Shin, S., Xie, Y., Kim, K. J., Khan, N. A., Di Cello, F. & Kim, K. S. (2005). Escherichia coli K1 RS218 interacts with human brain microvascular endothelial cells via type 1 fimbria bacteria in the fimbriated state. *Infect Immun* **73**, 2923-2931.

Teter, B., Goodman, S. D. & Galas, D. J. (2000). DNA bending and twisting properties of integration host factor determined by DNA cyclization. *Plasmid* **43**, 73-84.

Thaw, P., Sedelnikova, S. E., Muranova, T., Wiese, S., Ayora, S., Alonso, J. C., Brinkman, A. B., Akerboom, J., van der Oost, J. & Rafferty, J. B. (2006). Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Res* **34**, 1439-1449.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.

Thompson, J. F. & Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res* **16**, 9687-9705.

Titgemeyer, F., Reizer, J., Reizer, A. & Saier, M. H., Jr. (1994). Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology* **140 (Pt 9)**, 2349-2354.

Tobe, T., Yoshikawa, M., Mizuno, T. & 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**, 6142-6149.

Tong, S., Porco, A., Isturiz, T. & Conway, T. (1996). Cloning and molecular genetic characterization of the Escherichia coli gntR, gntK, and gntU genes of GntI, the main system for gluconate metabolism. *J Bacteriol* **178**, 3260-3269.

Traving, C. & Schauer, R. (1998). Structure, function and metabolism of sialic acids. *Cell Mol Life Sci* **54**, 1330-1349.

Tupper, A. E., Owen-Hughes, T. A., Ussery, D. W., Santos, D. S., Ferguson, D. J., Sidebotham, J. M., Hinton, J. C. & Higgins, C. F. (1994). The chromatin-associated protein H-NS alters DNA topology in vitro. *Embo J* **13**, 258-268.

Uehara, T., Suefuji, K., Valbuena, N., Meehan, B., Donegan, M. & Park, J. T. (2005). Recycling of the anhydro-N-acetylmuramic acid derived from cell wall murein involves a two-step conversion to N-acetylglucosamine-phosphate. *J Bacteriol* **187**, 3643-3649.

Ushida, C. & Aiba, H. (1990). Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. *Nucleic Acids Res* **18**, 6325-6330.

van der Woude, M. W. (2006). Re-examining the role and random nature of phase variation. *FEMS Microbiol Lett* **254**, 190-197.

van der Woude, M. W. & Low, D. A. (1994). Leucine-responsive regulatory protein and deoxyadenosine methylase control the phase variation and expression of the sfa and daa pili operons in Escherichia coli. *Mol Microbiol* **11**, 605-618.

van der Woude, M. W. & Baumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**, 581-611, table of contents.

Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130.

Varshavsky, A. J., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. & Georgiev, G. P. (1977). Histone-like proteins in the purified Escherichia coli deoxyribonucleoprotein. *Nucleic Acids Res* **4**, 2725-2745.

Vetsch, M., Puorger, C., Spirig, T., Grauschopf, U., Weber-Ban, E. U. & Glockshuber, R. (2004). Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431**, 329-333.

Vilar, J. M. & Leibler, S. (2003). DNA looping and physical constraints on transcription regulation. *J Mol Biol* **331**, 981-989.

Vilar, J. M. & Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* **15**, 136-144.

Vimr, E. & Lichtensteiger, C. (2002). To sialylate, or not to sialylate: that is the question. *Trends Microbiol* **10**, 254-257.

Vimr, E. R. (1994). Microbial sialidases: does bigger always mean better? *Trends Microbiol* **2**, 271-277.

Vimr, E. R. & Troy, F. A. (1985). Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol* **164**, 845-853.

Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. (2004). Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68**, 132-153.

Visca, P., Leoni, L., Wilson, M. J. & Lamont, I. L. (2002). Iron transport and regulation, cell signalling and genomics: lessons from Escherichia coli and Pseudomonas. *Mol Microbiol* **45**, 1177-1190.

Wang, H., Noordewier, M. & Benham, C. J. (2004). Stress-induced DNA duplex destabilization (SIDD) in the E. coli genome: SIDD sites are closely associated with promoters. *Genome Res* **14**, 1575-1584.

Wang, J. C. (1979). Helical repeat of DNA in solution. *Proc Natl Acad Sci U S A* **76**, 200-203.

Wang, J. C. & Giaever, G. N. (1988). Action at a distance along a DNA. *Science* **240**, 300-304.

Wang, Q. & Calvo, J. M. (1993a). Lrp, a major regulatory protein in Escherichia coli, bends DNA and can organize the assembly of a higher-order nucleoprotein structure. *Embo J* **12**, 2495-2501.

Wang, Q. & Calvo, J. M. (1993b). Lrp, a global regulatory protein of Escherichia coli, binds co-operatively to multiple sites and activates transcription of ilvIH. *J Mol Biol* **229**, 306-318.

Warne, S. E. & deHaseth, P. L. (1993). Promoter recognition by Escherichia coli RNA polymerase. Effects of single base pair deletions and insertions in the spacer DNA separating the -10 and -35 regions are dependent on spacer DNA sequence. *Biochemistry* **32**, 6134-6140.

Way, J. C., Davis, M. A., Morisato, D., Roberts, D. E. & Kleckner, N. (1984). New Tn10 derivatives for transposon mutagenesis and for construction of lacZ operon fusions by transposition. *Gene* **32**, 369-379.

Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L., Donnenberg, M. S. & Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete

genome sequence of uropathogenic Escherichia coli. *Proc Natl Acad Sci U S A* **99**, 17020-17024.

White-Ziegler, C. A. & Low, D. A. (1992). Thermoregulation of the pap operon: evidence for the involvement of RimJ, the N-terminal acetylase of ribosomal protein S5. *J Bacteriol* **174**, 7003-7012.

Willins, D. A., Ryan, C. W., Platko, J. V. & Calvo, J. M. (1991). Characterization of Lrp, and Escherichia coli regulatory protein that mediates a global response to leucine. *J Biol Chem* **266**, 10768-10774.

Wold, A. E., Thorssen, M., Hull, S. & Eden, C. S. (1988). Attachment of Escherichia coli via mannose- or Gal alpha 1----4Gal beta-containing receptors to human colonic epithelial cells. *Infect Immun* **56**, 2531-2537.

Wu, X. R., Sun, T. T. & Medina, J. J. (1996). In vitro binding of type 1-fimbriated Escherichia coli to uroplakins Ia and Ib: relation to urinary tract infections. *Proc Natl Acad Sci U S A* **93**, 9630-9635.

Xia, Y., Gally, D., Forsman-Semb, K. & Uhlin, B. E. (2000). Regulatory cross-talk between adhesin operons in Escherichia coli: inhibition of type 1 fimbriae expression by the PapB protein. *Embo J* **19**, 1450-1457.

Xie, Y., Yao, Y., Kolisnychenko, V., Teng, C. H. & Kim, K. S. (2006). HbiF regulates type 1 Fimbriation independently of FimB and FimE. *Infect Immun* **74**, 4039-4047.

Yamada, H., Yoshida, T., Tanaka, K., Sasakawa, C. & Mizuno, T. (1991). Molecular analysis of the Escherichia coli hns gene encoding a DNA-binding protein, which preferentially recognizes curved DNA sequences. *Mol Gen Genet* **230**, 332-336.

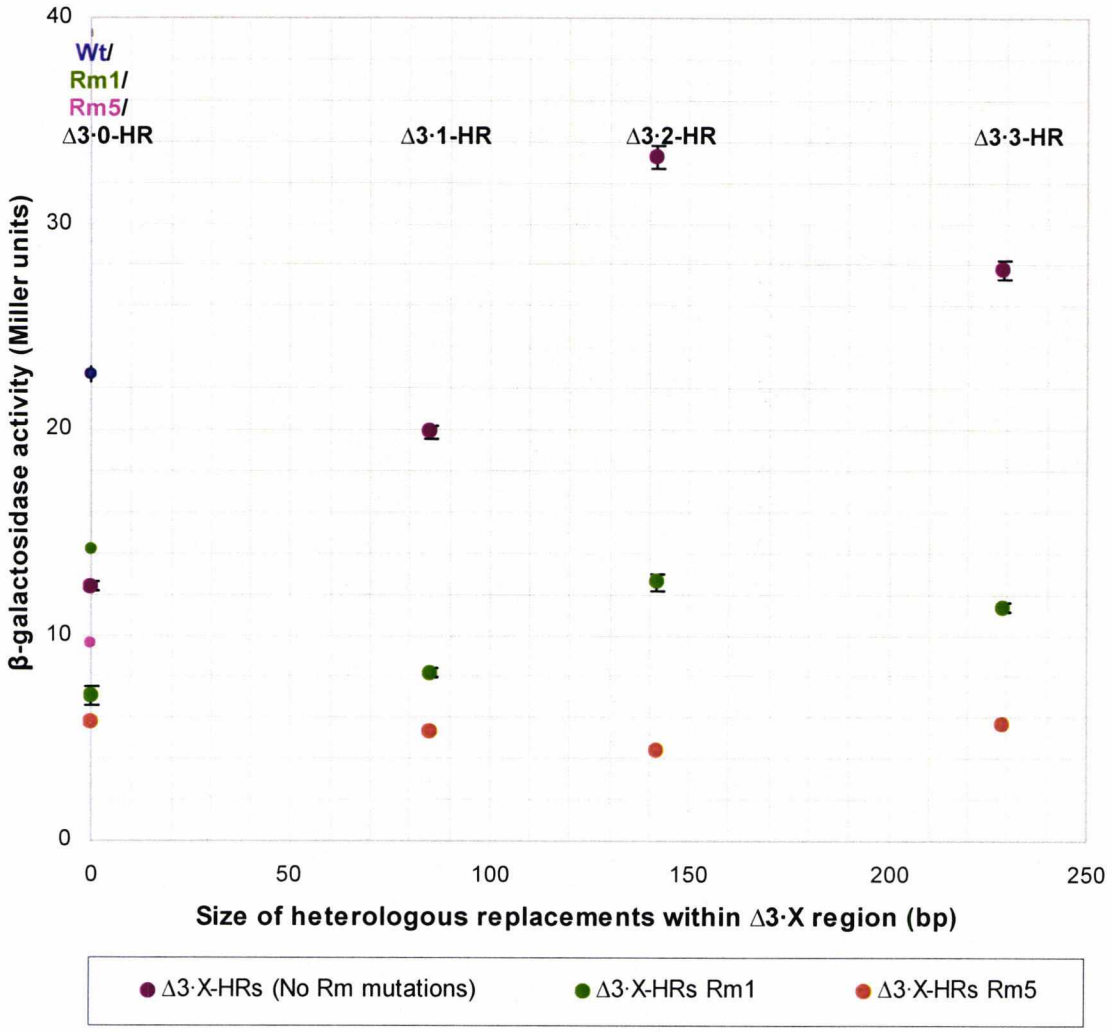
Zacharias, M., Goringer, H. U. & Wagner, R. (1992). Analysis of the Fis-dependent and Fis-independent transcription activation mechanisms of the Escherichia coli ribosomal RNA P1 promoter. *Biochemistry* **31**, 2621-2628.

Zechiedrich, E. L., Khodursky, A. B., Bachellier, S., Schneider, R., Chen, D., Lilley, D. M. & Cozzarelli, N. R. (2000). Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli. *J Biol Chem* **275**, 8103-8113.

Zhou, G., Mo, W. J., Sebbel, P., Min, G., Neubert, T. A., Glockshuber, R., Wu, X. R., Sun, T. T. & Kong, X. P. (2001). Uroplakin Ia is the urothelial receptor for uropathogenic Escherichia coli: evidence from in vitro FimH binding. *J Cell Sci* **114**, 4095-4103.

Zuber, F., Kotlarz, D., Rimsky, S. & Buc, H. (1994). Modulated expression of promoters containing upstream curved DNA sequences by the Escherichia coli nucleoid protein H-NS. *Mol Microbiol* **12**, 231-240.

APPENDIX



Appendix A1: Responsiveness of the heterologous Δ3·X replacement mutants to the Rm1 and Rm5 mutations (I). Each size-matched heterologous replacement, within the Δ3·X region was combined with the Rm1 or Rm5 mutation and measured for activity of a FimB-LacZ reporter fusion. Strains used were as follows: Wt, BGEC905; Rm1, KCEC118; Rm5, KCEC486; Δ3·0-HR, KCEC832; Δ3·0-HR Rm1, KCEC854; Δ3·0-HR Rm5, KCEC856; Δ3·1-HR, KCEC834; Δ3·1-HR Rm1, KCEC858; Δ3·1-HR Rm5, KCEC860; Δ3·2-HR, KCEC836; Δ3.2-HR Rm1, KCEC862; Δ3·2-HR Rm5, KCEC864; Δ3·3-HR, KCEC838; Δ3·3-HR Rm1, KCEC866; Δ3·3-HR Rm5, KCEC868.