Regulation of fimbrial phase transition frequencies in uropathogenic *Escherichia coli*

Makrina Totsika

B.Sc. (Edinburgh)

M.Sc. (Edinburgh)

Thesis presented for the degree of

Doctor of Philosophy

The University of Edinburgh

2007

Declaration

The research presented in this thesis is entirely my own work unless otherwise stated in the text. The material contained in this thesis has not been submitted for any other degree or professional qualification in any University.

Makrina Totsika

Acknowledgments

I would like to thank my supervisor Prof. David Gally for his continuous support, guidance and advice throughout the course of this study. I am grateful for the long hours devoted to our meetings, the challenging discussions and his genuine interest and support in my future career.

Many thanks also go to all the staff and students of the ZAP lab and the other Bacteriology groups based at the Medical School, whose help over these years has been greatly appreciated. I would like to separately thank Mrs Shonna Johnston for her expert advice on flow cytometry and Ms Denise Hutchison for being an excellent technical officer. Also special thanks to Luke Tysall, Dai Wang, Allen Flockhart, Tracy Rosser, Tracy Dransfield, Pablo Nart, Karla Sanchez, Allison Wroe and all the LBEP girls (Jeanette, Nouri, Caitriona, Bethan, Gill and Shruti) for all the fun times in and outside the lab. A great big THANK YOU goes to James Emmerson, for being the best 'PhD-friend' one could wish for! ☺ I couldn't have gone through the ups and especially the downs of this PhD without you.

I would also like to acknowledge everyone involved in running the Wellcome Trust 4-year PhD programme in the University of Edinburgh, particularly Dr. Margarete Heck and Dr. Karen Chapman for their advice and guidance. I am grateful to the Wellcome Trust for funding this research and for financially supporting me. Thanks are also due to all the 'Wellcome girls' in my year: Ita, Elaine, Effie, Natasha, Vicky, Jenn and Grainne for their friendship and for help with scientific and other matters over the years of this programme.

I would also like to thank all my friends who supported me throughout the nine years I have been studying in Edinburgh. It wouldn't be possible to list them all here, but I would like to specially thank Spiridoula, Jos and Marianna, Alex, Aileen, Rennos, Nikos, Ricardo, Neil and Uschi, Jules and Laura, Andrew, Leonidas, Andreas and Chrysa, Alexia, Manolis, Giorgos and Rania, Giorgos Sideris and Natasha, for their friendship and love.

This thesis would not have been possible without the love and friendship of my partner Dimitris Vagenas (not to mention the assistance with data analysis and manuscript editing!). I would like to thank him for his constant support and invaluable advice, especially during the hard and stressful times of this PhD, and for agreeing to follow me in my scientific pursuits to the other side of the world!!! I would also like to thank his family for the love and support they have given me throughout my studies.

Finally, my family, Apostolis, Sofi and Vaso, deserve a separate big thank you for believing in me and supporting me in every step of my life. I owe all my accomplishments and achievements to them.

Contents

Common Abbreviations

ABU	Asymptomatic bacteriuria
αCTD	α C-terminal domain
Ag43	Antigen 43
ANOVA	Analysis of variance
Ap^R	Ampicillin resistance
APS	Ammonium persulfate
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
cAMP	cyclic adenosine-3',5'-monophosphate
CAP	Catabolite activator protein
CFA	Colonisation factor antigen
cfu	colony forming units
Cm^R	Chloramphenicol resistance
CNF1	Cytotoxic necrotising factor type 1
CRP	cAMP receptor protein
CTAB	Hexadecyltrimethylammonium bromide
C-terminus	Carboxy-terminus
Dam	DNA adenine methylase
dH_2O	Distilled water
dN	Number of nonsynonymous substitutions per nonsynonymous site
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotise triphosphate
dS	Number of synonymous substitutions per synonymous site
DTT	Dithiothreitol
EagEC	Enteroaggregative *Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
eGFP	Enchanced GFP
EHEC	Enterohaemorrhagic *Escherichia coli*
EIEC	Enteroinvasive *Escherichia coli*

EMSA(s)	Electrophoretic mobility shift assay(s)
EPEC	Enteropathogenic *Escherichia coli*
ETEC	Enterotoxigenic *Escherichia coli*
ExPEC	Extraintestinal pathogenic *Escherichia coli*
FC	Flow cytometry
F primer	Forward primer
GEI(s)	Genomic island(s)
GFP	Green fluorescent protein
HA	Haemagglutination
HGT	Horizontal gene transfer
HRP	Horseradish peroxidase
IHF	Integration host factor
IgG	Immunoglobulin G
IL-	Interleukin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobases
KDa	KiloDaltons
Kn^R	Kanamycin resistance
LB	Luria-Bertani
LPS	Lipopolysaccharide
Lrp	Leucine-responsive regulatory protein
M	Molarity
μCi	microCurie
MLST	Multilocus sequence typing
MRHA	Mannose-resistant haemagglutination
MR/P	Mannose-resistant *Proteus*-like fimbriae
msa	multiple sequence alignment
NEB	New England biolabs
OD_{600}	Optical density at 600 nm
OM	Outer membrane
Pap	Pyelonephritis associated pili
PAI(s)	Pathogenicity island(s)

p_{BA}	*papBA* promoter
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
Pef	Plasmid-encoded fimbriae
PFA	Paraformaldehyde
p_I	*papI* promoter
Π	Nucleotide diversity
PMSF	Phenylmethylsulphonyl fluoride
Prf	P-related fimbriae
Prs	Pap-related sequence
qRT-PCR	Real-time reverse transcription PCR
RBCs	Red blood cells
RD	Rich defined
rfu	relative fluorescence units
R primer	Reverse primer
rpm	rounds per minute
RUTI	Recurrent UTI
SAP	Shrimp alkaline phosphatase
Sat	Secreted autotransporter toxin
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis
s.e.	Standard error of the mean
SNP(s)	Single nucleotide polymorphism(s)
SOC	Salt-optimised and carbon
SSM	Slip-stranded mispairing
T_a	Annealing temperature
TBE	Tris-borate/EDTA
Tc^R	Tetracycline resistance
TEMED	N,N,N',N'-Tetramethylethylene diamine
TFB	Transformation buffer
TGF	Tumour growth factor

TLR(s)	Toll-like receptor(s)
Tris	Trishydroxymethylaminomethane
U	Enzyme unit
UPEC	Uropathogenic *Escherichia coli*
UTI(s)	Urinary tract infection(s)
V	Volts
VUR	Vesicoureteral reflux
wHtH	winged helix-turn-helix
w.t.	Wild type
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
χ^2	Chi-square statistic

List of Figures

List of Tables

Abstract

Uropathogenic *Escherichia coli* (UPEC) are the etiologic agent of more than 80% of uncomplicated urinary tract infections (UTIs) in humans. Pyelonephritis-associated pili (Pap) are fimbrial adhesins that facilitate binding of UPEC to Galα-(1,4)-Galβ moieties contained in membrane glycolipids on human uroepithelial cells and are associated with acute kidney infection (pyelonephritis). Pap expression is phase variable and the frequency with which phase transition occurs determines the proportion of P-fimbriate bacteria in the population. In this study, *pap* phase transition frequencies were measured in clinical isolates for the first time and were shown to be markedly higher than the frequencies displayed by the same *pap* operons measured in *E. coli* K-12. In this relevant regulatory context, phase variation frequencies of homologous *pap* operons were found to be differentially affected by culture conditions, indicating a hierarchy of expression depending on environmental signals. Cross-talk between *pap* operons was also found to be dependent on culture conditions. The molecular mechanism leading to different phase variation frequencies between homologous *pap* operons was investigated by sequencing 82 *pap* regulatory regions and their regulators (*papI* and *papB*) from 54 UPEC isolates of different clinical origin (asymptomatic *vs.* symptomatic UTI) . One variable region identified was a high affinity binding site for the *pap* autoregulatory protein PapB. The site contained a variable number of 9 bp repeats with $(T/A)_3$ sequences, which affected PapB binding and the frequency of off-to-on phase transition, under particular environmental conditions. Sequence diversification via point mutation was also observed among *papI* genes, encoding for the *pap* transcriptional activator PapI, and were shown to be under positive selection (dN/dS > 1) for functionally adaptive

amino acid replacements. Certain PapI variants correlated with symptomatic disease and differed in their ability to activate *pap* operons. This study provided different lines of evidence supporting the hypothesis that UPEC isolates have evolved mechanisms to regulate the phase variation frequencies of homologous fimbrial operons, potentially to achieve their differential activation and sequential expression. The ability of UPEC to coordinate expression of multiple surface factors is critical for the successful colonisation of the varied complex micro-environments encountered in the human urinary tract.

Chapter 1

Introduction

1.1 Overview

During our life we are continually interacting with microorganisms. Although not visible to the naked eye, bacteria represent a major component of our bodies and outnumber our tissue cells by 10 to 1. Bacteria live on our skin, in our mouth, nose, throat, intestine and vaginal cavity without causing disease. Bacteria that are part of our normal flora are termed commensals. The work of the French chemist Louis Pasteur more than a century ago, provided strong support for the, until then controversial, 'germ theory of disease' allowing bacteria to be considered as causative agents of human disease. Further evidence was provided by the meticulous work of Robert Koch, who isolated different types of bacteria and demonstrated their direct role in known human diseases. Since then, much of scientific research has focused in delineating the molecular basis of microbial pathogenesis and disease by investigating both microbial and host factors that participate in these processes. This thesis describes work that aims to further our understanding of microbial disease by investigating the genetic regulation of pathogenicity factors utilised by *Escherichia coli* during infection of the human urinary tract. In an era of increased microbial resistance to antibiotics, understanding microbial pathogenicity is going to play a crucial role in the design of new effective strategies for disease prevention and treatment.

1.2 *Escherichia coli*

The bacterium *Escherichia coli* was first isolated in 1885 from the faeces of healthy individuals and was named after the German paediatrician, Theodore Escherich, who described it. *E. coli* are Gram negative rods and belong to the Family of

Enterobacteriaceae. They live as commensals in the large intestine, or colon, of humans and animals and represent approximately 1% of the total bowel flora of adult humans. Populations of commensal bacteria are now considered to be phenotypically and genetically diverse and to contain subpopulations that possess the genetic potential to cause disease (Foxman, 2007). *E. coli* have demonstrated the ability to act as multipotent pathogens that can cause disease in several body systems (Kaper *et al.*, 2004). Intestinal pathogenic *E. coli* strains with distinct genotypes, such as enteropathogenic (EPEC), enterohaemorrhagic (EHEC), enteroinvasive (EIEC), enterotoxigenic (ETEC), and enteroaggregative (EAgEC) are responsible for general outbreaks of diarrhoeal disease causing significant morbidity and mortality through several different mechanisms of pathogenesis (reviewed in Sussman, 1997). Extraintestinal *E. coli* (ExPEC) are another varied subset of life-threatening pathogens responsible for neonatal meningitis, septicaemia and urinary tract infections (UTIs) (reviewed in Sussman, 1997).

Pathogenic strains were recognised as early as the start of the 20th century and their discrimination from non-pathogenic strains was originally based on surface antigens. Systematic studies of *E. coli* virulence were made possible with the development of immunological serotyping originally proposed by Kauffman in 1944. According to his scheme strains are seroytped on the basis of three antigens, somatic lipopolysaccharide O-antigens, flagellar H-antigens and capsular K-antigens. The scheme has since then expanded to include the analysis of fimbrial F-antigens (Orskov and Orskov, 1990), which are however rarely included in the serological formula of a strain.

It soon became obvious that *E. coli* serogroups associate reproducibly with certain clinical syndromes. It is not, however, the serological antigens *per se* that confer virulence. Rather, they serve as easily identifiable chromosomal markers that correlate with specific bacterial virulence and 'fitness' factors. A group of *E. coli* strains sharing a defined set of virulence factors and causing a common disease constitutes a pathotype (Kaper *et al.*, 2004). Recent advances in whole-genome sequencing and comparative genomics revealed major genetic differences between commensal and pathogenic *E. coli* strains from different pathotypes, facilitating research in established and novel virulence determinants (Blattner *et al.*, 1997; Hayashi *et al.*, 2001; Perna *et al.*, 2001; Oelschlaeger *et al.*, 2002; Welch *et al.*, 2002; Glasner and Perna, 2004; Brzuszkiewicz *et al.*, 2006; Chen *et al.*, 2006; Willenbrock *et al.*, 2006; Johnson *et al.*, 2007; Lloyd *et al.*, 2007).

The focus of this thesis is uropathogenic *E. coli* (UPEC), which forms a group of extraintestinal pathogens equipped with the necessary virulence and 'fitness' factors to spread, adapt, and proliferate in the urinary tract, causing the majority of human urinary tract infections (UTIs).

1.3 Urinary Tract Infections (UTIs)

Urinary tract infections (UTIs) are defined by the presence of organisms in the urinary tract. Although normally sterile, the urinary tract is exposed to exogenous microorganisms via its opening to the body surface (urethra) making UTIs a very common problem, second in incidence only to bacterial respiratory infections.

1.3.1 Epidemiology

It is estimated that UTIs affect 150 million people worldwide and cost more than $6 billion annually in direct health care expenditure (Kucheria *et al.*, 2005). Just in the United States the magnitude of UTIs, as judged by visits to physicians annually, is estimated to be as high as 8 million and cost approximately $1.6 billion (Warren, 1996; Foxman, 2002). Most UTIs occur in otherwise healthy individuals with normal, unobstructed urinary tracts (uncomplicated or community-acquired UTIs), while infections occuring in patients with functional or anatomical abnormalities of the urinary tract are referred to as complicated UTIs. Infection prevalence of uncomplicated UTIs is influenced by age and sex. The incidence rate increases with age, and due to anatomical differences women are significantly more likely to be affected than men (Hooton *et al.*, 1996), except at early infancy (Zorc *et al.*, 2005). UTI incidence in young healthy men is very rare. In contrast, one in three women are estimated to have had at least one episode of acute uncomplicated UTI that requires antibiotic treatment by the age of 24 years (Foxman *et al.*, 2000) and 40-50% of all women will experience one UTI episode in their lifetime (Kunin, 1994). Moreover, of those affected, as many as 27% will suffer recurrent infections (RUTI) within six months, despite receiving appropriate antibiotic therapy (Foxman, 1990) and RUTI prevalence can increase to more than 40% within 12 months from the index UTI episode (Ikaheimo *et al.*, 1996).

Although acute uncomplicated UTIs are generally considered benign conditions that normally resolve without significant long-term medical sequelae, they can have a significant impact on quality of life, especially when recurrent (RUTI). Moreover,

their high prevalence and associated short-term morbidity result in significant financial ramifications (Foxman, 2002) signifying the need to understand the molecular pathogenesis of UTI and design effective prevention strategies for their clinical management.

1.3.2 Common uropathogens

Infecting organisms are most commonly bacteria however infections can also occur due to viruses, fungi and parasites. Principal bacterial pathogens are enterics from the host intestinal flora, predominantly *Escherichia coli*, which accounts for >80% of uncomplicated UTIs and ~30% of nosocomial infections (Bouza *et al.*, 2001). Other Gram negative species, such as *Proteus*, *Enterobacter*, *Klebsiella* and *Pseudomonas* are less common causes of uncomplicated cases, but are more prominent in complicated UTIs involving either compromised patients or individuals with functionally or anatomically defective urinary tracts. Gram positive organisms are also responsible for UTIs, with *Enterococcus* and *Staphylococcus* spp. being the commonest. In fact, *Staphylococcus saprophyticus* is the second most frequent cause of acute uncomplicated UTIs after *E. coli*, affecting mainly young female outpatients (Williams *et al.*, 1976; Jordan *et al.*, 1980; Latham *et al.*, 1983).

1.3.3 Clinical syndromes

A positive urine culture defines all UTIs (uncomplicated and complicated), but overall UTI clinical manifestations are diverse and typically range from asymptomatic bacteriuria (ABU), to infections of the bladder (cystitis) or the renal pelvis and kidneys (acute pyelonephritis). Asymptomatic bacteriuria is defined by the

isolation of $\geq 10^5$ colony forming units (cfu) per ml of urine and absence of local or systemic symptoms (Rubin *et al.*, 1992). The prevalence of ABU is estimated to be around 5% in young women (Hooton *et al.*, 2000), 1-2% in children (Jodal, 1987) and is highest in the elderly (Boscia *et al.*, 1986). Antibiotic treatment is not necessary for ABU, except for specific subpopulations such as pregnant women, in which if left untreated there is a 25% risk of progressing to acute pyelonephritis and leading to complications in pregnancy (Cunningham and Lucas, 1994). Cystitis accounts for the great majority of physician visits for UTIs and is characterised by $\geq 10^3$ cfu/ml of urine, frequency and urgency of urination, dysuria and occasionally subrapubic pain (Rubin *et al.*, 1992). In some cases, the infectious agent ascends from the lower urinary tract to the kidneys via the ureters causing acute pyelonephritis, a condition resulting from inflammation of the renal pelvis and kidney. Acute pyelonephritis is clinically characterised by isolation of $\geq 10^4$ cfu/ml of urine, fever and flank pain. If left untreated, about 30% of acute UTIs can progress to bacteraemia and become life threatening when the causative agent enters the blood stream (Agace *et al.*, 1996). Almost 25% of bacteraemia cases reported in an 18-year prospective hospital survey in England and Wales involved *E. coli* originating from the urinary tract (Gransden *et al.*, 1990). Recent evidence suggests that persistent and recurrent UTIs can also be associated with an invasive *E. coli* phenotype, even though the clinical presentation is very different to acute UTI. Persistence of invasive *E. coli* in the urinary tract is thought to be accomplished by strains being able to lie quiescent for long periods in the bladder/urethral epithelium (Martinez *et al.*, 2000; Mulvey *et al.*, 2001; Schilling *et al.*, 2002; Anderson *et al.*, 2003, 2004).

1.3.4 Host susceptibility factors

Many years of intensive UTI research has identified certain host susceptibility and risk factors associated with the varied clinical syndromes of the disease. Among healthy adults, women represent 90% of patients with UTIs. Behavioural factors, such as recent sexual intercourse and use of diaphragm with spermicide, have been associated with increased risk of UTI in sexually active young women (Hooton *et al.*, 1996), whereas a history of recurrent infections (RUTI) is also a predisposition factor (Foxman, 1990). Apart from demographic and behavioural factors, research is also focusing on identifying genetic factors that predispose to UTIs. One early observation was that nonsecretors of blood group antigens have a hereditary predisposition to RUTI compared to secretors (Kinane *et al.*, 1982; Navas *et al.*, 1994). Recently, acute pyelonephritis was suggested to be inherited through low expression levels of the human IL-8 receptor CXCR1 (Lundstedt *et al.*, 2007).

For the case of complicated UTIs, susceptibility mainly appears in two groups of patients. One group comprises of those with functional or anatomical abnormalities of the urinary tract. Instrumentation, such as urethral catheterisation, also causes an obstruction in the urinary tract and predisposes patients to UTIs (Warren, 2001). In a minority of children, structural defects of the bladder-ureter valve can lead to vesicoureteral reflux (VUR), a condition that allows uropathogens to ascend from the lower to the upper urinary tract. This condition was shown to be epidemiologically associated with renal scarring and severe reduction in renal function or even complete renal failure (Jakobsson *et al.*, 1994). The second group includes individuals with medical conditions that predispose them to a higher incidence or

severity of UTI. For example, patients with spinal cord injuries (Cardenas and Hooton, 1995) and neurogenic bladders (Matsumoto *et al.*, 2001), acute coronary syndrome (Sims *et al.*, 2005), diabetes (Ronald and Ludwig, 2001) and human immunodeficiency virus (HIV) (Schonwald *et al.*, 1999) have increased UTI risk.

In many of these patient groups *E. coli* is not the predominant UTI cause. Such epidemiological findings are important in the optimisation of risk assessment and treatment of individual cases. Therefore, the optimal approach to UTI clinical management requires not only the identification of susceptible individuals but most importantly, the investigation of bacterial factors that contribute to pathogenesis.

1.4 Uropathogenic *Escherichia coli* (UPEC)

As the major causative agent of UTIs, uropathogenic *E. coli* (UPEC) has been studied extensively. UPEC strains isolated from different UTI clinical syndromes have served as model organisms for the identification and study of bacterial factors that confer urovirulence, their evolution, epidemiology, interaction with the host, genetic organisation and regulation of expression.

1.4.1 Virulence factors

During the course of bacterial infection, including urinary tract infection, the pathogen is challenged to respond to several different environments within the host. Different niches have different requirements. In order for a bacterial pathogen to be successful, it has to possess and express different combinations of fitness factors. Genomic research has revealed that the chromosome of different pathogenic *E. coli*

strains is composed of a conserved 'core' genome, which is shared with commensal *E. coli* strains, and a 'flexible' gene pool, which contains additional traits that can be advantageous under certain conditions (Oelschlaeger *et al.*, 2002; Welch *et al.*, 2002; Glasner and Perna, 2004). This mosaic genome structure offers *E. coli* diverse disease-causing potential and reflects the lifestyle of the bacterium.

Intensive research over the last decades has led to the accumulation of knowledge on the virulence traits and mechanisms associated with uropathogenic *E. coli* (UPEC). Even though UPEC strains originate from the host's own bowel flora and therefore exist in the intestinal tract, they possess specific traits that promote their transition to and survival in the urinary tract. The set of virulence factors defining the UPEC pathotype can be divided into several functional groups, which include adhesins, toxins, iron-uptake systems, capsule, serum resistance and invasins (reviewed in Johnson, 1991). Adhesin factors facilitate attachment of bacteria to specific host cell receptors and drive the pathogen's tissue tropism. UPEC adhesin factors, particularly fimbriae, are the focus of this thesis and aspects of their biology are discussed more thoroughly in sections 1.4.3 and 1.4.4. Other virulence factors commonly produced by UPEC strains are toxins, which frequently associate with adhesin-encoding genes in pathogenicity islands (PAIs) (Blanco *et al.*, 1990, 1997; Blum *et al.*, 1995). A direct cytotoxic effect on host tissues has been demonstrated for haemolysin (Mobley *et al.*, 1990; Warren *et al.*, 1995), cytotoxic necrotising factor 1 (CNF1) (Mills *et al.*, 2000) and the secreted autotransporter toxin (Sat) (Guyer *et al.*, 2000, 2002; Maroncle *et al.*, 2006; Guignot *et al.*, 2007). UPEC strains also possess multiple iron acquisition systems to cope with the limited concentrations of soluble iron in the

urinary tract during infection (Carbonetti *et al.*, 1986; Snyder *et al.*, 2004). One of the most efficient siderophores, aerobactin, ensures sufficient iron solubilisation and uptake, and was shown to enhance bacterial survival (Montgomerie *et al.*, 1984). Other virulence-associated factors encoded by UPEC PAIs, such as carbohydrate capsules, outer membrane proteins and lipopolysaccharide, play a role in serum resistance and bacterial protection from the host immune system (Cross *et al.*, 1986; Weiser and Gotschlich, 1991; Jann and Jann, 1997; Schembri *et al.*, 2004). Interestingly, a plethora of uropathogen-specific coding sequences localised in UPEC PAIs have either no assigned function (hypothetical proteins) or have been assigned putative functions, strongly indicating the existence of novel urovirulence determinants (Welch *et al.*, 2002; Lloyd *et al.*, 2007).

1.4.2 Evolution of virulence

Most virulence traits are not encoded by the 'core' genome but are part of UPEC's 'flexible' gene pool. Early studies using UPEC isolates demonstrated a close linkage of genes encoding *E. coli* virulence factors (Low *et al.*, 1984; Welch *et al.*, 1983) and it was later shown that several of these linked virulence determinants were spontaneously lost from UPEC isolates due to large chromosomal deletions (Hacker *et al.*, 1983, 1990; Blum *et al.*, 1994, 1995; Ritter *et al.*, 1995). This led to the discovery of large genomic islands (GEIs), which were termed pathogenicity islands (PAIs) since they associated with virulence traits in pathogenic variants but were absent from closely-related non-pathogenic bacteria. PAIs are large mobile genetic elements that spread among bacterial genomes through horizontal gene transfer (HGT) (reviewed in Hacker *et al.*, 1999). Acquisition of multiple virulence factors

en-block by PAI HGT greatly contributes to the evolution of pathogenic microbes and together with genome reduction, genome rearrangements and generation of point mutations, HGT promotes bacterial genome plasticity. These mechanisms are considered essential for the ecological adaptation of both pathogenic and environmental microorganisms (reviewed in Dobrindt *et al.*, 2004).

The spread of certain PAIs among *E. coli* genomes explains how a clonally heterologous group of uropathogenic strains can be associated with the same type of disease. It is now well accepted that certain virulence factors encoded by PAIs occur more often in *E. coli* isolates from patients with UTI than among other *E. coli* strains (Bidet *et al.*, 2005; Parham *et al.*, 2005). UPEC PAIs are the most-extensively studied; they were originally discovered in UPEC strain 536 (Blum *et al.*, 1994) and have since been identified in the two other prototypic pyelonephritis strains J96 (Blum *et al.*, 1995; Swenson *et al.*, 1996; Melkerson-Watson *et al.*, 2001) and CFT073 (Kao *et al.*, 1997; Guyer *et al.*, 1998; Rasko *et al.*, 2001), as well as the recently sequenced cystitis isolate UTI89 (Chen *et al.*, 2006). The genetic traits encoded by UPEC PAIs enable bacterial adaptation to the hostile environment of the urinary tract and evasion of the host immune defenses (reviewed in Hacker *et al.*, 1999). However, not all *E. coli* strains causing uncomplicated UTIs possess every established virulence factor, indicating that the presence or absence of virulence genes determines the probability and not the ability to cause disease in the urinary tract. Moreover, although certain virulence determinants have been demonstrated to increase the pathogenic potential of *E. coli* strains, many more factors encoded by

genetic blocks unique to UPEC strains that could enhance urovirulence are yet to be identified and functionally characterised (Lloyd *et al.*, 2007).

1.4.3 Adhesin factors

Uropathogenic *Escherichia coli* strains carry the genetic potential to express several distinct types of fimbrial and afimbrial adhesins that permit adherence of the bacterium to a range of host cell receptors, thus determining host, tissue and target cell specificity for each strain (reviewed in Soto and Hultgren, 1999). Adherence is considered essential for initial establishment and persistence in the urinary tract, as adhesins allow 'staging post' colonisation of the perineum and vagina and attachment to the uroepithelium in the urethra and bladder, as well as the kidney (reviewed in Krogfelt, 1991). Moreover, adhesin factors mediate bacterial cell-cell interactions which are required in bacterial processes equally important for successful infection, such as microcolony and biofilm formation, cellular invasion and intracellular survival, as well as contact-dependent inhibition of growth (Pratt and Kolter, 1998; Danese *et al.*, 2000; Martinez *et al.*, 2000; Anderson *et al.*, 2003; Aoki *et al.*, 2005). Afimbrial adhesins are directly associated with the bacterial surface, whereas fimbrial adhesins are located within or at the distal tip of the fimbrial structure, which is a protein polymer extending from the bacterial surface. This thesis is concerned with fimbrial adhesins as virulence factors of UPEC isolates and in particular their regulation at the genetic level. For this reason, UPEC fimbriae are discussed in greater detail in the following sections.

1.4.4 Fimbriae

Fimbriae are hair-like protein fibers radiating from the bacterial cell surface (Fig. 1.1A). Their composite architecture consists of a rigid helical rod and a thinner flexible tip fibrillum that carries the adhesin protein at its distal end (Fig. 1.1B-C) (Kuehn *et al.*, 1992).

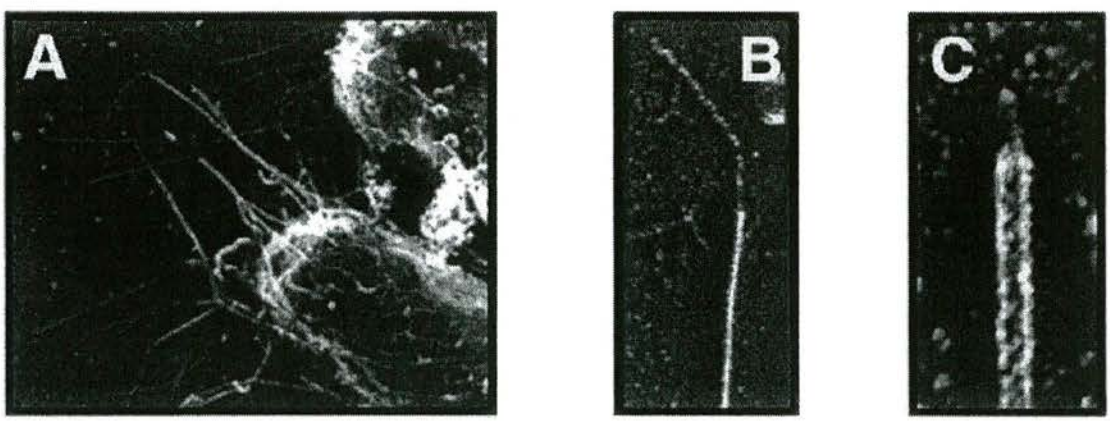


Fig. 1.1 Architecture of *E. coli* fimbriae. (*A*) Freeze-etch electron micrograph of *E. coli* cells expressing P fimbriae. Multiple hair-like appendages can be seen radiating from the bacterial surface. High-resolution quick-freeze deep-etch electron micrographs of (*B*) a single P and (*C*) type 1 fiber showing their composite architecture. Both fimbrial types consist of a rigid rod, composed of the main structural subunit (PapA for P and FimA for type 1 fimbriae), and a thin flexible fibrillum with the adhesin localised at the tip. A substantial difference in the fibrillum length of P and type 1 fimbriae is obvious (images adapted from Thanassi and Hultgren, 2000).

The term fimbriae originates from the Latin word for fringe or threads and is synonymous to pili, which was introduced later (Brinton, 1959) but is also in common use (Duguid and Anderson, 1967). Their discovery was directly linked to their capacity to agglutinate human erythrocytes and adhere to uroepithelial cells

(Svanborg-Eden and Hansson, 1978; Kallenius and Mollby, 1979; Kallenius *et al.*, 1980). The best-studied *E. coli* fimbriae are type 1 and P (or Pap for pyelonephritis-associated pili), which will be used as paradigms in the following sections to describe various aspects of fimbrial biology.

Type 1 fimbriae are by far the most prevalent fimbrial type among commensal and pathogenic *E. coli* strains, (Brinton, 1959; Buchanan *et al.*, 1985). They are encoded by the *fim* operon (Fig. 1.2) and the tip adhesin FimH interacts with a variety of mannose-containing glycoprotein receptors present in a number of different cell types (Ofek *et al.*, 1977; Connell *et al.*, 1996; Mulvey *et al.*, 1998; Martinez *et al.*, 2000). Studies over the last several years have identified natural FimH variants, with alterations as small as a single amino acid residue, that confer distinct adhesive properties to type 1 fimbriae and enhance the potential of *E. coli* to colonise the urinary tract and cause disease (Sokurenko *et al.*, 1994, 1995, 1997, 1998, 2004; Schembri *et al.*, 2000; Nilsson *et al.*, 2006; Weissman *et al.*, 2006).

P fimbriae are encoded by the *pap* (pyelonephiritis-associated pili) operon (Hull *et al.*, 1981) (Fig. 1.2) and are epidemiologically linked to pyelonephritis (Hagberg *et al.*, 1981; Plos *et al.*, 1990; Blanco *et al.*, 1997; Johnson, 1991; Guyer *et al.*, 1998). Several studies have indicated that expression of P fimbriae, and specifically the PapG adhesin, is associated with pyelonephritis in mouse models of UTI (Hagberg *et al.*, 1983; O'Hanley *et al.*, 1985; Svanborg-Eden *et al.*, 1987; Hull *et al.*, 1994; Tseng *et al.*, 2007), nonhuman primates (Roberts *et al.*, 1994), and humans (Wullt *et al.*, 2000; 2001a, b; Bergsten *et al.*, 2004), (reviewed in Lane *et al.*, 2007). Consistent

with this observation, the PapG adhesin was shown to bind to Gal(α1-4)Gal moieties present in the globoseries of glycolipids expressed by human erythrocytes and kidney cells (Leffler and Svanborg-Eden, 1980, 1981; Korhonen *et al.*, 1986b; Lund *et al.*, 1987). Glycolipid receptors with alterations in sugar composition (GbO3, GbO4, and GbO5) have been shown to be differentially recognised by distinct classes of PapG alleles (Stromberg *et al.*, 1990, 1991; Manning *et al.*, 2001), with *papG* allele II predominating among *E. coli* isolates from acute pyelonephritis (Johanson *et al.*, 1993; Dodson *et al.*, 2001; Sung *et al.*, 2001).

Fewer uropathogenic *E. coli* strains carry S fimbriae encoded by the *sfa* operon, which shares similar genetic organisation with both *fim* and *pap* operons (Fig. 1.2) (Schmoll *et al.*, 1990; Morschhauser *et al.*, 1993; Jones *et al.*, 1996). The SfaS tip adhesin mediates binding to sialic acid-containing receptors present in human kidney epithelium and vascular endothelium (Korhonen *et al.*, 1986b; Moch *et al.*, 1987; Morschhauser *et al.*, 1990), and S fimbriae are proposed to aid bacterial dissemination across host tissues, as S-fimbriate *E. coli* strains are often associated with pyelonephritis, newborn sepsis and meningitis (Korhonen *et al.*, 1985; Ott *et al.*, 1986; Hacker *et al.*, 1993). Studies using these strains have reported the presence of other fimbrial structures highly homologous to S fimbriae but with distinct receptor specificities, such as F1C fimbriae encoded by the *foc* operon (Ott *et al.*, 1987, 1988). F1C fimbriae were shown to bind to GalNac(1,4)Gal epitopes (Khan *et al.*, 2000) found in glycolipids expressed by epithelial cells of the distal tubules and collecting ducts of the kidney, as well as by bladder and kidney vascular endothelium (Virkola *et al.*, 1988). F1C fimbriae are encoded by approximately 14% of UPEC

isolates with highest occurrence among pyelonephritis strains (Pere *et al.*, 1985; 1987), including the sequenced pyelonephritis strain CFT073 (Welch *et al.*, 2002). The expression of F1C fimbriae has been implicated in UPEC colonisation and persistence in the murine urinary tract, as well as in the induction of the proinflammatory chemokine Interleukin (IL)-8 by human renal epithelial cells (Snyder *et al.*, 2005; Backhed *et al.*, 2002a).

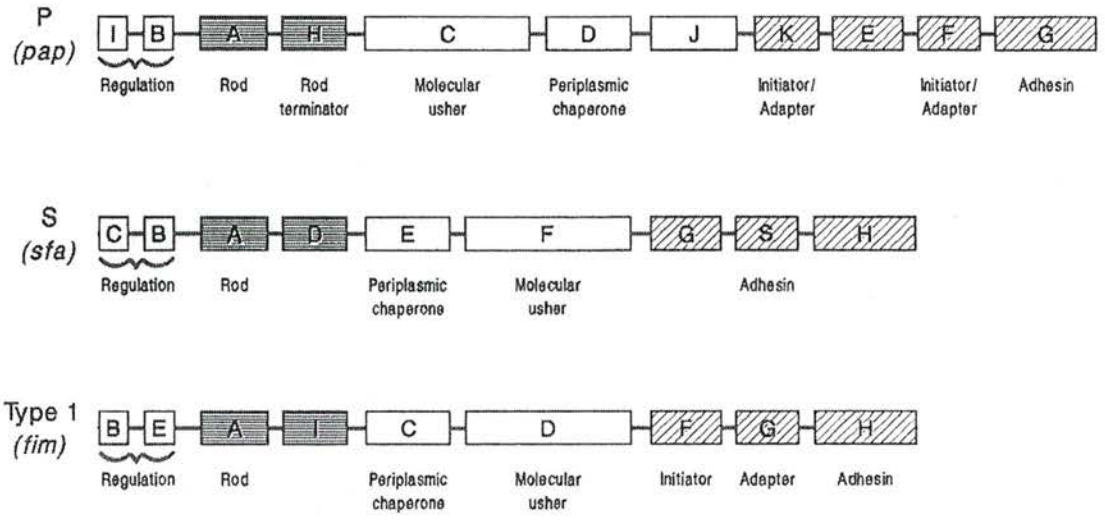


Fig. 1.2 Comparison of the genetic operons encoding for genes involved in the biogenesis of *E. coli* P, S, and Type 1 fimbriae. The *pap*, *sfa*, and *fim* operons share organisational and functional homologies. The location of genes is indicated by boxes. The first two genes in all three operons (white boxes) encode for regulatory proteins that act synergistically with *E. coli* global transcriptional regulators to determine the levels of fimbrial expression. Black boxes represent the main structural subunits composing the rigid fimbrial rod. Downstream white boxes represent genes encoding for the periplasmic chaperone and molecular usher that ensure the correct folding and translocation of fimbrial subunits across the outer membrane. Grey boxes represent minor fimbrial subunit genes encoding for the tip fibrillum, which carries the adhesin at its distal tip (image adapted from Jones *et al.*, 1996).

It is important to note that the fimbrial genetic load of UPEC strains does not necessarily associate with their fimbrial expression profile. In other words, a clinical *E. coli* isolate that has the genetic potential to express different types of fimbriae, does not necessarily express its entire repertoire during the course of infection. For example, *E. coli* strain 83972 that was isolated from a girl with asymptomatic bacteriuria who carried it for 3 years without adverse effects contains chromosomal fimbrial operons homologous to *fim, pap, uca,* and *foc* (Andersson *et al.*, 1991; Hull *et al.*, 1999). However, the strain does not express fimbriae or functional adhesins after *in vitro* culture or when recovered from the urinary tract (Andersson *et al.*, 1991; Hull *et al.*, 1999; Wullt *et al.*, 2000). Examples like that demonstrate that fimbrial genetic carriage in UPEC clinical isolates is different from fimbrial expression potential, which is important to realise in studies of bacterial pathogenicity *in vivo*.

1.4.4.1 Biogenesis of fimbriae

The biogenesis of the fimbriae described above, along with more than 30 other adhesive structures from different Gram negative pathogenic bacteria, relies on the synergistic action of highly conserved periplasmic chaperones and outer membrane usher proteins (Hultgren *et al.*, 1991; Thanassi *et al.*, 1998a; Thanassi and Hultgren, 2000). In the chaperone-usher protein secretion pathway the unfolded polypeptides are secreted through the cytoplasmic membrane in a Sec-dependent manner (Thanassi *et al.*, 2005) and are then bound by the periplasmic chaperone to allow proper subunit folding and prevent premature interactions between fimbrial subunits (Choudhury *et al.*, 1999; Hung *et al.*, 1999; Sauer *et al.*, 1999, 2002; Bann *et al.*,

2004; Vetsch *et al.*, 2004). Chaperone-subunit complexes next target the N-terminal region of the outer membrane (OM) usher (Nishiyama *et al.*, 2003; Ng *et al.*, 2004), which serves as the OM platform for ordered fiber assembly and secretion to the cell surface (Saulino *et al.*, 1998; Thanassi *et al.*, 1998b; Sauer *et al.*, 2002; Nishiyama *et al.*, 2005; Vetsch *et al.*, 2006) . The biogenesis of P fimbriae has served as a model system for elucidating the molecular and structural basis of chaperon-usher protein secretion and recent studies continue to unravel in impressive detail the mechanistic steps by which the chaperone and usher proteins (PapD and PapC) orchestrate fimbrial biogenesis (Li *et al.*, 2004; Mu *et al.*, 2005; Shu Kin So and Thanassi, 2006; Verger *et al.*, 2006; Vitagliano *et al.*, 2007).

1.4.4.2 Host immune responses to fimbriae

Fimbriae are surface organelles that mediate pathogen-host interactions; it is thus expected that they will be targeted by the host's defense mechanisms. Indeed, several studies have provided evidence suggesting that fimbriae are pro-inflammatory (Wullt *et al.*, 2001a, b; reviewed in Hedlund *et al.*, 2001a; Wullt *et al.*, 2002; Chassin *et al.*, 2006) and therefore high proportions of fimbriate bacteria during infection may be more likely to provoke a strong host immune response with resulting symptoms. Conversely, an isolate that persists without fimbriae or with very low numbers of fimbriate cells, is more likely to cause asymptomatic bacteriuria (Andersson *et al.*, 1991; Hull *et al.*, 1998, 1999; Graham *et al.*, 2001). This association between bacterial virulence, inflammation and disease severity is particularly clear in UTIs. Early epidemiological studies initially demonstrated that fimbriae-mediated attachment to uroepithelial cells is more prevalent among pyelonephritis than

asymptomatic *E. coli* isolates (Svanborg-Eden *et al.*, 1976), and subsequently showed an especially clear-cut association of P fimbriae with acute disease severity, with 90% of pyelonephritis and <20% of asymptomatic bacteriuria *E. coli* strains expressing P fimbriae (Leffler and Svanborg-Eden, 1981; Plos *et al.*, 1990; Blanco *et al.*, 1997). Since then, expression of P fimbriae and the binding of the PapG tip adhesin to glycosphingolipid receptors have been shown to trigger epithelial cell responses *in vitro* (Hedges *et al.*, 1992; Svensson *et al.*, 1994; Godaly *et al.*, 1998) and *in vivo*, both in the mouse and human urinary tract (Linder *et al.*, 1988, 1991; de Man *et al.*, 1989; Hedges *et al.*, 1991; Wullt *et al.*, 2001a, b). Human inoculation studies demonstrated a direct role of P fimbriae in promoting the early establishment but not persistence of *E. coli* in the human urinary tract and activating innate host responses (Hagberg *et al.*, 1983; Andersson *et al.*, 1991; Wullt *et al.*, 2000, 2001a, b; Bergsten *et al.*, 2004; Samuelsson *et al.*, 2004; Bergsten *et al.*, 2007). It was demonstrated that bacterial adherence through P fimbriae initiates mucosal inflammation at the site of infection through production and secretion of inflammatory mediators, including tumour growth factor (TGF), interleukin (IL)-1, IL-6, and IL-8/CXCL8, followed by recruitment of inflammatory cells (Shahin *et al.*, 1987; Linder *et al.*, 1991; Agace *et al.*, 1993a, b; Hedges *et al.*, 1991, 1994; Godaly *et al.*, 1997, 2000). The pattern recognition receptor Toll-like receptor 4 (TLR4) of the human innate immune system was proposed to control the signaling cascade triggered by P fimbriae, in a lipopolysaccharide (LPS)-independent manner (Hedlund *et al.*, 1999; Frendeus *et al.*, 2001; reviewed in Svanborg *et al.*, 2001). It is also possible that P fimbriae are inducing mucosal inflammation via other routes, involving for example the newest TLR family member, TLR11, which was shown to

be required for UPEC clearance from the mouse kidney (Zhang *et al.*, 2004). The ensuing neutrophil-mediated inflammation underlies disease pathology and links P fimbrial expression to UTI disease severity.

The emerging picture for type 1 fimbriae is suggestive of a slightly different role in induction of host immune responses. Firstly, there is no established epidemiological link between type 1 fimbriae and symptomatic disease, as this fimbrial type, in contrast to P fimbriae, occurs ubiquitously among pathogenic and commensal *E. coli* strains, with no frequency differences observed even among *E. coli* isolates of different UTI origin (Plos *et al.*, 1991). However, expression of type 1 fimbriae has been shown to enhance bacterial attachment and virulence in the urinary tract of several mouse models of UTI (Hagberg *et al.*, 1983; Hultgren *et al.*, 1985; Svanborg-Eden *et al.*, 1987; Connell *et al.*, 1996; Hedlund *et al.*, 2001b; Snyder *et al.*, 2004; Bergsten *et al.*, 2007). Moreover, type 1 fimbriae were shown to induce host responses *in vivo* by adhering to bladder mast cells and evoking vigorous exocytic and endocytic responses that were considered pivotal in determining both the duration and pathophysiology of infection (Abraham *et al.*, 2001). Lastly, interaction of type 1-fimbriate *E. coli* with bladder epithelial cells was shown to result in the rapid production of inflammatory mediators, including IL-6 and IL-8, through a mechanism dependent on LPS and TLR4 (Hedlund *et al.*, 2001b; Schilling *et al.*, 2001, 2003; Backhed *et al.*, 2002b), while a role of FimH in activation of caspase-dependent urothelial apoptosis has also been reported (Klumpp *et al.*, 2006). Recently however, a human inoculation study reported that type 1 fimbriae do not appear to play a pro-inflammatory role in the human urinary tract mucosa, as *E. coli*

asymptomatic strain 83972, reconstituted with a functional *fim* operon, failed to promote cytokine secretion and neutrophil influx in human patients compared to the isogenic *pap*+ 83972 variant (Bergsten *et al.*, 2007). The role of type 1 fimbriae in triggering the innate host immune response remains therefore less clear than for P fimbriae. It is important to note however, that the severity of disease symptoms and the ensuing tissue pathology are caused by the combined effect of *E. coli* virulence factors, such as fimbriae and toxins, together with the ability of the host to activate specific response pathways (reviewed in Svanborg *et al.*, 2001; Wullt *et al.*, 2003).

In addition to the inflammatory response, the host elicits antibody responses to invading uropathogens, but the role of anti-fimbrial antibodies in protection against UTI is not fully understood. Studies have shown that antibodies of the secretory IgA type and IgG with specific activity against O and K antigens of the infecting *E. coli* pathogen are produced locally in infected patients, while IgG and IgM are mainly found in the serum of UTI patients (Hanson *et al.*, 1977; Svanborg-Eden *et al.*, 1985). P-fimbriate *E. coli* was shown to induce production of specific anti-P antibodies in the serum of patients with acute pyelonephritis, which however failed to block fimbrial adherence *in vitro* (de Ree and van den Bosch, 1987). The role of these antibodies in resistance to human infection *in vivo* requires further study. Animal model studies have offered however some promising evidence by demonstrating that the presence of anti-P IgG antibodies in the serum and urine of mice immunized with purified Gal-Gal fimbriae correlated with their protection from subsequent experimental infection with fimbriate *E. coli* strains (Pecha *et al.*, 1989).

1.4.4.3 Regulation of fimbriae

The expression of the majority of UPEC fimbriae is not constitutive, as the energy cost associated with fimbrial biogenesis must be balanced with the benefits conferred by each particular fimbrial adhesin to the pathogen's fitness in each different niche colonised during UTI. Fimbrial expression in *E. coli* is in fact under tight regulatory control by phase variation, environmental signals and cross-regulation by other bacterial factors (reviewed in van der Woude and Baumler, 2004). The biological significance of regulating fimbrial expression via these mechanisms has generally been considered to facilitate bacterial evasion of host immune detection. It is also possible that the regulation of UPEC fimbriae by phase variation could be facilitating evasion of cross-immunity between *E. coli* serotypes, in a manner similar to that demonstrated for the phase-variable *lpf* operon encoding for long polar fimbriae in *Salmonella typhimurium* (Norris and Baumler, 1999). Another proposed role for the tight regulation of UPEC fimbrial expression is that it generates an ordered programme of gene expression events at the single-cell level to ensure optimal bacterial fitness in every spatial and temporal niche encountered during infection (reviewed in Holden and Gally, 2004; van der Woude, 2006).

1.4.4.3.1 Phase variation

Phase variation of fimbrial operons refers to a switch between on and off expressing phases. Transitioning between the two phenotypic states is reversible and heritable between generations and results in heterogeneous bacterial populations, in which only a subset of cells express the fimbrial adhesin. These cells have either inherited the on phase from the parent cell, or have switched from the off to the on expression

phase. Transitioning between expression phases occurs at certain frequencies, which are characteristic for each fimbrial operon, regulatory mechanism and bacterial species (reviewed in van der Woude and Baumler, 2004; van der Woude, 2006). These transition frequencies (off-to-on and on-to-off) determine the proportion of fimbriate cells in the bacterial population. The genetic or epigenetic switch mechanisms that dictate the heritable and reversible nature of phase variation have been extensively studied at the molecular level and mechanistically, more is known about the phase variation of type 1 and P fimbriae (reviewed in Blomfield, 2001). Expression of both fimbrial types is regulated at the level of transcription but by distinct phase variation mechanisms.

Phase variation of type 1 fimbriae is transcriptionally regulated and depends on the inversion of a short DNA sequence (*fimS*) (Eisenstein, 1981; Abraham *et al.*, 1985). The 314 bp invertible element is positioned at the *fimE-A* intergenic region of the *fim* operon (see Fig. 1.2) and contains the promoter pA directing transcription of the type 1 structural subunit *fimA*. The orientation of *fimS* and subsequently of the pA promoter dictates the expression phase of the *fim* operon. Inversion of *fimS* is an example of site-specific recombination and is catalysed by FimB and FimE recombinases, which are encoded by the *fimB* and *fimE* genes located adjacent to *fimA* (Klemm, 1986; Gally *et al.*, 1996). The activity and specificity of the two recombinases is markedly different: FimB promotes *fimS* inversion in both directions, whereas FimE promotes solely on-to-off switching at much higher frequencies than FimB (McClain *et al.*, 1991; Gally *et al.*, 1993). Recombination of the *fim* switch also requires the accessory global regulatory proteins integration host

factor (IHF) and leucine-responsive regulatory protein (Lrp) that are considered to directly contribute to the overall DNA architecture of the *fim* switch, facilitating inversion by the FimB and FimE recombinases (Eisenstein *et al.*, 1987; Blomfield *et al.*, 1993, 1997; Gally *et al.*, 1994). The abundant nucleoide-associated protein H-NS has also been implicated in the regulation of *fim* phase variation (Spears *et al.*, 1986; Schembri *et al.*, 1998).

Phase variation of P fimbriae is also under transcriptional control, but unlike type 1 fimbriae, the mechanism does not involve changes in DNA sequence. Expression of the *pap* operon was the first example of epigenetic gene regulation in bacteria (reviewed in Casadesus and Low, 2006). The *pap* epigenetic switch involves the formation of differential methylation patterns, which are generated by protein complexes on either one of two GATC methylation sites present in the *pap* regulatory region (*papI-B* intergenic region) (Blyn *et al.*, 1990; reviewed in van der Woude *et al.*, 1996). Both GATC sites, denoted distal ($GATC^{dist}$) and proximal ($GATC^{prox}$) to the main operon promoter p_{BA}, are methylation targets for DNA adenine methylase (Dam) and overlap two binding sites for the global *E. coli* regulator Lrp. Competition between Dam and Lrp for access to the overlapping sites results in two different methylation patterns that determine the expression phase (on/off) of the *pap* operon (Fig. 1.3) (Braaten *et al.*, 1991, 1992, 1994; Nou *et al.*, 1993, 1995; van der Woude *et al.*, 1992, 1995; Weyand and Low, 2000). Formation of the *pap* phase on state also requires the *pap*-encoded regulator PapI, which is transcribed from a separate divergent promoter (p_I) located upstream of the main operon promoter p_{BA}. PapI was shown to interact with Lrp-*pap* DNA complexes and

favour Lrp binding to promoter distal sites when $GATC^{prox}$ is methylated by Dam, thus promoting formation of the on phase (Kaltenbach *et al.*, 1995; Nou *et al.*, 1993, 1995; Hernday *et al.*, 2003). A second local regulator, PapB, is encoded by the first gene in the *pap* operon and has a dual role in the control of *pap* expression; it activates *papI* transcription and is thus required for off-to-on *pap* phase transition, but at high protein concentrations it acts as a transcriptional repressor of the p_{BA} promoter (Goransson *et al.*, 1989b; Forsman *et al.*, 1989). The autoregulatory function of PapB is achieved by hierarchical binding to three DNA sites: a high affinity activating site located near the p_I promoter and two lower-affinity classical repressor sites overlapping the p_{BA} promoter and the *papB* coding sequence (Forsman *et al.*, 1989). PapB was shown to bind to the high-affinity site 1 in an oligomeric fashion, through minor-grove interactions with AT-rich DNA (Xia *et al.*, 1998). PapI and PapB belong to separate families of homologous regulators encoded by various methylation-dependent fimbrial operons with the potential of complementation within the single cell (see section 1.4.4.3.3 below; Morschhauser *et al.*, 1994; Xia *et al.*, 2000; Holden *et al.*, 2001). The global *E. coli* regulator catabolite activator protein (CAP), also termed cyclic AMP (cAMP) receptor protein (CRP), is another essential regulator of *pap* transcription. Binding of CAP in a cAMP-dependent manner to a single site in *pap* regulatory DNA was shown to activate both p_I and p_{BA} divergent promoters (Goransson *et al.*, 1989b; Forsman *et al.*, 1992; Weyand *et al.*, 2001).

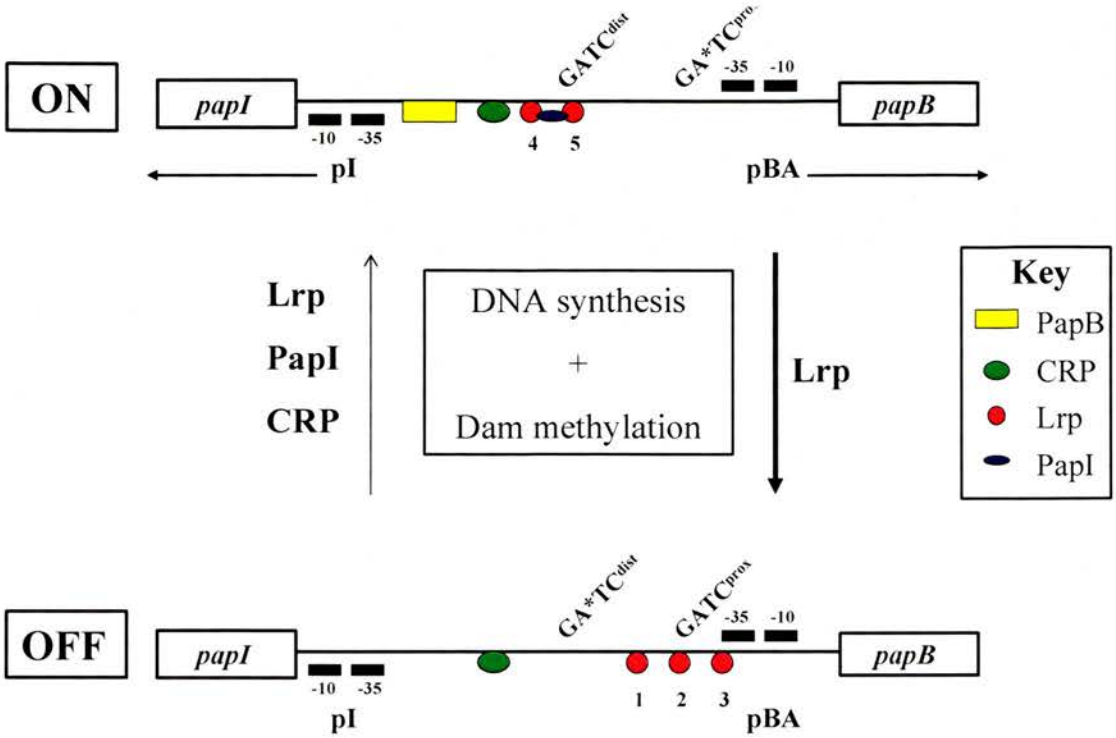


Fig. 1.3 DNA-protein complex formation in the *pap* regulatory region in phase ON and phase OFF bacteria. In phase ON cells a complex containing Lrp, PapI, CRP (or CAP) and PapB proteins is formed on the *pap* DNA region containing $GATC^{dist}$ methylation site, and transcription from both p_{BA} and p_I promoters is activated. $GATC^{dist}$ is protected from Dam methylation, whereas $GATC^{prox}$ is methylated (adenine marked with *). In phase OFF cells Lrp binding to sites 1, 2, 3 protects $GATC^{prox}$ site from Dam methylation and inhibits the initiation of transcription at the p_{BA} promoter probably by exclusion of RNA polymerase. In contrast, the adenine in $GATC^{dist}$ is methylated (*) by Dam. Loss of methylation requires DNA synthesis, which is therefore also required for *pap* phase variation. The on-to-off phase transition frequency (thick arrow) is higher than the off-to-on frequency (thin arrow). Off-to-on phase switching requires PapI and is stimulated by CRP, whereas Lrp plays a dual role in *pap* phase variation, binding as an activator at sites 4, 5 in phase ON cells and as a repressor at sites 1, 2, 3 in OFF cells. (Image adapted from Blomfield, 2001).

1.4.4.3.2 Environmental conditions

In order for bacterial pathogens to be successful in colonising varied environments it is important that gene expression is regulated in response to environmental conditions. Especially for fimbriae, which are both metabolically-expensive and immunogenic, it is crucial that expression occurs only under optimal growth conditions and is repressed promptly when fimbriation is detrimental to bacterial survival. Environmental signals that regulate fimbrial expression in *E. coli* include temperature, pH, osmolarity, carbon source, and nutrient availability, as well as the presence of iron and aliphatic amino acids (Blyn *et al.*, 1989; Goransson *et al.*, 1990; Gally *et al.*, 1993; White-Ziegler *et al.*, 2000; Schwan *et al.*, 2002; reviewed in Henderson *et al.*, 1999; Blomfield, 2001; van der Woude and Baumler, 2004). Environmental sensing is mediated by global regulatory proteins that modify fimbrial biogenesis either by influencing the frequency of phase variation or by having an epistatic effect on the phase variation mechanism. Temperature was shown to modulate phase variation of type 1 fimbriae by increasing the frequency of *fimE*-promoted on-to-off switching at 28°C, while the frequency of *fimB*-promoted recombination increased with temperature, being optimal within the range of physiological mammalian body temperatures (Gally *et al.*, 1993). The presence of aliphatic amino acids in growth media also modulated type 1 phase transition frequencies and the effect was shown to be dependent on the global regulator Lrp (Gally *et al.*, 1993; Blomfield *et al.*, 1993). Lrp is also essential for *pap* phase variation (van der Woude *et al.*, 1995), but the expression of P fimbriae seems to be unaffected by the presence of leucine in the growth medium (Braaten *et al.*, 1992). P fimbrial expression is also subject to thermoregulation with *pap* expression repressed

at low temperatures (below 26°C) by a mechanism involving the global regulator H-NS (Goransson *et al.*, 1989a, 1990; White-Ziegler *et al.*, 1990, 1998), which was subsequently shown to be an important environmental regulator of *pap* expression responsive to additional environmental signals, including osmolarity, carbon source and growth medium composition (White-Ziegler *et al.*, 2000). Expression of P fimbriae is also subject to repression in the presence of glucose (Baga *et al.*, 1985; Blyn *et al.*, 1989) with the phase variation mechanism being directly affected due to lack of the cAMP-CAP regulator, which is an essential activator of *papBA* and *papI* transcription (Goransson *et al.*, 1989b; Weyand *et al.*, 2001).

An important part of environmental sensing for bacteria involves being able to detect and respond to physical changes caused by bacterial processes, such as adhesion. For example, *E. coli* adherence mediated by P fimbriae was shown to be monitored by attachment-induced changes in the bacterial cell envelope, which activated the Cpx two-component regulatory system (Hung *et al.*, 2001). P fimbrial phase variation frequencies were directly modulated in response to bacterial adherence by the CpxR regulator, which inhibited *pap* off-to-on phase transition and expression of P fimbriae (Hernday *et al.*, 2004). A growing body of evidence is indicating that bacteria monitor fimbrial adherence to surfaces or specific receptors and regulate gene expression accordingly (Hung *et al.*, 2001; Otto and Silhavy, 2002; DiGiuseppe and Silhavy, 2003). It therefore appears that the Cpx system is not only modulating fimbrial expression in response to environmental cues, but also integrates the regulation of fimbrial expression into a wider regulatory network that includes other *E. coli* surface and virulence factors (Dorel *et al.*, 2006).

1.4.4.3.3 Cross-talk

Fimbrial expression in *E. coli* is integrated into the cell's regulatory network, which not only ensures that fimbrial production is responsive to the environmental signals described above, but also that it is subject to cross-regulation (also termed cross-talk) by other bacterial factors (Holden and Gally, 2004). Fimbrial operons can cross-talk to coordinate their expression and this can be advantageous for the uropathogen, as it limits physical interference between surface structures and permits sequential fimbrial expression, which is considered to aid colonisation of multiple niches. Early studies using *E. coli* strain KS71 demonstrated that although the strain has the potential to express type 1, P and type 1C fimbriae only a small proportion of cells in the bacterial population expressed two fimbrial types simultaneously (Nowicki *et al.*, 1984, 1985). Similarly, *E. coli* strain C1212 that contains two operons encoding for antigenically-distinct P fimbriae (*pap*-17 and *pap*-21) was shown to express only one P fimbrial variant on individual cells, with co-expression limited to only a small fraction of the bacterial population (Low *et al.*, 1987). A later study reported that a P-fimbriate uropathogenic *E. coli* strain was attenuated in the murine urinary tract due to inactivation of a second fimbrial type (Connell *et al.*, 1996). The molecular mechanism of regulatory cross-talk between type 1 and P fimbriae was the first example of fimbrial cross-talk to be elucidated (Xia *et al.*, 2000; Holden *et al.*, 2006). The mechanism involved the *pap*-encoded regulator PapB inhibiting type 1 expression by acting on the type 1 phase variation mechanism in a dual fashion: it increased FimE expression and therefore the on-to-off phase transition frequency and also blocked the activity of FimB on the *fim* switch preventing type 1 expression from switching on (Xia *et al.*, 2000). This regulatory cross-talk ensures that

expression of type 1 and P fimbriae in the bacterial surface is mutually exclusive, which is consistent with the distinct receptor specificities of the two fimbrial adhesins. A similar inhibitory effect on type 1 fimbrial expression was demonstrated by the PapB homologue SfaB of the *sfa* fimbrial operon, but not by DaaA, FaeB, FanA, FanB, ClpB, and PefB which are PapB homologues encoded by other *pap*-like operons found in *E. coli* or *Salmonella typhimurium* (Pef) (Holden *et al.*, 2001). The observation that the ability to cross-talk to a certain fimbrial operon is limited only to some members of a regulator family raises the possibility of a hierarchy in fimbrial expression. Expression of type 1 fimbriae was also shown to coordinately affect expression of P fimbriae in an inverse manner in a study investigating the transcriptome profile of type 1 phase-locked mutants in *E. coli* strain CFT073 (Snyder *et al.*, 2005).

Complementation has also been demonstrated between homologous fimbrial regulators of P and S fimbriae (Morschhauser *et al.*, 1994). Early work using the uropathogenic *E. coli* strain 536 demonstrated that deletion of pathogenicity island II (PAI II) resulted in failure of S fimbriae expression due to a block in the transcription of *sfa*-encoded genes (Knapp *et al.*, 1986). PAI II of *E. coli* 536 was later shown to contain the *prf* operon encoding for P-related fimbriae (Hacker *et al.*, 1990), while the *sfa* operon encoding for S fimbriae was localised within PAI III (Dobrindt *et al.*, 2001). The *prf*-encoded regulators *prfI* and *prfB* are homologous to *sfaC* and *sfaB* from the *sfa* operon (Goransson *et al.*, 1988) and were shown to complement mutations in *sfaC* and *sfaB* and have a positive effect on the expression of S fimbriae in *E. coli* strain 536. The same study demonstrated that mutants for *prfI* and *prfB*

repressed *sfaA* transcription *in vitro* and attenuated the virulence of *E. coli* 536 *in vivo* (Morschhauser *et al.*, 1994). Similarly, studies by other investigators demonstrated that mutations in the *pap*-encoded regulator PapI are complemented by the PapI homologues SfaC and DaaF encoded by the *sfa* and *daa* fimbrial operons found in *E. coli* (Goransson *et al.*, 1988; van der Woude and Low, 1994), as well as PefI, the regulator of Plasmid-encoded fimbriae (Pef) expressed by *Salmonella typhimurium* (Nicholson and Low, 2000). A possible outcome of this positive regulatory cross-talk is an increase in the probability of sequential fimbrial expression at the single cell level and possibly subtle changes in the total fimbriation level of UPEC isolates that contain multiple related fimbrial operons.

Regulatory cross-talk in *E. coli* is not restricted between fimbrial operons but expands to other surface structures that are important for bacterial fitness in varied specialised niches. Examples of cross-regulation have been described between fimbriae and the phase-varying antigen 43 (Ag43) autotransporter protein (Hasman *et al.*, 1999; Schembri and Klemm, 2001), as well as between phase-varying fimbriae and flagella. Ag43 is a self-recognising adhesin associated with *E. coli* aggregation and biofilm formation (Danese *et al.*, 2000) and was recently shown to promote UPEC long-term persistence in the murine urinary bladder (Ulett *et al.*, 2007). Ag43 expression is controlled by an epigenetic phase variation mechanism involving Dam and the global regulator OxyR, which represses *agn43* transcription by binding to three GATC methylation sites present in the promoter region (Henderson and Owen, 1999; Haagmans and van der Woude, 2000). The molecular mechanism of Ag43 repression by fimbrial expression in *E. coli* has been proposed to occur by a

fimbriation-induced increase in the reduced form of OxyR, which will repress *agn43* transcription (Schembri and Klemm, 2001), although recent evidence suggests that OxyR can act as *agn43* repressor in both the reduced and oxidised state (Wallecha *et al.*, 2003), posing a problem as to how, under the proposed model, Ag43 synthesis is relieved when fimbrial expression is switched off. A suggested alternative for a cross-talk mediator between the two phase-varying surface structures is the Cpx two-component signal-transduction pathway (van der Woude and Baumler, 2004). The Cpx system monitors stress in the bacterial cell envelope, including cell-surface interactions (Otto and Silhavy, 2002), the biogenesis of P fimbriae (Hung *et al.*, 2001) and its activation was recently shown to repress *pap* operon transcription by inhibiting phase transition from off-to-on (Hernday *et al.*, 2004). Another example of coordinated expression was recently identified between flagella and type 1 fimbriae in an invasive *E. coli* isolate (Barnich *et al.*, 2003). Deletion of the flagellin-encoding gene *fliC* resulted in downregulation of type 1 fimbriae and the molecular mechanism of this cross-talk appeared to act on type 1 phase variation. Additionally, a gene encoded terminally by some *E. coli* P and S fimbrial operons (*papX* and *sfaX*; Marklund *et al.*, 1992; Dobrindt *et al.*, 2001) reduced the motility of *Proteus mirabilis* by repressing flagella expression and was proposed to be a functional homologue of the MprJ regulator, which also represses transcription of the flagella regulon when expressed by the manose-resistant *Proteus*-like (MR/P) fimbrial operon in *Proteus mirabilis* (Li *et al.*, 2001). Several studies are reporting novel examples of cross-regulation between phase-varying or not phase-varying factors in *E. coli* and other bacterial species (Schembri *et al.*, 2004, 2005; Aoki *et al.*, 2005; Schwan *et al.*, 2005; Ulett *et al.*, 2006; Bryan *et al.*, 2006) and although some of the

molecular mechanisms still remain unknown, the data supports the presence of local regulatory networks that control expression of bacterial fitness and virulence factors in a coordinated or sequential manner (Fig. 1.4).

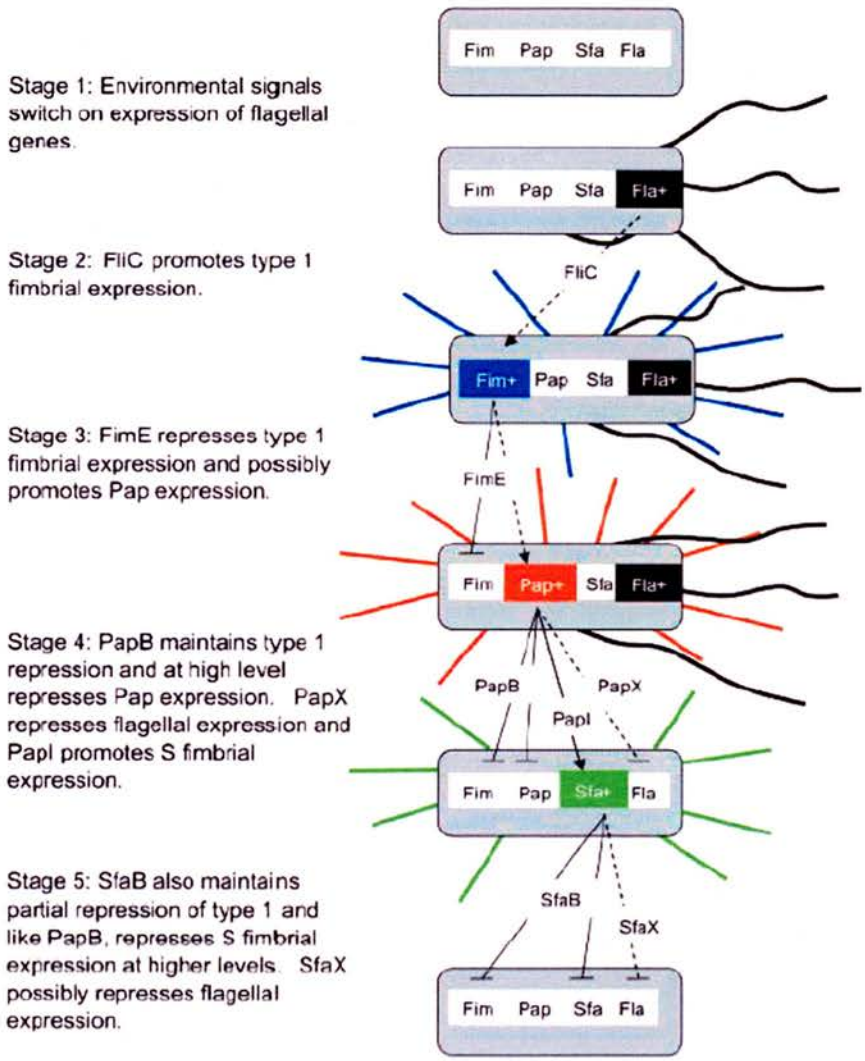


Fig. 1.4 Diagrammatic representation of one possible scenario of sequential virulence factor expression in a single *E. coli* bacterium. Type 1 fimbriae are denoted Fim and are coloured in blue, P fimbriae (Pap) in red, S fimbriae (Sfa) in green and flagella (Fla) in black. Arrows indicate positive regulation and blocked lines negative regulation. Filled lines represent cross-regulation examples supported by strong experimental evidence, while broken lines are used when some evidence is available. (Diagram adapted from Holden and Gally, 2004).

1.5 Research hypotheses and aims

As P fimbriae prove to be important in enhancing UPEC virulence and generating pro-inflammatory responses in the host, it is natural to hypothesise that the regulation of their expression must be critical for the progression of urinary tract infection and the clinical outcome. For example, fimbrial expression is more likely to be repressed in asymptomatic infections that persist for long periods in the host (Andersson *et al.*, 1991; Hull *et al.*, 1998; Hull *et al.*, 1999; Graham *et al.*, 2001), than in symptomatic infections, in which the proportion of fimbriated bacteria is likely to be higher. In this study it was hypothesised that the proportion of P-fimbriate bacteria, as determined by *pap* phase variation, is itself an important virulence determinant with the potential to affect the magnitude of host immune responses and disease severity. Therefore, the major aim was to investigate P fimbrial phase transition frequencies in uropathogenic *E. coli* isolates, as these frequencies will determine the proportion of P-fimbriate cells in the bacterial population.

To date, *pap* phase transition frequencies have been mainly measured using lysogenic reporter constructs integrated into the chromosome of *E. coli* K-12 (Blyn *et al.*, 1989, 1990; White-Ziegler *et al.*, 1990), but little is known about *pap* phase variation frequencies in clinical isolates. It is becoming increasingly evident that fimbrial expression in pathogenic *E. coli* strains is part of a cellular regulatory network controlling expression of several virulence determinants and bacterial fitness factors, suggesting that commensal *E. coli* strains do not provide a relevant genetic background for measuring phase variation frequencies of P fimbrial operons from UPEC clinical isolates. The initial research objective of this study was to establish a

system for measuring P fimbrial phase transition frequencies in a clinical isolate background.

In many isolates the combined transition frequencies (on-to-off and off-to-on) result in a predominately phase off population that allows a minor subset of bacteria to express fimbriae and colonise by binding to the uroepithelium, while the majority of bacteria are present in the urine without adhering (Hultgren *et al.*, 1985; Pere *et al.*, 1987). Previous findings in this laboratory however, suggest that some UPEC isolates have higher *pap* phase transition frequencies resulting in the majority of the population expressing P fimbriae (Miss L. Dixon, unpublished data). It was hypothesised that the differences in *pap* phase variation frequencies observed among different isolates reflect differences in *pap* regulatory DNA sequence and/or regulatory protein sequences. It was proposed that such mechanisms would allow modification of *pap* phase variation frequencies and play a role in coordinating the regulation and expression of different *pap* operons present in the same UPEC isolate. In order to investigate this hypothesis a number of *pap* regulatory regions belonging to UPEC isolates of distinct clinical origin were sequenced. Identified regions of nucleotide diversity would serve as candidates for molecular analyses in their effects on *pap* phase variation frequencies. Ultimately, the aim was to interpret sequence diversity and its effect on *pap* phase variation in the context of the clinical isolate, i.e., in a genetic background containing fimbrial operons and other virulence factors that can cross-talk.

It was finally hypothesised that if differences in *pap* regulation hold true *in vivo*, then *E. coli* strains capable of higher P-fimbriate proportions are likely to be far more inflammatory and aggressive in nature, not least because a much higher number will be interacting with epithelial cells lining the urinary tract. Host factors, such as receptor profiles and immune response, will certainly influence the clinical outcome of UTI but their effects are likely to be dependent on bacterial virulence factors. The final aim of this study was to assess the virulence potential of isogenic P fimbrial regulatory variants in a mouse model of UTI *in vivo*. The alternative hypothesis would predict that modulating *pap* phase variation frequencies and consequently P-fimbriate proportions in the bacterial population will not affect the symptoms and severity of infection and that isolates with differences in *pap* regulatory sequence and P-fimbriate proportions are equally likely to trigger the same host immune response.

In summary, the three original aims of this research study were:

1. To measure *pap* phase variation frequencies from different UPEC isolates in a clinical isolate genetic background.

2. To determine the genetic basis to differences in *pap* phase transition frequencies and assess whether this information can be predictive of cross-regulation scenarios coordinating virulence factor expression in UPEC isolates.

3. To determine whether UPEC isolates with different levels of P fimbrial expression alter the clinical outcome of infection in a mouse model of UTI.

Chapter 2

Materials and methods

2.1 Bacterial strains and plasmids

All bacterial strains and plasmids used in this study are listed in Tables 2.1A-C and 2.2. Unless otherwise stated, all bacterial strains and plasmids were constructed as part of this thesis.

2.1.1 Reference *E. coli* strains

Section (A) of Table 2.1 lists all reference *E. coli* K-12 strains and UPEC isolates used in this study. *E. coli* strain MG1655 (Guyer *et al.*, 1981) does not naturally contain *pap* operons and was frequently used as a negative control in P fimbriation experiments and PCR reactions. All genetic cloning steps were carried out with *E. coli* AAEC185 (Blomfield *et al.*, 1991).

UPEC isolates CFT073, J96, 536, and 83972 were included in this study as reference strains, as they have been studied extensively. CFT073 is a pathogenic *E. coli* strain isolated from the blood of a woman with acute pyelonephritis (Mobley *et al.*, 1990). The genome of CFT073 was the first complete UPEC genome to be sequenced and analysis revealed the presence of two *pap* operons, P1 and P2, encoding P fimbriae and one *foc* operon encoding F1C fimbriae (Guyer *et al.*, 1998; Welch *et al.*, 2002). Strain J96 belongs to an early extraintestinal *E. coli* isolate collection (Minshew *et al.*, 1978) and was isolated from a patient suffering from pyelonephritis. The virulence profile of this strain has been studied extensively (Hull *et al.*, 1981; Low *et al.*, 1984; Korhonen *et al.*, 1985; Blum *et al.*, 1995; Swenson *et al.*, 1996) and the presence of two *pap* operons, termed *pap* and *prs* (*pap*-related sequence) was demonstrated (Hull *et al.*, 1986; Karr *et al.*, 1989, 1990). *E. coli* 536 is another well-studied

pyelonephritis wild-type isolate, originally obtained from the Institute for Hygiene and Microbiology, Wurzburg, West Germany (Berger *et al.*, 1982). It contains P-related fimbrial adhesin-encoding genes (*prf*), as well as S-fimbriae-encoding genes (*sfa*), located in distinct pathogenicity islands (Hacker *et al.*, 1985; Janke *et al.*, 2001; Middendorf *et al.*, 2004) and its genome was recently sequenced (Brzuszkiewicz *et al.*, 2006). A collection of 536 isogenic fimbrial mutants (Holden *et al.*, 2006) was also included in this study. Strain 83972 is a prototype asymptomatic bacteriuria *E. coli* isolate identified by Andersson *et al.*, 1991 for its capacity to colonise the human bladder asymptomatically and persist for a long-time. Despite the presence of DNA sequences related to four fimbrial families (*fim*, *pap*, *foc*, *uca*), no functional adhesin expression was detected *in vitro* and *in vivo* (Hull *et al.*, 1999). This was later shown to be due to attenuating mutations in genes encoding type 1, P and F1C fimbriae (Klemm *et al.*, 2006; Roos *et al.*, 2006b). The strain has been used for prophylactic purposes and several trials have demonstrated its capacity to prevent other uropathogens from causing infection in the human urinary tract (Wullt *et al.*, 1998; Hull *et al.*, 2000; Darouiche *et al.*, 2001; Trautner *et al.*, 2002, 2003). The molecular mechanism of this inhibition was investigated by genome-wide expression profiling and was attributed to the strain's optimized growth properties in human urine (Roos *et al.*, 2006a, c; Roos and Klemm, 2006).

2.1.2 UPEC isolates

Section (B) of Table 2.1 lists all *E. coli* isolates from different clinical UTI cases that were used in this study.

2.1.2.1 Asymptomatic bacteriuria (ABU) isolates

In total, seventeen *E. coli* isolates from cases of asymptomatic bacteriuria (ABU) were obtained. Strains U43-U151 were isolated from pregnant women screened for significant urine cultures (>10^3 cfu per ml of urine) (Graham *et al.*, 2001). *E. coli* asymptomatic strain BR2-V1 was isolated from a female patient subjected to ileum enterocystoplasty (augmentation), BR10-V8 from a male patient that had undergone ileocaecum orthotopic bladder replacement, and BR20-V1 from a male patient subjected to ileocaecum enteroplasty (substitution) (Keegan *et al.*, 2003). *E. coli* strains HU1690-1761 were isolated from patients with asymptomatic bacteriuria and neuropathic bladders due to spinal cord and brain injuries (Hull *et al.*, 1998) and were kindly provided by R. Hull. All ABU strains included in this study were specifically selected as they contained at least one P fimbrial gene cluster (*pap*$^+$). However, it is important to note that *pap*$^+$ ABU strains represent the minority of ABU isolates, as epidemiological studies report that less than 20% express P fimbriae (Leffler and Svanborg-Eden, 1981; Blanco *et al.*, 1997).

2.1.2.2 Cystitis isolates

Fifteen *E. coli* strains isolated from cystitis cases were obtained from a Ph.D study performed by Dr. S. Keegan at the University of Newcastle (Keegan, 2001). Patients admitted to the Royal Victoria Infirmary and Freeman Hospital were diagnosed with community acquired (uncomplicated) cystitis when presenting with a combination of the following symptoms: dysuria, smelly urine, white blood cell count above >100 mm^{-3}, subrapubic pain and in more severe cases haematuria (Keegan, 2001). All

strains selected had been previously screened for the presence of various virulence determinants and were shown to be *pap*$^+$.

2.1.2.3 Pyelonephritis isolates

A total of 25 *E. coli* strains isolated from cases of pyelonephritis were collected for this study from distinct geographical locations. Strains AUTI 8, 11 and 20 were isolated from blood cultures of pyelonephritis cases (invasive strains) and strains AUTI 67-71 were isolated from urine cultures. All AUTI isolates were selected from the Keegan, 2001 study (Newcastle upon Tyne, UK) and have been characterised as *pap*$^+$. The rest of the pyelonephritis isolates were kindly provided by Prof. T. Korhonen and were collected in Helsinki, Finland. Isolate KS71 was originally used to demonstrate phase-variable expression of fimbriae in UPEC (Nowicki *et al.*, 1984). Strains AP4, AP5, AP7 and AP18 were selected from a study of P-fimbrial expression levels (Pere *et al.*, 1987). Isolates IHE 1041, 1049, 1086, 1106, 1152, 1190, 1268, 1402, and 1431 were used in a study of novel *E. coli* fimbriae by Pouttu *et al.*, 2001.

2.1.3 ZAP *E. coli* strains

The recombinant strains used in this study are listed in the order of ZAP number designations in Table 2.1C. They mainly consist of *E. coli* MG1655 and CFT073 derivatives with reporter gene fusions inserted in single-copy into the bacterial chromosome and were generally constructed by allelic exchange (see sections 2.3.14-16).

Table 2.1 Bacterial strains. (*A*) Reference *E. coli* strains (*B*) UPEC isolates (*C*) ZAP *E. coli* strains.

(A)

***E. coli* strains**	**Description/Relevant Genotype**	**Source/Reference**
MG1655	K-12 F^- λ^-	Guyer *et al.*, 1981
AAEC185	F^- λ^- *supE44 hsdR17 mcrA endA1 thi-1* Δ(*fimBEACDFGH*) Δ*recA*	Blomfield *et al.*, 1991
AAEC90A	Intermediate allelic exchange strain with *sacB-neo* cassette inserted in the *lac* locus of MG1655	Blomfield *et al.*, 1991
NEC11	BL21 (DE3) *him82*::Tn10 (*tet*^{R}) *lrp*::mTn10 (*kan*^{R})	Gally *et al.*, 1996
83972	Asymptomatic bacteriuria (ABU) strain (serotype OR:K5:H$^-$), *fim*$^+$, *pap*$^+$, *foc*$^+$, *uca*$^+$	Andersson *et al.*, 1991
CFT073	UPEC pyelonephritis strain (serotype O6:K2:H1), *fim*$^+$, *pap1*$^+$ (P1), *pap2*$^+$ (P2), *foc*$^+$ (F1C)	Mobley *et al.*, 1990
J96	UPEC pyelonephritis strain (serotype O4:K6), *fim*$^+$, *pap*$^+$, *prs*$^+$	Minshew *et al.*, 1978
536	UPEC pyelonephritis strain (serotype O6:H15:K31), *fim*$^+$, *prf*$^+$, *sfa*$^+$	Berger *et al.*, 1982
536 *prf*	536 strain with complete *prf* operon deleted	Holden *et al.*, 2006
536 *sfa*	536 strain with complete *sfa* operon deleted	"
536 *fim*	536 strain with complete *fim* operon deleted	"
536 *prf sfa*	536 strain with complete *prf* and *sfa* operons deleted	"
536 *prf fim*	536 strain with complete *prf* and *fim* operons deleted	"
536 *sfa fim*	536 strain with complete *sfa* and *fim* operons deleted	"
536 *prf sfa fim*	536 strain with complete *prf*, *sfa*, and *fim* operons deleted	"

(B)

Clinical origin	**UPEC isolates**	**Source/Reference**
Asymptomatic Bacteriuria (ABU)	U43, U68, U105, U113, U126, U151	Graham *et al.*, 2001
	BR2-V1, BR10-V8, BR20-V1	Keegan *et al.*, 2003
	HU1690, HU1691, HU1740, HU1751, HU1756, HU1757, HU1758, HU1761	Hull *et al.*, 1998
Cystitis	AUTI3, AUTI4, AUTI7, AUTI12, AUTI16, AUTI19, AUTI31, AUTI36, AUTI43, AUTI47, AUTI48, AUTI62, AUTI64, AUTI66, AUTI72	Keegan, 2001
Pyelonephritis	AUTI 8, AUTI 11, AUTI 20, AUTI 67, AUTI 68, AUTI 69, AUTI 70, AUTI 71	Keegan, 2001
	AP4, AP5, AP7, AP18	Pere *et al.*, 1987
	KS71	Nowicki *et al.*, 1984
	IHE1041, IHE1049, IHE1086, IHE1106, IHE1152, IHE1190, IHE1268, IHE1402, IHE1431	Pouttu *et al.*, 2001

Table 2.1 Bacteria strains (*continued*)

(C)

ZAP strains	Description/Relevant Genotype	Source/Reference
ZAP589	Allelic exchange of *papA* $_{J96\ F13}$*::egfp* from pLD2 in ZAP1164	Laboratory stocks
ZAP593	Allelic exchange of *papA* $_{J96\ F13}$*::lacZ* from pLD7 in ZAP1164	Holden *et al.*, 2007
ZAP594	Allelic exchange of *pap* $_{J96\ F13}$ complete operon from pLD1 in AAEC090A	"
ZAP595	Allelic exchange of *papA* $_{CFT073\ P2}$*::lacZ* from pMT06 in ZAP964	"
ZAP596	Allelic exchange of *papA* $_{AUTI72.1}$*::lacZ* from pMT07 in ZAP964	This study
ZAP597	Allelic exchange of *papA* $_{HU1761.2}$*::lacZ* from pMT08 in ZAP964	"
ZAP598	Allelic exchange of *papA* $_{IHE1086.2}$*::lacZ* from pMT09 in ZAP964	"
ZAP610	Allelic exchange of *papI* $_{J96\ F13}$*::gfp*$^+$ from pMT35 in ZAP1164	"
ZAP611	Allelic exchange of *papI* $_{CFT073\ P1}$*::gfp*$^+$ from pMT34 in ZAP1164	"
ZAP612	Allelic exchange of *papI* $_{CFT073\ P2}$*::gfp*$^+$ from pMT36 in ZAP1164	"
ZAP613	Allelic exchange of *papI* $_{AUTI72.1}$*::gfp*$^+$ from pMT37 in ZAP1164	"
ZAP614	Allelic exchange of *papI* $_{HU1761.2}$*::gfp*$^+$ from pMT38 in ZAP1164	"
ZAP615	Allelic exchange of *papA* $_{CFT073\ P2}$*::gfp*$^+$ from pMT19 in ZAP1164	"
ZAP616	Allelic exchange of *papI* $_{IHE1086.2}$*::gfp*$^+$ from pMT39 in ZAP1164	"
ZAP714	Allelic exchange of *papA* $_{CFT073\ P1}$*::egfp* from pMT01 in ZAP1164	Holden *et al.*, 2007
ZAP715	Allelic exchange of *focA* $_{CFT073}$*::egfp* from pMT03 in ZAP1164	"
ZAP833	Allelic exchange of *papA* $_{CFT073\ P1}$*::gfp*$^+$ from pMT32 in ZAP957	"
ZAP834	Allelic exchange of *papA* $_{CFT073\ P2}$*::gfp*$^+$ from pMT22 in ZAP957	"
ZAP838	Allelic exchange of *papA* $_{CFT073\ P1}$*::lacZ* from pMT31 in ZAP957	"
ZAP843	Allelic exchange of *papA* $_{CFT073\ P1}$*::lacZ* from pMT31 in ZAP964	"
ZAP955	Allelic exchange of *papA* $_{CFT073\ P2}$*::egfp* from pKC10 in ZAP1164	"
ZAP957	Intermediate allelic exchange strain with *sacB-kan* cassette inserted in the *lac* locus of CFT073	"
ZAP964	Intermediate allelic exchange strain with *sacB-kan* cassette inserted in the *lac* locus of ZAP965	"
ZAP965	CFT073 Δ*papI-B* $_{CFT073\ P1}$ Δ*papI-B* $_{CFT073\ P2}$	"
ZAP966	Intermediate allelic exchange strain with *sacB-kan* cassette inserted in the *papI-B* $_{CFT073\ P2}$ of ZAP969	"
ZAP969	CFT073 Δ*papI-B* $_{CFT073\ P1}$	"
ZAP972	Intermediate allelic exchange strain with *sacB-kan* cassette inserted in the *papI-B* $_{CFT073\ P1}$ of CFT073	"
ZAP992	Allelic exchange of *papA* $_{CFT073\ P1}$*::lacZ* from pKC41 in ZAP1164	"
ZAP996	Allelic exchange of *papA* $_{J96\ F13}$*::lacZ* from pKC37 in ZAP957	"
ZAP1000	Allelic exchange of *papA* $_{AUTI72.1}$*::lacZ* from pMT07 in ZAP957	This study
ZAP1002	Allelic exchange of *papA* $_{HU1761.2}$*::lacZ* from pMT08 in ZAP957	"

ZAP1003	Allelic exchange of *papA* $_{IHE1086.2}$*::lacZ* from pMT09 in ZAP957	This study
ZAP1137	Allelic exchange of *papA* $_{CFT073\ P2}$*::lacZ* from pMT06 in ZAP957	Holden *et al.*, 2007
ZAP1164	Intermediate allelic exchange strain, with *sacB-kan* cassette inserted in the *lac* locus of MG1655	Porter *et al.*, 2004

2.1.4 Plasmids

All plasmids used in this study are listed in Table 2.2. Commercial vectors were obtained from laboratory stocks, except for pCR®4-TOPO® vector that was supplied with the TOPO TA Cloning® Kit for sequencing (Invitrogen). Plasmid pHMG98 was kindly provided by Prof. B. E. Uhlin and plasmids pIB307 and pIB462 were a gift from Dr. I. Blomfield. All remaining plasmids were either constructed as part of this study (pMTs) or were kindly provided by other investigators in the research laboratory that this study was carried out in (pAJRs were constructed by Dr. A. J. Roe, pKCs by Miss K. Catherwood, pLDs by Miss L. Dixon and pNJHs by Dr. N. J. Holden).

Table 2.2 Plasmids

Plasmids	Relevant features	Source/Reference
pACYC184	Commercial cloning vector, Cm^R Tc^R (NEB)	Rose, 1988
pBAD18	P_{BAD} expression vector, Ap^R	Guzman *et al.*, 1995
pBR322	Commercial cloning vector, Ap^R Tc^R (NEB)	Bolivar *et al.*, 1977
pET11	Commercial expression vector, Ap^R (NEB)	Studier *et al.*, 1990
pUC18	commercial cloning vector, Ap^R (NEB)	Yanisch-Perron *et al.*, 1985
pCR®4-TOPO®	Commercial cloning vector for sequencing, TOPO® cloning site, T3 and T7 priming sites, Ap^R Kn^R,	Invitrogen™
pPap5	pBR322 with *pap* $_{J96\ F13}$ operon on 9.6 kb *Eco*RI-*Bam*HI fragment, Ap^R	Lindberg *et al.*, 1984
pAJR25	pIB307 with MG1655 *lacIA* flanking regions	Porter *et al.*, 2004

Table 2.2 - *continued*

pAJR28	pAJR25 with promoterless *egfp* cloned between MG1655 *lacIA*	Holden *et al.*, 2006
pAJR36	pAJR25 with promoterless *lacZ* cloned between MG1655 *lacIA*	Porter *et al.*, 2004
pAJR145	pACYC184 with *rpsM::gfp*$^{+}$ transcriptional fusion, Cm^{R}	Roe *et al.*, 2004
pHMG98	pACYC184 with inducible $papI_{J96}$, Tc^{R}	Forsman *et al.*, 1989
pIB307	Allelic exchange temperature-sensitive vector, pSC101 replicon, Cm^{R}	Blomfield *et al.*, 1991
pIB462	pIB307 with MG1655 *lac* flanking regions	"
pKC1	pBAD18 with inducible $papI_{CFT073\ P1}$	Laboratory stocks
pKC2	pBAD18 with inducible $papI_{CFT073\ P2}$	"
pKC8	pIB307 with CFT073 *lacIA* flanking regions	Holden *et al.*, 2006
pKC10	pAJR28 with *papI-A'* $_{CFT073\ P2}$*:: egfp* translational fusion cloned between MG1655 *lacIA*	Holden *et al.*, 2007
pKC11	pKC8 with *sacB-kan* cassette cloned between CFT073 *lacIA*	Holden *et al.*, 2006
pKC26	pAJR145 with *rpsM* removed	Holden *et al.*, 2007
pKC29	pKC8 with promoterless *lacZ* cloned between CFT073 *lacIA*	"
pKC37	pKC29 with *papI-A* $_{J96\ F13}$*::lacZ* translational fusion cloned between CFT073 *lacIA*	"
pKC41	pAJR36 with *papI-A* $_{CFT073\ P1}$*::lacZ* translational fusion cloned between MG1655 *lacIA*	"
pKC47	pKC8 with promoterless *gfp*$^{+}$ cloned between CFT073 *lacIA*	"
pKC48	pKC8 with promoterless *gfp*$^{+}$ cloned between MG1655 *lacIA*	Laboratory stocks
pLD1	pIB462 with *pap* $_{J96\ F13}$ operon from pPap5 sub-cloned into *Eco*RI-*Bam*HI	Holden *et al.*, 2007
pLD2	pAJR28 with *papI-A'* $_{J96\ F13}$*::egfp* translational fusion cloned between MG1655 *lacIA*	Laboratory stocks
pLD7	pAJR36 with *papI-A'* $_{J96\ F13}$*::lacZ* translational fusion cloned between MG1655 *lacIA*	Holden *et al.*, 2007
pMT01	pAJR28 with *papI-A'* $_{CFT073\ P1}$*::egfp* translational fusion cloned between MG1655 *lacIA*	"
pMT03	pAJR28 with *sfaC-focA'* $_{CFT073}$*::egfp* translational fusion cloned between MG1655 *lacIA*	This study
pMT06	pKC29 with *papI-A'* $_{CFT073\ P2}$*::lacZ* translational fusion cloned between CFT073 *lacIA*	Holden *et al.*, 2007
pMT07	pKC29 with *papI-A'* $_{AUTI72.1}$*::lacZ* translational fusion cloned between CFT073 *lacIA*	This study

pMT08	pKC29 with *papI-A'* $_{HU1761.2}$*::lacZ* translational fusion cloned between CFT073 *lacIA*	This study
pMT09	pKC29 with *papI-A'* $_{IHE1086.2}$*::lacZ* translational fusion cloned between CFT073 *lacIA*	"
pMT10	pIB307 with P1 *papIA* (UP and DOWN) flanking regions from CFT073	Holden *et al.*, 2007
pMT11	pMT10 with *sacB-kan* cassette cloned between CFT073 P1 *papIA*	"
pMT16	pIB307 with P2 *papIA* (UP and DOWN) flanking regions from CFT073	"
pMT17	pMT16 with *sacB-kan* cassette cloned between CFT073 P2 *papIA*	"
pMT19	pKC48 with *papI-A'* $_{CFT073\ P2}$*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	This study
pMT22	pKC47 with *papI-A'* $_{CFT073\ P2}$*::gfp*$^+$ transcriptional fusion cloned between CFT073 *lacIA*	Holden *et al.*, 2007
pMT24	pKC26 with *papI* $_{CFT073\ P1}$ promoter*::gfp*$^+$ transcriptional fusion	"
pMT31	pKC29 with *papI-A* $_{CFT073\ P1}$*::lacZ* translational fusion cloned between CFT073 *lacIA*	"
pMT32	pKC47 with *papI-A* $_{CFT073\ P1}$*::gfp*$^+$ transcriptional fusion cloned between CFT073 *lacIA*	"
pMT34	pKC48 with with *papI* $_{CFT073\ P1}$ promoter*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	This study
pMT35	pKC48 with with *papI* $_{J96\ F13}$ promoter*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	"
pMT36	pKC48 with with *papI* $_{CFT073\ P2}$ promoter*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	"
pMT37	pKC48 with with *papI* $_{AUTI72.1}$ promoter*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	"
pMT38	pKC48 with with *papI* $_{HU1761.2}$ promoter*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	"
pMT39	pKC48 with with *papI* $_{IHE1086.2}$ promoter*::gfp*$^+$ transcriptional fusion cloned between MG1655 *lacIA*	"
pNJH20	pBAD18 with $papB_{J96}$ cloned at *Eco*RI, Amp^R	Holden *et al.*, 2001
pNJH36	pET11 *papB*	Laboratory stocks
pNJH107	pUC18 with P1 *papIA* (UP and DOWN) flanking regions from CFT073	Holden *et al.*, 2007
pNJH117	pUC18 with P2 *papIA* (UP and DOWN) flanking regions from CFT073	"

All chemicals were obtained from Sigma-Aldrich, UK, unless otherwise stated.

2.2 Bacterial culture conditions and media

Bacterial strains were routinely cultured at 37°C in Luria-Bertani (LB) broth[1] or agar[2] (BDH Merck) supplemented with 25 μg ml^{-1} tetracycline, 25 μg ml^{-1} chloramphenicol, 25 μg ml^{-1} kanamycin, or 50 μg ml^{-1} ampicillin when necessary. All UPEC isolates were originally confirmed by plating on McConkey agar[3]. Colonisation factor antigen (CFA) medium[4,5] was used for promoting expression of P fimbriae (Evans *et al.*, 1977). In switching assays, M9 medium was used containing M9 salts supplemented with 2 mM magnesium sulphate ($MgSO_4.7H_2O$), 0.1 mM calcium chloride ($CaCl_2.H_2O$), 20 mM thiamine ($C_{12}H_{17}N_4OS$), and 0.2% glycerol or glucose. Essential and non-essential amino acids (Sigma-Aldrich) were added to generate rich defined (RD) medium. Human urine was obtained from eight healthy volunteers, with no history of UTI or antibiotic usage in the last six months. The urine was pooled, filtered sterilised, stored in aliquots at -20°C and used within two weeks. Urine was added to agar to make plates at a final concentration of 0.5-fold, or used neat for static growth experiments.

[1]*LB broth: For 1 L, 10 g tryptone, 5 g yeast extract, 5 g NaCl.*

[2]*LB agar: as for LB broth with 15 g bacto agar added.*

[3]*McConkey agar: prepared according to manufacturer's instructions.*

[4]*CFA broth: For 1 L, 10 g casein hydrolysate, 1.5 g yeast extract. Following sterilisation by autoclaving, filter-sterile solutions of magnesium sulphate ($MgSO_4.7H_2O$) and manganese chloride ($MnCl_2.4H_2O$) were added to a final concentration of 400 μM and 40 μM, respectively.*

[5]*CFA agar: as for CFA broth with 15 g bacto agar added.*

2.3 DNA analysis and genetic manipulation methods

Unless otherwise stated, all methods were performed as for Sambrook *et al.*, 1989.

2.3.1 Genomic DNA extraction from *E. coli* strains

Genomic DNA was extracted from *E. coli* strains using the CTAB method as described by Murray and Thompson, 1980. This method is based on the selective precipitation properties of hexadecyltrimethylammonium bromide (CTAB), which removes cell debris, proteins and polysaccharides and leaves high-molecular weight DNA that can be recovered by isopropanol precipitation. Briefly, a 1.5 ml aliquot of an overnight bacterial culture was microcentrifuged (13,400 g, 1 min) and the cell pellet was suspended in 567 µl TE buffer [10 mM Tris (pH 8.0), 1 mM EDTA]. Three µl proteinase K (20 mg/ml) and 30 µl 10% sodium dodecyl sulphate (SDS) were added, and the mix was incubated at 37°C for 1 hour. A volume of 100 µl of 5 M solution of sodium chloride (NaCl) was added and mixed thoroughly, followed by addition of 80 µl 10% CTAB solution in 0.7 M NaCl. The mix was incubated at 65°C for 10 min. An equal volume of chloroform:isoamyl alcohol (24:1) was added and the samples were microcentrifuged (13,400 g, 5 min). The aqueous supernatant was then subjected to two rounds of phenol-chloroform extraction. DNA was precipitated with 0.6 volumes of isopropanol, pelleted by microcentrifugation at 4°C (13,400 g, 10 min), and the resulting DNA pellet was washed with 0.5 ml 70% ethanol. Lastly, DNA pellets were suspended in 50 µl TE buffer and stored at -20°C.

2.3.2 Preparation of crude DNA extracts from *E. coli* strains

Crude DNA preparations were used as templates for PCR screens mainly in cloning experiments. A single bacterial colony from overnight growth on LB agar was suspended in 100 µl MQ water (Sigma-Aldrich). Bacterial suspensions were heated to 100°C for 5 min, microcentrifuged (13,400 g, 1 min) and stored at -20°C for one week. One µl supernatant volume was used as DNA template in PCR screens.

2.3.3 Plasmid DNA extraction from *E. coli* strains

Medium and high copy-number plasmids (eg. pACYC184, pBAD18, pBR322, pET11, or pCR4®-TOPO and derivatives) were extracted from bacteria cultured overnight in 5 ml LB broth supplemented with the appropriate antibiotic. The QIAprep™ spin plasmid miniprep kit (QIAGEN) was used according to manufacturer's instructions and all buffers were supplied with the kit. Plasmid DNA was eluted in 50 µl MQ water, yielding approximately 300 ng μl^{-1}, as determined by quantitative analysis on an agarose gel. For low copy-number plasmids (eg. pIB307 and derivatives) the plasmid contents of four 5 ml culture aliquots were pooled through a single QIAGEN spin column of the QIAGEN miniprep kit and eluted in 35 µl MQ water to achieve a yield of ~150 µg ml^{-1}. Alternatively, the QIAGEN plasmid midi and maxi kit was used according to manufacturer's instructions with expected yields of 20-100 ng low-copy plasmid DNA.

2.3.4 DNA amplification by PCR

Polymerase chain reaction (PCR) was carried out in sterile thin-wall tubes in 50 µl reaction volumes, containing 1 µl template DNA (1:50 dilution for genomic DNA),

0.5 μM primer (Table 2.3), 0.5 U *Taq* DNA polymerase (Roche) and 5 μl 10 × PCR reaction buffer (Roche) [100 mM Tris-HCl (pH 8.3 @ 20°C), 15 mM $MgCl_2$, 500 mM KCl], 0.4 mM mixture of dNTPs (Roche) and filter-sterile MQ water to the final volume. Amplification was performed in a Thermo-Hybaid PCR Express cycler programmed for an initial 4-minute DNA denaturing step at 94°C, followed by 25 cycles of 45 sec at 94°C, 45 sec at primer-optimal annealing temperature, and then 1 min 30 sec at 72°C, with a final 10-min extension step at 72°C. Amplification products were resolved by agarose gel electrophoresis, as described in the following section. Negative controls, in which all reagents but template DNA were added (DNA blank), were performed in every PCR run and all reactions were repeated if the controls contained amplification products (faint bands). Positive and negative controls were also performed for every primer combination using DNA template from positive and negative reference *E. coli* strains. Test strains were regarded as positive if they contained amplification products of the correct size (bp) and negative if they failed to give correct-size products in two independent PCR repeats.

For cloning and sequencing purposes, high-fidelity amplification of genomic DNA fragments was achieved using the Expand Long Template PCR System (Roche) according to manufacturer's instructions. This system is based on the same PCR principles as described above, but contains an enzyme mix of *Taq* DNA polymerase and *Tgo* DNA polymerase with inherent proofreading activity, ensuring a lower error rate than PCR using *Taq* DNA polymerase alone.

Table 2.3 Oligonucleotide primers for PCR. The primer name and associated sequence (5'-3') is listed, together with the optimal annealing temperature (T_a) for each primer pair. Primer tail sequence is shown in lower case and any included restriction sites are underlined. The table is sectioned to group primers used for similar purposes: top – amplification of *pap* regulatory regions, second – construction of *E. coli* strain ZAP965, third – amplification of *papI* promoter regions, last – gene screening and other.

Primer ID	DNA Sequence (5'-3')	T_a (°C)	Comment	Reference
P1_480	cgcggatccGAAGTTTATGGCGTTTGTATTTTG	66	$P1_{CFT073}$ regulatory region	Holden *et al.*, 2007
papAP1_1910	cgggtaccCCCCTGAGGAATAGTTGG			
P2_52	cgcggatccCTGATTCGTCATTCTATTCTTATTGA	66	$P2_{CFT073}$ regulatory region	"
papAP2_1507	cgggtaccGCCTTGAGGGATAGATGCA			
F1C_74	cgcggatccGGAGGTTACAGGATTTGCTTTTGTA	66	$F1C_{CFT073}$ regulatory region	This study
focA_1689	cgggtaccGTTAACCGTGGTGACAGCAG			
J96 P for	cgcggatccGGCCATGCAGTAAAACCGG	53	Pap_{J96} regulatory region	Holden *et al.*, 2007
J96 P rev	cgggtaccCCCCTGTGGAATAGTTGGAG			
CFTP1+2 for	cgcggatccGTTTCAGTGAAGCATGCCCAC	58	Generic *pap* regulatory region	"
CFTP1+2 rev	cgggtaccCATAAATAACAACCTCTTTTTCATTAC			
P1UPfor	gctgagctcCGGTTCAGTAATATCTGA	56 (P1)	$P1,2_{CFT073}$ upstream flanking region	Holden *et al.*, 2007
P2UPfor	gctgagctcGTGCCGACGATCCCCTGA	57 (P2)		
P12SUPrev	gctggatccTTCACTGAAACAGATAAAWGT[a]			
P12DNfor	cgggatccatggtaccTCGGTTATTGCCGGTGCG	59 (P1)	$P1,2_{CFT073}$ downstream flanking region	"
P1DNrev	cgctctagaCCATCTTTTCTGACGGCAGC	55 (P2)		
P2DNrev	cgctctagaCTATTATCTTTCTTAACAAATGC			
papIP1for	cgctctagaCATATATTCACTCATCTCACTG	55	P1 *papI* promoter region	Holden *et al.*, 2007
papIP1rev	cgctctagaGTTTCCCCCTTCTGTCGGGC			
papI 15_KpnI(F)	cgggtaccCATATATTCACTCATCTCACTG	53	Generic *papI* promoter region	This study
papI prom_CAP_ BglII(R)	cgagatctCAGAAAAATAAMACRCAAAATGTG[a]			
PfamA	GTGCAGATTAACATCAGGGG	60	*papF* screen	Keegan, 2001
PfamB	ATGCTCATACTGGCCGTGGT			
papB1	CCGGAATTCTATATCCAGGGGCCC	54	*papB* screen	Holden *et al.*, 2001
papB2	CCAGAATTCCCGGAATGCCATAAT			
sacB 5'	CAGCTCTTTGAACATCAACGG	56	*sacB* screen	Emmerson *et al.*, 2006
sacB 3'	CTTGGTAGCCATCTTCAGTTCC			
pIB 5'	AGACAAATGGATCTCGTAAGCG	55	pIB307 backbone sequence screen	"
pIB 3'	GCTGTAACAAGTTGTCTCAGGTGT			
PapB site1_F	CTCACTGTAACAAAGTTTCTTCG	58	PapB binding site 1 variants	This study
PapB site1_R	CAGCATAAAAGATCGTCTAAATG			
seq papB (F)	GCGAGAAATACCAGATGAATAATGG	53	Internal primers for sequencing	"
seq papB (R)	CCATTATTCATCTGGTATTTCTCGC			

[a] W: A+T, M: A+C, R: A+G

2.3.5 DNA electrophoresis

DNA fragments were resolved by electrophoresis in gels containing 0.8-1.5% agarose in 1 × Tris-borate/EDTA (TBE) buffer [89 mM trishydroxymethyl aminomethane (Fisher), 89 mM boric acid, 2 mM ethylenediamine tetraacetic acid (EDTA)] and 1 µl of 1 M ethidium bromide for DNA staining. Electrophoresis was typically performed in 1 × TBE buffer at constant voltage (100 V) using Bio-Rad systems. DNA samples contained 10 × DNA loading buffer (Invitrogen) and fragment sizes were determined by comparison to molecular markers (1 Kb, 100 bp or λ *Hind*III ladder; Invitrogen). DNA fragment visualisation was performed with UV illumination in a Flowgen MultiImage™ light cabinet (Shenstone, England) and gel images were captured and analysed using ChemiImager 4000i v.4.04 software.

2.3.6 DNA enzymatic digestion

All restriction endonucleases were purchased from New England BioLabs (NEB, UK) and were used according to manufacturer's instructions. Typically, an appropriate concentration of genomic or plasmid DNA was mixed with 4 µl 10 × appropriate NEB buffer, 2 µl restriction endonuclease (NEB), 4 µl 10 × bovine serum albumin (BSA; NEB), if required, and filter-sterilised water to a final volume of 40 µl. Incubation at the recommended temperature was carried out for approximately 2 hours.

2.3.7 DNA fragment purification

DNA fragments from PCR (100 bp – 10 Kb) and other enzymatic reactions were purified from contaminants using the QIAquick PCR purification kit (QIAGEN)

following manufacturer's instructions. Extraction of DNA fragments from agarose gels in TBE buffer was performed using the QIAquick gel extraction kit (QIAGEN) according to manufacturer's instructions. All buffers were supplied with the kits. Purified DNA fragments were reconstituted in 30-50 µl filter-sterilised MQ water.

2.3.8 DNA fragment ligation

To generate recombinant DNA plasmids, vector and insert DNA was digested with appropriate restriction endonucleases and purified as described in section 2.3.7. If required, the linearised vector was treated with shrimp alkaline phosphatase (SAP; NEB) to dephosphorylate 5' restricted termini and prevent vector religation. The DNA concentration of the digested vector and insert were determined by quantitative analysis on an agarose gel and the amount of insert (ng) used in a ligation reaction with 50-100 ng of vector DNA was calculated using the following equation:

$$\text{Insert (ng)} = \frac{\text{ng of vector} \times \text{bp of insert}}{\text{bp of vector}} \times \text{Insert:Vector ratio}$$

where Insert: Vector ratio used was 3:1. The DNA fragments were mixed with 1 µl T4 DNA ligase (NEB) and 1 µl 10 × ligase reaction buffer (NEB) [50 mM Tris-HCl (pH 7.5 @ 25°C), 10 mM $MgCl_2$, 1 mM ATP, 10 mM dithiothreitol (DTT), 25 µg/ml BSA] in a final volume of 10 µl. For ligation of complementary ends the mixture was incubated at 16°C overnight.

2.3.9 Southern blotting and hybridisation

The number of *pap* operons present in each *E. coli* isolate used in this study was detected by Southern hybridisation. Reference *E. coli* strains CFT073, J96, 536 and

AAEC185 (Table 2.1A), and pPap5 plasmid (Table 2.2), were used as controls for probe specificity. Genomic DNA from each isolate and plasmid DNA were digested overnight with *Hind*III or *Bgl*II restriction endonucleases, resolved on 1% agarose gels and blotted onto a positively-charged nylon membrane (Hybond^{TM}-N^{+}, Amersham Biosciences) using the capillary transfer method described in Sambrook *et al.*, 1989. Briefly, gels were prepared for DNA transfer by firstly soaking in 0.25 M HCl for 10 min with gentle agitation at room temperature, to depurinate DNA, and then washed in dH_2O followed by soaking in 0.5 M NaOH, 1.5 M NaCl for 25 min, to denature DNA. Finally, gels were washed in dH_2O and neutralised by soaking in 0.5 M Tris-HCl (pH 7.5), 1.5 M NaCl for 30 min. Treated gels were then placed upside down onto a strip of 3MM Whatman paper that was pre-soaked in transfer buffer and rested on a solid support with its ends dipped into a reservoir of transfer buffer [10 × SSC buffer: 1.5 M NaCl, 0.15 M trisodium citrate ($Na_3C_6H_5O_7$)]. A piece of nylon membrane (Hybond^{TM}-N^{+}, Amersham Biosciences) was placed on top of the gel and covered by 3 pieces of 3MM Whatman paper. Lastly, a stack of dry paper towels was secured with a weight on top of the blotting papers to facilitate the upward flow of liquid necessary for the capillary transfer of DNA from the gel onto the nylon membrane. Transfer was allowed to occur overnight and the retrieved membranes were exposed to UV light for 2 min to cross-link DNA onto the membrane. Membranes were then pre-hybridised in 0.3 ml cm^{-2} pre-heated hybridisation buffer [5 x SSC buffer, 0.1% SDS, 1:20 liquid block (Amersham Biosciences), 5% dextran sulphate] for 2 hours at 60°C. Fifty ng of denatured hybridisation probes were added and hybridisation was allowed to occur overnight at 60°C. The hybridisation probe used for detecting all *pap* operons present in each

isolate was complementary to the *papB* gene and was amplified using primers papB1 and papB2 (Table 2.3). The amplified product was labelled with fluorescein-dUTP using the ECF Random Prime Labelling kit (Amersham Biosciences) according to manufacturer's instructions. Following hybridisation, the membrane was washed with high stringency buffers [Wash Solution 1: 1 x SSC buffer, 0.1% SDS and Wash Solution 2: 0.5 x SSC buffer, 0.1% SDS] at 60°C and the signal was amplified using an ECF Signal Amplification System (Amersham Biosciences) according to manufacturer's instructions. The fluorescence signal was detected with a Fuji FLA-2000 scanner with a 580 nm filter.

2.3.10 Transformation of competent *E. coli*

E. coli strains were made competent for transformation with plasmid DNA following methods described in Sambrook *et al.*, 1989.

2.3.10.1 Preparation and transformation of chemically-competent *E. coli* cells

E. coli K-12 strains were cultured in 5 ml LB broth shaking at 200 rpm at 37°C overnight. The following day the culture was diluted 1:100 in LB broth and cultured under the same conditions until bacterial growth reached an optical density OD_{600} = 0.4-0.6. Cultures were incubated on ice prior to centrifugation at 4°C (3220 g, 8 min). Cell pellets were gently suspended in 0.4 volumes (of the original culture volume) of ice-cold transformation buffer I [TFBI: 30 mM potassium acetate ($KC_2H_3O_2$), 10 mM $CaCl_2.2H_2O$, 100 mM KCl, 15% glycerol, 45 mM $MnCl_2$] and incubated on ice for 15 min. Cells were harvested as above and gently suspended in 0.04 volumes of ice-cold TFBII [75 mM $CaCl_2$, 10 mM KCl, 15% glycerol, 10 mM Na-MOPS (pH

7.0)]. Cells were incubated on ice for 15 min prior to aliquoting in 100 µl volumes and storing at -70°C.

Typically, 100 µl of chemically-competent *E. coli* cells were transformed with 1 µl plasmid DNA or 10 µl ligation reactions (see section 2.3.8) by mixing, incubating on ice for 30 min and heat-shocking in a 42°C waterbath for 45 sec. Reactions were incubated on ice for 2 min immediately after heat-shocking and then 350 µl of salt-optimised and carbon medium (SOC; Invitrogen) was added. Transformants containing pIB307-based temperature-sensitive vectors were cultured at 28°C, shaking at 80 rpm for 2 hours. For all other plasmids, transformants were recovered after one hour incubation at 37°C. A 200 µl culture aliquot was plated on LB agar supplemented with the appropriate antibiotics and incubated overnight at the appropriate temperature.

2.3.10.2 Preparation and transformation of electro-competent *E. coli* cells

All plasmid transformations in wild-type UPEC isolates and derivatives were performed by electroporation. *E. coli* cultures for preparation of electro-competent cells were prepared in a similar manner to those for chemically-competent cells. Culture pellets were initially washed with 0.5 volumes of ice-cold 10% glycerol and harvested by centrifugation at 4°C (3220 g, 8 min). A second wash in ice-cold 10% glycerol followed and cells were finally suspended in 0.005 volumes of ice-cold 10% glycerol and incubated on ice prior to aliquoting in 100 µl volumes and storing at -70°C.

Typically, 100 µl of electro-competent *E. coli* cells were transformed with 1 µl purified plasmid DNA by gentle mixing, incubating on ice for 30 min and electroporating at 2.5 KV in a 1.5 ml electroporation cuvette (Flowgen, UK). A volume of 1 ml of SOC medium was immediately added, and the suspension transferred to a 1.5 ml Eppendorf tube. Transformants containing pIB307-based temperature-sensitive vectors were cultured at 28°C, shaking at 200 rpm for 2 hours. For all other plasmids, transformants were recovered after one hour incubation at 37°C. A 200 µl culture aliquot was plated on LB agar supplemented with the appropriate antibiotics and incubated overnight at the appropriate temperature.

2.3.11 TOPO TA cloning

The TOPO TA cloning® kit for sequencing (Invitrogen) provides a highly efficient cloning strategy for the direct insertion of PCR products into the pCR4®-TOPO® vector and their sequencing. PCR products amplified by *Taq* DNA polymerase contain poly-A 3' overhangs, created by the addition of single deoxyadenosine (A) to the 3' end of PCR products during amplification, by *Taq*'s non-template dependent terminal transferase activity. Amplification products generated using the Expand Long Template system (Roche) are lacking poly-A 3' overhangs, as they are removed by *Tgo* DNA polymerase. For inserting these products into the pCR4®-TOPO® vector, PCR reactions were incubated post-amplification with 1 U of *Taq* DNA polymerase at 72°C for 10 min. TOPO TA cloning® relies on ligating the linearised pCR4®-TOPO® vector containing poly-T 3' overhands with the PCR products carrying poly-A 3' overhangs and was carried out according to

manufacturer's instructions. Ligation reactions were subsequently transformed into One shot® TOP10 chemically-competent *E. coli* cells (Invitrogen).

All *pap* regulatory regions, amplified by Expand Long Template PCR (Roche) using primers [CFTP1+2 for and rev] (Table 2.3), were cloned into pCR4®-TOPO® vector according to manufacturer's instructions. The use of generic primers ensured amplification of all *pap* regulatory regions present in the same UPEC isolate and meant that PCR reactions potentially contained more than one type of amplification products. Preliminary evidence for this was provided by screening several TOPO-clones from the same cloning reaction for insert size (bp) by *Eco*RI diagnostic digestion (see Fig. 4.7). Typically, two TOPO®-clones were screened for UPEC isolates with one *pap* operon, five for isolates with two *pap* operons and eight for isolates with three *pap* operons (the number of *pap* operons per isolate was determined by Southern blotting, as described in section 2.3.9). The high efficiency of TOPO® cloning ensured that almost all selected clones contained inserts and hence, the excess number of screened clones was sufficient for obtaining different amplification products.

2.3.12 DNA Sequencing

All DNA sequencing for this study was performed by MWG-Biotech (Single Read DNA sequencing services), according to the company's guidelines. Each sequencing reaction was guaranteed to contain up to 900 bp of good-quality sequence (PHRED 20).

All *pap* regulatory regions examined in this study were cloned into pCR4®-TOPO® vector by TOPO TA cloning® (section 2.3.11) and sequenced using 1 μg plasmid DNA as template, and T3 and T7 primers that initiate sequencing reactions at priming sites flanking the TOPO® cloning site. A second pair of internally-situated primers [seq papB (F) and (R), Table 2.3] was used in order to achieve 100% double-stranded sequence coverage of good-quality bases for the ~1.2 Kb inserts. To ensure sequencing of all *pap* regulatory regions present in each UPEC isolate, two TOPO-clones were sequenced for isolates with one *pap* operon, five for isolates with two *pap* operons and eight for isolates with three *pap* operons.

A double-stranded DNA sequence for each *pap* regulatory region was assembled by aligning four contigs obtained from the four individually-primed sequencing reactions and covering both directions. Contig alignment was performed using ContigExpress, a component of Vector NTI Suite 8.0. Sequence chromatograms were viewed with CodonCode Aligner v.1.3.4 and were visually inspected for all singleton mutations to ensure consistency between contigs and strands.

2.3.13 DNA sequence analysis

The assembled DNA sequences for all *pap* regulatory regions were aligned using ClustalX 1.83 (Thompson *et al.*, 1997). The multiple sequence alignment (msa) of all sequenced clones aided in two tasks: firstly, identifying clones from each UPEC isolate that carried *pap* regulatory sequences from the same or different operons, and secondly, identifying polymorphic base positions across the complete sequenced region. For this study, all point mutations were considered real sequence variation if

they occurred in all sequenced clones representing the same operon from the same isolate or in several clones from different UPEC isolates. Singleton mutations that occurred in only one clone out of the ~200 sequenced clones were assumed to be errors, potentially incorporated by PCR, and were converted to the base observed in 99.5% of the sequences.

DNA polymorphism analyses and mean diversity estimates were performed with DnaSP 4.0 software (Rozas *et al.*, 2003). Construction of phylogenetic trees of PapI protein variants was performed with MEGA v.3.1 software (Kumar *et al.*, 2004) using the neighbor-joining distance method (Saitou and Nei, 1987) and support for predicted tree nodes was provided by the bootstrap method (1000 replicates) (Felsenstein, 1988). Synonymous (dS) and nonsynonymous (dN) mean pairwise diversities were calculated with MEGA v.3.1 software (Kumar *et al.*, 2004) using the Nei-Gojobori method (Nei and Gojobori, 1986) and differences between synonymous and nonsynonymous distances were compared with a codon-based Z-test to test for positive selection (hypothesis tested: dN>dS).

2.3.14 Allelic exchange

All chromosomal deletions and gene replacements in *E. coli* K-12 and CFT073 were performed by a temperature-sensitive allelic exchange plasmid system, as described in Blomfield *et al.*, 1991, Porter *et al.*, 2004 and Emmerson *et al.*, 2006.

The method relies on the initial construction of an intermediate strain in which the *sacB-kan* cassette is inserted into the chromosome replacing the desired wild-type

chromosomal sequence (eg. ZAP1164, ZAP957 and ZAP964 in Table 2.1C). The *sacB-kan* cassette contains the *sacB* gene from *Bacillus subtilis* and the neomycin (kn^R) gene from Tn5. Chromosomal insertion of the *sacB-kan* cassette firstly involves cloning of the *sacB-kan* cassette between homologous flanking regions of DNA in a temperature-sensitive, chloramphenicol-resistant plasmid (pIB307-based). This plasmid is transformed into the desired *E. coli* genetic background and selection for plasmid chromosomal integration occurs at 42°C in the presence of chloramphenicol, while plasmid backbone excision occurs at 28°C in the presence of kanamycin. Successful integrants are identified by growth in kanamycin-containing media but not chloramphenicol. PCR screens for the presence of *sacB* gene and absence of pIB307 plasmid backbone and/or wild-type sequence replaced by the *sacB-kan* cassette were also carried out for confirmation of successful integrants.

For generation of clean chromosomal deletions or gene replacements the intermediate allelic exchange strain carrying the *sacB-kan* cassette was transformed with the temperature-sensitive plasmid containing either the homologous flanking regions alone or DNA cloned between the flanking regions. Growth at 42°C in the presence of chloramphenicol selected for plasmid chromosomal integration, and was followed by growth at 28°C in the absence of antibiotics to promote excision of the plasmid together with the *sacB-kan* cassette that was exchanged with the desired target sequence. Elimination of unsuccessful integrants was achieved by culturing on LB agar containing 6% sucrose, which is cleaved by the product of the *sacB* gene (levansucrase) and results in levels of levans that are toxic for many *E. coli* strains. Successful integrants were sucrose tolerant, kanamycin and chloramphenicol

sensitive, and were confirmed by PCR screens for absence of the *sacB* gene and pIB307 backbone sequence (for primers see Table 2.3) and presence of the exchanged sequence or a 'scar' region resulting from the clean chromosomal deletion.

2.3.15 Construction of *E. coli* strains containing chromosomal single-copy reporter fusions

Plasmids and strains for all single copy reporter fusions were constructed using a similar strategy. For reporter fusions to the p_{BA} promoter (main *pap* operon promoter), the *pap* regulatory region including *papI* and *papB* genes was firstly amplified by PCR using primers listed in the top panel in Table 2.3. For reporter fusions to the p_I promoter, the regulatory region downstream the first proposed translational start site of *papI*, including PapB site 1 and the CAP binding site, was amplified by PCR using primers [papI 15_KpnI (F)] and [papI promoter_CAP_BglII (R)] (Table 2.3). The amplified promoter regions were subsequently cloned into *Bam*HI-*Kpn*I sites in allelic exchange vectors containing either *lacZ* (pAJR36 and pKC29), *egfp* (pAJR28) or *gfp*$^+$ (pKC47 and pKC48) reporter genes, flanked by DNA sequences homologous to K-12 (pAJR36, pAJR28 and pKC48) or CFT073 (pKC29 and pKC47) chromosomal *lacI* and *lacA*. This directional cloning step placed the reporter gene under the control of either p_{BA} or p_I promoter. The plasmids were then transformed into intermediate allelic exchange strains ZAP1164, ZAP597 and ZAP964 that contained the *sacB-kan* cassette in the chromosomal *lac* locus of MG1655, CFT073 wild-type and ZAP965, respectively. Using the strategy of allelic exchange (section 2.3.14) the reporter fusions were integrated into the *E. coli*

chromosome in single-copy, generating the majority of the ZAP strains listed in Table 2.1C.

2.3.16 Construction of *E. coli* CFT073 fimbrial mutants

Strain ZAP965 was generated through a series of intermediate strains (see Table 2.1C) to delete the regulatory region from the 3' end of *papI* to the 5' end of *papA* of both P1 and P2 operons of *E. coli* CFT073. The flanking regions incorporating *papI* were amplified by PCR using primers P1UPfor (P1) or P2UPfor (P2) together with P12SUPrev (both) and the flanking regions incorporating *papA* were amplified by PCR using P12DNfor (both) and P1DNrev (P1) or P2DNrev (P2) (for primer sequences see Table 2.3). The flanking regions were sequentially cloned by Dr. N. J. Holden into the *Sac*I-*Bam*HI (*papI*) and *Bam*HI-*Xba*I (*papA*) sites in pUC18 to generate pNJH107 (P1) and pNJH117 (P2). Both upstream and downstream flanking regions were then sub-cloned into the *Sac*I-*Xba*I sites in pIB307 generating pMT10 (P1) and pMT16 (P2). Subsequently, the *sacB-kan* cassette was cloned in the *Bam*HI site of pMT10 and pMT16 to generate the allelic exchange vectors pMT11 (P1) and pMT17 (P2) (for plasmids see Table 2.2). *E. coli* CFT073 was transformed with pMT11 by electroporation and the *sacB-kan* cassette was integrated into the regulatory region of P1 operon by homologous recombination and counter selection, as described in section 2.3.14. The resulting strain ZAP972 was transformed with pMT10 and the chromosomal *sacB-kan* cassette was excised from the P1 region, resulting in strain ZAP969 with *papIB* (P1) deleted. ZAP969 was subsequently transformed with pMT17 to insert the *sacB-kan* cassette in the regulatory region of the P2 operon. The resulting strain ZAP966 was transformed with pMT16 to excise

the *sacB-kan* cassette, deleting the *papIB* P2 region and generating strain ZAP965, a CFT073-derivative with both *pap* regulatory regions deleted (for confirmation of strain ZAP965 see Fig. 3.4).

2.4 Indirect immunofluorescence staining of P fimbriae

For labelling P-fimbriae on the bacterial cell surface, CFA agar streak plates of UPEC isolates cultured overnight at 37°C were washed with 3 ml of phosphate buffered saline [137 mM NaCl, 10 mM Na_2HPO_4, 2.7 mM KCl, 1.4 mM K_2HPO_4, adjusted to pH 7.4] (PBS; Oxoid). Bacterial suspensions adjusted to OD_{600} = 0.6 were prepared from the bacterial slurry and washed twice in 1 ml PBS and then mixed with the appropriate primary anti-P rabbit polyclonal serum diluted 1:50 in PBS and incubated for 30 min at room temperature. For detection of J96 F13 P fimbriae an anti-PapA antibody supplied by Prof. B.E. Uhlin was used (Nilsson *et al.*, 1996), whereas for detection of the *E. coli* CFT073 P1 and P2 fimbriae polyclonal OM12 serum provided by Prof. T. Korhonen was used (Pere *et al.*, 1987). Different primary antibodies were used reflecting the different P fimbrial variants being detected. Excess primary antibody was removed with three washes in 1 ml PBS and cells were then mixed with the secondary goat Alexa Fluor 488 conjugate to anti-rabbit IgG (1: 500 in PBS; Molecular Probes) for 30 min at room temperature. Excess secondary antibody was removed with three washes in 1 ml PBS.

2.5 Methods for fluorescence detection

2.5.1 Flow cytometry (FC)

Fluorescence either from serum-labelled P-fimbriate bacteria (section 2.4) or strains possessing a single-copy fluorescent reporter fusion was detected using a FACSCalibur flow cytometer (Becton Dickinson). Samples of antibody-labelled P-fimbriate bacteria were prepared as described in 2.4, whereas fusion strains were collected with 3 ml PBS from CFA agar plates, adjusted to $OD_{600} = 0.6$ and washed twice in 1 ml PBS. Acquisition and analysis of flow cytometry data was performed using the CELLQuest software. The selected R1 region was optimized in order to exclude small particles and debris using sized beads (Becton Dickinson) and multiple *E. coli* strains. The gate for the detection of fluorescence signals was set such that cells under investigation were considered positive when their fluorescence intensity (FL-1 height) exceeded that of all but a very small fraction (0.5%) of the negative control population of the same UPEC strain grown under the same conditions but labelled only with the secondary antibody or the non-fluorescent parent strain of a fusion strain. Typically, a population of 20,000 bacterial cells (events) was assessed in every sample run.

2.5.2 Fluorescence microscopy

Visualisation of serum-labelled P-fimbriate bacteria and strains containing fluorescent reporter fusions was performed using fluorescence microscopes by Leica Microsystems and image acquisition and analysis was performed using Improvision OpenLab software. Bacterial samples were prepared as described above, fixed onto

glass slides with 4% paraformaldehyde (PFA) and mounted with fluorescence mounting medium (Dako).

2.5.3 Whole population fluorescence measurements

To investigate *papI::gfp*$^+$ expression in different genetic backgrounds, *E. coli* strains AAEC185, CFT073 and ZAP965 (Table 2.1A&C) were transformed with plasmids pKC26 and pMT24 (Table 2.2) and single transformants cultured in CFA medium supplemented with 25 μg ml^{-1} chloramphenicol at 37° overnight. Samples were diluted to OD_{600} = 1.0 and three replicate aliquots of 200 μl per well were assayed in a Flurostar fluorimeter (FLUOstar Optima), using 485 nm absorbance and 520 nm emission, at a gain of 1500. The results, expressed in relative fluorescence units (rfu), were corrected for background auto-fluorescence by subtraction of pKC26 levels at the equivalent optical densities.

2.6 Switching assays

To determine phase transition frequencies of *pap* operons fused to the *lacZ* reporter under a number of different conditions the following strategy was used: fusion strains were cultured on CFA agar at 37°C from which at least three individual colonies were picked, serially diluted in PBS and plated onto the appropriate medium containing 40 μg ml^{-1} 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal). After overnight growth on the relevant induction medium, at least six colonies, either blue (mostly phase on) or white (mostly phase off), were then picked, serially diluted in PBS and plated onto RD M9 medium with glycerol and X-gal to allow quantification of the proportions of phase on and phase off colonies (measuring

plates). Colonies containing blue colour in >50% of their surface area were scored as on and all others were scored as off. The frequency of phase transition was calculated using the formula $(M/N)/g$, as described in Blyn *et al.*, 1989. 'M/N' is the ratio of cells that underwent phase transition to total cells, and 'g' is the number of generations of growth from a single cell to the number of cells in the colony assessed. In all frequency calculations it was assumed that an on or off colony resulted from a single on or off parent cell. The various induction conditions tested were as follows: for phase transition on rich undefined media, CFA agar containing X-gal was used; for different environmental conditions minimal M9 medium with glycerol and X-gal, RD M9 medium with glycerol and X-gal or RD M9 medium with glucose and X-gal were used; for temperature effects, induction on RD M9 medium with glycerol and X-gal was performed at 37°C or 28°C; for phase transition in urine, urine agar plates containing X-gal were used. To determine off-to-on phase transition in static urine single off colonies were selected from RD M9 medium containing glucose and X-gal, diluted to approximately 5 cfu ml^{-1} in PBS and paired samples cultured statically in 5 ml urine or 5 ml RD M9 medium with glycerol at 37°C for 24 hours. Bacterial cultures were then serially diluted and plated onto RD M9 medium with glycerol and X-gal to determine the proportions of phase on and phase off bacteria (measuring plates). The starting phase of the initial off colonies was confirmed on measuring plates and colonies containing >2% phase on cells were excluded from the analysis, as they were probably not true phase off colonies but rather phase on under catabolite repression.

2.7 Electrophoretic Mobility Shift Assays (EMSAs)

EMSA analysis of PapB on PapB site 1 variants was performed as described in Forsman *et al.*, 1989. Radiolabelled DNA fragments containing different site 1 variants were obtained by PCR using 20 μCi [^{32}P]-dATP, sequenced TOPO®-clones as templates and primers [PapB site 1 (F) and (R)] (Table 2.3). The relative concentrations of the radiolabelled amplification products were determined by electrophoresis in 4% polyacrylamide (acrylamide:bisacrylamide 37.5:1) gels in TBE buffer. Binding reactions contained 1 μl radiolabelled DNA of appropriate dilution, 0.5 μg poly-(dI-dC) as non-specific competitor DNA and 50 mM KCl in buffer B [30 mM Tris-HCl (pH 7.9), 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 5% glycerol]. One μl of whole cell protein extract prepared as described in section 2.8.1 was added in serial two-fold dilutions and binding reactions were incubated at room temperature for 20 min. Free DNA and protein-DNA complexes were separated by electrophoresis in 4% polyacrylamide-bis (37.5:1) gels in TBE buffer at constant voltage (160 V) for 2 hours. Gels were transferred onto 3MM Whatman paper, dried at 80°C for 45 min and autoradiographed using BioMax MR Film (KODAK). Radioactive signals were also detected using the Molecular Imager® system (Bio-Rad) and quantified using MULTI-ANALYST software v.1.1 (Bio-Rad).

2.8 Methods for protein analysis

2.8.1 Whole-cell protein extracts

In *papI* expression experiments, *E. coli* strains containing single-copy p_I::*gfp*$^+$ fusions (ZAP610-614/616, Table 2.1C) were cultured in 50 ml CFA broth at 37°C. For strains containing pBAD18 or pNJH20 plasmids, CFA broth was supplemented

with ampicillin and 0.2% arabinose. Bacterial cells were harvested at stationary phase by centrifugation (3220 g, 30 min at 4°C) and pellets were frozen-thawed once and suspended in 0.5 ml sonication buffer [100 mM NaCl, 50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 0.1 mM DTT, 1 μM phenylmethylsulphonyl fluoride (PMSF)]. Cell suspensions were subjected to three 15-sec sonication rounds at 5 microns amplitude using a Soniprep 150. Cell debris was removed by microcentrifugation (13400 g, 15 min at 4°C) and the volume of the supernatant containing the cytoplasmic proteins was adjusted with sonication buffer to account for OD_{600} differences in the original bacterial cultures. Protein samples were aliquoted and stored at -20°C.

For PapB and control protein preparations used in EMSAs, *E. coli* NEC11 transformed with pNJH36 or pET11 plasmids was cultured in 25 ml LB broth supplemented with ampicillin at 37°C. Cultures were induced during mid-exponential growth with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours and subsequently cells were harvested, sonicated and protein extracts prepared as described above. Cleared lysates were stored with 50% glycerol at -20°C.

2.8.2 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples obtained from whole-cell lysis were separated by SDS-PAGE using the BIO-RAD Mini-protean II gel apparatus (Bio-Rad, UK). A sample volume of 12 μl was mixed with 8 μl SDS loading buffer [50 mM Tris-HCl (pH 6.8), 4% sodium dodecyl sulphate (SDS), 12% glycerol, 2% β-mercaptoethanol, 0.1% bromophenol blue] and heated to 100°C for 5 min to denature the proteins. For analysis of p_I::*gfp*$^+$

fusion expression, protein samples were loaded in a 5% stacking gel [5% acrylamide-bisacrylamide (37.5:1), 0.125 M Tris-HCl (pH 6.8), 0.1% SDS, 0.1% ammonium persulfate (APS), 0.001% N,N,N',N'-Tetramethylethylene diamine (TEMED)] and resolved in a 12% resolving gel [12% acrylamide-bisacrylamide (37.5:1), 0.375 M Tris-HCl (pH 8.8), 0.1% SDS, 0.1% APS, 0.001% TEMED] by electrophoresis in running buffer [25 mM Tris-HCl (pH 8.3), 250 mM glycine, 0.1% SDS] at constant voltage (100 V) for approximately 2 hours. Molecular weight markers (Rainbow marker, Amersham Phamacia Biotech) were used for protein size (KDa) determination.

2.8.3 Colloidal Blue Staining of SDS-PAGE Gels

Proteins separated in polyacrylamide gels were stained with Colloidal blue (Invitrogen) according to manufacturer's instructions. Gels were de-stained in dH_2O overnight and transferred onto 3MM Whatman paper. Gel images were captured using a Flowgen MultiImage^{TM} Light cabinet and subsequently gels were dried at 50^{o}C for 90 min.

2.8.4 Western blotting

Detection of reporter proteins expressed from p_I promoter fusions was performed using the method of Western hybridisation. Separated proteins were transferred from polyacrylamide gels onto nitrocellulose membranes (Amersham Phamacia Biotech) using the Bio-Rad transfer apparatus (Bio-Rad, UK). Briefly, the gel and membrane were placed between 4 pieces of 3MM Whatman paper, which were in turn placed between two sponge pads, pre-soaked in Transfer buffer [39 mM glycine, 48 mM

Tris, 20% methanol] and held together in a plastic clamp. The transfer of proteins was performed in cold transfer buffer at constant voltage (60 V) for 100 min. Nitrocellulose membranes were subsequently blocked in 8% dried milk (Marvel) solution in PBS at 4°C overnight. Excess blocking buffer was removed by three 10-min washes with washing buffer [1% dried milk (Marvel), 0.05% Tween® 20 (polyoxyethylenesorbitan monolaurate) in PBS]. Membranes were mixed with primary antibody solution (mouse anti-GFP, clone JL-8, BD Biosciences), prepared at a 1:2000 dilution in washing buffer, and incubated for 1 hour at room temperature on a rocking platform. Excess primary antibody was removed with three 10-min washes in washing buffer. Membranes were then suspended in secondary antibody solution [Rabbit anti-mouse HRP (horseradish peroxidase) conjugate, Dako], prepared at 1:4000 in washing buffer, and incubated for 1 hour at room temperature on a rocking platform. Excess secondary antibody was removed with three 10-min washes in washing buffer. Membranes were finally incubated in 10 ml ECL Solution 1 [100 mM Tris-HCl (pH 8.5), 2.5 mM Luminol ($C_8H_7N_3O_2$) , 0.4 mM *p*-Coumaric acid ($C_9H_8O_3$)] mixed with 10 ml ECL Solution 2 [100 mM Tris-HCl (pH 8.5), 0.02% H_2O_2] for 5 minutes at room temperature on a rocking platform. Chemiluminescence was detected using Hyperfilm™ ECL chemiluminescence film (Amersham Phamacia Biotech), developed in a Protec automatic film processor (Optimax).

2.8.5 Protein structure

The structure of the PapI protein (Kawamura *et al.*, 2007) was obtained from the RCSB Protein Data Bank (PBD ID 2HTJ) (Berman *et al.*, 2000) and molecular

viewing and analysis was performed using the MBT Protein Workshop software (Moreland *et al.* 2005).

2.9 Statistical methods

Associations between observation data were tested using the chi-square test or Fisher's exact test if data sets contained a small number of observations. Regression analysis was used to determine the relationship between explanatory and dependent variables. Equality of mean values between populations of normally distributed data was tested by analysis of variance (ANOVA). For non-normally distributed continuous data, median values were compared by the non-parametric Kruskal-Wallis test. All statistical analyses were performed using MINITAB (r.12) and Genstat 5 (r. 4.1) packages.

Chapter 3

Regulation of P fimbrial phase transition frequencies in uropathogenic *Escherichia coli* CFT073

Adhesion of uropathogenic *Escherichia coli* (UPEC) to host tissues via P fimbriae is required for colonisation and infection of the urinary tract (Wullt *et al.*, 2000; Tseng *et al.*, 2007). Expression of P fimbriae is regulated by phase variation and its transition frequencies are important as they determine the proportion of P-fimbriate bacteria in the population and govern expression of other surface factors at the single-cell level through positive and negative cross-talk (Xia *et al.*, 2000; Holden *et al.*, 2001, 2004, 2006). To date, measurements of P fimbrial phase variation frequencies have been performed in the *E. coli* K-12 background using lysogenic reporter constructs (Blyn *et al.*, 1989, 1990; White-Ziegler *et al.*, 1990). Laboratory *E. coli* strains do not naturally possess *pap* operons and are potentially less suitable genetic backgrounds for studying P-fimbrial phase variation than clinical UPEC strains. Moreover the reported frequencies in *E. coli* K-12 are not predictive of the high proportion of P-fimbriate bacteria reported for several UPEC clinical isolates (Nowicki *et al.*, 1984; Low *et al.*, 1987; Pere *et al.*, 1987). The aim of this work was to measure P-fimbrial phase variation frequencies in their relevant genetic background for the first time. This chapter describes the methods employed for measuring P fimbriation and associated *pap* phase transition frequencies in two UPEC clinical isolates, including the sequenced strain CFT073, and comparing them to levels and frequencies for homologous *pap* constructs in *E. coli* K-12. Phase variation frequencies in the isolate background were also investigated under different environmental conditions, including human urine, the closest physiological environment to human UTI.

N.B. The results of this study were published in a peer-reviewed journal (Holden *et al.*, 2007, Appendix A). The work presented in this chapter was performed as part of this thesis and contribution from co-authors is stated in the appropriate sections.

3.1 Comparison of the proportion of P-fimbriate bacteria in populations of UPEC clinical isolates with a cloned *pap* operon in *E. coli* K-12

The proportion of P-fimbriate bacteria was measured in populations of two well-characterised UPEC clinical isolates that both contain two homologous *pap* operons; isolates J96 (Hull *et al.*, 1981) and CFT073 (Mobley *et al.*, 1990), and compared to P-fimbriate proportions in an *E. coli* K-12 derivative containing a complete *pap* operon. Measurements were obtained from single colonies cultured on CFA medium, a culture condition known to promote expression of P fimbriae (Evans *et al.*, 1977; Holden *et al.*, 2006). Colonies of the pyelonephritis *E. coli* isolate J96 (Fig. 3.1A-C) were shown by flow cytometry to contain an average of 18% ± 2% P-fimbriate bacteria, as determined by immunofluorescence labelling with a polyclonal anti-PapA serum specific for F13 P fimbriae (Nilsson *et al.*, 1996). In *E. coli* CFT073 (Fig. 3.1D-F), the positive proportion determined by flow cytometry was 4.8% ± 0.4% using the polyclonal anti-P serum OM12 (Pere *et al.*, 1987). This percentage is however an underestimate as this polyclonal serum caused a degree of bacterial aggregation which resulted in a substantial fraction of labelled cells being incorporated into large particles and hence being missed by flow cytometry due to the imposed size limitations (see section 2.5.1). To obtain a more accurate proportion for CFT073 colonies, P-fimbriate bacteria were counted by immunofluorescence microscopy and produced proportions of between 10% and 20% depending on the colony measured (Fig. 3.1D-E). As a comparison to assessing fimbriation proportions in *E. coli* clinical isolates, the complete F13 *pap* operon (*papI-papG*) from *E. coli* J96 (Lindberg, *et al.*, 1984; Marklund *et al.*, 1992) was exchanged into the *E. coli* K-12 chromosome at the *lac* locus, generating strain ZAP594 (Table

2.1C). In this strain, only 0.26% ± 0.02% of the bacterial population was P-fimbriate as determined by flow cytometry using the anti-PapA serum (Fig. 3.1I - the associated micrograph is depicted in Fig. 3.1G-H). The results demonstrated an effect of genetic background on proportion of P-fimbriate cells.

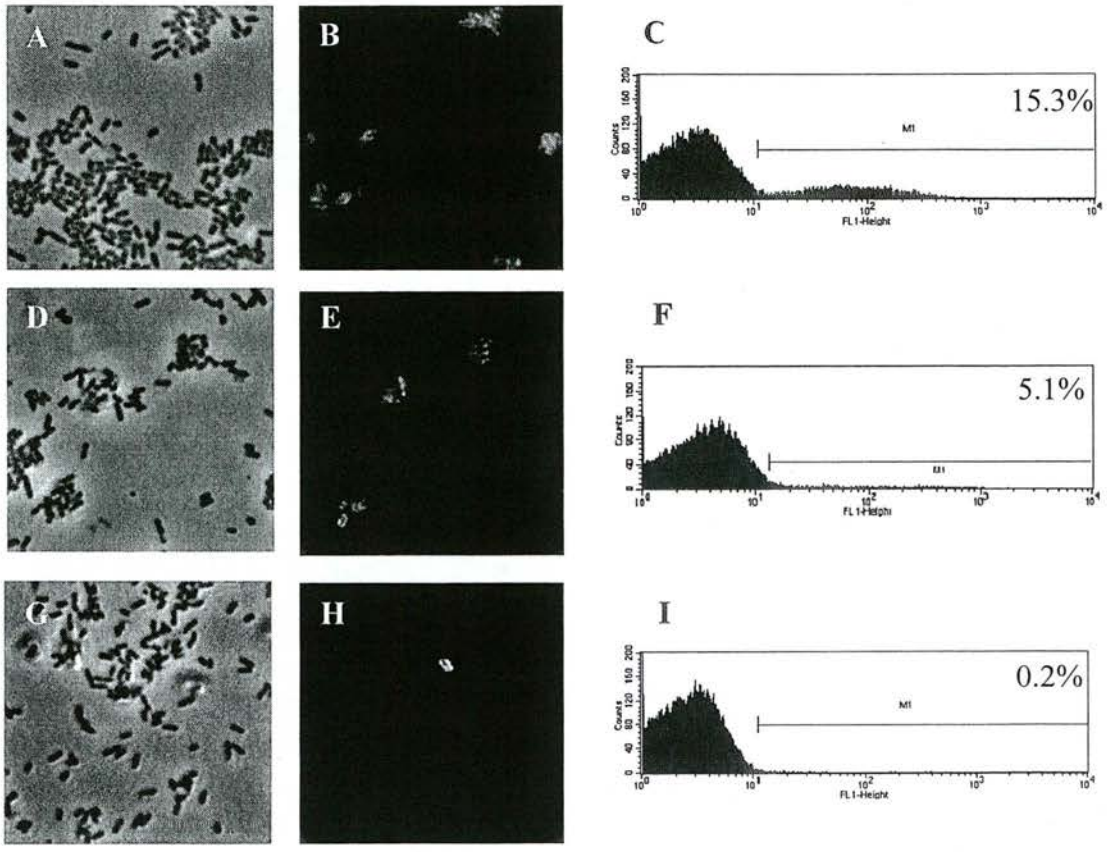


Fig. 3.1 P fimbriation in different *E. coli* strains examined by immunofluorescence microscopy and flow cytometry. Phase-contrast images in panels A, D & G and fluorescence images in panels B, E & H. Example flow cytometry histograms (C, F & I). The gate for fluorescence detection is indicated by M1 and was set as defined in section 2.5.1. The proportion of fluorescent events is indicated in the upper right corner of each histogram. *E. coli* J96 (A, B & C); *E. coli* CFT073 (D, E & F); *E. coli* ZAP594 (G, H & I). For panel H multiple fields had to be scanned to detect a single P-fimbriate bacterium.

3.2 Reporter fusions to GFP reflect fimbriation proportions in clinical isolates and the *E. coli* K-12 backgrounds

Green fluorescent protein (GFP) reporter fusions were constructed to both *pap* operons (P1 and P2) from *E. coli* CFT073 and inserted into both *E. coli* CFT073 and K-12 chromosomes at *lac*, generating strains ZAP833/834 and ZAP714/955, respectively (Table 2.1C). The *pap* regulatory regions used in the fusions included both *papI* and *papB* and the first codon of *papA*. The proportions of GFP-expressing bacteria under P-optimal growth conditions (CFA medium, 37°C) were then determined by flow cytometry. The proportions of CFT073 expressing P1 and P2 operon fusions were 24% ± 1.9% and 32% ± 3.6%, respectively. In contrast, the proportions of K-12 expressing P1 and P2 operon fusions were 0.23% ± 0.04% and 0.37% ± 0.13%, respectively. These results confirmed marked differences in P-fimbriate proportions between the CFT073 and K-12 genetic backgrounds.

3.3 Measurement of *pap* phase transition frequencies in *E. coli* CFT073

Differences in P-fimbriate proportions were proposed to result from different *pap* phase variation frequencies in the CFT073 and K-12 backgrounds. In order to measure phase transition frequencies in a clinical isolate background, fusions of the CFT073 P1, P2 and J96 F13 *pap* regulatory regions were made to *lacZ* and exchanged into the CFT073 chromosome at *lac*, generating strains ZAP838/1137/996, respectively (Table 2.1C). Phase transition frequencies were determined under P-optimal growth conditions (CFA medium, 37°C) using switching assays, as described in section 2.6. The J96 *pap* operon fusion phase varied at a high off-to-on frequency, $>1 \times 10^{-2}$ per cell per generation. The off-to-on phase transition

frequencies of the CFT073 P1 and P2 operon fusions were $4.4 \times 10^{-3} \pm 0.3$ and $6.1 \times 10^{-3} \pm 0.6$ per cell per generation, respectively (Table 3.1). As a comparison, a fusion to P1 was constructed in *E. coli* K-12 (ZAP992) and the off-to-on frequency determined was $<1 \times 10^{-5}$ per cell per generation (data contributed by Miss K. Catherwood), confirming the observed differences in P expression and proportions of P-fimbriate bacteria between the isolate and K-12 backgrounds. Phase transition was also measured in the opposite (on-to-off) direction and was $3.3 \times 10^{-2} \pm 0.08$ and $3.5 \times 10^{-2} \pm 0.02$ per cell per generation for P1 and P2, respectively, in the CFT073 background (Table 3.1). These on-to-off frequencies were found to be consistent when measured under a number of different conditions and were at least 10-fold higher than off-to-on transition frequencies measured for the same operon fusions.

3.4 Correlation of P-fimbriate proportions and *pap* phase transition frequencies

In the current study the frequencies of *pap* phase variation were determined together with the proportion of P-fimbriate bacteria in both the *E. coli* K-12 and clinical isolate backgrounds. As these two parameters are linked, the proportion of P-fimbriate bacteria in the population can be calculated from phase transition frequencies and *vice versa*. Figure 3.2 (contributed by Prof. D. L. Gally) shows the expected proportions of phase on (fimbriate) bacteria in colonies plotted against the number of generations. These were calculated using a range of off-to-on transition frequencies and a single on-to-off frequency of 3×10^{-2} per cell per generation that was based on the measurements obtained in this study. From this correlation, it can be estimated that the two *pap* operons in J96 switch on at a combined frequency of $>1 \times 10^{-2}$ per cell per generation, while the *pap* operons in CFT073 vary from off-to-

on at a frequency between 3×10^{-3} and 1×10^{-2} per cell per generation. In comparison, the estimated frequency for the single J96-derived operon in the *E. coli* K-12 background is 1×10^{-4} per cell per generation based on 0.26% fimbriation. These frequency estimates based on P-fimbriate proportions are in agreement with the actual transition frequencies measured using the *papA*::*lacZ* fusions.

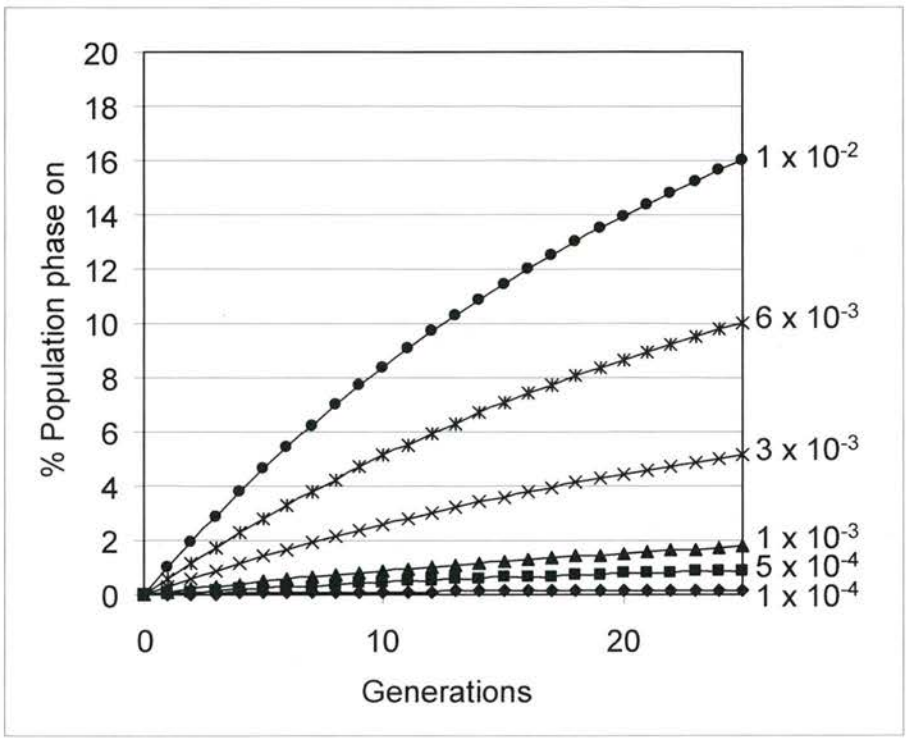


Fig. 3.2 Correlation between the proportion of phase on bacteria in a population and phase transition frequencies. Different frequencies of phase off-to-on transition were plotted against the number of generations, using a constant frequency of on-to-off transition of 3×10^{-2} per cell per generation. The resultant proportions of the population that are phase on are expressed as percentages of the total population. The phase on transition frequencies (per cell per generation) used for the correlation are indicated as follows (with the expected percentage of phase on bacteria at equilibrium shown in parentheses): ●, 1×10^{-2} (25%); ✳, 6×10^{-3} (16.7%); x , 3×10^{-3} (9.09%); ▲, 1×10^{-3} (1.2%); ■, 5×10^{-4} (1.64%); ♦, 1×10^{-4} (1.33%). All cultures would reach an apparent steady state in more than 200 generations. (Figure contributed by Prof. D. L. Gally).

3.5 Environmental signals have differential effects on off-to-on transition frequencies of P1 and P2 *pap* operons from *E. coli* CFT073

Previous research has shown that *pap* regulation is responsive to a number of environmental signals in an *E. coli* K-12 background, such as carbon source and temperature (Blyn *et al.*, 1989). In addition, recent studies have suggested differential regulation of these *pap* clusters in human urine and during murine UTI (Snyder *et al.*, 2004). In light of the genetic background effect on *pap* expression, it was interesting to investigate the effect of environmental signals on *pap* transition frequencies in the context of the clinical isolate background. Frequencies were determined using the P1 and P2 *lacZ* reporter fusions in the CFT073 background (ZAP838 and ZAP1137 respectively, Table 2.1C). Table 3.1 shows the phase transition frequencies for CFT073 P1 and P2 *pap* operons in the off-to-on orientation cultured under different conditions (part of the data was contributed by Dr. N. J. Holden). In contrast to the results obtained with a complex medium (CFA), off-to-on transition frequencies were lower in defined media. As expected from previous studies (Baga *et al.*, 1985; Blyn *et al.*, 1989), both operons were catabolite repressed to similar extents; the decrease was highly significant for both P1 and P2 *papA::lacZ* fusions ($p < 0.001$). The frequencies were also higher in the presence of amino acids, although P1 was more responsive (four-fold) than P2 (two-fold) ($p < 0.001$ and $p = 0.03$, respectively). A temperature decrease from 37°C to 28°C only moderately affected the P1 transition frequency and had no effect on P2.

The effect of human urine on phase transition frequencies of *E. coli* CFT073 P1 and P2 was measured on plates at 37°C and in static liquid urine at 37°C. On urine plates

the off-to-on frequencies for P1 were equivalent to those on CFA, whereas the frequency for P2 was approximately five-fold lower. The frequency was significantly higher for P1 ($p < 0.001$). In liquid urine the frequencies were slightly lower than on plates, although the differential between P1 and P2 was still evident. On-to-off phase transition frequencies were also determined on urine plates and were similar to those measured on CFA medium (~ 3×10^{-2} per cell per generation). Phase off transition on urine plates was marginally higher for P2 (1.5×10^{-2}) than for P1 (1×10^{-2}). Taken together these results indicate that P1 is expressed at higher levels in human urine compared to P2.

Table 3.1 Phase transition frequencies of P1 and P2 *pap* operons from *E. coli* CFT073, measured under different environmental conditions. Switching assays were performed using strains ZAP838 (P1) and ZAP1137 (P2) and several growth media, as described in section 2.6. Mean frequencies (10^{-3} per cell per generation) are listed, together with the estimated standard error of the mean (± s.e.). ND: not determined.

Condition	Mean **off-to-on** transition frequency x 10^{-3} (±s.e.)		Mean **on-to-off** transition frequency x 10^{-3} (±s.e.)	
	P1	P2	P1	P2
RD M9 glycerol, 37°C	1.92 (0.14)	1.01 (0.12)	ND	ND
RD M9 glucose, 37°C	0.94 (0.11)	0.32 (0.05)	"	"
minimal M9 glycerol, 37°C	0.79 (0.19)	0.60 (0.11)	"	"
RD M9 glycerol, 28°C	1.25 (0.20)	1.02 (0.08)	"	"
static urine, 37°C	2.55 (0.31)	1.25 (0.14)	"	"
urine plates, 37°C	4.31 (0.53)	1.28 (0.22)	10.16 (2.02)	14.74 (1.00)
CFA, 37°C	4.4 (0.3)	6.1 (0.6)	32.60 (0.79)	35.30 (1.77)

3.6 PapI *in trans* can increase the *pap* off-to-on transition frequency in the *E. coli* K-12 background

PapI is an established positive regulator of *pap* off-to-on phase transition (Baga *et al.*, 1985; Kaltenbach *et al.*, 1995; Nou *et al.*, 1995; Hernday *et al.*, 2003). Changes in the amount of PapI may explain the background differences reported in this study and in a clinical isolate background this may arise as a consequence of either the presence of multiple *pap* operons (i.e., gene copy number), and/or differences in the regulatory network that differentially affect expression from each *pap* operon. To determine whether increasing *papI* expression in *E. coli* K-12 could increase *pap* off-to-on phase transition frequencies and fimbriation proportions, plasmid-based *papI* (pHMG98; Forsman *et al.*, 1989) was transformed into two *E. coli* K-12 strains; one containing the integrated single-copy fusion of the J96 F13 *pap* regulatory region to *lacZ* (ZAP593) and the other containing the complete J96 F13 *pap* operon (ZAP594; for strains see Table 2.1C). The off-to-on transition frequency increased 34-fold in the presence of induced PapI (Fig. 3.3A). This correlated well with a 53-fold increase in the proportion of P-fimbriate bacteria detected using immunofluorescence staining with anti-papA serum and flow cytometry (Fig. 3.3B). Taken together these data suggest that low PapI levels in the *E. coli* K-12 background could in part explain the lower off-to-on phase transition frequencies in this background.

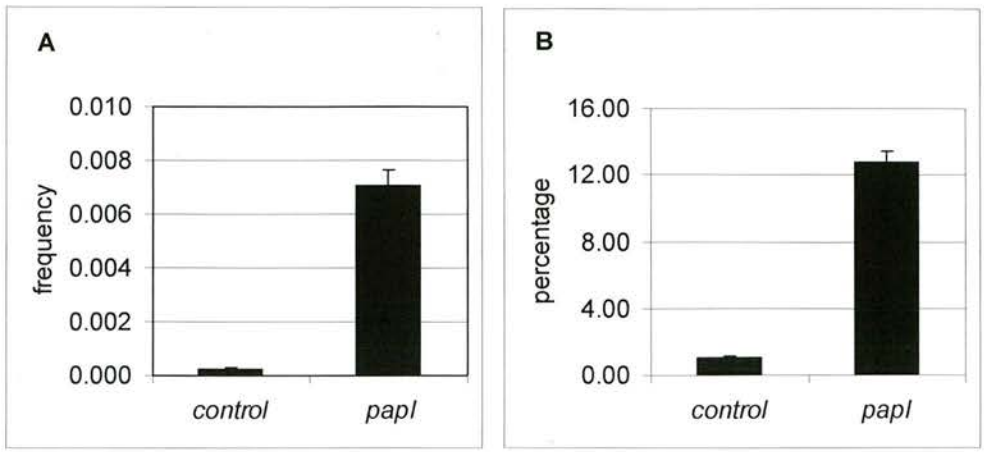


Fig. 3.3 Effect of *papI* addition *in trans* on *pap* phase variation and P fimbriation in the *E. coli* K-12 genetic background. (*A*) Off-to-on phase transition frequency (per cell per generation) of strain ZAP593. (*B*) Percentage of P-fimbriate bacteria in ZAP594 populations stained with anti-PapA antibody and determined by flow cytometry. In both cases the strains were transformed with pACYC184 (*control*) or pHMG98 (*papI*). Mean values (± s.e.) are shown.

3.7 Effect of multiple *pap* operons on *pap* phase transition frequencies in *E. coli* CFT073

To investigate whether the presence of two *papI* gene copies in the CFT073 chromosome is responsible for the higher *pap* expression observed in this isolate, a CFT073 mutant strain was constructed, in which both P1 and P2 regulatory regions were deleted (ZAP965; Table 2.1C). *E. coli* strain ZAP965 was constructed as described in section 2.3.16 and deletions were confirmed genotypically by PCR and phenotypically by immunofluorescence labelling using the OM12 anti-P polyclonal serum (Fig. 3.4).

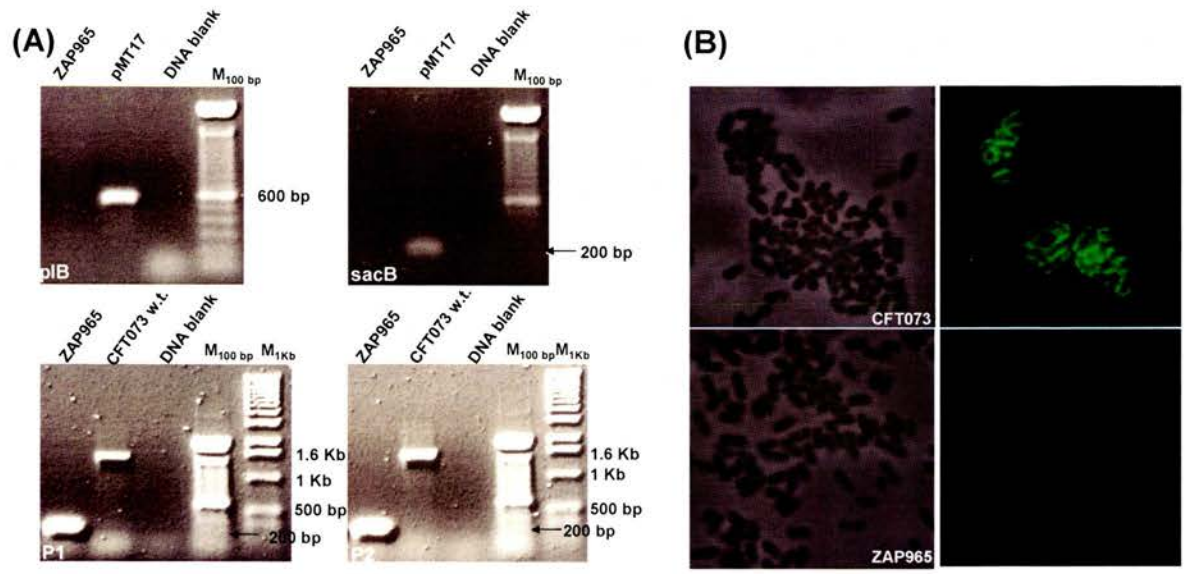


Fig. 3.4 Genotypic and phenotypic characterisation of *E. coli* strain ZAP965. (*A*) PCR screens using ZAP965 genomic DNA to confirm the absence of pIB307 backbone sequence and of the *sacB* gene, as well as deletion of the *papIB* region from both P1 and P2 operons. The sequence of associated primers is listed in Table 2.3 Plasmid pMT17 (Table 2.2) served as a positive control in pIB and sacB PCR screens, and genomic DNA from CFT073 w.t. was used as a positive control for amplification of the 1.43 Kb P1 and 1.45 Kb P2 *papIBA*' regions. Deletion of the *papIB* region in strain ZAP965 left a 200-300 bp 'scar' in both P1 and P2 operons. PCR negative controls (DNA blank samples) were performed in each amplification reaction. (*B*) Phase-contrast and fluorescence microscopy images of *E. coli* CFT073 (*top*) and ZAP965 (*bottom*) populations stained for P fimbriae using the OM12 anti-P polyclonal serum (Pere *et al.*, 1987) and the secondary Alexa Fluor 488 IgG-conjugate (Molecular Probes), as described in sections 2.4 and 2.5.2. No P-fimbriate ZAP965 cells were detected in several scanned fields or under longer fluorescence exposure times.

The P1 and P2 *lacZ* reporter fusions were then assayed in the ZAP965 background (ZAP843 and ZAP595, respectively; Table 2.1C). The off-to-on transition frequency on CFA medium for P1 was reduced slightly from 4.4×10^{-3} per cell per generation in the CFT073 wild-type background to 3.3×10^{-3} per cell per generation in the CFT073 mutant background, while for P2 there was a 2.4-fold reduction from 6.1×10^{-3} to 2.6×10^{-3} per cell per generation ($p < 0.001$). When the P1 and P2 *lacZ* reporter fusions in the CFT073 mutant background were assayed under alternative conditions (28°C, RD M9 medium containing glycerol) the impact of the genetic background was again more evident in the P2 fusion. The off-to-on transition frequency for the P2 fusion was reduced eight-fold, while the P1 frequency was reduced two-fold. While these data indicate some cross-regulation occurring due to homologous regulators, they cannot account for the marked background differences between the *E. coli* K-12 and CFT073 strains demonstrated in this study.

To investigate whether the background affected *papI* expression *per se*, a plasmid-based transcriptional fusion of the P1 *papI* promoter region to *gfp* (pMT24; Table 2.2) was transformed in *E. coli* K-12, wild-type CFT073 and the CFT073 *pap* regulatory region mutant strain (ZAP965). Expression of *papI* was higher (4.4-fold) in both wild-type strain CFT073 and the *pap* regulatory region mutant of CFT073 than in *E. coli* K-12 (Fig. 3.5). It was concluded that increased activation of *papI* in the clinical isolate background in combination with cross-regulation from homologous operons accounts for the observed P-fimbriation in *E. coli* isolates CFT073 and J96.

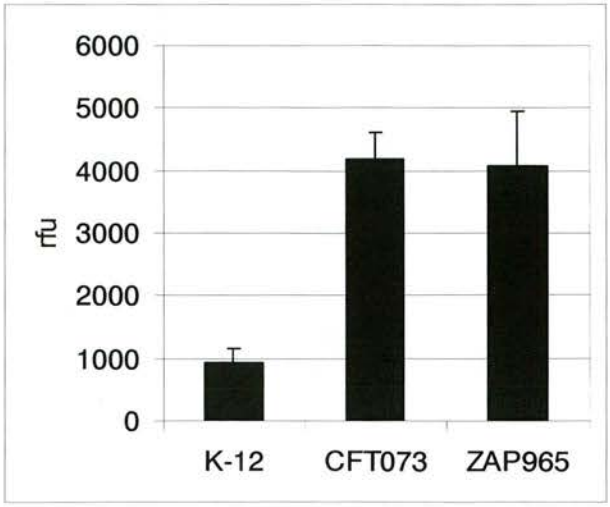


Fig. 3.5 Expression of *papI*::*gfp* plasmid-based fusion in different *E. coli* genetic backgrounds. The pACYC184-based construct pMT24 was transformed into *E. coli* K-12, CFT073 and ZAP965 strains and whole population GFP expression from the plasmid-based fusion (pMT24) was measured, as described in section 2.5.3. The background fluorescence levels determined from transformation of pKC26 (control) in each strain was subtracted from the pMT24 corresponding level. Mean relative fluorescence units (rfu) are presented (± s.e.).

3.8 Discussion

P fimbriae are an established virulence factor utilised by uropathogenic *E. coli* to promote colonisation of urinary and gastrointestinal tracts (Pere *et al.*, 1987; Tullus *et al.*, 1992; Hull *et al.*, 2002; Roberts *et al.*, 2004). Like expression of most fimbrial adhesins produced by *E. coli*, expression of the *pap* operon is phase variable. The mechanism underlying this variation and the genetic and environmental inputs that affect it have been the focus of many years of research. The majority of these studies have been carried out on regulatory regions cloned into the *E. coli* K-12 chromosome and the frequencies of phase transition have been determined using *lacZ* fusions in

this non-native background, (i.e., one that does not normally support expression of a *pap* operon). The current study reports for the first time the regulation of P-fimbrial phase transition frequencies in a uropathogenic *E. coli* strain and demonstrates the importance of the genetic background for these frequencies. In particular, high off-to-on transition frequencies may be accounted for by elevated *papI* expression levels.

Early studies using immunoelectron microscopy and indirect immunofluorescence reported high proportions of P-fimbriate bacteria in UPEC isolate populations cultured *in vitro* and demonstrated that phase variation was extremely rapid, although no actual frequency measurements were performed. The pyelonephritis *E. coli* isolate KS71 possesses two *pap* operons and when cultured on CFA agar produced a P-fimbriate subpopulation of 17% (Nowicki *et al.*, 1984). A different UPEC strain, C1212, also carrying two *pap* operons, produced colonies in which 84% of the bacteria expressed P-fimbriae, although in this case one cluster (pilin-21) was dominant (Low *et al.*, 1987). Pere and colleagues showed that P-fimbriate proportions in populations of *E. coli* strains found in urine from patients with symptomatic UTI were between 0.01 and 95% (Pere *et al.*, 1987). Overall, the reported high proportions of P-fimbriate cells would predict that *pap* off-to-on transition frequencies in UPEC strains are higher than the frequencies determined in the *E. coli* K-12 background (Blyn *et al.*, 1989). These studies report *pap* frequencies of one cell in 10,000 transitioning from phase off to phase on in each generation. This frequency would generate a phase on sub-population of less than 1% (Fig. 3.2) which is not sufficiently high to account for the P-fimbriate subpopulations reported in the studies described above (Nowicki *et al.*, 1984; Low *et al.*, 1987).

To address this question in detail, phase transition frequencies were measured for the two *pap* operons present in *E. coli* CFT073 using *lacZ* and *gfp* fusions to the *papI-papBA* regulatory regions. As anticipated (Baga *et al.*, 1985; Blyn *et al.*, 1989; Goransson *et al.*, 1989a; White-Ziegler *et al.*, 1990), the frequencies were dependent on culture conditions but ranged from 6×10^{-4} per cell per generation on minimal M9 glycerol medium to 6×10^{-3} per cell per generation on CFA medium. On-to-off transition frequencies were also measured and remained fairly constant at 3×10^{-2} per cell per generation. Based on these frequencies the expected proportion of fimbriate bacteria cultured on CFA medium would be between 10% and 30% depending on the amount of cross-regulation between the two operons and the rate with which bacteria that turn off expression lose fimbriae from their surface. The predicted proportion of fimbriate bacteria, based on the measured frequencies, was entirely consistent with that detected for *E. coli* CFT073 in the present study (10-20%, Fig. 3.1D-E). It should be noted that construction of the chromosomally-integrated reporter fusions does introduce an extra copy of the regulatory region into *E. coli* CFT073 but that the phase transition frequencies and levels of expression were consistent with fimbriation under the conditions tested.

P1 and P2 are homologous *pap* operons in *E. coli* CFT073 both encoding a PapG class II adhesin. Here it was demonstrated that these operons are differentially regulated by certain environmental conditions. The P1 operon was more responsive to amino acids and temperature. By contrast, both operons showed the same level of catabolite repression. Urine has been shown to contain levels of amino acids sufficiently high to support bacterial growth, and the increased off-to-on phase

transition frequency of P1 relative to P2 may be due to an amino acids response (Asscher *et al.*, 1966). In addition, the P2 on-to-off transition frequency was higher, indicating that the overall proportion of P1-fimbriate cells was higher in human urine. This is the first time that phase transition frequencies have been measured in a UPEC isolate cultured in human urine, the closest physiological environment to human UTI.

The present study has shown that *pap* operons exhibit significantly reduced frequencies of phase variation when measured in an *E. coli* K-12 background. This background difference was confirmed using a complete *pap* operon exchanged into the *E. coli* K-12 chromosome at *lac*. Fimbriate proportions in populations of this strain were 100-fold lower than for the parental isolate strain, J96 (Fig. 3.1). Based on the current understanding of the *pap* phase variation mechanism (van der Woude *et al.*, 1996; Blomfield, 2001) all the required regulators are present in both backgrounds. The higher frequency in the clinical isolate must be explained by altered levels of one or more of the established regulators or novel regulators of the system. While off-to-on transition frequencies were much higher in the isolate background, transition frequencies in the opposite direction were equivalent. This indicates that the difference is due to a factor or factors that promote only phase off-to-on transition. A key candidate is the positive regulator, PapI. Higher total levels of PapI could be achieved by increased expression from individual *papI* promoters and/or by the presence of multiple homologous *pap* operons and therefore an increased *papI* copy number. To determine the significance of cross-activation in the clinical isolate background, the two *papIB* regulatory regions from P1 and P2 in *E. coli* CFT073 were deleted and expression of *pap* was examined from a regulatory

region fusion engineered into the *lac* locus. Deletion of the *pap* regulatory regions made no significant difference to the off-to-on phase transition frequency of P1 but did reduce the P2 transition frequency. These results indicate that cross-activation between P-related operons alone cannot account for the frequency differences in the different backgrounds. Subsequently, *papI* expression was measured in the two backgrounds and was shown to be higher in the clinical isolate than in *E. coli* K-12. Furthermore, *papI* expression was the same in the *papIB* double deletion strain which suggests that signals over and above native *pap* regulators influence *papI* expression in the clinical isolate. While plasmid copy number or GFP stability differences may account for the higher apparent expression level in *E. coli* CFT073, elevated *papI* expression would explain the higher phase off-to-on transition frequencies and P-fimbriate proportions observed in the clinical isolate background. Not surprisingly, the results also indicate that the regulatory network controlling *papI* expression is different in a clinical isolate and this is likely to be due to both the evolution of common signal transduction systems as well as regulatory cross-talk with other specifically acquired virulence determinants and their regulators. At present there is little published research on the regulation of *papI* and further work is required in this area.

Phase variable expression of P-fimbriae has two proposed roles in aiding UPEC colonisation of the human urinary tract. As P-fimbriae are surface antigens, phase variation is considered to limit immune exposure of the specific antigen and assist in evasion of innate and adaptive host responses (Kantele *et al.*, 2003; Snyder *et al.*, 2006). At the same time, phase variation generates a heterogeneous bacterial

population with a subpopulation of P-fimbriate bacteria that are equipped to colonise cells expressing Galα-(1,4)-Galβ-containing receptors (Leffler and Svanborg-Eden, 1981; Lindberg *et al.*, 1984; Korhonen *et al.*, 1986b). Another proposed function of phase variation is to coordinate expression of surface factors that are required in complex micro-environments, such as the colonised epithelium (Holden and Gally, 2004). Surface factors can block or hinder each other and so operons expressing surface factors such as fimbrial adhesins, non fimbrial adhesins, flagella, outer-membrane proteins and capsule are connected via a regulatory network that is driven in part by phase variation (Hasman *et al.*, 1999; Xia *et al.*, 2000; Schembri *et al.*, 2001; Barnich *et al.*, 2003). Phase variation frequencies, therefore, are not only important in controlling expression levels of a specific local factor but also impact on the regulation of a wide spectrum of other surface components expressed during infection.

Chapter 4

Genotypic and phenotypic characterisation of UPEC isolates for P fimbriae

The finding that homologous *pap* operons are differentially expressed under the same or different environmental conditions prompted the investigation of the regulatory sequence of *pap* operons from several UPEC clinical isolates, with the aim to identify sequence variation that has an effect on the regulation and expression of P fimbriae. This chapter describes the characterisation of selected UPEC clinical isolates for the presence of *pap* operons and P fimbrial expression. The final part presents the strategy for obtaining the sequence of *pap* regulatory regions and identifying areas of nucleotide conservation or divergence.

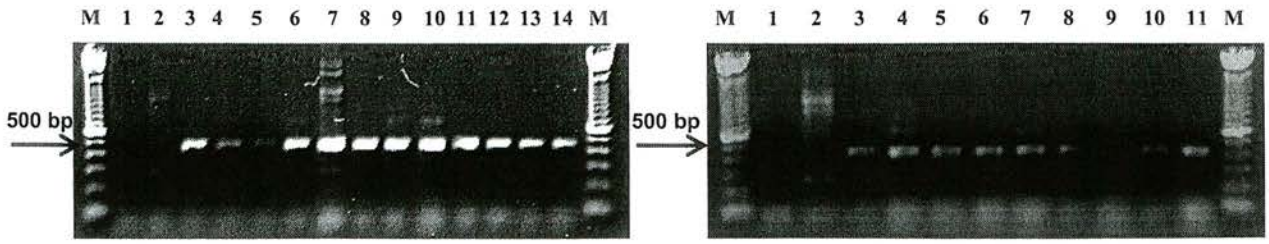
4.1 Pap genotypic and phenotypic characterisation of clinical UPEC isolates

Fifty seven *E. coli* isolates from different UTI clinical syndromes were collected for this study. The collection contained UPEC strains isolated from 17 cases of asymptomatic bacteriuria (ABU), 15 cases of cystitis and 25 cases of pyelonephritis, including the three prototype strains: CFT073 (Mobley *et al.*, 1990), J96 (Hull *et al.*, 1981), and 536 (Berger *et al.*, 1982). For a detailed description of all isolates used in this study see section 2.1.2.

4.1.1 PCR detection of *pap* genes

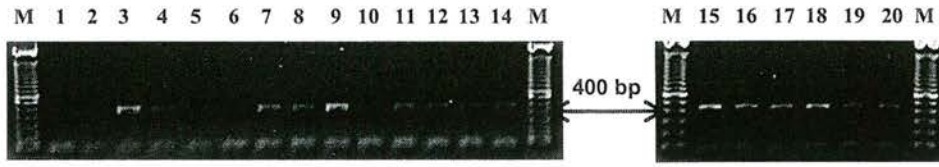
Source information suggested all isolates were positive for genetic carriage of *pap*. This was confirmed for 54 of the 57 isolates by PCR screening using PfamA and B primers (Table 2.3) for the detection of *papF* gene (Fig. 4.1). Two cystitis isolates, AUTI3 and AUTI16, and the pyelonephritis isolate AP5 were negative for the presence of *papF* (Fig. 4.1 II and III).

I. Asymptomatic Bacteriuria (ABU) isolates



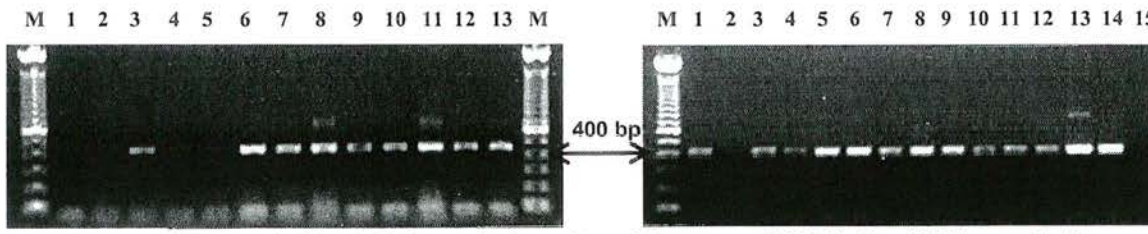
Lane M:	DNA ladder (500 bp marked)	Lane M:	DNA ladder (500 bp marked)
Lane 1:	DNA blank (PCR control)	Lane 1:	DNA blank (PCR control)
Lane 2:	AAEC185 (negative control)	Lane 2:	AAEC185 (negative control)
Lane 3 - 5:	CFT073, J96, 536 (positive controls)	Lane 3:	CFT073 (positive control)
Lane 6 - 11:	U43, U68, U105, U113, U126, U151	Lane 4 - 11:	HU1690, HU1691, HU1740, HU1751, HU1756, HU1757, HU1758, HU1761
Lane 12 - 14:	BR2-V1, BR10-V8, BR20-V1		

II. Cystitis Isolates



Lane M:	DNA ladder (400 bp marked)
Lane 1:	DNA blank (PCR control)
Lane 2:	AAEC185 (negative control)
Lane 3 - 5:	CFT073, J96, 536 (positive controls)
Lane 6 - 14:	AUTI3, AUTI4, AUTI7, AUTI12, AUTI16, AUTI19, AUTI31, AUTI36, AUTI43
Lane 15 - 20:	AUTI47, AUTI48, AUTI62, AUTI64, AUTI66, AUTI72

III. Pyelonephritis isolates



Lane M:	DNA ladder (400 bp marked)	Lane M:	DNA ladder (400 bp marked)
Lane 1:	DNA blank (PCR control)	Lane 1 - 4:	AP4, AP5, AP7, AP18
Lane 2:	Negative control: AAEC185	Lane 5:	KS71
Lane 3 - 5:	CFT073, J96, 536	Lane 6 - 14:	IHE1041, IHE1049, IHE1086, IHE1108, IHE IHE1152, IHE1268, IHE1402, IHE1431
Lane 6 - 13:	AUTI8, AUTI11, AUTI20, AUTI67, AUTI68, AUTI69, AUTI70, AUTI71	Lane 15:	DNA blank (PCR control)

Fig. 4.1 PCR screening for *papF* gene in 57 *E. coli* isolates. (*I*) 17 asymptomatic bacteriuria (ABU), (*II*) 15 cystitis and (*III*) 25 pyelonephritis isolates were screened for the amplification of 431 bp of *papF* gene using primers PfamA and PfamB (Table 2.3). Isolates AUTI3, AUTI16 (*II*) and AP5 (*III*) were negative. Prototype isolates CFT073, J96 and 536 were used as positive controls and *E. coli* K-12 strain AAEC185 served as a negative control. A PCR negative control (DNA blank) was also performed in every screen.

Further screening using primers papB1 and papB2 (Table 2.3) for the amplification of *papB* gene also suggested a *pap*-negative status for these three isolates and subsequently, they were excluded from the study. Reference *E. coli* isolates CFT073, J96 and 536 were used as positive controls in all PCR screens and the K-12 strain AAEC185 that does not contain a *pap* operon was used as a negative control. Weak, non-specific amplification products were observed in AAEC185 PCR reactions, but no amplification product of the correct size was present (431 bp).

4.1.2 Determination of the number of *pap* operons present in UPEC isolates by Southern hybridisation

E. coli isolates causing UTI have been shown to possess multiple P fimbrial operons. For instance, the first sequenced pyelonephritis isolate CFT073 was found to contain two *pap* operons, P1 and P2 (Welch *et al.*, 2002). In order to determine the number of *pap* operons present in each of the 54 *E. coli* isolates Southern hybridisation analysis was performed, as described in section 2.3.9. Genomic DNA digests (*Hind*III or *Bgl*II) were electrophoretically separated, transferred to a nylon membrane, and hybridised to fluorescein-dUTP labelled DNA probes complementary to *papB*. Detection of multiple *pap* operons was verified with reference UPEC isolates CFT073, J96 and 536, since the number of *pap* operons present in each is known: two, two and one, respectively (Welch *et al.*, 2002; Hull *et al.*, 1986; Hacker *et al.*, 1990). Plasmid pPap5 contains the complete *pap* operon (*papI-G*) from isolate J96 (Lindberg, *et al.*, 1984) and was used as a positive control, whereas genomic DNA from strain AAEC185 was used as a negative control. Southern blot images are shown in Figure 4.2 and the number of *pap* operons

detected in each isolate is listed in Table 4.1A.

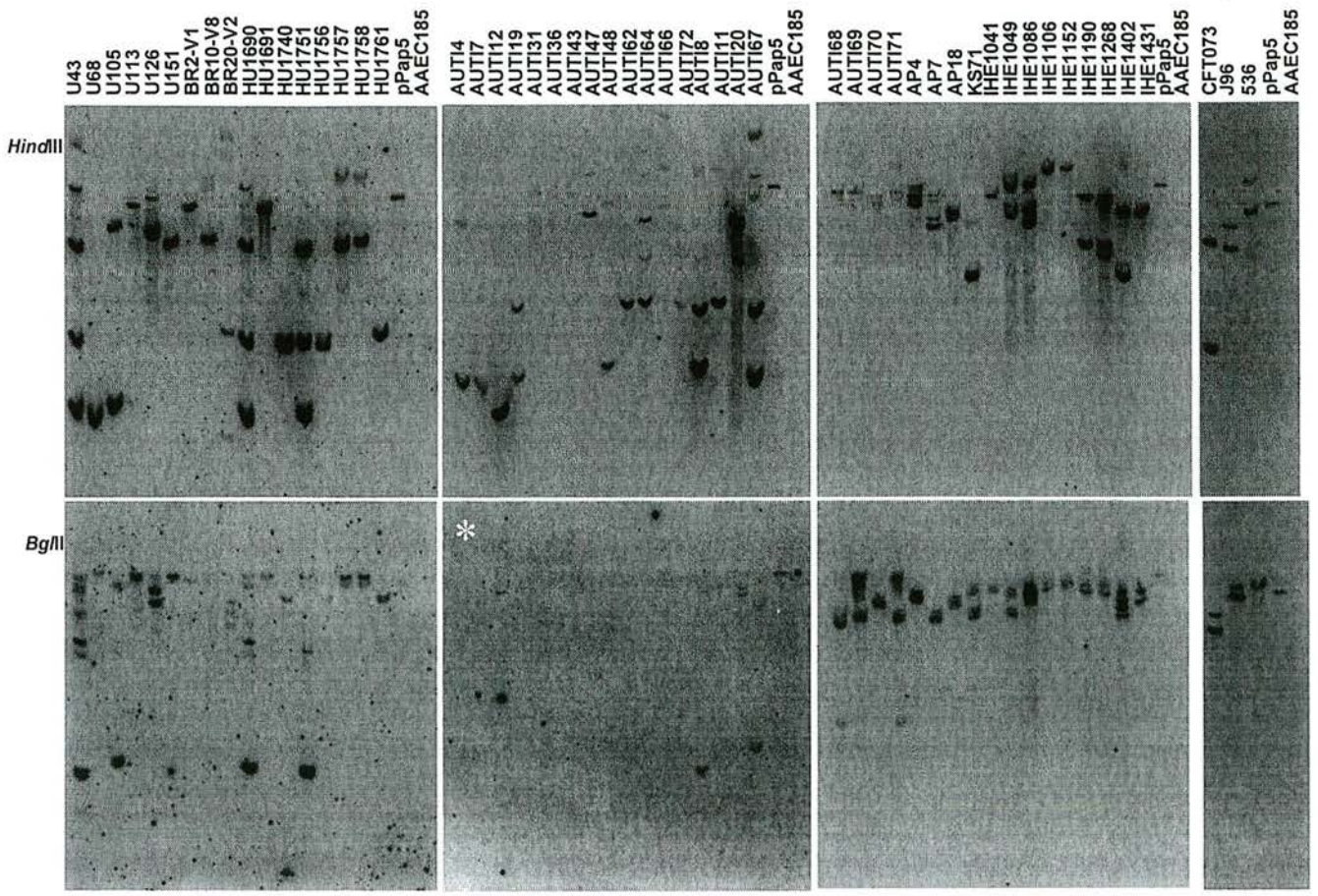


Fig. 4.2 Southern hybridisation for the detection of *pap* operons in 54 UPEC isolates. Genomic DNA digests (*Hind*III *top* –*Bgl*II *bottom*) from 54 clinical isolates were screened for the number of *papB* gene copies. Detection specificity was verified with reference strains CFT073, J96 and 536, shown in the final blots, for which 2, 2 and 1 copies were detected, respectively. Control plasmid DNA (pPap5) and AAEC185 genomic DNA digests were also included in each blot. Fainter bands were not counted as they probably represent *papB* homologues, such as *sfaB* and *focB* encoded by S and F1C fimbrial operons, respectively. *Note*: the second *Bgl*II blot (marked with the white asterisk) failed to develop in the majority of lanes and results were obtained from the corresponding *Hind*III blot.

Table 4.1 Summary of Southern hybridisation analysis. (*A*) Number of *pap* operons present in 54 *E. coli* isolates from cases of ABU, cystitis and pyelonephritis, as determined by Southern hybridisation analysis. (*B*) Frequency of asymptomatic and symptomatic isolates with one or multiple (2 or 3) *pap* operons. Evidence of association between isolate clinical origin and the number of *pap* operons it contains was obtained by chi-square test (p=0.04).

(A)

ABU		Cystitis		Pyelonephritis	
U43	3	AUTI4	2	CFT073	2
U68	1	AUTI7	1	J96	2
U105	2	AUTI12	1	536	1
U113	1	AUTI19	2	AUTI8	2
U126	2	AUTI31	1	AUTI11	1
U151	1	AUTI36	1	AUTI20	2
BR2-V1	1	AUTI43	1	AUTI67	3
BR10-V8	1	AUTI47	1	AUTI68	2
BR20-V1	1	AUTI48	2	AUTI69	2
HU1690	3	AUTI62	1	AUTI70	1
HU1691	1	AUTI64	3	AUTI71	2
HU1740	1	AUTI66	1	AP4	2
HU1751	3	AUTI72	2	AP7	2
HU1756	1			AP18	2
HU1757	1			KS71	2
HU1758	1			IHE1041	1
HU1761	1			IHE1049	2
				IHE1086	3
				IHE1106	1
				IHE1152	1
				IHE1190	2
				IHE1268	2
				IHE1402	2
				IHE1431	1

(B)

Isolate clinical origin	Number of *pap* operons		
	One	Multiple (2 or 3)	Total
Asymptomatic	12	5	17
Symptomatic	15	22	37
Total	27	27	**54**

The frequency of isolates with one or multiple (two or three) *pap* operons was determined with respect to clinical origin (asymptomatic or symptomatic) and is presented in Table 4.1B. Analysis by chi-square test provided evidence of association between the clinical origin of an isolate and the number of *pap* operons it contains ($p<0.05$). Symptomatic isolates were significantly more likely to contain multiple operons than isolates from asymptomatic cases, despite the selective bias for the presence of at least one *pap* operon in each UPEC isolate examined. Generally, a smaller proportion of asymptomatic isolates contain P fimbrial operons than symptomatic isolates (Plos *et al.*, 1989, 1990).

4.1.3 Quantification of P fimbrial expression in UPEC isolates by immunofluorescence and flow cytometry.

The observation that symptomatic isolates tend to possess multiple *pap* operons suggests that they may contain a higher proportion of P-fimbriate bacteria in their populations than asymptomatic isolates. This could be responsible for the generation of disease symptoms and acute pathology, due to the inflammatory potential of P fimbriae. In order to investigate whether the number of *pap* operons carried by an isolate affects the proportion of P-fimbriate bacteria in its population, P fimbriation in UPEC isolate populations was examined using polyclonal anti-P sera. *E. coli* isolates were grown overnight at 37°C on CFA agar to promote P fimbriation (Evans *et al.*, 1977). The bacterial cells were subsequently incubated with a mix of three anti-P rabbit polyclonal sera for primary detection of P fimbriae, followed by secondary detection using Alexa Fluor 488, a commercial green-fluorophore protein conjugate to goat anti-rabbit IgG (Molecular Probes). Three anti-P sera (OM12, OM5

and OM15) were kindly provided by Prof. T. Korhonen and each was raised against purified P fimbriae from three test strains (IH11002, IH11024, IH11062, respectively) that were selected to represent common UPEC serotypes and contained only one P fimbrial variant (Pere *et al.*, 1987). All three antisera were shown to be broadly cross-reactive and stained in total 98% of 92 P-fimbriate isolates from major pyelonephritis-associated serogroups (Pere, 1989). In this study we used an equimolar mix of the three anti-P sera to ensure recognition of a broad spectrum of P fimbrial variants encoded by the 54 UPEC isolates examined. The proportion of P-fimbriate bacteria in the population of twelve asymptomatic isolates was determined by immunofluorescence staining and flow cytometry (FC), as described in sections 2.4 and 2.5.1, respectively, and is graphically presented in Figure 4.3. The extent of P fimbriation in the examined set of ABU isolates ranged from 3% to 90%. A proportion of P-fimbriate bacteria as high as 90%, suggests that *pap* phase variation occurs at extremely high frequencies ($>10^{-2}$ $cell^{-1}generation^{-1}$; see Fig. 3.2) and would argue against the role of phase-variation in immune evasion and conservation of bacterial metabolic energy. It could be argued that isolates with multiple *pap* operons could generate high P-fimbriate bacterial proportions, if all operons phase-vary at high frequencies and if all P fimbrial types are recognised by the anti-P mix used. However, the P-fimbriate proportion measured in each isolate population does not seem to correlate with the number of *pap* operons it contains. It therefore seemed possible that the observed high fimbriate proportions are due to a staining artefact.

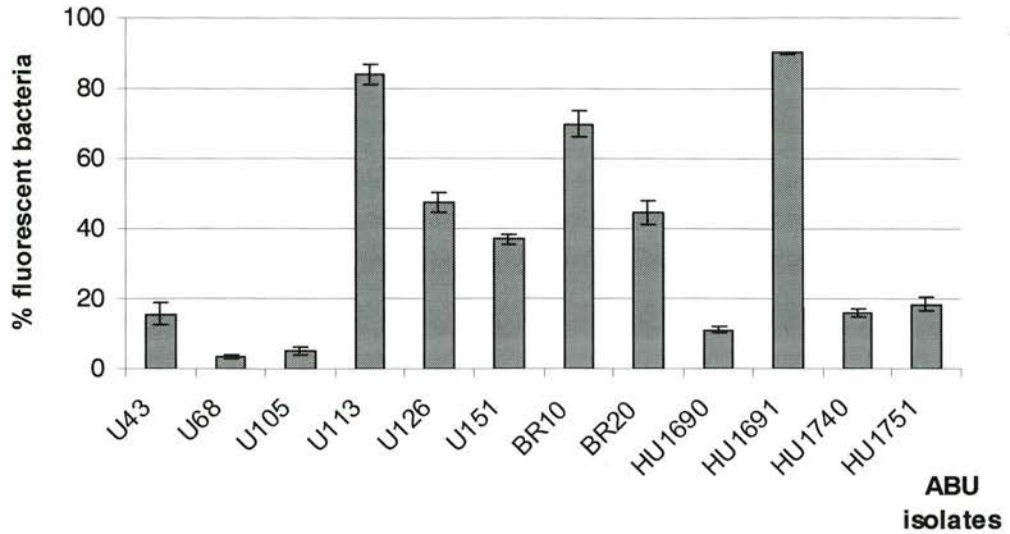


Fig. 4.3 Proportion of P-fimbriate bacteria in populations of 12 asymptomatic UPEC isolates. Immunofluorescence staining of P fimbriae was performed using a mix of three anti-P sera and the secondary fluorophore-IgG conjugate Alexa Fluor 488. Quantification of fluorescent bacteria was performed by flow cytometry in populations of constant size (20,000 cells), as described in section 2.5.1. Results are presented as mean proportions of fluorescent bacteria (± s.e.) of three biological replicates.

Examination of the FC data and fluorescence microscopy images revealed that the high number of stained cells observed for some UPEC isolates is due to cross-reaction of the anti-P antibody mix, probably with non-P surface antigens present on some but not all UPEC isolates tested. Figure 4.4A compares the fluorescence histograms obtained by FC for one UPEC population with low proportion of P-fimbriate cells (U68 - 3%) and one with high (BR10 - 67%). In the case of isolate U68 the majority of the bacteria are not stained by the anti-P mix and only 3% of the population appears positive for fluorescence (counts within marker M1). In contrast, the majority (67%) of bacteria in the population of isolate BR10 appear to be stained.

Most stained BR10 cells however, show low fluorescence intensities (FL1-Height $<10^2$) with only the right-hand skew of the distribution demonstrating high FL1-Height values. It therefore appears that most bacteria in the BR10 population react 'mildly' with the primary antibody mix and do not fluoresce as brightly as the positive U68 cells. This observation was confirmed by examination of fluorescence microscopy images of the two isolate populations (Fig. 4.4B). For U68 the majority of cells are unstained and appear as dark background, with the minority of stained cells fluorescing brightly. In contrast, in the BR10 population most cells fluoresce moderately creating a low-fluorescence background and only some of the cells fluoresce as brightly as the stained U68. A similar pattern was observed for other isolates with high numbers of stained cells, which suggests that the anti-P mix binds to non-P surface structures in an isolate-specific manner.

Non-specific binding of the primary antibody mix to surface structures other than P fimbriae was further tested by immunofluorescence staining of isolates that either do not contain P fimbrial operons (P^-) or are isogenic fimbrial knock-out mutants. Staining of two P^- cystitis isolates, AUTI1 and AUTI73 (Keegan, 2001) with the anti-P antibody mix resulted in 10% and 0% P-fimbriate subpopulations, respectively. This result provides supportive evidence for the lack of antibody specificity to P fimbriae and suggests that non-specific binding is isolate-specific.

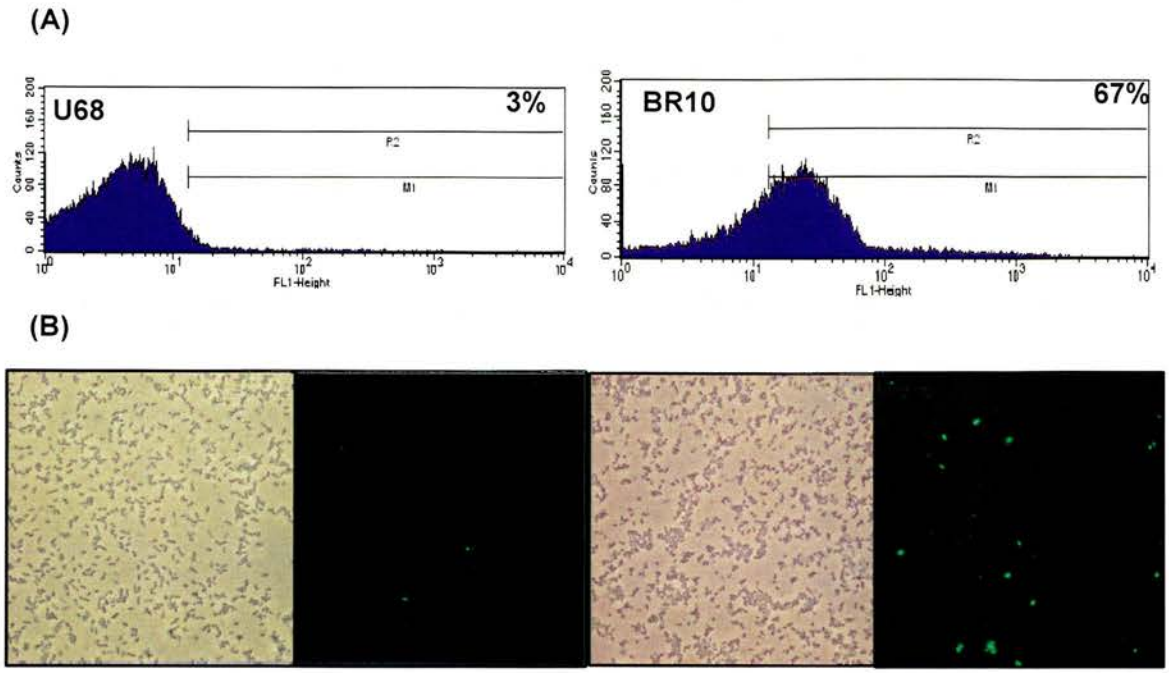


Fig. 4.4 Comparison of flow cytometry data and fluorescence microscopy images for two asymptomatic UPEC isolates with different proportions of P-fimbriate bacteria. Asymptomatic isolates U68 and BR10 were stained with a mix of three anti-P sera and the secondary fluorophore-IgG conjugate Alexa Fluor 488. (*A*) Flow cytometry histograms showing the fluorescence intensity of each cell in the examined population. The proportion of fluorescent bacteria is shown at the top right corner. (*B*) Phase contrast and immunofluorescence microscopy of the two isolates employing the anti-P sera mix. Different staining patterns and higher background staining for isolate BR10 were observed.

Further evidence was provided by comparing immunofluorescence staining of isolate 536 and isogenic fimbrial knock-out mutants (Table 2.1A). This reference pyelonephritis isolate encodes for P-related fimbriae (*prf*), S fimbriae (*sfa*), and type 1 fimbriae (*fim*) (Brzuszkiewicz *et al.*, 2006). Each strain was labelled with the anti-P mix (Fig. 4.5A) and with each anti-P serum separately (Fig. 4.5B), in order to

determine which of the three accounts for the non-specific binding. The anti-P mix resulted in high numbers of antibody-labelled bacteria in the 536 isolate population and all the fimbrial mutants, including the *prf* deletion strain, suggesting non-specific binding to a non-P structure on the surface of 536 (Fig. 4.5A). Anti-P serum OM5 appeared to be mainly responsible for the high levels of non-specific binding, as it stained an equal proportion of bacteria with the anti-P mix (Fig. 4.5B). Anti-P sera OM12 and OM15 resulted in much lower proportions of stained bacteria, but both failed to specifically detect the 536 P fimbrial variant, as similar proportions were found in populations of 536 wild-type and 536 *prf*$^-$ (Fig. 4.5B). Fluorescence microscopy confirmed a similar staining pattern between 536 wild-type and the fimbrial mutants with the anti-P mix or each of the anti-P sera separately.

Reactivity of each anti-P serum was finally tested using reference isolates CFT073 and J96. Immunofluorescence staining results from FC and microscopy further supported the observed trend that each anti-P serum reacts in a different way with the surface of different uropathogenic *E. coli* isolates. Interestingly, anti-P serum OM12 was found to be specific for P fimbriae of the model isolate CFT073 as it failed to cross-react with an isogenic P fimbrial mutant strain (ZAP965). It was therefore concluded that the three anti-P sera are unsuitable for measuring the P-fimbriate proportions in diverse UPEC isolate populations, and specific detection of P fimbriae should be confirmed using isogenic P fimbrial mutants.

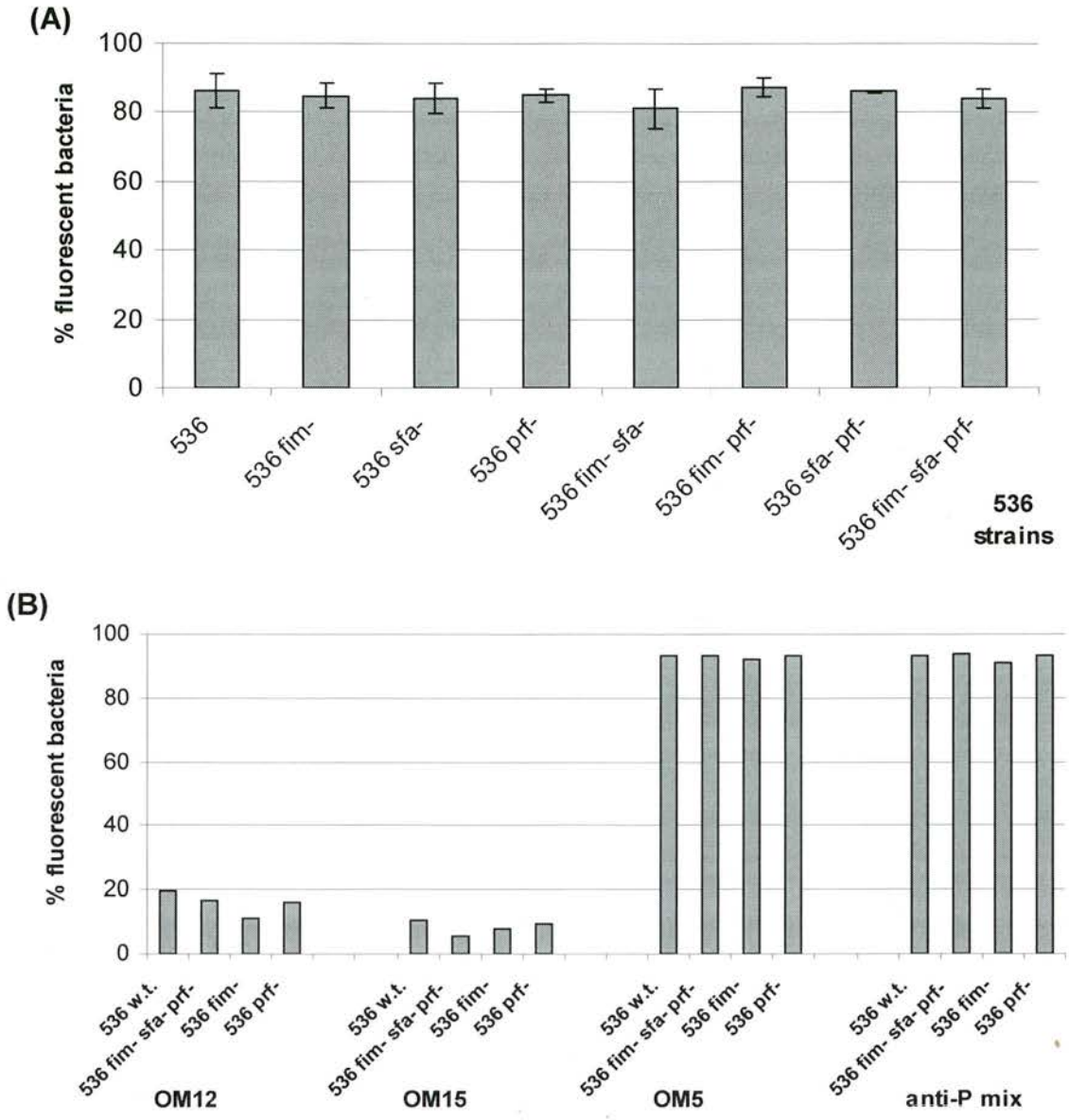


Fig. 4.5 Proportion of antibody-labelled bacteria in populations of *E. coli* 536 and isogenic fimbrial knock-out mutants. (*A*) Seven isogenic fimbrial mutant strains were compared to 536 wild-type (w.t.) for staining with a mix of three anti-P sera. Results are presented as mean proportions (± s.e.) of three biological replicates. (*B*) Three isogenic fimbrial mutants were compared to 536 wild-type for staining with three separate anti-P sera (OM12, OM15, OM5) or an equimolar mix of the three (anti-P mix).

4.1.4 Quantification of P fimbrial expression in UPEC isolates by mannose-resistant haemagglutination (MRHA).

Due to the lack of a specific anti-P serum, P-fimbriate bacterial proportions for each UPEC isolate were assessed by MRHA of human red blood cells (RBCs), a measure of P-fimbrial binding capacity (Norgern *et al.*, 1984). Assays were performed by Miss E. Mahler in a parallel study investigating cross-talk between P and type-1 fimbriae in UPEC clinical isolates (Holden *et al.*, 2006, Appendix B). P fimbrial adhesins (PapG) recognise the P bloodgroup antigens present on human erythrocytes and binding leads to agglutination. Mixing of equal concentrations of RBCs with serially-diluted bacteria allowed the quantification of P fimbrial expression for each isolate. Assays were performed in the absence and presence of mannose to differentiate between agglutination caused by P and type-1 fimbriae, as the latter is sensitive to mannose. The agglutination titre was determined as the highest bacterial dilution that prevented formation of a defined RBC pellet and the reciprocal value determined for each UPEC isolate is listed in Table 4.2.

Levels of MRHA were investigated with respect to isolate clinical origin and number of *pap* operons and are presented as boxplots in Figure 4.6. Analysis of variance (ANOVA) suggested that symptomatic isolates exhibit significantly higher levels of MRHA than asymptomatic isolates ($p<0.05$). There was also statistically significant evidence suggesting that the number of *pap* operons present in an isolate has a positive effect on the level of MRHA ($p<0.001$).

Isolate ID	Total HA	MRHA	No. of *pap* operons
HU1690	32	16	3
HU1691	0	0	1
HU1740	8	2	1
HU1751	4	2	3
HU1756	16	16	1
HU1757	2	1	1
HU1758	1	1	1
HU1761	4	2	1
U43	64	8	3
U68	0	0	1
U105	64	64	2
U113	2	2	1
U126	16	16	2
U151	0	0	1
BR2	8	4	1
BR20	16	16	1
AUTI4	64	64	2
AUTI7	16	8	1
AUTI12	0	0	1
AUTI19	32	32	2
AUTI31	16	16	1
AUTI36	2	2	1
AUTI43	16	16	1
AUTI47	16	16	1
AUTI48	32	32	2
AUIT62	8	8	1
AUTI64	16	16	3
AUTI66	4	4	1
AUTI72	32	32	2
AUTI8	16	8	2
AUTI11	8	8	1
AUTI20	16	16	2
AUTI67	16	16	3
AUTI68	16	16	2
AUTI69	16	16	2
AUTI70	64	64	1
AUTI71	16	16	2
AP4	8	8	2
AP7	64	64	2
AP18	0	0	2
IHE1041	8	8	1
IHE1049	2	2	2
IHE1086	32	32	3
IHE1106	0	0	1
IHE1152	0	0	1
IHE1190	8	8	2
IHE1268	8	8	2
IHE1402	64	64	2
IHE1431	8	8	1
KS71	64	64	2

Table 4.2 Levels of MRHA of human RBCs induced by UPEC isolates. Total agglutination and mannose-resistant haemagglutination (MRHA) titres are expressed as the reciprocal value of the highest bacterial dilution that prevented formation of a defined RBC pellet. The number of *pap* operons present in each isolate is also listed. The table is divided in three sections containing isolates from distinct clinical UTI syndromes: ABU (*top*), cystitis (*middle*) and pyelonephritis (*bottom*).

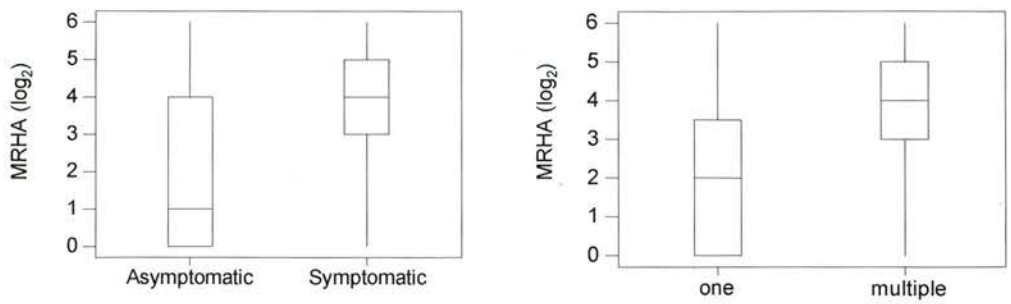


Fig. 4.6 MRHA levels of UPEC clinical isolates. Boxplots showing the level of MRHA for clinical isolates as a function of their clinical origin (asymptomatic *vs.* symptomatic; *left*) or the number of *pap* operons they contain (one *vs.* multiple; *right*). MRHA levels are expressed as log_2 of the maximum agglutination titre and the mean was significantly higher in the symptomatic group of isolates (p= 0.01) and the group containing multiple *pap* operons and (p<0.001).

Taken together the results demonstrate that symptomatic UPEC isolates are more likely to contain multiple *pap* operons in their chromosome and higher proportions of P-fimbriate cells in their populations than asymptomatic isolates. Moreover, these two attributes appear to be linked, suggesting that the genetic carriage and functional expression of multiple *pap* operons could impact on the isolate's pathogenic potential and fitness in the urinary tract.

4.2 Cloning of the regulatory region of *pap* operons from UPEC isolates

Having identified the number of *pap* operons present in each of the 54 UPEC isolates, the aim was to obtain and examine the sequence of their regulatory regions. PCR was used to amplify the *pap* regulatory regions present in each isolate with primers designed complementary to highly conserved regions at the 3' end of *papI* and the 5'

end of *papA* [CFTP1+2 for and rev] (Table 2.3). An enzyme mix containing *Taq* and the proof-reading DNA polymerase *Tgo* (Roche) was used for high fidelity amplification (section 2.3.4). PCR products were subsequently cloned into the pCR4®-TOPO® vector as described in section 2.3.11. Recombinant plasmid clones were screened for the presence of an inserted *pap* regulatory region by digestion with *Eco*RI restriction endonuclease. Screening of several clones from each reaction revealed the presence of inserts with small size differences, confirming the successful amplification and cloning of different *pap* regulatory regions from the same isolate. An example is shown in Figure 4.7 for isolate IHE1086 that was shown to contain three *pap* operons by Southern blot analysis (Table 4.1A). Six *Eco*RI-digested clones are shown with small differences in insert size.

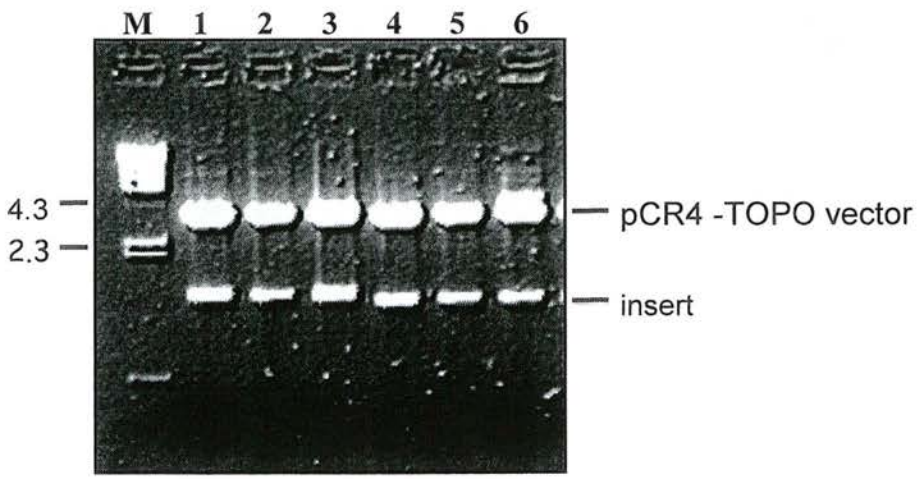


Fig. 4.7 Screening of six IHE1086 TOPO® clones for the presence of insert by *Eco*RI digestion. Differences in insert sizes are visible between clones 1, 2, 3 and 4, 5, 6. Lane M: λ *Hind*III DNA ladder with 4.3 and 2.3 Kb reference bands shown.

4.3 Sequencing of *pap* regulatory regions from UPEC isolates

The TOPO®-cloned *pap* regulatory regions from each isolate were sequenced by MWG-Biotech, as described in section 2.3.12. The number of sequenced clones per isolate was determined by and was in excess of the number of *pap* operons detected in each isolate by Southern hybridisation (Table 4.1A); two clones were sequenced for isolates containing one *pap* operon, five for isolates containing two, and eight for isolates containing three operons. Reference strain CFT073 served as a control for the reliability of this rule, as the sequence of each *pap* regulatory region (P1 and P2) was represented by two TOPO®-clones. Double-stranded sequences were assembled for each regulatory region using high-quality sequence data from four individually-primed sequencing reactions that covered both directions. Base quality was thoroughly checked particularly for singletons.

All sequenced clones per isolate were compared by multiple sequence alignment (msa) to identify regulatory regions belonging to the same or different *pap* operons. The number of *pap* operons identified in each isolate by sequencing of their regulatory regions is listed in Table 4.3A. One caveat of sequencing only the regulatory region of each *pap* operon was that it would be impossible to identify two operons with identical regulatory regions, so the total number of identified operons could be underestimated. Nevertheless, there was little discrepancy between the number of *pap* operons identified in each isolate by Southern hybridization analysis and by sequencing of the regulatory regions (compare Tables 4.1A and 4.3A). For the majority of isolates, sequence information was obtained for every *pap* regulatory region detected by Southern hybridization. The regulatory region sequence of 8 *pap*

operons identified by Southern blotting was not obtained (n/a; AUTI4, AUTI8, AUTI20, AUTI71, AP4, AP18, IHE1268, KS71), and there were two cystitis isolates for which an extra *pap* operon was identified by sequencing (AUTI43, AUTI66). Table 4.3B lists the number of *pap* operons present in each of the 54 UPEC isolates, as determined by combination of Southern hybridisation and sequencing analyses. These data were used in all subsequent analyses. It is important to note that the association between isolate clinical origin and number of *pap* operons was still evident in the revised dataset (p=0.042).

Table 4.3 Number of *pap* operons present in 54 *E. coli* isolates from cases of ABU, cystitis and pyelonephritis, as determined (*A*) by sequencing of regulatory regions and (*B*) by combining the results of Southern hybridisation and sequencing analyses.

(A)

ABU		Cystitis		Pyelonephritis	
U43	3	AUTI 4	1	CFT073	2
U68	1	AUTI 7	1	J96	2
U105	2	AUTI 12	1	536	1
U113	1	AUTI 19	2	AUTI 8	1
U126	2	AUTI 31	1	AUTI 11	1
U151	1	AUTI 36	1	AUTI 20	1
BR10-V8	1	AUTI 43	2	AUTI 67	3
BR20-V1	1	AUTI 47	1	AUTI 68	2
83972	2	AUTI 48	2	AUTI 69	2
HU1690	3	AUTI 62	1	AUTI 70	1
HU1691	1	AUTI 64	3	AUTI 71	1
HU1740	1	AUTI 66	2	AP 4	1
HU1751	3	AUTI 72	2	AP 7	2
HU1756	1			AP 18	1
HU1757	1			KS71	1
HU1758	1			IHE 1041	1
HU1761	1			IHE 1049	2
				IHE 1086	3
				IHE 1106	1
				IHE 1152	1
				IHE 1190	2
				IHE 1268	1
				IHE 1402	2
				IHE 1431	1

Table 4.3 – *continued*

(B)

ABU		Cystitis		Pyelonephritis	
U43	3	AUTI 4	2	CFT073	2
U68	1	AUTI 7	1	J96	2
U105	2	AUTI 12	1	536	1
U113	1	AUTI 19	2	AUTI 8	2
U126	2	AUTI 31	1	AUTI 11	1
U151	1	AUTI 36	1	AUTI 20	2
BR10-V8	1	AUTI 43	2	AUTI 67	3
BR20-V1	1	AUTI 47	1	AUTI 68	2
83972	2	AUTI 48	2	AUTI 69	2
HU1690	3	AUTI 62	1	AUTI 70	1
HU1691	1	AUTI 64	3	AUTI 71	2
HU1740	1	AUTI 66	2	AP 4	2
HU1751	3	AUTI 72	2	AP 7	2
HU1756	1			AP 18	2
HU1757	1			KS71	2
HU1758	1			IHE 1041	1
HU1761	1			IHE 1049	2
				IHE 1086	3
				IHE 1106	1
				IHE 1152	1
				IHE 1190	2
				IHE 1268	2
				IHE 1402	2
				IHE 1431	1

4.4 Sequence analysis of *pap* regulatory regions from UPEC isolates

In total, sequence information was obtained for 82 *pap* regulatory regions from 54 UPEC clinical isolates. The sequenced regions included the *papI* and *papB* genes, their intergenic region (*pap* switch) and the *papB-A* intergenic sequence up to the first codon of *papA*. Comparison by msa revealed conserved and variable regions, which are depicted in Figure 4.8. Conserved regions, not surprisingly, contained sites known to play a key role in the reversible operation of the *pap* epigenetic switch, suggesting the importance of preserving their function. In contrast, the presence of sequence variation, both in sites of the *pap* switch and the two regulators *papI* and

papB, suggested potential changes in their function that could affect *pap* expression. In this chapter the sequence of important features of the *pap* regulatory region is described. Observed sequence variation was investigated in the context of the clinical isolate combining information on isolate clinical origin, number of *pap* operons and level of P fimbriation. Sequence variation in the *papI* gene and the PapB binding site 1 is described in greater detail separately in chapters 5 and 6, respectively, together with the major findings from analysing their effect on *pap* expression.

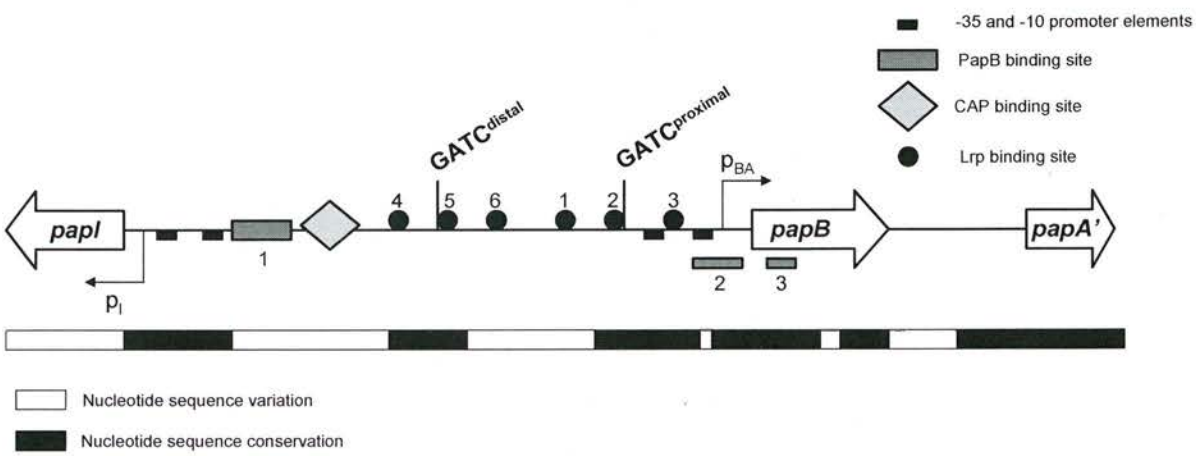


Fig. 4.8 The *pap* regulatory region. The ~1.2 Kb sequenced region from *pap* operons present in 54 UPEC isolates is shown schematically, including the *papI*, *papB* and beginning of *papA* genes. The three PapB, one CAP and six Lrp binding sites present in the *papI-B* intergenic region (*pap* switch) are shown along with the -35 and -10 hexamers of the two divergent promoters (p_I and p_{BA}). Two Dam methylation sites ($GATC^{proximal}$ and $GATC^{distal}$) are shown overlapping Lrp sites 2 and 5, respectively. A sequence variation bar below the *pap* regulatory region marks roughly areas of nucleotide conservation (black bar) and variation (white bar) identified by msa of 82 sequenced *pap* regulatory regions. The length of genes, intergenic regions and the distance between protein binding sites are not drawn to scale.

4.4.1 p_I and p_{BA} promoters

The *pap* switch contains two divergent promoters, p_I and p_{BA} driving transcription of *papI* and the *pap* operon, respectively. The sequence of the putative -10 and -35 p_I elements was conserved in all sequenced switches (Fig. 4.9), suggesting that expression of the PapI regulator is not modulated through sequence variation in regions interacting with the σ^{70} subunit of RNA polymerase. The -35 and -10 hexamers of the divergent p_{BA} promoter were also highly conserved (Fig. 4.9). In particular, the -35 element was conserved in 100% of the sequenced switches, whereas the sequence of the -10 box contained one variable base: $TA^{A}/_{C}CTT$. Cytosine at position -11 was only observed in 5 *pap* switches (6%) that also had a thymine, instead of guanosine, at the transcription initiation site (+1).

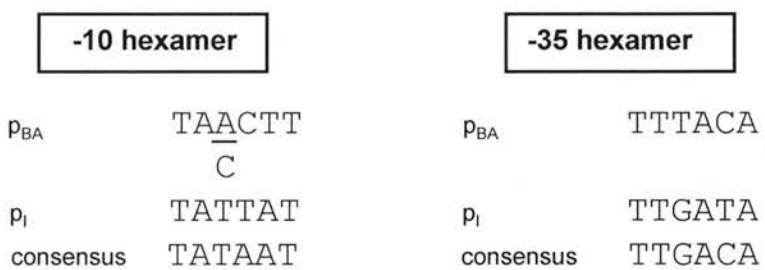


Fig. 4.9 Sequence of p_I and p_{BA} promoter elements. The conserved hexamer sequence of -10 and -35 promoter elements found in 82 p_I and p_{BA} promoters from UPEC isolates is shown together with the consensus sequences found in strong *E. coli* promoters (Busby and Ebright, 1994). The -10 element sequence of p_{BA} contained one variable base 11 bp upstream of the transcription start site that was observed in 6% of the promoter sequences examined. Both p_I promoter elements were 84% identical to the consensus sequences, whereas -10 and -35 elements of p_{BA} promoter were 50% and 84% identical to consensus, respectively.

4.4.2 Dam and Lrp binding sites

The mechanism of *pap* phase variation relies on the competition between Dam and Lrp for access to two overlapping binding regions present within the *pap* switch: one proximal to the main operon promoter p_{BA} and one distal (Fig. 4.8). The on and off methylation patterns of the *pap* switch are generated by Dam recognition and adenine methylation of one of the two GATC signal sequences and are established by Lrp binding to the other, non-overlapping region. Both GATC signal sequences are therefore important for the reversible operation of the *pap* switch and mutations in both have been shown to lock the switch to either phase (Braaten *et al.*, 1994; Nou *et al.*, 1995). This was reflected by the conservation observed for these sequences among UPEC isolates. Of the 82 sequenced *pap* switches only two were found to contain mutations in the GATC methylation signals. In both cases $GATC^{proximal}$ was mutated to GATT (Fig. 4.10), a transition that will abort recognition of the site by Dam and prevent subsequent adenine methylation. This point mutation is also predicted to disrupt Lrp binding to site 2 since Lrp was shown to come to close contact with the first and last guanosine residues within $GATC^{proximal}$ (one in the top and one in the bottom DNA strand) (Nou *et al.*, 1995). Moreover, a 6 bp substitution mutation over $GATC^{proximal}$ that disrupted Lrp binding to site 2 was shown to lock the *pap* switch on (Nou *et al.*, 1995). Interestingly, in this study both *pap* switches found to contain the $GATC^{proximal}$ mutation belonged to UPEC isolates that demonstrated high MRHA levels (see AUTI48 and AP7 in Table 4.2), indicating high levels of P fimbriation. Both isolates are of symptomatic clinical origin and contained a second *pap* operon with both Dam signal sequences intact. In contrast to $GATC^{proximal}$, the sequence of $GATC^{distal}$ overlapping Lrp binding site 5 was conserved in all *pap*

switches examined (Fig. 4.10), indicating that Dam methylation and Lrp binding to site 5 can occur and allow transition of the *pap* switch from on-to-off.

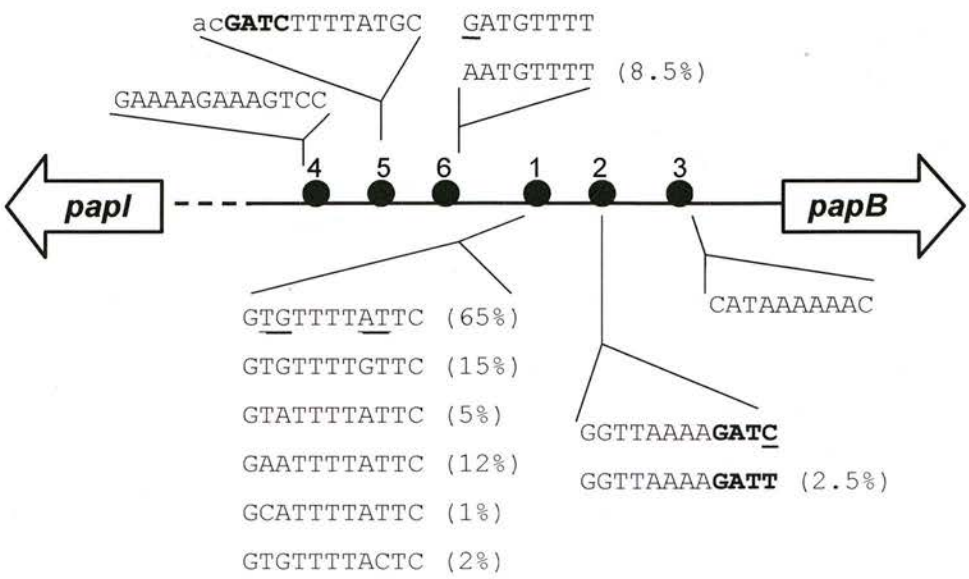


Fig. 4.10 Sequence of Lrp and Dam binding sites in the *pap* regulatory region. The sequence of Lrp sites 3, 4, and 5 was identically conserved in 100% of the *pap* switches examined. Variable bases identified in Lrp sites 1, 2 and 6 are underlined and the sequence of every variant is shown together with the frequency with which it was observed. The two GATC signal sequences for Dam methylation overlapping Lrp sites 2 and 5 and are shown in bold. The sequence upstream of Lrp site 5 is part of the PapI 'response element' and is shown in lower case. Sequence variation between sites 6 and 1 is not shown.

The flanking sequences of both *pap* GATC sites are important in determining Dam's methylation ability and processive catalysis of the two sites (Peterson and Reich, 2006). All *pap* switches examined contained identical GATC flanking sequences (gacGATC^{dist}ttt and aaaGATC^{prox}gtt) which are poorly preferred for methylation by

Dam (Peterson and Reich, 2006), indicating that Dam methylation kinetics are conserved among *pap* operons and do not serve as targets for modulating *pap* expression. The conserved sequence immediately upstream of $GATC^{distal}$ (Fig. 4.10) is also part of the PapI 'response element' ACGATC, which was shown to be required for PapI-dependent binding of Lrp to site 5 and therefore off-to-on *pap* transition (Hernday *et al.*, 2003).

Each of the two Lrp binding regions in *pap* DNA contains three separate binding sites (sites 1-3 are located proximal to the p_{BA} promoter and sites 4-6 distal, see Figure 4.8). All six Lrp binding sites present in the *pap* switch are important for the reversible operation of the switch and were found to be highly conserved. The sequence of each site is shown in Figure 4.10 together with sequence variation observed in sites 1 and 6 and the frequency with which they occurred. Variation in site 6 involved the guanosine residue of the consensus Lrp binding sequence for *pap* DNA 'Gnn(n)TTTt' (Nou *et al.*, 1995). A transition to adenine was observed in 7 *pap* operons, all of symptomatic clinical origin. It is difficult to predict the impact of this point mutation on *pap* expression as the role of Lrp site 6 in phase variation is not completely clear. Even though the consensus 'G' residue of all six binding sites has been shown to come in close contact with Lrp, the complete substitution of site 6 sequence barely affected *pap* off-to-on transition frequency (Nou *et al.*, 1995), suggesting that site 6 does not play a significant role in the regulation of *pap* expression. Variation in site 1 was more commonly observed, but did not involve any of the conserved residues in the consensus Lrp binding site 'Gnn(n)TTTt'. Nevertheless, the observed sequence variation could affect Lrp binding to site 1,

which is essential for *pap* transcription since the translocation of Lrp from proximal to distal binding sites has been proposed to require the interaction of PapI with Lrp bound at site 1 (Nou *et al.*, 1995; Kaltenbach *et al.*, 1995). The region between sites 6 and 1 was also found to contain some variable residues that could have a modulatory effect on the regulation of *pap* expression. Evidence indicating that this region could play a positive role in regulation of *pap* transcription was provided by a 6 bp substitution mutation between sites 6 and 1 that resulted in a locked off phenotype (Nou *et al.*, 1995).

4.4.3 CAP binding site

Expression of P fimbriae is repressed by glucose metabolism (Baga *et al.*, 1985). During glucose starvation *E. coli* synthesises cyclic AMP (cAMP) which is bound by catabolite activator protein (CAP). CAP was shown to bind *pap* DNA in a cAMP-dependent manner at a site centred at 215.5 bp upstream of the p_{BA} transcription start site and binding was essential for activation of *pap* transcription (Goransson *et al.*, 1989b; Forsman *et al.*, 1992; Weyand *et al.*, 2001). The region of CAP binding was proposed to function as an upstream activating sequence for both p_I and p_{BA} promoters (Goransson *et al.*, 1989b) and was conserved among the *pap* switches examined in this study. Figure 4.11 shows the sequence variants of the CAP binding site found in 82 *pap* switches together with their associated frequencies. All variable bases identified were not part of the consensus CAP binding site sequence 5'-AAnTGTGAnnTnnnTCAnATW-3' (Ebright *et al.*, 1984) suggesting that the observed sequence variation is unlikely to alter CAP binding affinity. Interestingly, the first variable base was observed with high frequency (35%) and associated with

other sequence variation upstream and downstream of the CAP site. This could indicate a potentially altered interaction between CAP and RNA polymerase via the C-terminal domains of its α subunits (αCTD), since possible αCTD binding sites in *pap* DNA could be the AT-rich sequences flanking the CAP site. The second variable base identified within the CAP site was found in almost 10% of the sequenced switches, whereas the final variable base was only observed in the *pap* switch of the reference isolate J96.

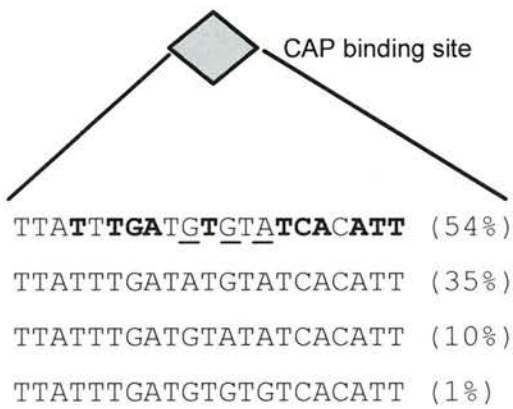
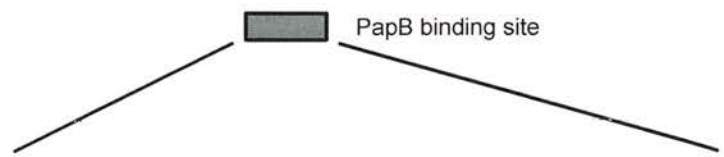


Fig. 4.11 Sequence of the CAP binding site in the *pap* regulatory region. The variable bases identified are underlined and the sequence of each variant site is shown together with the frequency it was observed among the 82 *pap* switches examined. Bases identical to the consensus CAP binding site 5'-AAnTGTGAnnTnnnTCAnATW-3' are shown in bold (W is either A or T; Ebright *et al.*, 1984).

4.4.4 PapB binding sites

Binding studies demonstrated the presence of three *pap* DNA regions with affinity for the *pap*-encoded regulator PapB (Forsman *et al.*, 1989). Named in the order of their PapB affinity (from highest to lowest), site 1 mapped >200 bp upstream of the

papBA transcription start site and directly adjacent to the CAP binding site, site 2 overlapped the -10 p_{BA} promoter element and site 3 was localised within the *papB* gene (see Fig. 4.8). The position of each site, its affinity for PapB, and the finding that high PapB protein levels repressed *pap* expression, led to the proposition that PapB acts both as an activator and a repressor of *pap* transcription and that its autoregulatory role is achieved by differential binding to the three sites (Forsman *et al.*, 1989). The sequence of the PapB binding sites is AT-rich and contains conserved bases (Forsman *et al.*, 1989; Xia *et al.*, 1998). PapB site 1 was the most variable region identified within the *pap* switch and analysis of its sequence variation is the research focus of the following chapter. In contrast, the sequence of PapB sites 2 and 3 was highly conserved. In particular, site 3 was 100% identical in all the examined switches and site 2 contained a few variable bases at non-conserved positions. The sequences of the three different PapB site 2 variants and of site 3 are shown in Figure 4.12 together with their associated frequencies.



Site 2: A**A**TTT<u>A</u>AC**TTATTGC**<u>G</u>T<u>G</u>**A**AGAGT<u>A</u>TTT (81%)
A**A**TT**T**CAC**TTATTGC**GTT**A**AGAGTGTTT (6%)
A**A**TT**T**AAC**TTATTGC**ATG**A**AGAGTATTT (13%)
Site 3: G**A**AA**T**GCG**TTTTTGC**TGA**A**TATACGCGA (100%)

Fig. 4.12 Sequence of PapB binding sites 2 and 3 in the *pap* regulatory region. The variable bases identified are underlined and the sequence of site 2 variants is shown together with their occurrence frequency among the 82 *pap* switches examined. Identical bases between sites 2 and 3 are shown in bold (Forsman *et al.*, 1989).

4.4.5 *papB* gene

The first cistron to be transcribed from the main *pap* operon promoter is *papB*. The gene encodes for one of the two transcriptional regulators of *pap* expression that was shown to have autoregulatory function (Forsman *et al.*, 1989). Comparing the sequence of 82 *papB* genes obtained from *pap* operons belonging to 54 UPEC isolates by msa revealed the presence of mutations. In 22% of the genes examined were the same frameshift mutation was found that introduced a premature stop codon resulting in a truncated PapB protein. The resulting 35 amino acid polypeptide is predicted to be non-functional as it is lacking both the DNA binding and oligomerisation regions required for PapB to function as a transcriptional regulator (Xia and Uhlin, 1999, see Fig. 4.13). Investigation of the isolates encoding for the mutant *papB* genes revealed that the great majority contained multiple *pap* operons and therefore wild-type copies of *papB* that could complement the mutant operon (Table 4.4). In fact, five of the six isolates found to have 3 *pap* operons contained at least one with a mutant *papB* gene, but no isolate had all its copies of *papB* mutated. Isolates with a single *pap* operon were rarely found to carry a mutated *papB* gene ($p<0.05$), as this would abolish P fimbrial expression, unless other fimbrial operons encoding for PapB homologues, such as SfaB, were present in the same isolate. Indeed, three of the four isolates that contained a mutated *papB* gene in their only *pap* operon exhibited very low MRHA levels (see HU1740, U68 and AUTI12 in Table 4.2). Interestingly, these isolates although rare, tended to be of asymptomatic clinical origin, suggesting that the *papB* gene mutation could be part of the general 'taming' of pathogenicity demonstrated by *E. coli* isolates causing ABU, such as isolate 83972 (Hull *et al.*, 1999; Graham *et al.*, 2001; Klemm *et al.*, 2006). This

prototype asymptomatic UPEC strain was shown to encode for a non-functional PapG (Klemm *et al.*, 2006) and in this study, a second *pap* regulatory region was identified containing a mutant *papB* gene.

Table 4.4 Frequency of mutant *papB* genes in 54 UPEC isolates. UPEC isolates with multiple *pap* operons (2 or 3) were found to be more likely to contain at least one mutant *papB* gene by a chi-square test of association (p=0.036).

	Number of UPEC isolates with:			
	1 *pap* operon	2 *pap* operons	3 *pap* operons	**Total**
mutant *papB*	4	8	5	17
wild-type *papB*	20	16	1	37
Total	24	24	6	**54**

The sequence of the remaining majority of *papB* genes was intact and highly conserved. Thirteen single nucleotide polymorphisms (SNPs) were observed generating 11 haplotypes with an estimated nucleotide diversity (Π) of 0.007% ±0.002%. Eight of the SNPs led to synonymous amino acid substitutions and the remaining 5 in replacements. Translation of the *papB* alleles resulted in three protein variants designated B.1-B.3 (Fig. 4.13). Interestingly, variant B.3 was only found in the operon of prototype UPEC strain 536 and contained four of the five observed amino acid replacements. Similarly, variant B.2 was present only in one operon and differed in a single amino acid residue from variant B.1, which represents the sequence encoded by the great majority (96%) of operons (Fig. 4.13). The amino acid sequence of the three variants is homologous and the observed amino acid

substitutions are not predicted to change the overall structure of the protein. Moreover, the observed amino acid changes involved non-conserved residues of PapB family members (Xia and Uhlin, 1999; Holden *et al.*, 2001).

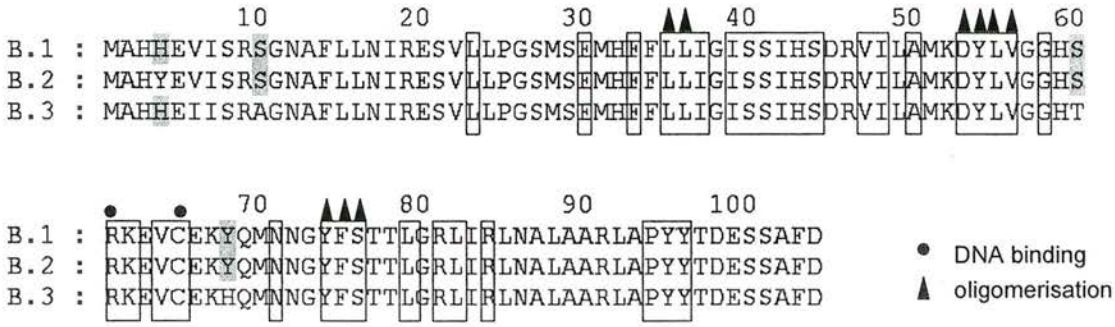


Fig. 4.13 Amino acid sequence alignment of PapB protein variants. All *pap* operons containing intact *papB* genes were found to encode for protein variant B.1, except for one operon from isolate IHE1268 (B.2) and the *prf* operon from reference UPEC 536 (B.3). The observed amino acid substitutions (shaded) did not involve residues conserved among members of the PapB regulator family (boxed) or consensus residues required for DNA binding (black circle) and oligomerisation (black triangle) previously identified (Xia and Uhlin , 1999).

In summary, the sequence of regulatory features of the *pap* epigenetic switch was conserved among UPEC isolates with little sequence variation tolerated suggesting that the reversible function of the switch is highly conserved. The sequence of *papB* gene, encoding one of the operon's transcriptional regulators, was also found to be highly conserved, signifying the importance of maintaining PapB's regulatory function, as well as the limited tolerance for amino acid replacements, since most SNPs resulted in synonymous changes and the most frequent mutation resulted in

loss of protein function. Sequence variation that could result in modulation of *pap* regulation and expression was identified in the *papI* regulator and PapB binding site 1, described in chapters 5 and 6.

4.5 Discussion

In this study a selection of 54 UPEC clinical isolates was used as a source of 82 *pap* operons with the aim to investigate the sequence of their regulatory regions. The isolates used originated from asymptomatic or symptomatic UTI cases and their clinical origin was found to correlate with the number of *pap* operons they contained, as well as the level of functional expression of P fimbriae. Specifically, symptomatic isolates were significantly more likely to possess two or more *pap* operons compared to asymptomatic isolates and displayed higher levels of P fimbriation, as assessed by MRHA. These findings were in agreement with previous studies that report increased *pap* operon carriage in symptomatic UPEC isolates, as well as lower functional expression of P fimbriae in asymptomatic isolates (Plos *et al.*, 1989, 1990; Andersson *et al.*, 1991; Blanco *et al.*, 1997). Moreover, the results from this phenotypic and genotypic characterisation provided the basis for investigating for the first time regulatory cross-talk between P and type 1 fimbriae in clinical UPEC isolates (Holden *et al.*, 2006, Appendix B). Type 1 fimbrial expression correlated in an inverse manner with P fimbrial expression and was significantly reduced in the group of symptomatic UPEC isolates that exhibited the highest levels of P fimbrial expression.

Despite chromosomal carriage of at least one *pap* operon in all *E. coli* isolates examined, the asymptomatic group had the highest proportion of isolates displaying no clear MRHA, implying either that expression of P fimbriae is repressed in these isolates or that mutations preventing expression have occurred in operon genes. This is consistent with asymptomatic strains not expressing factors that can trigger an inflammatory response in the host (Hull *et al.*, 1999; Graham *et al.*, 2001; Lim *et al.*, 1998). Attenuation of pathogenicity through inactivating mutations has recently been demonstrated in the prototype asymptomatic *E. coli* isolate 83972 for several of its fimbrial operons: type 1 and P (Klemm *et al.*, 2006), and F1C (Roos *et al.*, 2006b). Several lines of evidence, including multilocus sequence typing (MLST) and the possession of FimH and PapA alleles associated with virulent UPEC strains, suggest that strain 83972 originated from a pyelonephritis UPEC isolate (Klemm *et al.*, 2006). In this study strain 83972 was found to possess a novel *pap* regulatory region with several sequence differences to the published *pap* sequence (Klemm *et al.*, 2006). Interestingly, nucleotide variation was observed in both *pap*-encoded regulators; *papI* contained unique base changes resulting in two amino acid replacements in the protein sequence and *papB* contained the commonly-observed frameshift mutation resulting in PapB protein truncation and inactivation. This finding further supports the notion that *E. coli* 83972 has evolved from an ancestral pyelonephritis UPEC isolate as it contains two *pap* operons, while the presence of inactivating mutations in both operons reflects the strain's lifestyle transition and need to inactivate genes that are no longer compatible with the novel lifestyle. Assessing this possibility for each of the asymptomatic isolates used in this study would require extensive sequence analysis of complete *pap* operons, which would be outside the scope of the current

analysis. It is therefore possible that any of the 82 *pap* operons investigated contains mutations in operon genes that could abolish functional expression of P fimbriae.

The research focus of this study was directed in the regulatory region of the detected *pap* operons, with the aim of examining modulation of *pap* regulation through sequence changes. The most prominent areas of sequence variation identified in the 82 sequenced *pap* regulatory regions involved the *papI* gene, encoding for the *pap* transcriptional activator, and PapB site 1, the binding site with the highest affinity for the *pap* autoregulator PapB (Forsman *et al.*, 1989). Investigating the effect of these variable sequence regions on *pap* regulation and expression constitutes the remainder of this thesis and is presented and discussed in chapters 5 and 6.

In this chapter regions of sequence conservation were described and were found to include sites known to be important for the phase-variable expression of the *pap* operon. In particular, the two GATC sites targeted by Dam and required for the formation of differential methylation patterns that govern *pap* expression were highly conserved. In addition, *pap* autoregulation seemed to be highly conserved through sequence conservation of PapB sites 2 and 3 that are responsible for repressing *pap* expression (Forsman *et al.*, 1989). In contrast, the sequence of site 1 was highly variable, involving insertion/deletion of 9 bp sequences, which suggested that activation of *pap* expression is subject to modulation, whereas repression is highly conserved and occurs at much higher frequencies (Blyn *et al.*, 1989; Holden *et al.*, 2007). This signifies the importance of tightly controlling *pap* expression, as P fimbriae are metabolically expensive for the bacterium, as well as targets for the host

immune system.

Minor sequence variation was also identified in other features of the *pap* regulatory region, such as protein binding sites and the *papB* gene. These observations provide insights into the extent of sequence variability that naturally occurs in *pap* regulatory regions and could be part of future research. For example, the effect of sequence variation in Lrp binding sites 1 and 6 (see Fig. 4.10) on the affinity of Lrp binding could be investigated using electrophoretic mobility shift assays. Changes in Lrp affinity could alter the sensitivity of *pap* transcription to leucine (Ernsting *et al.*, 1993), potentially explaining the differential effect of amino acid-containing medium on the off-to-on transition frequency of P1 and P2 operons (see section 3.5 and Table 3.1). Additionally, future work could investigate the *pap* operons observed to contain mutant *papB* genes and integrate them in single-copy into the chromosome of CFT073 wild-type and its isogenic P fimbrial mutant ZAP965 to test for complementation by wild-type *papB* copies in a clinical isolate genetic background. Previous studies suggested that complementation of *papB* mutants can occur upon expression of wild-type PapB from plasmid-based systems (Baga *et al.*, 1985; Goransson *et al.*, 1989b; Morschhauser *et al.*, 1994), but it will be interesting to determine whether phase-variable expression of *pap* operons containing mutant *papB* genes can be restored by wild-type PapB levels found in the same UPEC isolate.

The finding that *pap* operons with mutant *papB* sequences frequently occur in UPEC isolates containing intact *papB* genes in other *pap* operons, together with the sequence conservation observed among intact *papB* genes, is particularly interesting

when considering regulatory cross-talk within UPEC isolates, as it provides a possible mechanism for modulating expression of homologous *pap* operons found in the same isolate. Expression of a *pap* operon carrying a mutant *papB* gene will rely on the regulatory action of wild-type PapB protein expressed from other operons that have to be higher in the regulatory hierarchy. Assuming that the concentration of PapB regulator required for stimulating self as opposed to related *pap* operons is different, then this mechanism would preclude co-expression of homologous *pap* operons and increase the probability of them being sequentially expressed at the single-cell level and under identical environmental conditions. The presence of sequence variation in other sites of the *pap* switch as well as the *pap* transcriptional activator *papI*, could be serving a similar role in modulating *pap* expression, and together with *papB* mutations it could provide a mechanism for coordinating expression of homologous P fimbrial operons in UPEC isolates. These possibilities are explored further in the remaining chapters.

Chapter 5

Sequence variation in PapB binding site 1 and effects on *pap* phase variation

This chapter describes the sequence variation observed in a high-affinity PapB binding site present in *pap* regulatory regions from UPEC isolates of different clinical origin (asymptomatic *vs.* symptomatic UTI). The sequence variation was examined in the context of each UPEC isolate, the number of *pap* operons it contains and its clinical origin. Formulated hypotheses regarding the role of PapB site 1 sequence variation in *pap* operon phase-variable expression are described, together with the *in vitro* and *in vivo* methods used to test them and the associated findings.

5.1 Variation in PapB site 1 sequences from UPEC isolates

The most variable region identified in the 82 *pap* switches sequenced in this study (see sections 4.3 and 4.4) involved a protein binding site for the *pap*-encoded regulator PapB. This is the first of three PapB binding sites present in the regulatory region of the *pap* operon and was shown to have the highest affinity for PapB and to act as an activating site for *pap* transcription (Forsman *et al.*, 1989). The sequence of PapB site 1 is AT-rich and contains multiple 9 bp sequences with $(T/A)_3$ at conserved positions and repeating with 9 bp periodicity (Xia *et al.*, 1998). PapB protein has been proposed to recognise these repeats via minor groove interactions and bind in an oligomeric fashion. Site 1 mutagenesis studies have demonstrated the importance of the repeated $(T/A)_3$ sequences and their periodicity in PapB binding capacity and *pap* transcriptional activation (Xia *et al.* 1998). Sequencing of 82 *pap* regulatory regions from 54 UPEC strains as described in chapter 4 of this thesis, revealed variation in the number of 9 bp sequence repeats present in PapB site 1. Five naturally-occurring site 1 sequence variants containing 4, 6, 7, 8 or 9 repeats were identified (Fig. 5.1).

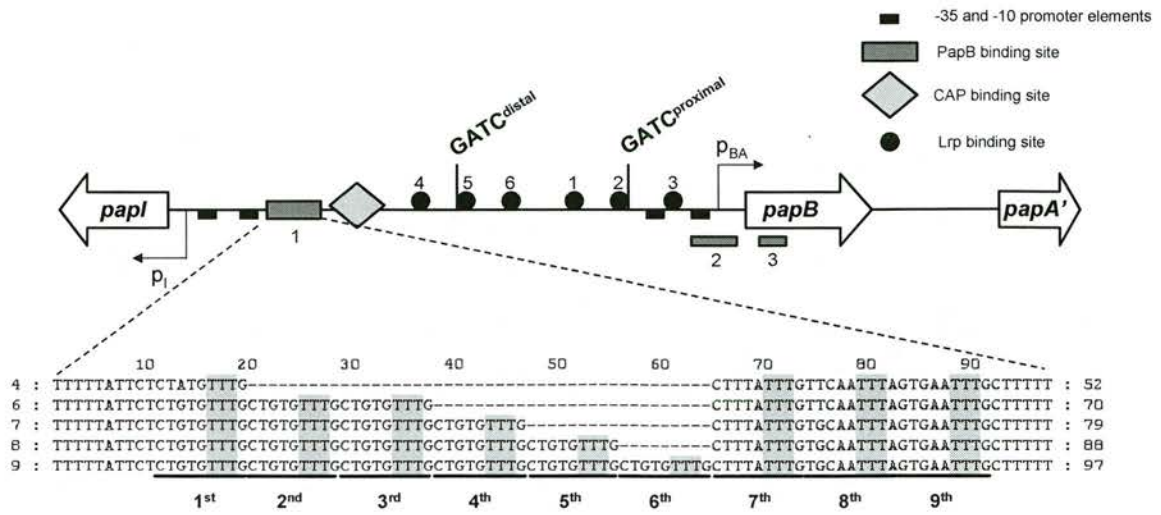


Fig. 5.1 Alignment of PapB site 1 sequence variants from UPEC isolates. Five natural sequence variants of PapB site 1 were identified in 82 sequenced *pap* regulatory regions from 54 UPEC isolates. Sequence variation involved the number of 9 bp repeats (underlined) containing $(T/A)_3$ at conserved positions (shaded grey). The variants are designated 4, 6-9 according to the number of 9 bp repeats they contain. The minimum site 1 identified was 52 bp long and contained four 9 bp repeats, and the maximum was 97 bp with nine repeats. Note that the second, third, fourth, fifth and sixth 9 bp repeats have identical sequence: 5'-CTGTGTTTG-3', which is also shared by the first repeat present in sites with >4 repeats (6, 7, 8, or 9). Sites with 4 repeats were found to contain base differences in non-conserved positions and one variable base can be seen in the first 9 bp repeat of variant 4.

Longer site 1 variants appeared to result from insertion of additional 9 bp sequences between the first and second 9 bp repeats of the smallest site 1 variant (4 repeats). The sequence of the additional 9 bp repeats was identical in 100% of the sites examined (5'-CTGTGTTTG-3') and identical to the sequence of the first 9 bp repeat. A few variable nucleotides were also identified in the first and penultimate 9 bp sequence repeats, but involved non-conserved positions and did not alter the important $(T/A)_3$ feature or the 9 bp periodicity.

Each *pap* operon in the UPEC isolate collection was characterised for the number of 9 bp repeats present in its PapB site 1 sequence (Table 5.1A). The majority (62%) of site 1 sequences examined contained 4 repeats, whereas sites with > 4 repeats (6, 7, 8, and 9) were less frequent among the *pap* operons, with each variant occurring in approximately 10% of the sequences (Table 5.1B).

Table 5.1 PapB site 1 sequence variants in *pap* operons from UPEC clinical isolates. (*A*) Fifty four UPEC isolates divided in three clinical groups (ABU-top, cystitis-middle, pyelonephritis-bottom) are listed together with the number of *pap* operons contained in each one, as determined by Southern hybridisation and sequencing of regulatory regions (see Table 4.3B). Every *pap* operon present in each isolate was scored for the number of 9 bp repeats (4, 6, 7, 8, or 9) within its PapB site 1 sequence. (*) Isolates containing a *pap* operon detected by Southern hybridisation, for which the sequence of the regulatory region was not obtained. (*B*) Frequency of each PapB site 1 sequence variant (with 4, 6, 7, 8 or 9 repeats) among the 82 *pap* regulatory regions examined.

(A)

Isolate ID	No. of *pap* operons	No. of 9 bp repeats in PapB site 1				
		4	6	7	8	9
HU1690	3	2[a]		1		
HU1691	1	1				
HU1740	1				1[a]	
HU1751	3	1[a]		2		
HU1756	1					1
HU1757	1	1				
HU1758	1	1				
HU1761	1				1	
U43	3	3[a]				
U68	1	1[a]				
U105	2	1[a]			1	
U113	1	1				
U126	2	1	1			
U151	1	1				
BR10	1	1				
BR20	1	1[a]				
83972	2	2[a]				
AUTI4*	2	1[a]				
AUTI7	1	1				
AUTI12	1	1[a]				
AUTI19	2	1	1			

Table 5.1 - *continued*

AUTI31	1	1				
AUTI36	1	1				
AUTI43	2	1				1
AUTI47	1	1				
AUTI48	2	1[a]			1	
AUITI62	1			1		
AUTI64	3	2[a]		1		
AUTI66	2	1				1
AUTI72	2	1		1		
AUTI8*	2		1			
AUTI11	1			1		
AUTI20*	2	1				
AUTI67	3	3[a,a]				
AUTI68	2	2[a]				
AUTI69	2	1				1
AUTI70	1	1				
AUTI71*	2				1	
AP4*	2				1	
AP7	2	1[a]			1	
AP18*	2	1				
IHE1041	1					1
IHE1049	2		1	1		
IHE1086	3	1				2
IHE1106	1	1				
IHE1152	1	1				
IHE1190	2	1[a]				1
IHE1268*	2					1
IHE1402	2	1[a]	1			
IHE1431	1	1				
KS71*	2		1			
CFT073	2	1	1			
J96	2	2				
536	1	1				

[a]*pap* operon with mutant *papB* gene encoding for truncated PapB (see section 4.4.5)

(B)

No. of 9 bp repeats in PapB site 1	Frequency among *pap* operons
4	51
6	7
7	8
8	7
9	9
Total	**82**

The majority of isolates with a single *pap* operon (75%) contained a site 1 variant with 4 repeats, although this observation was not statistically significant (Table 5.2A). It is important to note though that in the UPEC collection studied, asymptomatic isolates are significantly more likely to contain a single *pap* operon ($p<0.05$; see Table 4.3B). These observations are suggestive of a trend between asymptomatic isolates, a single *pap* operon and a site 1 sequence with 4 repeats. Similarly, the majority of isolates with two *pap* operons (75%) contained a combination of site 1 variants with 4 and >4 repeats (Table 5.2B). Alternative combinations (4 and 4, or >4 and >4) were rare; only one isolate was found to contain a >4-repeat site variant in both operons, and three isolates contained two 4-repeat site variants. Interestingly, two of the latter carried mutant *papB* gene copies in one of the two *pap* operons. Furthermore, the great majority (87.5%) of isolates with two *pap* operons in the UPEC collection studied are of symptomatic clinical origin.

Table 5.2 Frequency of PapB site 1 sequence variants in (*A*) 24 UPEC isolates with a single *pap* operon and (*B*) 16 UPEC isolates with 2 *pap* operons, for which sequence information was obtained for both regulatory regions. PapB site 1 variants were grouped into sequences with 4 or >4 repeats (6, 7, 8 or 9). The expected frequencies were estimated from the overall frequency of site 1 variants (see Table 5.1B) and were compared to the observed frequencies using the chi-square goodness of fit test. (*A*: $\chi^2 = 1.67$ on 1 degree of freedom, $p>0.05$; *B*: $\chi^2 = 5.025$ on 2 degrees of freedom, $p>0.05$).

(A)

No. of 9 bp repeats in PapB site 1	Observed frequency among isolates with one *pap* operon	Expected frequency
4 repeats	18	14.927
>4 repeats	6	9.073

(B)

Combination of PapB site 1 variants	Observed frequency among isolates with 2 *pap* operons	Expected frequency
4, 4 repeats	3	6.189
4, >4 repeats	12	7.524
>4, >4 repeats	1	2.28

To test for a potential association between the number of repeats in PapB site 1 sequence and the extent of P fimbriation in an *E. coli* isolate, the level of MRHA of UPEC isolates with a single *pap* operon (see Table 4.2) was examined in relation to the number of repeats present in its PapB site 1 sequence (4 or >4). It appeared that isolates with a single *pap* operon containing sites with 4 repeats tended to display slightly lower levels of MRHA, although this observation was not statistically significant (Fig. 5.2).

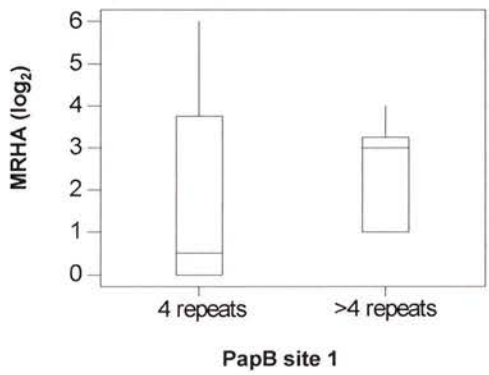


Fig. 5.2 MRHA levels of UPEC clinical isolates with one *pap* operon. Boxplots showing the distribution of MRHA levels of isolates with a single *pap* operon containing 4 or >4 repeats in their PapB binding site 1 sequence. MRHA levels are expressed as log_2 of the maximum agglutination titre. Equality of median values (middle line in boxplots) was tested using the Kruskal-Wallis non-parametric test and median MRHA levels were not found significantly different (p= 0.2).

Based on the observed trends in UPEC isolates and the current knowledge on the function of PapB site 1 (Forsman *et al.*, 1989; Xia *et al.*, 1998), the observed sequence variation was hypothesised to alter the mode of PapB interactions with *pap* DNA and subsequently influence *pap* phase variation and P fimbrial expression.

5.2 Effect of PapB site 1 sequence variation on PapB protein binding

PapB protein was shown to bind to site 1 with high affinity (Forsman *et al.*, 1989). To investigate whether the observed sequence variation in site 1 altered the PapB binding ability, gel mobility shift assays were carried out, as described in section 2.7. Due to the tendency of PapB protein to form large, insoluble aggregates, whole-cell extracts containing over-expressed PapB were used. These were generated by transformation of *E. coli* NEC11 (Table 2.1A) with pNJH36, a pET11-based expression vector containing the *papB* gene (Table 2.2). Control extracts were obtained similarly from transformation of *E. coli* NEC11 with vector pET11. DNA fragments containing the natural site 1 sequence variants with 4, 6, 7, 8 or 9 repeats were labelled with ^{32}P-dATP during PCR amplification using primers papBsite1_F and papBsite1_R (Table 2.3). Sequenced TOPO®-clones containing the corresponding *pap* regulatory regions were selected for PCR amplification templates. Figure 5.3 depicts a sequence alignment of the different DNA fragments used in the electrophoretic mobility shift assays (EMSAs). Isogenic DNA fragments for site 1 variants 7, 8 and 9 were amplified from the TOPO®-cloned regulatory regions AUTI72.1, HU1761.2 and IHE1086.2 respectively, but no naturally isogenic regulatory regions were found for site 1 variants 4 and 6. Instead, PapB site 1 variants with 4 and 6 repeats were amplified from P1 and P2 *pap* operons of the sequenced UPEC strain CFT073. Sequence variation outside PapB site 1 can be seen in Figure 5.3.

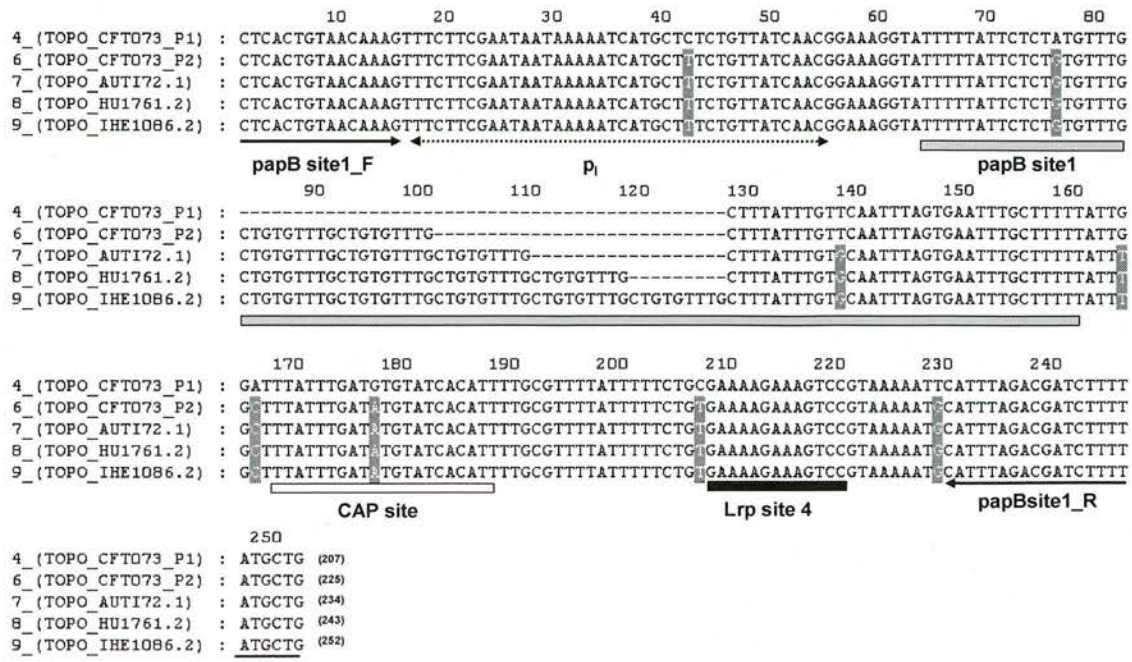


Fig. 5.3 Alignment of DNA fragments used in EMSAs. The sequence of five DNA fragments containing different PapB site 1 variants (grey bar) is shown together with some other important sequence features: PCR primers papBsite1_F and papBsite1_R (black arrows), the region containing p_I promoter elements (double dotted arrow), the CAP binding site (white bar), and Lrp binding site 4 (black bar). The fragments are designated 4, 6, 7, 8 or 9 according to the number of 9 bp repeats present in their PapB site 1 sequence and the plasmid template used for their amplification is shown in parenthesis. Sequence differences outside PapB site 1 are shaded in grey. *Note*: Fragments 7, 8 and 9 are isogenic PapB site 1 variants, whereas fragments 4 and 6 represent the PapB site 1 sequence found in the two *pap* operons (P1, P2) from *E. coli* CFT073 and contain further base differences.

For all EMSA reactions an equal amount of each DNA fragment was incubated with serial dilutions of PapB cell extract (NEC11pNJH36) and reactions were subjected to gel electrophoresis. Negative reactions were performed for each DNA fragment using either control cell extract (NEC11pET11) or no cell extract (free DNA). Figure 5.4 illustrates representative autoradiographs of EMSA reactions using DNA fragments with 4 (minimum) and 9 (maximum) repeats.

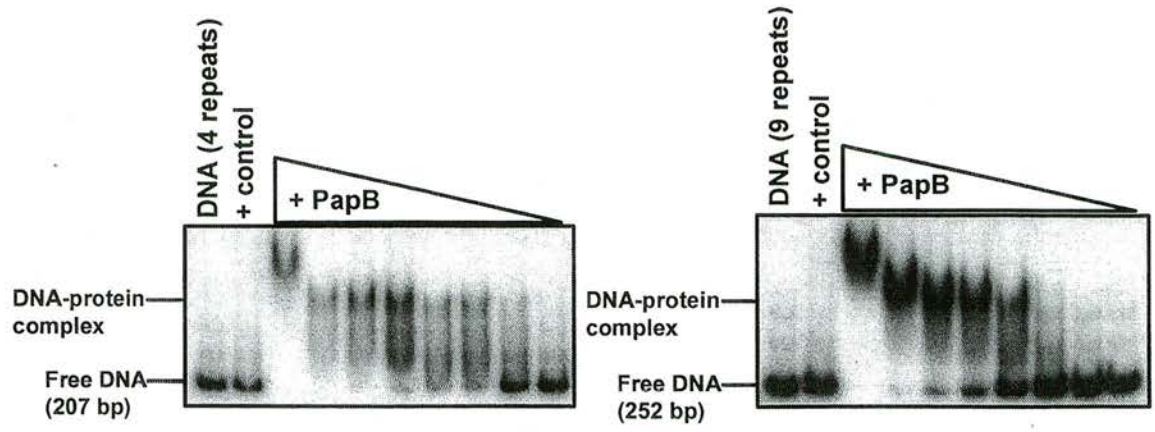


Fig. 5.4 Electrophoretic mobility shift assays (EMSAs) of PapB binding to site 1 variants with 4 or 9 repeats. DNA control samples (no cell extract) are shown in the first lane of both gels and a sample mix of DNA and the control cell extract NEC11pET11 is shown in the second lane. The 8 final lanes contain sample mixes of DNA with serial dilutions of cell extract containing over-expressed PapB (NEC11pNJH36). The free DNA and DNA-protein complex bands are marked in each gel. Images are autoradiographs of representative experiments.

The loss of free DNA was quantified and relative PapB binding affinities for each site 1 variant are shown in Figure 5.5. The results demonstrate that higher protein concentrations are required for 50% occupancy of longer site 1 variants. Similarly, comparison of the proportion of shifted DNA at mid-range protein concentrations (e.g. protein dilutions 0.0313 and 0.0625) suggested that longer site 1 variants (8 and 9) have lower relative affinities for PapB compared to shorter sites (4 and 6), as they demonstrated a lower proportion of shifted DNA. At high relative PapB concentrations, the majority (90-100%) of shorter site 1 variants was shifted, whereas >20% of DNA fragments with 8 and 9 repeats remained free. Interestingly, investigation of binding at low protein concentrations (0.0078) showed a higher

proportion of shifted DNA for longer site 1 variants compared to sites with 4 and 6 repeats. This small difference could be indicative of a potential difference in initial PapB binding to site 1 variants. Since two PapB molecules have been shown to occupy each 9 bp sequence present in site 1 (Xia *et al.* 1998), it was hypothesised that longer site 1 variants will have a higher probability of interacting with PapB due to the higher number of 9 bp repeats present in their sequence. Consequently, site 1 variation could be affecting the sensitivity of activation of *pap* operons and their phase variation frequencies.

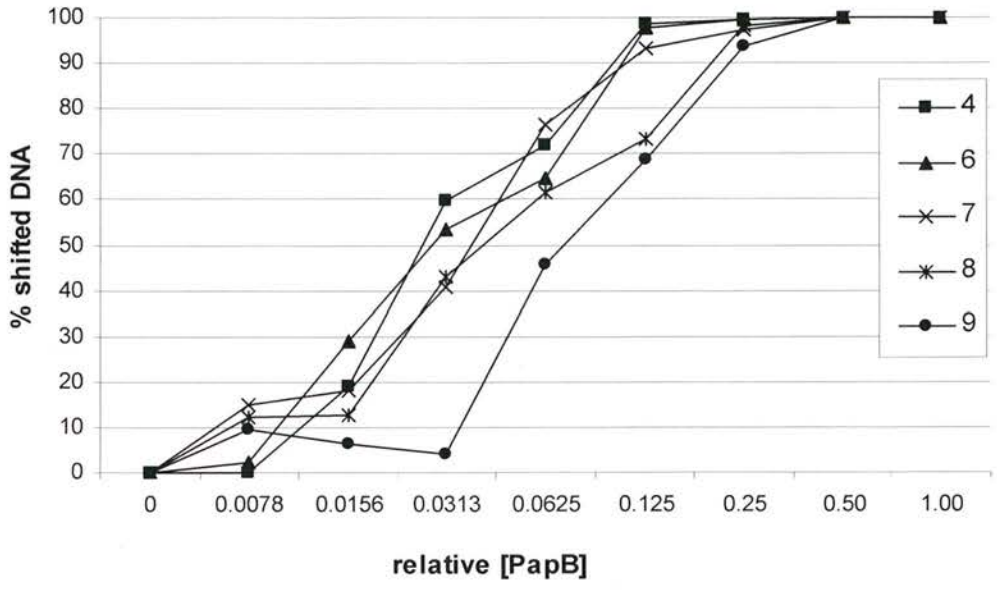


Fig. 5.5 Relative PapB binding affinity of natural PapB site 1 sequence variants. Binding of five DNA fragments containing PapB site 1 sequences with variable number of 9 bp repeats (4, 6-9) was carried out using serial dilutions of whole cell extract containing over-expressed PapB (NEC11pNJH36). The relative amount of DNA shifted at every relative PapB concentration was estimated for each DNA fragment and the mean of 4-6 independent experimental replicates is graphically presented.

5.3 Effect of PapB site 1 sequence variation on *pap* phase transition frequencies

The requirement for different PapB concentrations suggested that sequence variation in site 1 can have an effect on *pap* phase variation and regulation of P fimbrial expression. In order to investigate the effect of site 1 sequence variation on *pap* phase transition frequencies, *pap* regulatory regions containing natural site 1 variants were fused to the *lacZ* reporter gene. Pap regulatory regions containing site 1 variants with 4, 6, 7, 8, or 9 repeats were selected from sequenced TOPO®-clones and sub-cloned into *Bam*HI-*Kpn*I of the pIB307-based vector pKC29 (Table 2.2) for construction of translational *lacZ* fusions. The reporter gene was placed under the control of the main operon promoter p_{BA} and both *papI* and *papB* regulators were included in the fusion. Regulatory regions with 7, 8 and 9 repeats were isogenic site 1 variants, whereas P1 and P2 regulatory regions from UPEC CFT073, which were used to represent natural site 1 variants with 4 and 6 repeats, contained further single-base changes. All fusions were subsequently exchanged into the *lac* chromosomal locus of the pyelonephritis *E. coli* isolate CFT073 generating strains ZAP838 (4), ZAP1137 (6), ZAP1000 (7), ZAP1002 (8) and ZAP1003 (9) (Table 2.1C; for construction of *E. coli* strains see section 2.3.15). *E. coli* CFT073 was selected as the appropriate genetic background for regulatory integration of the fusions, as it provided a relevant context in which to assess *pap* expression. Work presented in chapter 1 of this thesis has shown that reporter expression levels from p_{BA} promoter fusions in an *E. coli* K-12 background are lower than the clinical isolate background (Holden *et al.*, 2007). Pap phase transition frequencies of CFT073 strains carrying single-copy *lacZ* fusions were determined by switching assays, as described in section 2.6.

In rich defined medium the off-to-on transition frequencies of strains containing fusions with longer PapB site 1 variants were higher than the on-to-off frequency displayed by the strain containing a site 1 sequence with 4 repeats ($p<0.001$, Fig. 5.6-top). In addition, off-to-on switching of *pap* regulatory regions with 7 repeats in site 1 was occurring at significantly lower frequencies than switching of isogenic variants with 8 or 9 repeats ($p<0.001$). Switching frequencies in the opposite direction (on-to-off transition) were measured under identical culture conditions and were found to be slightly lower in CFT073 strains containing isogenic fusions with 7, 8 or 9 site 1 variants ($p=0.001$, Fig. 5.6-bottom). On-to-off frequencies were more than 100-fold higher than off-to-on for all CFT073 strains tested.

Switching frequencies under different culture conditions however were less affected by sequence variation in PapB site 1. Off-to-on phase transition frequencies on a rich complex medium considered to be optimal for P fimbrial expression (CFA; Evans *et al.*, 1977), appeared to be high for all tested fusions (Fig. 5.7-top). Significant differences were still observed in the mean frequency displayed by the different CFT073 strains ($p<0.001$, ANOVA), but they were more prominent between strains with 4 and 9 repeats (minimum and maximum site 1 variants). The on-to-off transition frequencies were also determined on CFA medium and were found to be similarly high to rich-defined M9 medium (Fig. 5.7-bottom). The CFT073 strains containing longer site 1 variants (6, 7, 8 or 9 repeats) had equally high mean on-to-off transition frequencies, whereas the strain with 4 repeats in site 1 switched off at a slightly lower mean frequency ($p<0.05$).

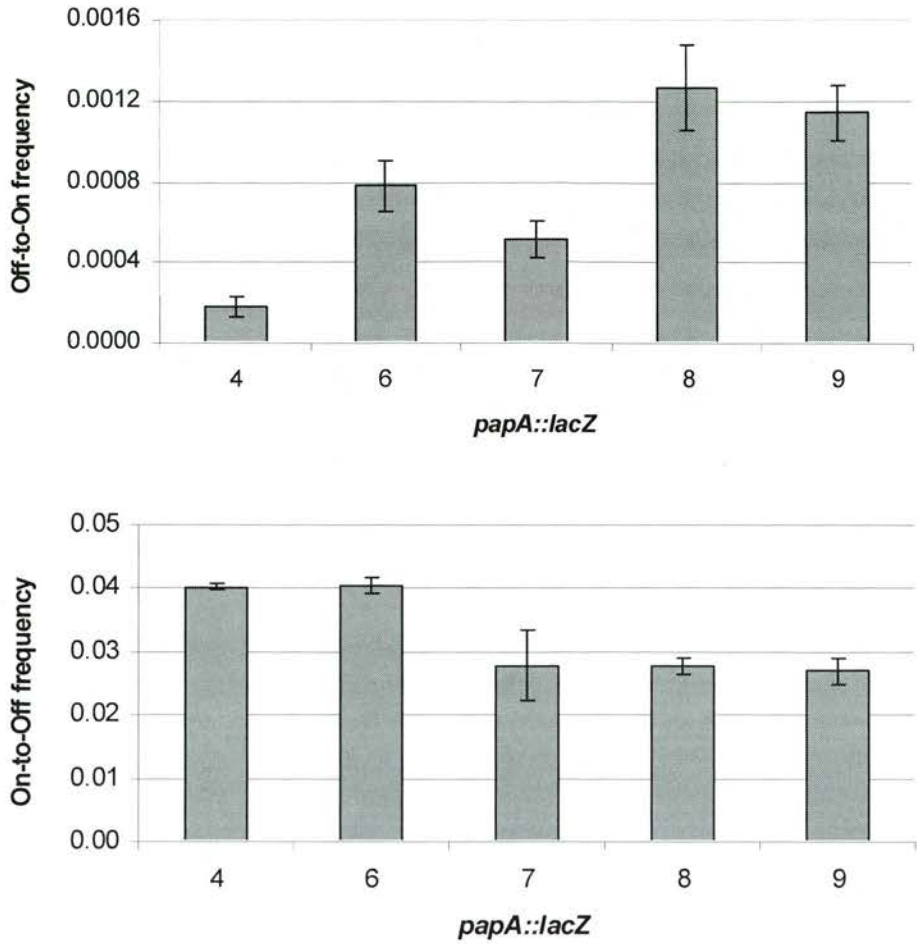


Fig. 5.6 Phase transition frequencies of *E. coli* CFT073 strains containing single-copy *papA::lacZ* fusions with natural PapB site 1 variants in rich defined medium (M9 AA glycerol). CFT073 strains ZAP838/1137/1000/1002/10003 contain single-copy *lacZ* fusions to *pap* regulatory regions with 4, 6, 7, 8, or 9 repeats in PapB site 1, respectively. Off-to-on (*top graph*) and on-to-off (*bottom graph*) transition frequencies (cell^{-1} generation^{-1}) were determined under identical culture conditions (M9 RD glycerol, 37°C) by switching assays, as described in section 2.6, and are presented as mean frequencies (± s.e.). On average, 6500 colonies of each CFT073 strain were assayed for estimating the off-to-on transition frequency and 1000 colonies for the on-to-off frequency. Due to lack of normality in the distribution of the observed frequencies (both directions), the conservative non-parametric Kruskal-Wallis test was used to compare the median frequency displayed by each CFT073 strain. There was supportive evidence for differences among the median frequencies (both directions) of the tested strains (off-to-on, $p<0.001$ and on-to-off, $p=0.001$).

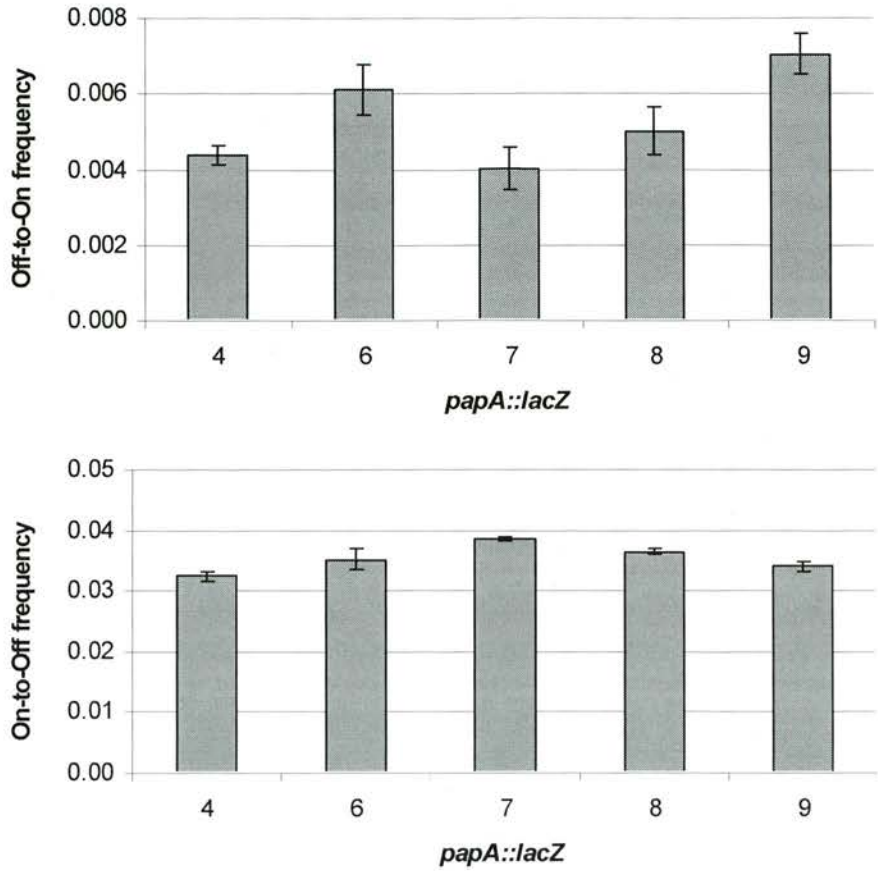


Fig. 5.7 Phase transition frequencies of *E. coli* CFT073 strains containing single-copy *papA::lacZ* fusions with natural PapB site 1 variants in rich complex medium (CFA). CFT073 strains ZAP838/1137/1000/1002/10003 contain single-copy *lacZ* fusions to *pap* regulatory regions with 4, 6, 7, 8, or 9 repeats in PapB site 1, respectively. Off-to-on (*top graph*) and on-to-off (*bottom graph*) transition frequencies ($cell^{-1}$ $generation^{-1}$) were determined under identical culture conditions (CFA, 37°C) by switching assays, as described in section 2.6, and are presented as mean frequencies (± s.e.). On average, 1400 colonies of each CFT073 strain were assayed for estimating the off-to-on transition frequency and 650 colonies for the on-to-off frequency. Equality of mean frequencies (both directions) was tested by ANOVA (normally distributed data) and significant evidence of differences among the tested strains was found for both off-to-on and on-to-off mean frequencies ($p<0.001$ and $p=0.004$, respectively).

In summary, it appears that under certain conditions *pap* operons with longer site 1 sequences switch on and potentially remain on more frequently than operons containing site 1 sequences with 4 repeats. Since *pap* phase variation frequencies determine the proportion of P-fimbriate bacteria in the population, it was hypothesised that the overall fimbrial output of a UPEC isolate can vary depending on the PapB site 1 sequence present in the *pap* operons it contains. The differential in sensitivity of activation among the tested regulatory regions was less apparent under growth conditions considered optimal for P expression (CFA agar, 37°C), suggesting that this mechanism might be of relevance in limiting conditions for *pap* expression. In the system described above, phase variation of each fusion was potentially regulated by PapI and PapB encoded by the regulatory region fused to the *lacZ* reporter, as well as the regulators present in the two CFT073 native *pap* operons (P1 and P2). Both PapI and PapB regulators have been shown to activate *pap* operons *in trans*, suggesting that the *lacZ* fusions in the CFT073 wild-type background were subject to higher levels of PapI and PapB regulators, especially under culture conditions that will stimulate P expression. Assuming that sequence variation in site 1 is affecting the sensitivity of transcriptional activation of *pap* operons, it was interesting to assess the variant fusions in a genetic background with limiting concentrations of regulators. In order to investigate the differential sensitivity of regulatory regions containing site 1 variants, the same *papA::lacZ* fusions were introduced in single-copy into the *lac* chromosomal locus of the CFT073 double *papIB* deletion strain (ZAP964), generating strains ZAP843 and ZAP595-598 (Table 2.1C). Deletion of the regulators and regulatory regions of both native *pap* operons in CFT073 permitted the regulation of *lacZ* fusions only by the *pap*-encoded regulators

present in the promoter fusion, while the fusions were still integrated in a relevant genetic background of a clinical UPEC isolate.

Switching assays performed in rich defined medium revealed a step-wise increase in the off-to-on transition frequency of strains containing fusions with increasing number of repeats in site 1 ($p<0.001$, Fig. 5.8-top). Isogenic strains containing site 1 variants with 7, 8 and 9 repeats displayed significantly higher off-to-on transition frequencies. Overall, off-to-on frequencies were lower for all tested fusions in the CFT073 P^- background than in the CFT073 wild-type background, with the highest-fold decrease observed for the fusion containing the P2 regulatory region with 6 repeats. In contrast, on-to-off transition frequencies were not affected by genetic background but displayed significant differences among the strains tested (Fig. 5.8-bottom). Strains containing fusions with 8 and 9 repeats switched off at lower frequencies than the strain carrying the fusion with 4 repeats ($p<0.001$). The results confirmed the frequency differences observed for the same *lacZ* fusions in the CFT073 wild-type background.

Taken together, the results suggest that sequence variation in PapB site 1 has an effect on *pap* phase transition frequencies, with longer site 1 variants switching on at higher frequencies than shorter site 1 variants and potentially remaining in the on transcriptional phase for longer.

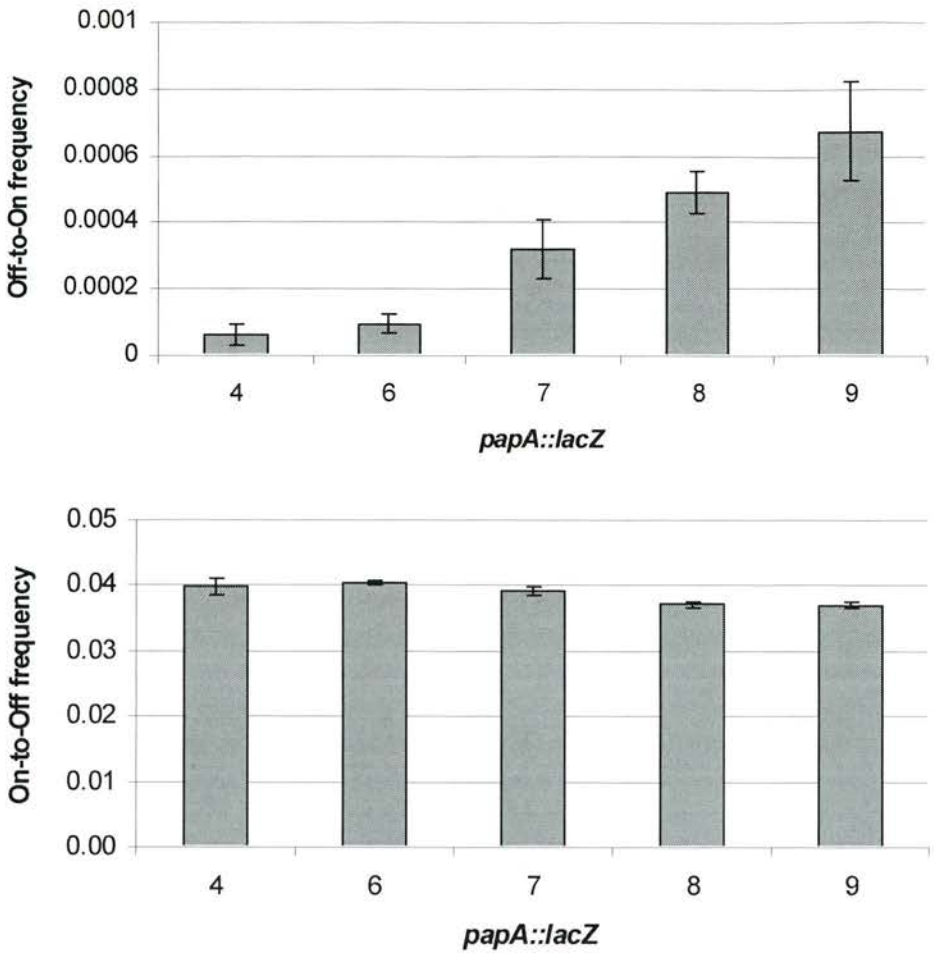


Fig. 5.8 Phase transition frequencies of *E. coli* CFT073 P⁻ strains containing single-copy *papA::lacZ* fusions with natural PapB site 1 variants in rich defined medium (M9 AA glycerol). CFT073 P⁻ strains ZAP843/595/596/597/598 contained single-copy *lacZ* fusions to *pap* regulatory regions with 4, 6, 7, 8, or 9 repeats in PapB site 1, respectively. Off-to-on (*top graph*) and on-to-off (*bottom graph*) transition frequencies ($cell^{-1}$ $generation^{-1}$) were determined under identical culture conditions (M9 RD glycerol, 37°C) by switching assays, as described in section 2.6, and are presented as mean frequencies (± s.e.). On average, 4200 colonies of each CFT073 P⁻ strain were assayed for estimating the off-to-on transition frequency and 1900 colonies for the on-to-off frequency. The conservative, non-parametric Kruskal-Wallis test was used to compare the median frequency displayed by each CFT073 P⁻ strain. There was supportive evidence for differences in the median frequencies of the tested strains (off-to-on, $p<0.001$ and on-to-off, $p<0.001$).

5.4 Effect of PapB site 1 sequence variation on *papI* expression

The effect of site 1 sequence variation on *pap* phase transition frequencies was proposed to be the result of a direct effect of this site on expression of the positive regulator PapI. PapB binding site 1 lies approximately 70 bp upstream from the transcription start site of *papI*, a reasonable distance for a transcriptional activator. Moreover, there is evidence to suggest that PapB binding to site 1 activates *papI* transcription (Goransson *et al.*, 1989b). In order to investigate whether sequence variation in site 1 affects expression of *papI*, p_I promoter regions containing the five observed site 1 variants were amplified by PCR using primers [papI 15_KpnI (F)] and [papI prom_CAP_BglII (R)] (Table 2.3) and fused transcriptionally to the *gfp+* reporter gene. The p_I promoter regions included the upstream CAP binding site, the PapB site 1 variant and the sequence up to the second proposed *papI* translation start site, but excluded PapB sites 2 and 3 to prevent any possible repressor effects from PapB binding. The p_I promoter fusions to *gfp+* were constructed using vector pKC48 (Table 2.2), which is a low-copy temperature-sensitive vector used for allelic exchange into the *E. coli* K-12 chromosomal *lac* locus. The fusions were subsequently introduced in single-copy into the chromosome of *E. coli* K-12 by allelic exchange generating strains ZAP611 (4), ZAP612 (6), ZAP613 (7), ZAP616 (8) and ZAP614 (9) (Table 2.1C; for construction of *E. coli* strains by allelic exchange see sections 2.3.14-15). Expression of GFP was examined under P-promoting growth conditions (CFA, 37°C) in the absence or presence of PapB. *E. coli* K-12 does not contain *papB* genes, so each fusion strain was transformed with the inducible vector pNJH20, which is a pBAD18-based construct carrying *papB* under the control of the pBAD promoter (Table 2.2; Holden *et al.*, 2001). Control

transformations with the pBAD18 vector were also performed for all strains. Fluorescence levels were relatively low for all *E. coli* K-12 strains and their plasmid-containing transformants, as demonstrated by fluorescence microscopy and flow cytometry. Examination of the bacterial populations after induction of PapB expression revealed however a small but reproducible increase in the mean population fluorescence intensity of K-12 strains containing promoter fusions with >4 repeats, but not in the mean fluorescence intensity of the K-12 population carrying the 4-repeat fusion (Fig. 5.9).

For more sensitive detection of GFP the K-12 strains were examined by immunoblotting using a commercial anti-GFP antibody (BD Biosciences), as described in section 2.8.4. Each *E. coli* K-12 strain and its pBAD18 (control) and pNJH20 (PapB) transformants were cultured in CFA medium at 37°C. In the absence of PapB no GFP protein could be detected. Expression of PapB *in trans* activated GFP expression only in strains with fusions containing the longer PapB site 1 variants (Fig. 5.10). These findings were consistent with the flow cytometry data and taken together, the results suggest that *pap* operons with longer PapB site 1 sequences switch on at higher frequencies partly due to higher levels of *papI* expression.

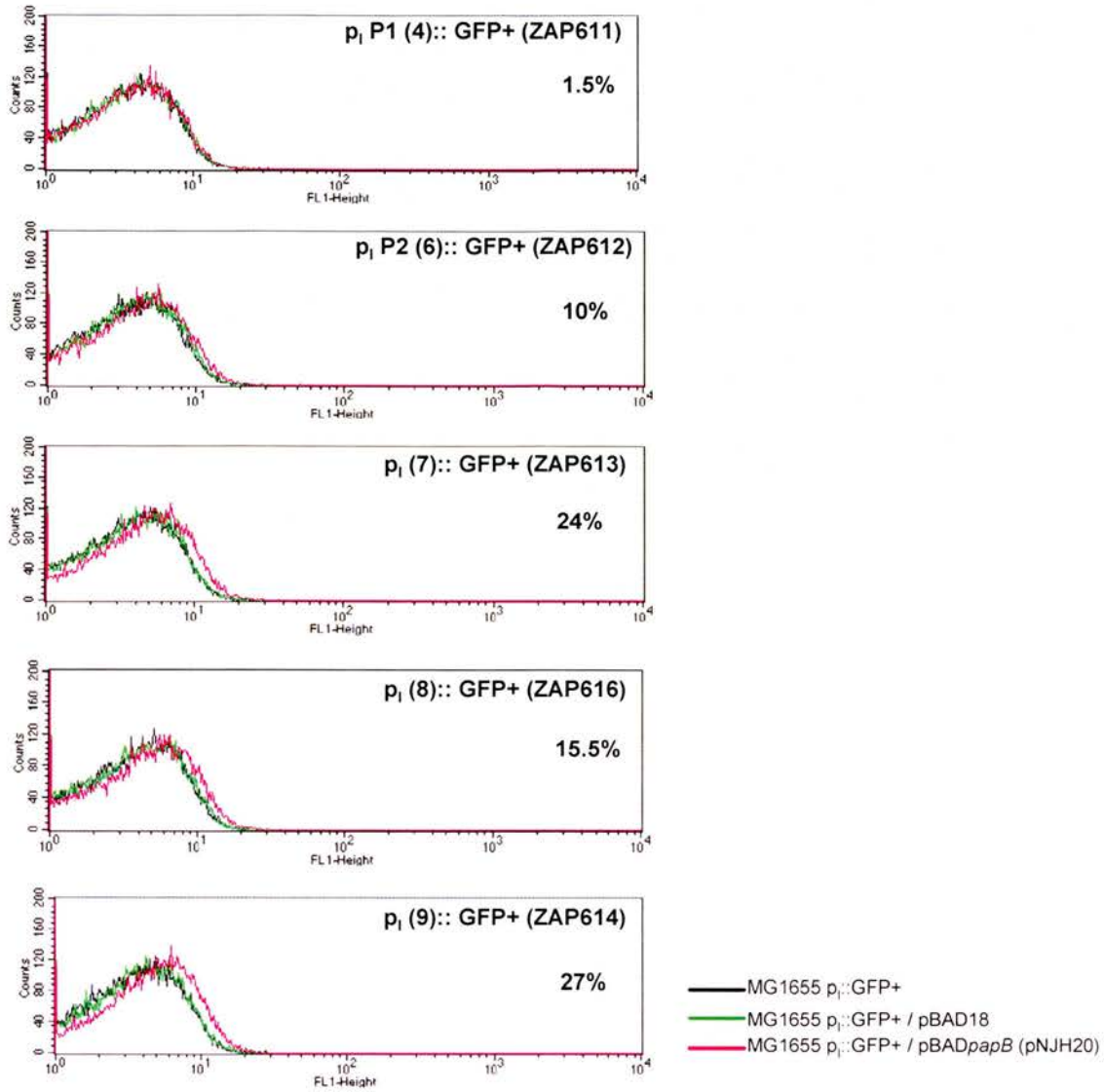


Fig. 5.9 Population fluorescence intensity histograms of *E. coli* K-12 strains containing single-copy *papI::gfp*+ fusions with natural PapB site 1 variants in the presence or absence of PapB. *E. coli* K-12 strains ZAP611-614 and ZAP616 carry single-copy *gfp*+ fusions to p_I promoter regions containing PapB site 1 sequence variants with 4, 6, 7, 9 and 8 repeats, respectively. Fluorescence intensity (FL1-Height) histograms obtained by flow cytometry were overlayed for populations of each K-12 strain (black line) and its plasmid-containing transformants (pBAD18: green line/ pNJH20: pink line). The PapB-induced increase in mean population fluorescence intensity is shown for each strain proportionally (%).

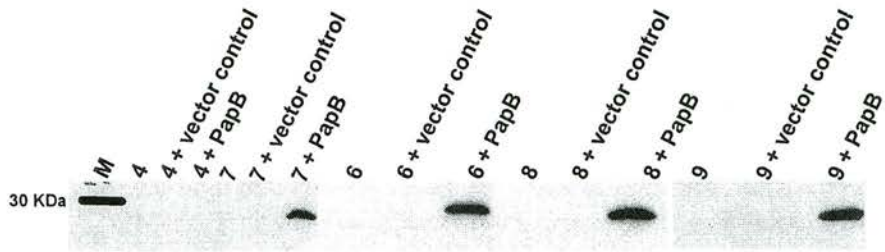


Fig. 5.10 Western hybridisation analysis of GFP expression in *E. coli* K-12 strains containing single-copy *papI::gfp*+ fusions with natural PapB site 1 variants in the presence or absence of PapB. *E. coli* K-12 strains ZAP611/612/613/614 and ZAP616 are designated 4, 6, 7, 9, and 8, respectively, according to the number of 9 bp repeats present in the p_I promoter region fused to *gfp*+. All strains were cultured in CFA medium overnight with 0.2% (w/v) arabinose added to pBAD18 (vector control) and pNJH20 (PapB) transformants. Protein extracts were prepared for Western hybridisation analysis, as described in section 2.8.1 and immunoblotting was carried out using a mouse anti-GFP monoclonal antibody (JL-8; BD Biosciences Clontech) and an HRP-conjugate rabbit anti-mouse IgG secondary antibody (DAKO), as described in section 2.8.4. GFP bands (~27 KDa) were detected for strains 6, 7, 8 and 9 in the presence of PapB.

In order to determine whether the complete lack of GFP expression in *E. coli* strain ZAP611 is a direct consequence of the 4-repeat PapB site 1 variant and not other sequence changes present in the promoter region, an *E. coli* K-12 strain containing a single copy *gfp*+ fusion to the p_I promoter from the *pap* operon of *E. coli* isolate J96 (ZAP610) was constructed in a similar manner. The regulatory region of the pap_{J96} operon also has 4 repeats in PapB site 1, but contains different sequence variation to that present in the P1 regulatory region (Fig. 5.11).

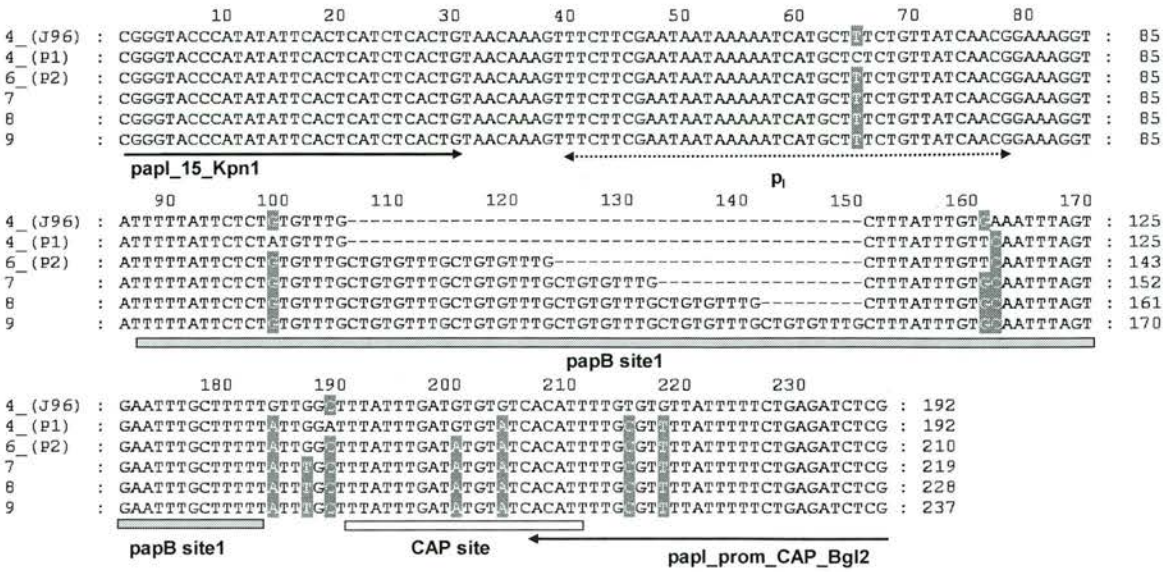


Fig. 5.11 Alignment of p_I promoter regions containing natural PapB site 1 variants. The DNA sequence of the six p_I promoter regions used for the construction of *E. coli* K-12 strains ZAP610-614/616 is shown, with important sequence features marked: PCR primers [papI 15_KpnI (F)] and [papI prom_CAP_BglII (R)] (black arrows), the region containing p_I promoter elements (double dotted arrow), PapB binding site 1 (grey bar), and the CAP binding site (white bar). The DNA fragments are designated 4, 6, 7, 8 or 9 according to the number of 9 bp repeats present in their PapB site 1 sequence. The two fragments designated 4, represent DNA sequences found in the *pap* operon of *E. coli* isolate J96 and the P1 operon of *E. coli* isolate CFT073, and while both contain 4 repeats in PapB site 1, base changes can be seen both in site 1 and in the rest of the fragment. Fragment 6 represents the DNA sequence from the P2 operon of *E. coli* isolate CFT073, and fragments 7, 8 and 9 are isogenic PapB site 1 variants. Single base differences are shaded in grey.

It was hypothesised that if the level of *papI* activation is solely determined by the number of repeats present in PapB site 1, *E. coli* strain ZAP610 would be predicted not to express GFP, similarly to strain ZAP611 that also contains 4 repeats in PapB site 1. Analysis of GFP expression in strain ZAP610 in a similar manner performed

for strains ZAP611-614 and ZAP616, revealed however a different outcome. GFP expression was stimulated by the J96 p_I promoter in the presence of PapB, to a similar extent demonstrated by promoters with longer PapB site 1 variants. This was shown both as an increase in the mean fluorescence intensity of a ZAP610pNJH20 population and by immunodetection of GFP protein production (Fig. 5.12). It therefore appears that variation in repeat number in PapB site 1 is not solely responsible for determining the levels of *papI* expression and that sequence changes in site 1 and/or other regulatory sites, may also play a role.

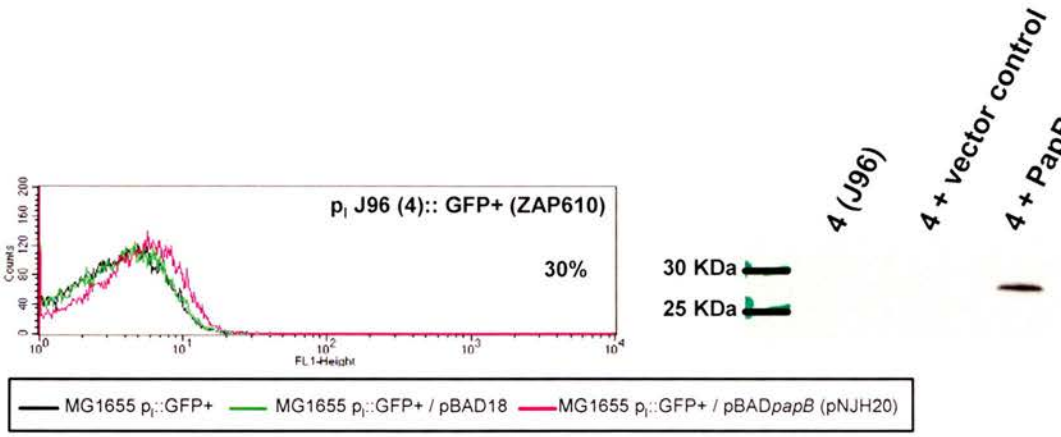


Fig. 5.12 Population fluorescence intensity and GFP expression in *E. coli* K-12 strain ZAP610 in the presence or absence of PapB. *E. coli* strain ZAP610 contains a single-copy *gfp*+ reporter fusion to the *papI* promoter region from *E. coli* isolate J96, which contains 4 repeats in PapB site 1. Fluorescence intensity (FL1-Height) histograms obtained by flow cytometry are shown (*left*), and were overlayed for population comparisons between strain ZAP610 (black line), pBAD18-containing transformants (green line) and pNJH20-containing transformants (pink line). The PapB-induced increase in mean population fluorescence intensity was 30%. The GFP immunoblot (*right*) shows GFP expression in *E. coli* ZAP610 only in the presence of PapB.

In conclusion, the results suggest that a longer site 1 sequence could increase the probability of initial PapB binding, potentially increase the sensitivity of *papI* expression, and subsequently increase the frequency of off-to-on phase transition. A longer site 1 sequence can also occupy a higher total number of PapB molecules before reaching saturation, which under limiting PapB concentrations could restrict PapB from binding to autoregulatory sites 2 and 3 that will ultimately repress transcription from the p_{BA} promoter.

5.5 Discussion

The aim of this work was to investigate the effect of a highly variable regulatory site in the *pap* epigenetic switch on *pap* regulation and expression. The sequence of the high-affinity PapB binding site 1 was found to differ greatly among the 82 sequenced *pap* regulatory regions, mainly due to multiple insertion/deletion events of a 9 bp sequence. This variation explains the size differences observed in TOPO®-clones containing *pap* regulatory regions from UPEC isolates, described in chapter 4 (see Fig. 4.7). In total, five natural site 1 variants were observed with 4, 6, 7, 8 or 9 repeats of 9 bp sequences, increasing the extent of sequence variation previously reported for this site (Blyn *et al.*, 1989; Xia *et al.*, 1998). Interestingly, no site variants with 5 repeats were observed in this study, suggesting that operons similar to the *pap*-21 operon from *E. coli* isolate C1212 (Blyn *et al.*, 1989) are rare and potentially unstable. Earlier work has demonstrated that PapB recognises a DNA structure containing $(T/A)_3$ repeats with 9 bp periodicity rather than a specific sequence in *pap* DNA, and mutagenesis experiments suggested that these repeats as well as their periodicity are important for PapB binding and *pap* expression (Xia *et*

al., 1998). The sequence variation in site 1 identified in this study did not alter these important features, suggesting that it can function as a modulator of *pap* expression.

Binding of PapB to natural site 1 variants was investigated *in vitro* and longer sites were shown to require higher protein concentrations for 50% site occupancy. These findings were in agreement with previous work reporting that site 1 variants with 6 repeats require four more PapB molecules to cover the binding site than sites with 4 repeats (Xia *et al.*, 1998). It was therefore hypothesised that longer site 1 variants have a higher probability of initially interacting with PapB molecules, simply because they contain a higher number of repeats that can be recognised by PapB. The hypothesis was partly supported by the observation that at low protein concentrations a slightly higher proportion of the longer site variants was protein-bound.

The differential PapB affinity demonstrated for site 1 variants was shown to have an effect on *pap* phase transition frequencies. Under certain conditions, *pap* switches with longer site 1 variants displayed higher on-to-off phase transition frequencies than switches containing the short site 1 variant with 4 repeats. This differential sensitivity in activation of *pap* expression was however less apparent in rich complex medium for reasons that remain unclear. It is highly possible that other sequence changes present in the regulatory region of the P1 operon, which was used to represent a natural minimum site 1 variant, is responsible for the media-induced increase in off-to-on switching frequency. The construction of isogenic strains with 4 and 9 repeats in site 1 could aid in resolving this. It is important to note, however, that investigating the regulation of such isogenic switches might not be biologically

relevant, as no naturally-occurring isogenic switches containing site 1 variants with 4 and >4 repeats were observed among the 82 switch sequences examined.

On-to-off phase transition frequencies were also investigated and were overall higher than off-to-on frequencies (increase range of 5 to 400-fold, depending on the fusion strain and growth conditions). This was consistent with previous studies reporting 100-fold frequency differences (Blyn *et al.*, 1989) and supports the general notion that fimbrial expression is promptly repressed when it does not provide an advantage to bacterial fitness. Variation in the number of repeats in site 1 did not appear to affect off-to-on transition frequencies greatly. A small decrease in switching off frequencies was observed for longer site 1 variants in both CFT073 wild-type and *papIB* mutant backgrounds, but this subtle differential was less apparent in assays performed in rich complex medium (CFA). The reason for this effect is unclear, but the observation that *pap* phase transition frequencies in both directions were generally higher on CFA than other growth media could indicate that at high switching frequencies, cues over and above local *pap* regulators are responsible for *pap* repression. One potential candidate could be the cytoplasmic response regulator CpxR of the two-component CpxAR envelope stress system, which when phosphorylated was shown to specifically bind to the *pap* regulatory region and inhibit *pap* transcription and P fimbriation (Hernday *et al.*, 2004).

The effect of site 1 sequence variation on *papI* expression was also investigated, as previous studies have demonstrated that binding of PapB to site 1 activates transcription from both divergent *pap* promoters (Forsman *et al.*, 1989; Goransson *et*

al., 1989b). Moreover, increased PapI expression from regulatory regions with longer site 1 variants could explain the higher off-to-on frequencies observed in this study, as PapI is essential for the translocation of Lrp to distal sites in the p_{BA} promoter and activation of *pap* operon transcription. Promoter regions containing site 1 variants with >4 repeats stimulated PapI expression at low levels, but no expression was detected from a promoter containing 4 repeats in site 1 (P1 from UPEC CFT073). It is possible that a very low level of GFP expression from the P1 promoter fusion failed to be detected by the employed methods and this could be investigated further with more sensitive transcript-detection methods, such as qRT-PCR. Investigation of a different natural promoter with 4 repeats (*pap* from UPEC J96) demonstrated that the number of repeats present in PapB site 1 does not affect the level of PapI expression and suggested that other sequence changes may be limiting *papI* expression from the P1 operon. Sequence comparison of p_I promoter regions from $P1_{CFT073}$ and Pap_{J96} identified single-base changes in CAP binding site and its upstream sequence as potential candidates. The finding however that *papI* encoded by the P1 operon from *E. coli* CFT073 failed to be activated in the presence of PapB, raised questions as to how this operon transitions from the off-to-on phase to allow expression of P1 fimbriae. Moreover, the lack of *papI* activation from the P1 p_I promoter does not explain how the divergent p_{BA} promoter in the same operon can switch on, even at relatively low frequencies (Fig. 5.6-8). Comparison of the results however requires caution, as activation of the two divergent promoters was assayed in different genetic backgrounds (*E. coli* K-12 for p_I *vs.* CFT073 for p_{BA}), which were shown in chapter 3 of this thesis, to provide a different regulatory context for *papI* gene expression. This is further supported by the fact that the very low off-to-on

transition frequency presented in section 3.3 for the P1 p_{BA} promoter in the *E. coli* K-12 background (ZAP992) is consistent with the lack of *papI* expression from the P1 p_I promoter in the same background that was described here. Alternatively, it is possible that differences in *papI* expression levels exist between different *pap* operons in the same genetic background and that they are linked to *papI* sequence variation. For example, $P1_{CFT073}$ and pap_{J6} operons that differ greatly in *papI* activation levels contain sequence changes in *papI* resulting in different PapI proteins. These findings are analysed and discussed further in the following chapter.

The observation that *pap* operons display differential sensitivity of activation is proposed to be physiologically significant for UPEC isolates that contain multiple copies of *pap* and *pap*-related operons. Sequence variation in PapB site 1 is hypothesised to provide a mechanism for differential responsiveness of homologous *pap* operons to activation and possibly coordinated timing of P fimbrial expression in single cells. This hypothesis is supported by the finding that almost all UPEC isolates with two *pap* operons examined in this study were found to contain operons with different site 1 variants, with the combination of minimum (4) and maximum (9) length sites being the most common. Based on this finding, PapB site 1 sequence variation is proposed to have the potential of a diagnostic tool. PCR amplification over the variable site 1 sequence would be a fast and simple method to gain good indication of the presence of multiple *pap* operons in a UPEC isolate and in turn of its pathogenic potential.

Chapter 6

Sequence variation in *papI* genes from UPEC isolates and effects on regulation of *pap* expression

This chapter describes the extent of natural sequence divergence observed in *papI* genes from UPEC clinical isolates associated with different UTI clinical syndromes. Sequence divergence was investigated with respect to isolate clinical origin and *pap* operon copy number, in order to link allelic variation to isolate pathogenic potential. Sequence variation in *papI* was also investigated with respect to other variable regions identified in the *pap* epigenetic switch and described in chapters 4 and 5. Lastly, the function of different PapI protein variants as transcriptional activators was examined on different *pap* operons and homologous fimbrial operons from *E. coli*.

PapI is the positive transcriptional regulator of the *pap* operon. It is a small protein (8.8 KDa) encoded by the *papI* gene, which is located ~ 400bp upstream of *papB* and is divergently transcribed from its own promoter (p_I). PapI is essential for *pap* phase variation. Early studies investigated the possibility of PapI exerting its regulatory function by binding *pap* DNA sites specifically, but obtained no evidence to suggest that the protein was interacting with *pap* DNA independent of Lrp (Forsman *et al.*, 1989; Nou *et al.*, 1993; Kaltenbach *et al.*, 1995). It was later shown that PapI was required for the transition of the *pap* switch from off to on by a mechanism involving the *E. coli* global regulators Lrp and Dam and specifically, that PapI favoured Lrp binding to distal promoter sites when proximal sites were methylated by Dam, thus promoting formation of the on transcriptional phase (Kaltenbach *et al.*, 1995; Hernday *et al.*, 2003). Recently, the structure of PapI has been solved and the proposed fold resembled that of the winged helix-turn-helix (wHtH) family of DNA-binding proteins (Kawamura *et al.*, 2007). In the same study the ability to bind DNA

non-specifically was also demonstrated in the absence of Lrp, providing new insights into PapI's mode of action as a regulator of *pap* transcription.

6.1 Sequence variation in *papI* genes from UPEC isolates

The sequence of *papI* genes present in 82 *pap* operons from 54 UPEC isolates was obtained as described in section 4.3. Comparison of *papI* sequences by multiple sequence alignment (msa) revealed novel polymorphism within the population of UPEC clinical isolates (Fig. 6.1). Twenty one single nucleotide polymorphisms (SNPs) were identified in the 234 bp coding region resulting in 16 *papI* alleles. Unlike *papB*, no *papI* gene was found to contain frameshift mutations that led to protein truncation and higher nucleotide diversity was displayed by *papI* than full-length *papB* sequences ($\Pi_{papI} = 0.034 \pm 0.008\%$ *vs.* $\Pi_{papB} = 0.007 \pm 0.002\%$). Twelve of the observed SNPs were synonymous (S) substitutions, whereas the remaining nine were nonsynonymous (N) and resulted in amino acid replacements. The number of synonymous substitutions per synonymous site (dS) and the number of nonsynonymous substitutions per nonsynonymous site (dN) were determined and compared by a codon-based Z-test using the Nei-Gojobori method (Nei and Gojobori, 1986). The test revealed significant evidence for positive selection acting on *papI* genes ($dN>dS$, $p<0.05$).

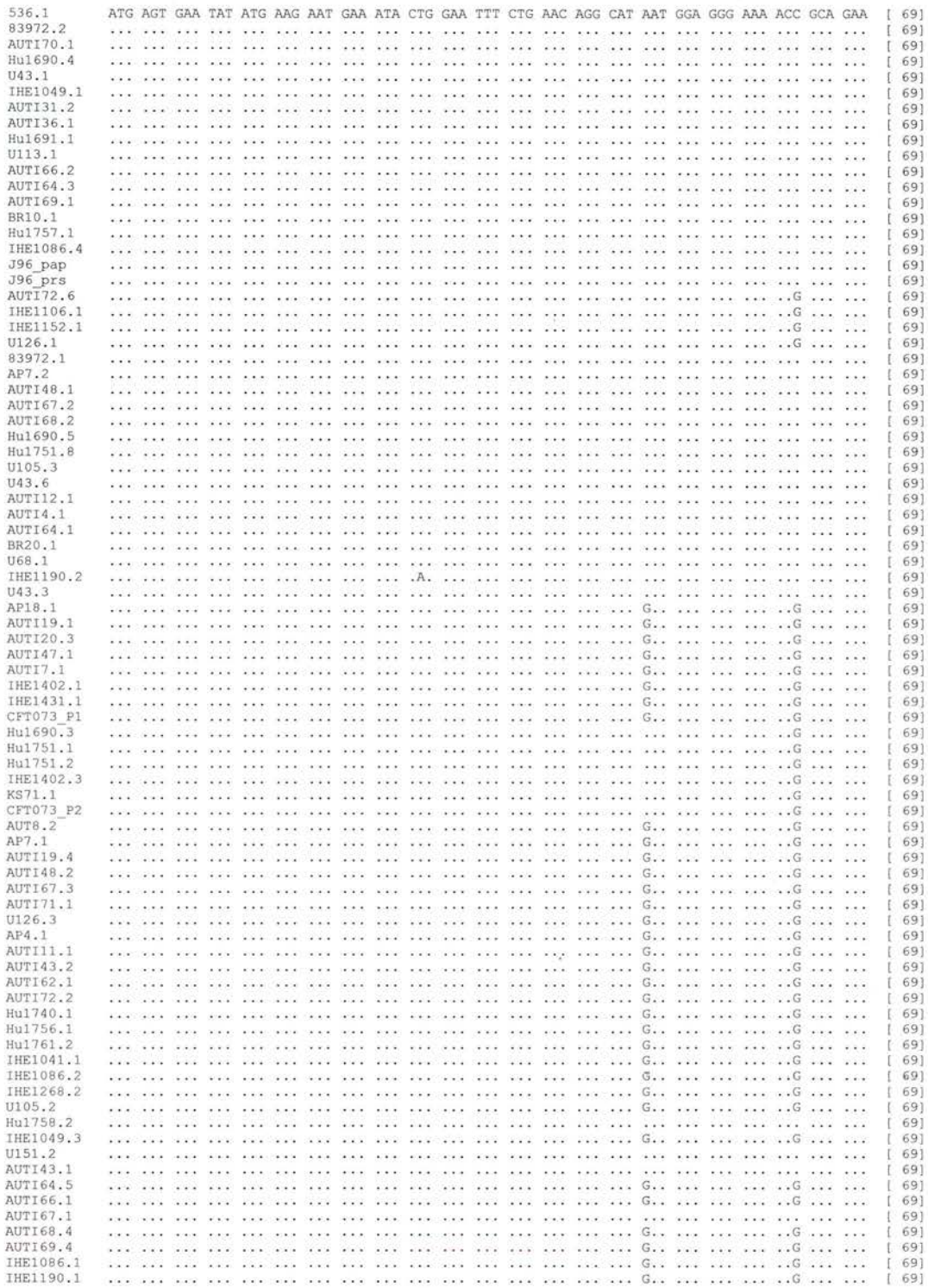


Fig. 6.1 Multiple sequence alignment (msa) of 82 *papI* genes from 54 UPEC clinical isolates. Dots represent nucleotides identical to the first DNA sequence listed, while polymorphic bases are shown. DNA sequences were aligned using ClustalX 1.83 (Thompson *et al.*, 1997).

```
536.1       ATT GCG GAG GCG CTG GCG GTA ACG GAT TAT CAG GCA CGC TAT TAC CTG TTA TTA CTG GAA AAA GAA GGT  [138]
83972.2     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI70.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
Hu1690.4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
U43.1       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
IHE1049.1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI31.2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI36.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
Hu1691.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
U113.1      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI66.2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI64.3    ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI69.1    ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
BR10.1      ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
Hu1757.1    ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
IHE1086.4   ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
J96_pap     ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
J96_prs     ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI72.6    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... ... ...  [138]
IHE1106.1   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... ... ...  [138]
IHE1152.1   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... ... ...  [138]
U126.1      ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... ... ...  [138]
83972.1     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AP7.2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI48.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI67.2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI68.2    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
Hu1690.5    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
Hu1751.8    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
U105.3      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
U43.6       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI12.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI4.1     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI64.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
BR20.1      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
U68.1       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
IHE1190.2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
U43.3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AP18.1      ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI19.1    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI20.3    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI47.1    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI7.1     ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1402.1   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1431.1   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
CFT073_P1   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1690.3    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1751.1    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1751.2    ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1402.3   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
KS71.1      ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
CFT073_P2   ... ..A ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUT8.2      ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AP7.1       ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI19.4    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI48.2    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI67.3    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI71.1    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
U126.3      ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AP4.1       ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI11.1    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI43.2    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI62.1    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI72.2    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1740.1    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1756.1    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1761.2    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1041.1   ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1086.2   ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1268.2   ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
U105.2      ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
Hu1758.2    ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
IHE1049.3   ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
U151.2      ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI43.1    ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI64.5    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI66.1    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI67.1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [138]
AUTI68.4    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
AUTI69.4    ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1086.1   ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
IHE1190.1   ... ... ... ... ... ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... .C. ...  [138]
```

Fig. 6.1 – *continued* (*papI* bases: 70-138)

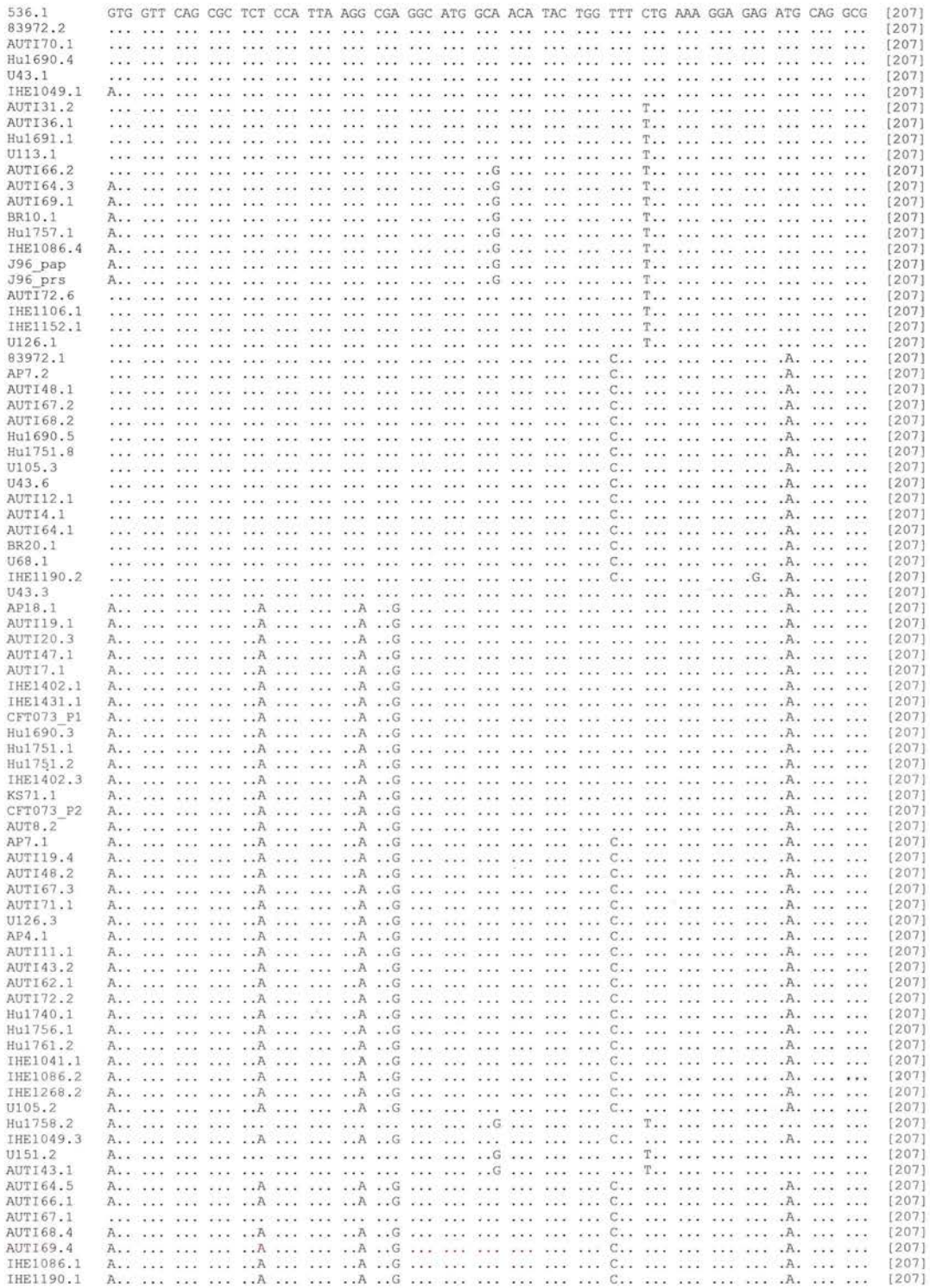


Fig. 6.1 – *continued* (*papI* bases: 139-207)

```
536.1       GGA CAG AGC TGT TCT TCC ACA ACT TAA  [234]
83972.2     ... ..A .A. ... ... ... ... ... ...  [234]
AUTI70.1    ... ..A .A. ... ... ... ... ... ...  [234]
Hu1690.4    ... ..A .A. ... ... ... ... ... ...  [234]
U43.1       ... ..A .A. ... ... ... ... ... ...  [234]
IHE1049.1   ... ..A .A. .C. ... ... ... ... ...  [234]
AUTI31.2    ... ..A .A. ... ... ... ... ... ...  [234]
AUTI36.1    ... ..A .A. ... ... ... ... ... ...  [234]
Hu1691.1    ... ..A .A. ... ... ... ... ... ...  [234]
U113.1      ... ..A .A. ... ... ... ... ... ...  [234]
AUTI66.2    ... ..A .A. ... ... ... ... ... ...  [234]
AUTI64.3    ... ..A .A. ... ... ... ... ... ...  [234]
AUTI69.1    ... ..A .A. ... ... ... ... ... ...  [234]
BR10.1      ... ..A .A. ... ... ... ... ... ...  [234]
Hu1757.1    ... ..A .A. ... ... ... ... ... ...  [234]
IHE1086.4   ... ..A .A. ... ... ... ... ... ...  [234]
J96_pap     ... ..A .A. ... ... ... ... ... ...  [234]
J96_prs     ... ..A .A. ... ... ... ... ... ...  [234]
AUTI72.6    ... ..A .A. ... ... ... ... ... ...  [234]
IHE1106.1   ... ..A .A. ... ... ... ... ... ...  [234]
IHE1152.1   ... ..A .A. ... ... ... ... ... ...  [234]
U126.1      ... ..A .A. ... ... ... ... ... ...  [234]
83972.1     ..G ... .A. ... ... ... ... ... ...  [234]
AP7.2       ..G ... .A. ... ... ... ... ... ...  [234]
AUTI48.1    ..G ... .A. ... ... ... ... ... ...  [234]
AUTI67.2    ..G ... .A. ... ... ... ... ... ...  [234]
AUTI68.2    ..G ... .A. ... ... ... ... ... ...  [234]
Hu1690.5    ..G ... .A. ... ... ... ... ... ...  [234]
Hu1751.8    ..G ... .A. ... ... ... ... ... ...  [234]
U105.3      ..G ... .A. ... ... ... ... ... ...  [234]
U43.6       ..G ... .A. ... ... ... ... ... ...  [234]
AUTI12.1    ..G ... ... ... ... ... ... ... ...  [234]
AUTI4.1     ..G ... ... ... ... ... ... ... ...  [234]
AUTI64.1    ..G ... ... ... ... ... ... ... ...  [234]
BR20.1      ..G ... ... ... ... ... ... ... ...  [234]
U68.1       ..G ... ... ... ... ... ... ... ...  [234]
IHE1190.2   ..G ... ... ... ... ... ... ... ...  [234]
U43.3       ..G ... ... ... ... ... ... ... ...  [234]
AP18.1      ..G ... ... ... ... ... ... ... ...  [234]
AUTI19.1    ..G ... ... ... ... ... ... ... ...  [234]
AUTI20.3    ..G ... ... ... ... ... ... ... ...  [234]
AUTI47.1    ..G ... ... ... ... ... ... ... ...  [234]
AUTI7.1     ..G ... ... ... ... ... ... ... ...  [234]
IHE1402.1   ..G ... ... ... ... ... ... ... ...  [234]
IHE1431.1   ..G ... ... ... ... ... ... ... ...  [234]
CFT073_P1   ..G ... ... ... ... ... ... ... ...  [234]
Hu1690.3    ..G ... ... ... ... ... ... ... ...  [234]
Hu1751.1    ..G ... ... ... ... ... ... ... ...  [234]
Hu1751.2    ..G ... ... ... ... ... ... ... ...  [234]
IHE1402.3   ..G ... ... ... ... ... ... ... ...  [234]
KS71.1      ..G ... ... ... ... ... ... ... ...  [234]
CFT073_P2   ..G ... ... ... ... ... ... ... ...  [234]
AUT8.2      ..G ... ... ... ... ... ... ... ...  [234]
AP7.1       ..G ... ... ... ... ... ... ... ...  [234]
AUTI19.4    ..G ... ... ... ... ... ... ... ...  [234]
AUTI48.2    ..G ... ... ... ... ... ... ... ...  [234]
AUTI67.3    ..G ... ... ... ... ... ... ... ...  [234]
AUTI71.1    ..G ... ... ... ... ... ... ... ...  [234]
U126.3      ..G ... ... ... ... ... ... ... ...  [234]
AP4.1       ..G ... ..T ... ... ... ... ... ...  [234]
AUTI11.1    ..G ... ..T ... ... ... ... ... ...  [234]
AUTI43.2    ..G ... ..T ... ... ... ... ... ...  [234]
AUTI62.1    ..G ... ..T ... ... ... ... ... ...  [234]
AUTI72.2    ..G ... ..T ... ... ... ... ... ...  [234]
Hu1740.1    ..G ... ..T ... ... ... ... ... ...  [234]
Hu1756.1    ..G ... ..T ... ... ... ... ... ...  [234]
Hu1761.2    ..G ... ..T ... ... ... ... ... ...  [234]
IHE1041.1   ..G ... ..T ... ... ... ... ... ...  [234]
IHE1086.2   ..G ... ..T ... ... ... ... ... ...  [234]
IHE1268.2   ..G ... ..T ... ... ... ... ... ...  [234]
U105.2      ..G ... ..T ... ... ... ... ... ...  [234]
Hu1758.2    ... ..A .A. ... ... ... ... ... ...  [234]
IHE1049.3   ..G ... ..T ... ... ... ... ... ...  [234]
U151.2      ... ..A .A. ... ... ... ... ... ...  [234]
AUTI43.1    ... ..A .A. ... ... ... ... ... ...  [234]
AUTI64.5    ..G ... ... ... ... ... ... ... ...  [234]
AUTI66.1    ..G ... ..T ... ... ... ... ... ...  [234]
AUTI67.1    ..G ... .A. ... ... ... ... ... ...  [234]
AUTI68.4    ..G ... ... ... ... ... ... ... ...  [234]
AUTI69.4    ..G ... ... ... ... ... ... ... ...  [234]
IHE1086.1   ..G ... ..T ... ... ... ... ... ...  [234]
IHE1190.1   ..G ... ..T ... ... ... ... ... ...  [234]
```

Fig. 6.1 – *continued* (*papI* bases: 139-234)

Translation of the 16 *papI* alleles resulted in 11 protein variants, denoted I.1 - I.11 (Fig. 6.2). Six of the identified variants were novel and the remaining five matched PapI sequences available in public sequence databases belonging to either sequenced *E. coli* isolates (CFT073, UTI89, 536) or prototype strains (J96, 83972).

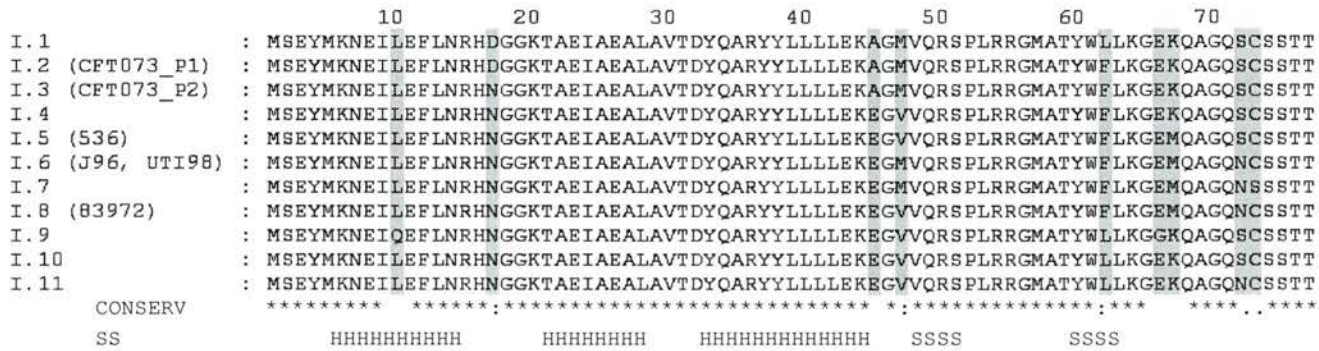


Fig. 6.2 Multiple sequence alignment (msa) of 11 PapI protein variants identified in 54 UPEC clinical isolates. The identified PapI variants were designated I.1-I.11. The sequence of five variants is identical to PapI protein sequences from sequenced *E. coli* isolates (CFT073, 536, UTI89) and prototype strains (J96, 83972). Variable residue positions are shaded grey. The CONSERV line indicates the degree of conservation in each column, with '*'=identically conserved, ':'=contains only very conservative substitutions and '.'=contains mostly conservative substitutions. The SS line shows the secondary structure elements of PapI (H, α-helix; S, β-strand) as proposed by Kawamura and colleagues (Kawamura *et al.*, 2007).

Colour-mapping of the 9 observed amino acid replacements onto the protein structure of PapI (PBD ID 2HTJ; Kawamura *et al.*, 2007) was performed using the MBT Protein Workshop (Moreland *et al.*, 2005) and is illustrated in Figure 6.3.

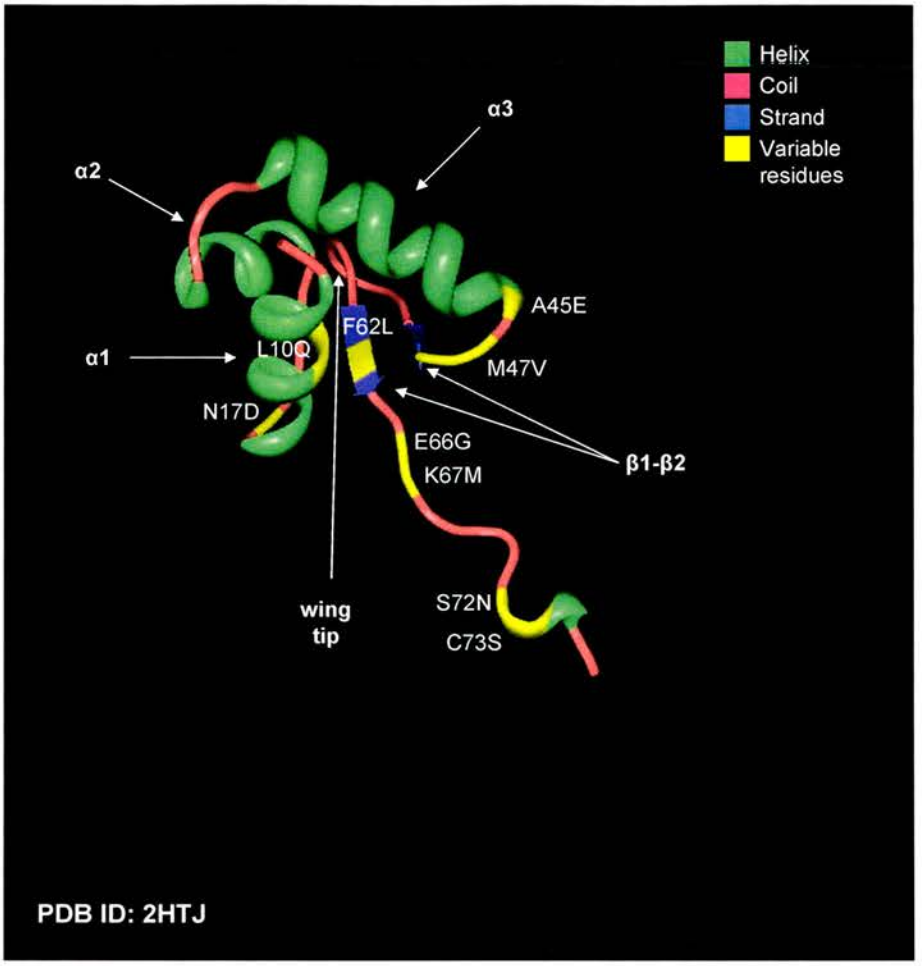


Fig. 6.3 Diagram showing the location of identified amino acid replacements in PapI structure. The winged helix-turn-helix (wHtH) structure of PapI consists of a HtH (α1-α2-turn-α3), followed by the wing (β1-loop-β2) (Kawamura *et al.*, 2007). Helices are shaded in green and are marked by arrows, coils are shaded in red and strands are shaded in blue and are also marked by arrows. The loop between the two β strands is marked by an arrow and labelled 'wing tip'. Variable residues are shaded in yellow with the position and residue replacement labelled. The ribbon diagram of PapI structure (PDB ID: 2HTJ; Kawamura *et al.*, 2007) was obtained from the Protein Data Bank (RCSB PDB; Berman *et al.*, 2003) and editing was performed using the MBT Protein Workshop (Moreland *et al.*, 2005).

Most changes were located within coil structures and the C-terminal tail. One change was found in the first helix (α1) and was only observed in variant I.9, which in turn was found in a single *pap* operon (see Table 6.1 below). A phenylalanine replacement by leucine was more commonly observed and localised in the highly hydrophobic second strand (β2) of the antiparallel β-sheet, but it is not predicted to change the overall hydrophobicity of the strand. Interestingly, the sequence of the third helix (α3), known as the recognition helix and proposed to be the major surface contacting DNA in wHtH proteins (Kawamura *et al.*, 2007), was identically conserved. A similar role was proposed for the tip of the wing, which is the loop between the β-strands, which was also conserved among the identified variants.

To examine the evolutionary relationships of the *papI* gene sequences phylogenetic analysis was performed using the neighbor-joining method, which uses distances between pairs of taxa (eg. number of nucleotide differences between sequence pairs in msa) to construct an evolutionary tree (Fig. 6.4; Saitou and Nei, 1987). Distance methods, such as the neighbor-joining, are the most suitable for phylogenetic analysis of the large number of sequences analysed in this study, whereas parsimony or maximum-likelihood methods are limited to smaller datasets. To evaluate the reliability of branch topology in the constructed phylogenetic tree the bootstrap method was used (Felsenstein, 1988). This method operates by randomly resampling vertical columns from the aligned sequences to produce several sequence alignments of the same size. In each new dataset some of the original data could be represented more than once and others not at all. A minimum-length tree is produced for each alignment of resampled sequences and the predicted topology of the original tree is

significant when the same branches occur consistently in most of the trees from the resampled datasets.

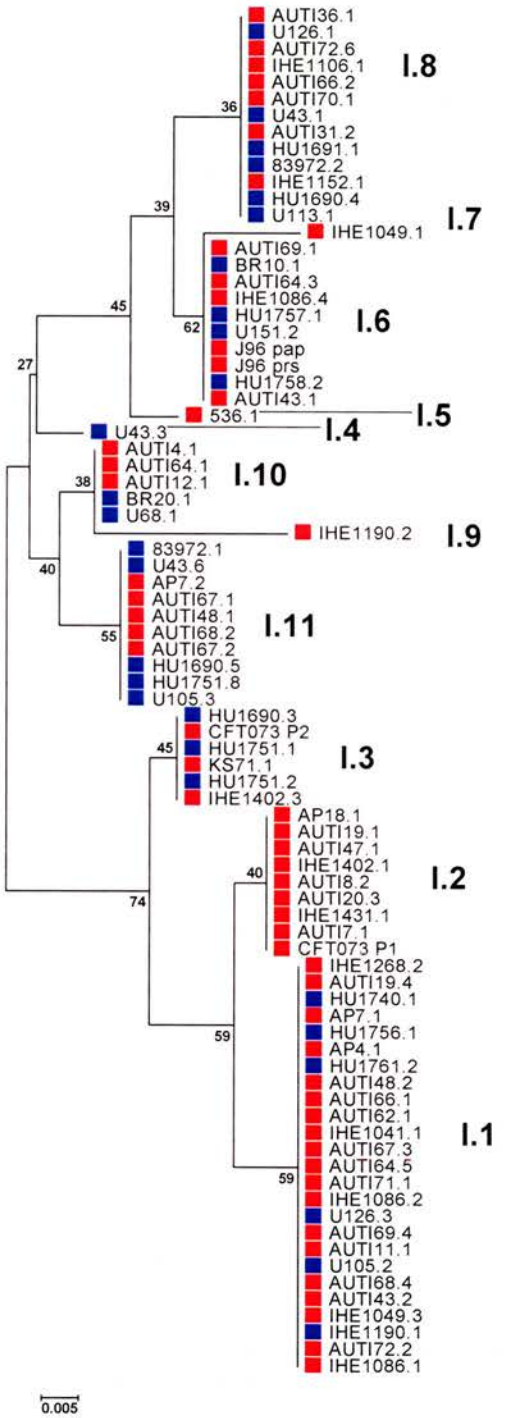


Fig. 6.4 Phylogenetic tree of PapI amino acid sequences from UPEC isolates. Translated *papI* gene sequences from 54 UPEC isolates were compared by msa (BLOSUM62 scoring matrix) and evolutionary relationships were predicted using the neighbor-joining distance method. The PapI protein variant designations I.1 - I.11 corresponding to each group of sequences is shown with the clinical origin of every sequence colour-coded (blue=asymptomatic, red=symptomatic). The predicted tree topology was tested using the bootstrap method (1000 replicates) and branch frequencies are shown in bold. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.1 (Kumar *et al.*, 2004).

Table 6.1 lists the PapI protein variant encoded by each *pap* operon from the UPEC isolates examined. The information on isolate clinical origin, number of *pap* operons and PapB site 1 repeat number is also included in the table to enable examination of *papI* polymorphism in the context of the clinical isolate.

Table 6.1 PapI protein variants in *pap* operons from UPEC clinical isolates. The *papI* alleles present in the *pap* operons from each UPEC isolate were translated and assigned a protein variant number (I.1-I.11). Each of the 54 UPEC isolates is listed along with the number of *pap* operons it contains and the PapI protein variant found in each operon. The PapB site 1 repeat number found in the sequence of each corresponding operon is also listed in the final column.

(*) Isolates containing a *pap* operon that was detected by Southern hybridisation but for which the sequence of the regulatory region was not obtained (N/O).

(a) PapI protein variants that were found only in one *pap* operon.

(b) PapI protein variants encoded by a *pap* operon with a mutant *papB* gene (see section 4.4.5).

Isolate ID	No. of *pap* operons	PapI variant	PapB site 1 repeat number
HU1690	3	I.11[b], I.3, I.8	4, 7,4
HU1691	1	I.8	4
HU1740	1	I.1[b]	8
HU1751	3	I.3, I.3, I.11[b]	7, 7, 4
HU1756	1	I.1	9
HU1757	1	I.6	4
HU1758	1	I.6	4
HU1761	1	I.1	8
U43	3	I.8, I.4[a], I.11[b]	4, 4, 4
U68	1	I.10[b]	4
U105	2	I.11[b], I.1	4, 8
U113	1	I.8	4
U126	2	I.8, I.1	4, 6
U151	1	I.6	4
BR10	1	I.6	4
BR20	1	I.10[b]	4
83972	2	I.11[b], I.8	4, 4
AUTI4*	2	I.10[b], N/O	4, N/O
AUTI7	1	I.2	4
AUTI12	1	I.10[b]	4
AUTI19	2	I.2, I.1	4, 6
AUTI31	1	I.8	4
AUTI36	1	I.8	4
AUTI43	2	I.6, I.1	4, 9

Table 6.1 - *continued*			
AUTI47	1	I.2	4
AUTI48	2	I.11[b], I.1	4, 8
AUTI62	1	I.1	7
AUTI64	3	I.10[b], I.6, I.1	4, 4, 7
AUTI66	2	I.8, I.1	4, 9
AUTI72	2	I.8, I.1	4, 7
AUTI8*	2	I.2, N/O	6, N/O
AUTI11	1	I.1	7
AUTI20*	2	I.2, N/O	4, N/O
AUTI67	3	I.11[b], I.11[b], I.1	4, 4, 4
AUTI68	2	I.11[b], I.1	4, 4
AUTI69	2	I.6, I.1	4, 9
AUTI70	1	I.8	4
AUTI71*	2	I.1, N/O	8, N/O
AP4*	2	I.1, N/O	8, N/O
AP7	2	I.11[b], I.1	4, 8
AP18*	2	I.2, N/O	4, N/O
IHE1041	1	I.1	9
IHE1049	2	I.7[a] , I.1	6, 7
IHE1086	3	I.6, I.1, I.1	4, 9, 9
IHE1106	1	I.8	4
IHE1152	1	I.8	4
IHE1190	2	I.9[a,b], I.1	4, 9
IHE1268*	2	I.1, N/O	9, N/O
IHE1402	2	I.2[b], I.3	4, 6
IHE1431	1	I.2	4
KS71*	2	I.3, N/O	6, N/O
CFT073	2	I.2, I.3	4, 6
J96	2	I.6, I.6	4, 4
536	1	I.5[a]	4
Total	**54**	**90**	**82**	**82**

The occurrence of each PapI protein variant in the *pap* operons of the isolate collection was recorded and then analysed on the basis of *i*) isolate clinical origin (asymptomatic or symptomatic), *ii*) isolate *pap* copy number (one or multiple) and *iii*) PapB site 1 repeat number (4 or >4) (Table 6.2). Regression analysis provided evidence for the association of certain protein variants with the clinical origin of the isolate (p=0.03). Interestingly, the most common PapI sequence found in *pap* operons, I.1, is a novel variant and was found to associate with isolates causing

symptomatic disease. Variant I.2 also associated with symptomatic disease but generally occurred less frequently in the examined operons. The sequence of this variant is identical to PapI1 from the P1 fimbrial operon of *E. coli* CFT073 and was not found in any *pap* operons of asymptomatic clinical origin. PapI variants were also significantly associated with the copy number of *pap* operons in each isolate ($p<0.001$). Protein variants I.1, I.2 and I.6 were predominantly found in isolates that contained multiple *pap* operons, and PapI variants I.3 and I.11 were never observed in lone *pap* operons. These results are supportive of the association of certain variants, such as I.1, I.3 and I.6, with symptomatic disease since in the studied UPEC collection isolates of symptomatic disease origin are more likely to possess multiple *pap* operons (see Table 4.3B). A very strong association was also detected between PapI variants and the repeat number in PapB site 1 sequence present in the same *pap* operon ($p=10^{-10}$). In particular, the commonest variant I.1 was almost exclusively found in operons with longer site 1 sequences (>4 repeats). In fact, all operons with site 1 sequences containing 8 or 9 repeats encoded for PapI variant I.1 (see Table 6.1). Similarly, I.3 was always associated with operons containing longer site 1 sequences, although this protein variant was fairly uncommon among the isolates studied. In contrast, variants I.2, I.6, I.8, I.10 and I.11 were associated with operons containing a PapB site 1 sequence with 4 repeats. The results indicate that certain combinations of PapI variants will be commonly encountered in isolates with multiple *pap* operons. Interestingly, the occurrence of mutations in *papB* genes also appeared to correlate with certain PapI variants. In particular, all *pap* operons encoding for PapI variants I.9, I.10 and I.11 also encoded for truncated PapB proteins (Table 6.1), due to the

frameshift mutation described in section 4.4.5. These operons also had the minimum (4) number of 9 bp repeats in their PapB site 1 sequence (Table 6.2).

Table 6.2 Frequency of PapI protein variants in 82 *pap* operons from 54 UPEC clinical isolates. The total number of *pap* operons encoding for each PapI protein variant is listed in the first frequency column (total number of examined *pap* operons: 82). The frequency of each PapI variant was then investigated depending on whether the encoding *pap* operons belonged to (*1*) an asymptomatic or symptomatic isolate, (*2*) an isolate with a single or multiple *pap* operons and (*3*) depending on whether the encoding *pap* operon contained a PapB site 1 sequence with 4 or >4 repeats. Logistic regression analysis on variants with more than 6 occurrences provided significant evidence that variants I.1 and I.2 are found more often in *pap* operons from symptomatic isolates (marked with *, p=0.03). Similarly, variants I.1, I.2, I.3, I.6 and I.11 were encoded more frequently in *pap* operons from isolates that contained multiple *pap* operons (marked with *, p<0.001). A Fisher exact test provided very strong evidence of an association between PapI variants and PapB site 1 repeat number, with variants I.2, I.6, I.8, I.10 and I.11 frequently coinciding with PapB site 1 sequences containing 4 repeats and variants I.1 and I.3 associating with >4-repeat sites (marked with *, $p=10^{-10}$).

		Frequency (no. of *pap* operons) examined by:					
		1. Isolate clinical origin		**2.** Isolate *pap* copy number		**3.** No. of repeats in PapB site 1	
PapI variant	**Frequency (no. of *pap* operons)**	*asymptomatic*	*symptomatic*	*one*	*multiple*	*4*	*>4*
I.1	25	5	20*	6	19*	2	23*
I.2	9	0	9*	3	6*	8*	1
I.3	6	3	3	0	6*	0	6*
I.4	1	1	0	0	1	1	0
I.5	1	0	1	1	0	1	0
I.6	10	4	6	4	6*	10*	0
I.7	1	0	1	0	1	0	1
I.8	13	6	7	7	6	13*	0
I.9	1	0	1	0	1	1	0
I.10	5	2	3	3	2	5*	0
I.11	10	5	5	0	10*	10*	0

The presence of positive Darwinian selection acting on *papI* genes found in UPEC isolates suggests that the observed sequence variation could be a mechanism for modulating the protein's regulatory function on the *pap* operon. PapI has been shown to complement fimbrial regulators from homologous operons (Goransson *et al.*, 1988; van der Woude and Low, 1994; Morschhauser *et al.*, 1994; Nicholson and Low, 2000), so the observed mutations could also be affecting the protein's cross-talk potential. If the hypothesis holds true then the presence of certain PapI combinations in isolates could result in different fimbrial expression outputs due to different regulatory interactions between the isolate's homologous fimbrial operons.

6.2 Effect of *papI* sequence variation on P fimbriation and *pap* expression

In order to investigate the effect of PapI polymorphism on P fimbriation inducible plasmids encoding for different PapI variants were transformed into an *E. coli* K-12 strain (ZAP594; see Table 2.1C) engineered to chromosomally contain a single copy of the complete F13 *pap* operon (*papI-papG*) from *E. coli* J96 (Linberg *et al.*, 1984). Bacteria expressing P fimbriae on their surface were detected by immunofluorescence using a PapA-specific polyclonal antiserum supplied by Prof. B.E. Uhlin (Nilsson *et al.*, 1996) and the Alexa Fluor 488 IgG conjugate (Molecular Probes) and were subsequently enumerated by flow cytometry, as described in sections 2.4 and 2.5.1. The specificity of the antibody for P fimbriae of the F13 serotype was tested using the *E. coli* K-12 strain MG1655 as a negative control and MG1655pPap5 as a positive control. The pPap5 plasmid is a pBR322-based construct that contains the complete F13 *pap* operon from J96 (Linberg *et al.*, 1984). Transformation of *E. coli* MG1655 with this multi-copy plasmid and staining with

the anti-P serum resulted in a mean of 86% (± 3.5%) fluorescent bacteria, of which only a negligible fraction (0.07%) was due to non-specific binding to MG1655 surface components (Fig. 6.5). It was therefore concluded that the tested anti-P serum recognised F13 P fimbriae specifically.

The proportion of P-fimbriate bacteria detected in a population of *E. coli* ZAP594 was very low (0.26 ± 0.02%) in the absence of PapI added *in trans* (Fig. 6.5). Induced expression of three different PapI variants (I.2, I.3 and I.6) increased the proportion of P-fimbriate bacteria in the population to 2.1%, 9.5% and 12.8%, respectively (Fig. 6.6). The *papI* genes encoding variants I.2 and I.3 were cloned under the control of the pBAD promoter in the pBAD18 vector generating plasmids pKC1 and pKC2, respectively (Table 2.2), whereas the *papI* gene for I.6 was cloned into pACYC184 under the control of the *lac* promoter, making vector pHMG98 (Table 2.2; Forsman *et al.*, 1989). PapI variants I.3 and I.6 had a significant positive effect on ZAP594 P fimbriation compared to their corresponding vector controls pBAD18 and pACYC184 ($p<0.001$), whereas variant I.2 had a marginal effect. The three PapI variants tested differed in one (I.2 and I.3), four (I.2 and I.6) and three (I.3 and I.6) amino acid residues (Fig. 6.2). These results indicate that sequence variation in PapI affects P fimbriation and the mechanism is predicted to involve changes in the protein's ability to regulate *pap* operon expression.

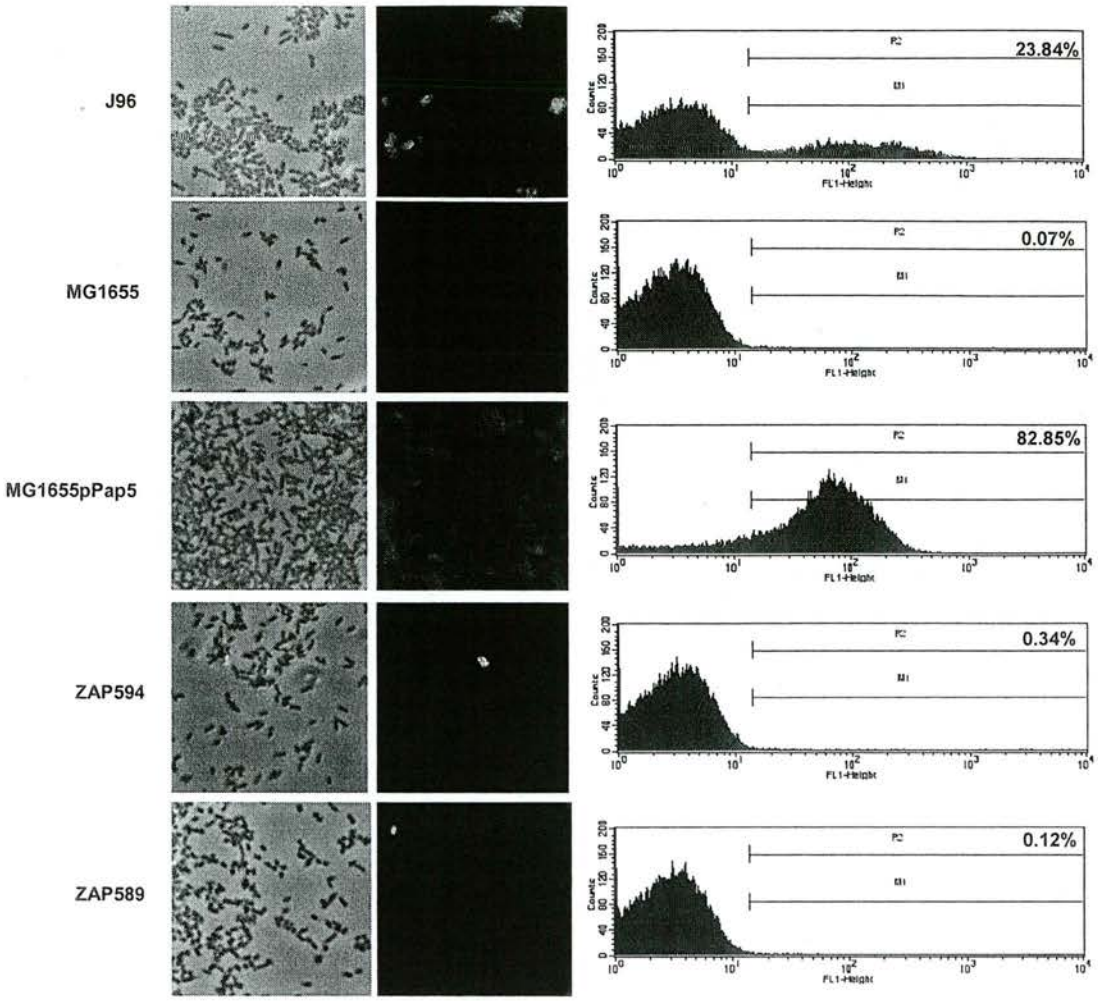


Fig. 6.5 Proportion of P-fimbriate bacteria in populations of *E. coli* J96 and K-12 strains. Phase contrast and immunofluorescence microscopy images of bacterial cells stained with anti-P (F13) serum and the secondary fluorophore-IgG conjugate Alexa Fluor 488 (Molecular Probes) are shown for *E. coli* isolate J96 and K-12 strains MG1655, MG1655pPap5, ZAP594 and ZAP589. Fluorescence histograms obtained by flow cytometry are shown for each population (left), with the proportion of fluorescent bacteria indicated in the top right-hand corner. The gate for fluorescence detection is shown by M1 and was set as defined in section 2.5.1. Microscopy images and fluorescent histograms are representatives of one immunofluorescence experiment. At least three biological replicates were performed. For ZAP594 and ZAP589 images, multiple fields had to be scanned to detect a single fluorescent bacterium, indicating the low frequency of F13 *pap* off-to-on transition in the *E. coli* K-12 background.

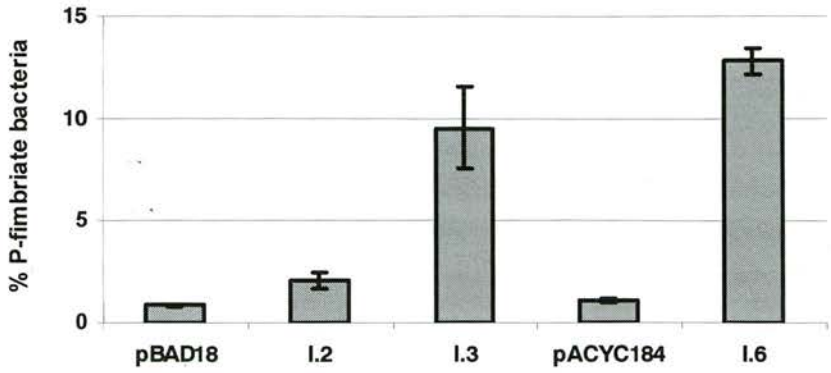


Fig. 6.6 Effect of different PapI variants on P fimbriation in *E. coli* ZAP594. Inducible expression of PapI variants I.2 and I.3 in ZAP594 by transformation with plasmids pKC1 and pKC2 resulted in 2.1% and 9.5% P-fimbriate bacteria, respectively. The effect of variant I.2 was marginal compared to pBAD18 (control vector), whereas the fimbriate proportion induced by I.3 was significantly higher ($p<0.01$). Similarly, expression of PapI variant I.6 from the inducible plasmid pHMG98 significantly increased the proportion of P-fimbriate bacteria (12.8%) compared to the pACYC184 (control vector; $p<0.001$). P-fimbriate bacteria were detected by immunofluorescence using an anti-P serum specific for the F13 serotype (Nilsson *et al.*, 1996) and Alexa Fluor 488 IgG conjugate (Molecular Probes). Proportions of fluorescent bacteria in the population were quantified by flow cytometry, as described in section 2.5.1. Results are presented as mean proportions (± s.e.) of three biological replicates.

The same activating trend was observed for the three PapI variants when tested on an *E. coli* K-12 strain engineered to contain a chromosomal single-copy fusion of the F13 *pap* regulatory region to the *egfp* reporter (ZAP589; see Table 2.1C). This translational fusion placed the *egfp* reporter gene under the control of the p_{BA} promoter and included *papI*, *papB* and the first 23 codons of *papA*. Fluorescence microscopic examination of a ZAP589 population grown under P-promoting conditions (CFA medium, 37°C) revealed that F13 *pap* expression in the K-12

background is phase variable but at very low frequencies (Fig. 6.5). The mean proportion of fluorescent bacteria in ZAP589 populations was 0.15% (± 0.5%), as determined by flow cytometry and was consistent with the low proportion of P-fimbriate bacteria detected by immunofluorescence in ZAP594 populations grown under identical conditions (see Fig. 6.5). This suggested that *pap* expression in the two systems is regulated in a similar manner.

Induced expression of PapI variants I.2, I.3 and I.6 significantly increased the proportion of fluorescent bacteria in the population compared to control vectors pBAD18 and pACYC184, respectively (Fig. 6.7). Similar to their effect on P fimbriation, PapI variant I.6 caused a ~470-fold increase in *pap* expression, while I.3 induced a higher-fold increase than variant I.2 (~85-fold and ~35-fold, respectively). The results confirmed that sequence variation altered PapI's capacity to activate *pap* operon expression. Moreover, the activation trend of the three PapI variants was the same for both the complete F13 *pap* operon strain and that containing the translational reporter fusion to the operon's regulatory region. This suggested that expression of p_{BA} reporter fusions reflects P-fimbriate bacterial proportions in the *E. coli* K-12 background and therefore single-copy reporter fusions to *pap* regulatory regions can be used instead of the complete operons for studying *pap* expression.

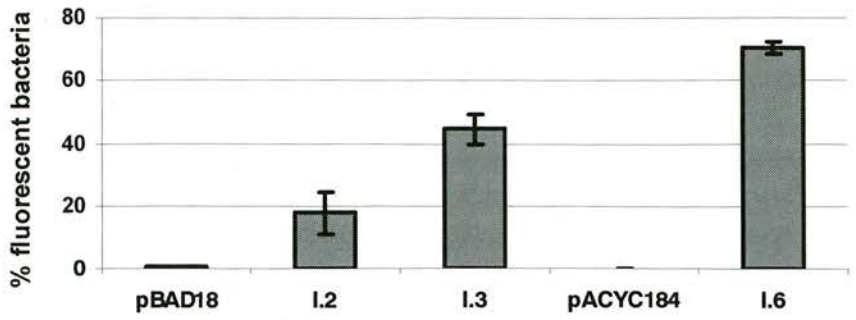


Fig. 6.7 Effect of different PapI variants on *pap* expression in *E. coli* ZAP589. Induced expression of PapI variants I.2 and I.3 in ZAP589 by transformation with plasmids pKC1 and pKC2 resulted in 17.7% and 44.6% *egfp*-expressing bacteria, respectively. Both variants resulted in a significant increase compared to pBAD18 (control vector; $p<0.001$). Similarly, expression of PapI variant I.6 from the inducible plasmid pHMG98 significantly increased the proportion of *egfp*-expressing bacteria (70.3%) compared to pACYC184 (control vector; $p<0.001$). Quantification of fluorescent bacteria was performed by flow cytometry, as described in section 2.5.1. Results are presented as mean proportions (± s.e.) of three biological replicates.

Taken together the results suggest that PapI sequence variation has an effect on the proportion of P-fimbriate bacterial in the population and *pap* expression. There is evidence to suggest that such an effect is due to altered *pap* phase variation frequencies and not elevated transcription from the p_{BA} operon promoter. Firstly, the fluorescence intensity of strain ZAP589 remained the same in the absence and presence of multi-copy PapI but the number of *egfp*-expressing bacteria increased after addition of PapI. This differentiation was possible by using flow cytometry instead of other population-based fluorescence detection systems, since the former allows fluorescence examination at the single-cell level. Moreover, additional laboratory findings (Miss K. Catherwood, unpublished data) using an *E. coli* K-12

strain (ZAP593) that contains a single-copy *lacZ* reporter fusion to the regulatory region of the F13 *pap* operon, showed that different PapI variants differentially increased the off-to-on transition frequency of the *pap* operon. These findings were in agreement with the effect observed on P fimbriation and *pap* expression of strains ZAP594 and ZAP589, respectively.

6.3 Effect of *papI* sequence variation on positive cross-talk between homologous fimbrial operons

PapI has been shown to cross-activate homologous fimbrial operons, such as *sfa* and *daa* (van der Woude and Low, 1994; Morschhauser *et al.*, 1994). The difference in activator potential observed for PapI variants I.2 and I.3 using the F13 *pap* operon and the fact that these two variants are present in the same *E. coli* strain (the first sequenced UPEC, CFT073) indicated that different cross-talk scenarios can take place in UPEC isolates. In order to investigate the ability of different PapI variants to cross-regulate homologous fimbrial operons, three *E. coli* K-12 strains were constructed, containing chromosomal single-copy *egfp* fusions to three homologous *pap* regulatory regions. The selected homologous operons were *pap1*, *pap2* and *foc* encoding for P1, P2 and F1C fimbriae, respectively and belonged to the same *E. coli* strain (CFT073), in order to investigate the PapI cross-regulatory network present in a single well-characterised UPEC isolate. The *pap1*, *pap2*, and *foc* regulatory regions were fused to e*gfp* and then introduced in single-copy into the *E. coli* K-12 *lac* chromosomal locus generating strains ZAP714, ZAP955 and ZAP715, respectively (Table 2.1C), in a manner similar to ZAP589. The cross-activating potential of the two CFT073 PapI variants (I.2 and I.3) was investigated on the three operon fusions

contained in strains ZAP714, ZAP955 and ZAP715. Transformation of plasmids pKC1 and pKC2, carrying the two CFT073 PapI variants into each K-12 strain and induction of PapI expression resulted in different proportions of *egfp*-expressing bacteria in each population (Fig. 6.8). PapI variant I.3 was able to cross-activate all three CFT073 operons to an equal extent (~10-fold). In contrast, PapI variant I.2 was able to activate the native *pap1* operon and *foc*, but failed to cross-regulate the *pap2* operon fusion.

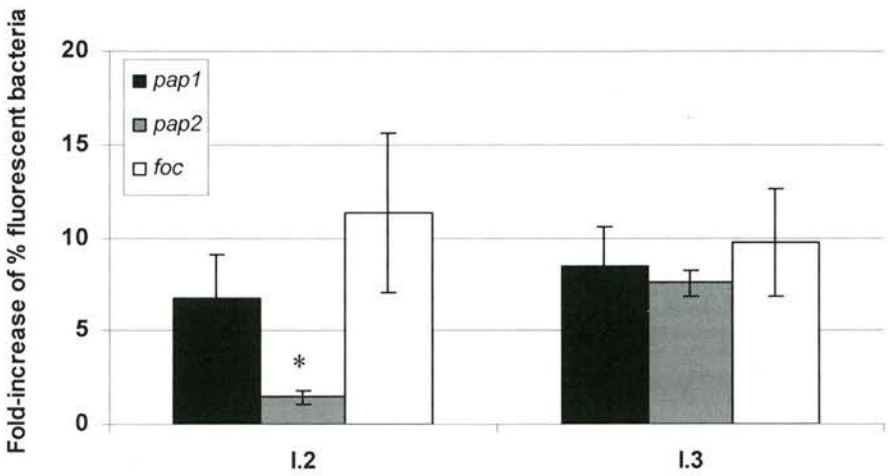


Fig. 6.8 Effect of CFT073 PapI variants on cross-regulation of CFT073 homologous fimbrial operons in *E. coli* ZAP714 (*pap1*), ZAP955 (*pap2*) and ZAP715 (*foc*). Expression of PapI variant I.2 from inducible plasmid pKC1 resulted in a mean 6.7-fold, 1.8-fold and 11.4-fold increase of *egfp*-expressing bacteria in strains ZAP714 (*pap1*), ZAP955 (*pap2*) and ZAP715 (*foc*), respectively. Expression of PapI variant I.3 from inducible plasmid pKC2 resulted in a mean 8.5-fold, 4.9-fold and 9.8-fold increase of *egfp*-expressing bacteria in strains ZAP714 (*pap1*), ZAP955 (*pap2*) and ZAP715 (*foc*), respectively. The effect of I.2 on the *pap2* operon was significantly lower than its effect on the other two CFT073 operons (p=0.016, marked as *). In addition, the effect of I.2 on the *pap2* operon was significantly lower than the effect of I.3 variant on the same operon (p=0.012). Quantification of

fluorescent bacteria was performed by flow cytometry, as described in section 2.5.1. Results are presented as mean proportions (± s.e.) of at least three biological replicates.

The protein sequence of PapI variant I.2 differs from I.3 by a single amino acid residue at position 17 (D17N). In order to confirm the profound effect of this replacement change on PapI's ability to activate the *pap2* operon fusion, the *pap1*, *pap2* and *foc* regulatory regions, as well as the *papI* genes cloned in plasmids pKC1 and pKC2 were checked by sequencing and no sequence errors were detected. To further verify the observed difference in PapI function, the *pap2* operon fusion was remade using a sequenced TOPO®-clone for CFT073 carrying the *pap2* regulatory region (see sections 4.2 and 4.3) and the brighter reporter variant *gfp+*. The fusion was introduced in single-copy into the chromosome of *E. coli* K-12 generating strain ZAP615 (Table 2.1C). Transformation of plasmids pKC1 and pKC2 into *E. coli* ZAP615 and induced expression of the different PapI variants confirmed the inability of variant I.2 to cross-activate the *pap2* operon from UPEC CFT073 (Fig. 6.9).

In summary, the results demonstrated that naturally-occurring PapI variants differ in their ability to activate transcription of the same *pap* operon leading to different proportions of P-fimbriate bacteria in the population. Moreover, it was shown that *pap* operons have different sensitivities to activation by the same PapI variant, as demonstrated by comparison of the J96 and CFT073 *pap* operons.

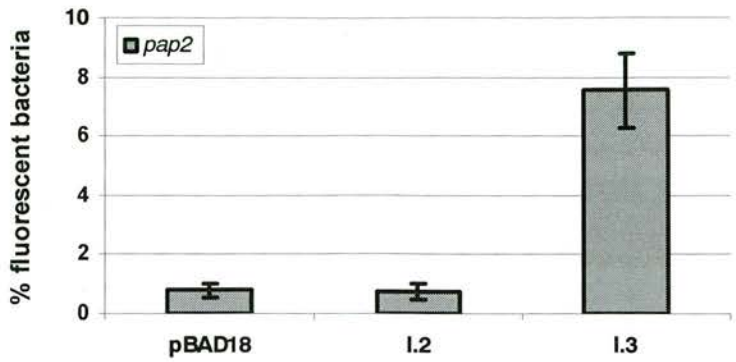


Fig. 6.9 Effect of CFT073 PapI variants on CFT073 *pap2* operon expression in *E. coli* ZAP615. Expression of PapI variants I.2 and I.3 from inducible plasmids pKC1 and pKC2 in ZAP615 resulted in 0.76% and 7.5% *gfp*-expressing bacteria, respectively. Variant I.2 failed to affect *pap2* expression, while I.3 resulted in a significant increase of expression compared to the pBAD18 (control vector; $p<0.001$). Quantification of fluorescent bacteria was performed by flow cytometry, as described in section 2.5.1. Results are presented as mean proportions (± s.e.) of seven biological replicates.

6.4 Discussion

The aim of this work was to investigate the sequence variability in *papI* genes from multiple UPEC isolates. PapI plays a central role in the regulation of P fimbriae expression in uropathogenic *E. coli* through a phase variation mechanism that relies on the synergistic action of Dam methylase, Lrp and PapI regulatory proteins on *pap* DNA (Hernday *et al.*, 2002). More specifically, PapI was shown to be required for the transition of the *pap* operon to the transcriptionally 'on' phase by binding to Lrp-*pap* DNA complexes and increasing the affinity of Lrp for p_{BA} promoter distal binding sites (Kaltenbach *et al.*, 1995; Nou *et al.*, 1995; Hernday *et al.*, 2003). Further insight into PapI's mode of action as a *pap* transcriptional activator was

gained from the recent solution structure, which identified PapI as a winged helix-turn-helix (wHtH) DNA-binding protein (Kawamura *et al.*, 2007).

The extent of nucleotide diversity identified in this study using 82 *papI* gene sequences from 54 UPEC isolates was greater than that predicted from the few *papI* sequences previously available in public databases. In total, eleven PapI protein variants were identified, of which six were novel. The point amino acid replacements in PapI were acquired at a high rate and driven by positive (adaptive) selection (dN/dS>1). Recently, the adhesive P fimbrial subunit PapG was also reported to be under strong positive selection and to acquire structural mutations (Weissman *et al.*, 2006). The authors proposed that the identified PapG mutations were pathoadaptive functional modifications of the adhesin's receptor binding properties, similar to the functionally adaptive amino acid replacements discovered in FimH, the tip adhesin of type-1 fimbriae (Sokurenko *et al.*, 2004; Weissman *et al.*, 2006). P fimbriae are surface structures with demonstrated immunogenic potential and the finding that their regulation is under adaptive selective pressure, targeted to the transcriptional activator PapI, hinted that the observed PapI mutations reflect functional differences.

Moreover, certain PapI variants were found to associate with cases of symptomatic infections and specific PapI variant combinations were frequently occurring in isolates of symptomatic clinical origin. This association also suggested a functionally adaptive role for the observed amino acid replacements. An association was also observed between PapI variants and other variable regions of the *pap* switch identified in this study and described in previous chapters, but this was partly

expected since the *pap* operons under investigation are present in pathogenicity islands that spread as complete genomic units among *E. coli* strains through horizontal gene transfer (Hacker *et al.*, 1999).

PapI acts as a transcriptional activator of the *pap* operon that encodes for it, but has also been shown to affect expression of homologous fimbrial operons, such as *sfa* (S fimbriae) and *daa* (F1845 fimbriae) (Goransson *et al.*, 1988; van der Woude and Low,1994; Morschhauser *et al.*,1994). The effect of PapI sequence variation in the protein's role as a fimbrial transcriptional activator has not been examined. Here, certain PapI variants were shown to differ in their ability to activate *pap* expression. The effect of sequence variation on PapI's mode of action was examined using K12 strains containing a chromosomal copy of a complete *pap* operon or different p_{BA} promoter fusions to the *lacZ* reporter gene. Differences in regulatory function between selected PapI protein variants were demonstrated using the same *pap* regulatory sequence fusion in the same genetic background, indicating that the mechanism by which amino acid replacements in PapI affect *pap* phase variation is likely to involve altered PapI interactions with the Lrp-*pap* DNA complex. There is only one amino acid difference between variants I.2 and I.3 (D17N), yet variant I.2 is much less efficient in activating transcription of the same *pap* operon promoter. The amino acid replacement is located within the structural loop joining the first two helices (α1 and α2) in PapI, which is not proposed to interact with DNA (Kawamura *et al.*, 2007). However, loops tend to lie on the protein surface and interact with the surrounding aqueous environment and other proteins, so it is likely that this residue is interacting with Lrp. Variant I.6 contains further amino acid replacements, which

could be playing a role in the elevated activating potential observed for this protein compared to variants I.2 and I.3. However, this result should be interpreted with caution as I.6 was expressed from a different inducible vector than I.2 and I.3 (pACYC184-based *vs.* pBAD18-based), which could be the reason for the observed difference in activation levels. Future work could investigate the interaction kinetics of different PapI variants with Lrp on *pap* regulatory DNA using competition electrophoretic mobility shift assays, in order to identify specific amino acid residues that play a role in this interaction.

The ability of PapI to cross-activate homologous fimbrial operons was also shown to be affected by observed amino acid replacements. This aspect of PapI function was investigated using homologous operons from *E. coli* CFT073, in order to investigate the PapI cross-regulatory network present in a single well-characterised UPEC isolate. Variant I.2 failed to cross-activate the P2 operon from CFT073, while it had a positive effect on the isolate's two other operons (P1 and F1C), as well as the *pap* operon from UPEC J96. Variant I.3 differs in a single amino acid residue from I.2 (N17D) and was able to cross-activate all operons tested. This difference was again suggestive of residue 17 playing a role in PapI interactions with Lrp-DNA complexes. Interestingly, the results showed that the extent of *pap* operon activation is not only dependent on the PapI protein sequence, but the DNA sequence of the regulatory region as well. The *pap* operon from J96 appeared more sensitive to activation than P1 and P2 operons from CFT073, as evident from the higher fold-increase of ZAP589 fluorescent bacteria upon I.2 and I.3 expression compared to ZAP714 and ZAP955. The regulatory region of the J96 *pap* operon is 95.8% and 94.9% identical

to P1 and P2 regulatory regions, respectively. Sequence differences exist in the region of Lrp binding sites and although PapI alone was shown to lack DNA substrate specificity (Kawamura *et al.*, 2007) the nucleotide differences could be accounting for differences in affinity of PapI-Lrp complexes, which were reported to have specificity for DNA sequence (Hernday *et al.*, 2004). It is therefore proposed that sequence changes in all three interacting factors can impact on their ability to activate *pap* expression (Fig. 6.10), without however precluding the possibility that other unknown factor(s), also required for *pap* phase switching, might play a role too.

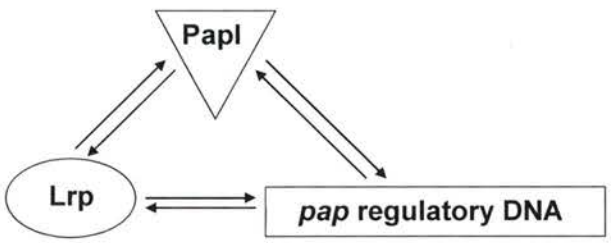


Fig. 6.10 Diagrammatic representation of interdependent sequence relationships between PapI, Lrp and the *pap* regulatory region. The depicted triangle is composed of three known factors involved in *pap* off-to-on phase transition: PapI, Lrp and *pap* regulatory DNA. The double arrows indicate interdependent relationships between the sequence (DNA or protein) of all factors. Evidence supporting that PapI and *pap* regulatory DNA sequence changes affect *pap* activation was presented in this study, while Lrp mutations affecting *pap* off-to-on transition were demonstrated in Kaltenbach *et al.*, 1998. The role of Dam is not included in this diagram.

In conclusion, the results indicate that UPEC clinical isolates can encode for PapI regulators that may or may not cross-activate homologous fimbrial operons present in the same isolate. Moreover, the extent of activation can differ between different

PapI regulators and different *pap* operons. Depending on the PapI combinations and the responsiveness of each operon to activation, it is predicted that each isolate can have different fimbrial outputs and that *papI* sequence variation could be part of the mechanism by which uropathogens achieve different surface-presentations required for their survival in different niches.

Chapter 7

General discussion

P fimbriae are important virulence factors of uropathogenic *Escherichia coli* (UPEC) for bacterial colonisation of the human urinary tract and play a key role in the development of symptomatic urinary tract infections (UTIs). Expression of P fimbriae, like the majority of *E. coli* adhesins, is regulated by phase-variation and the frequencies at which this mechanism occurs are linked to the proportion of P-fimbriate bacteria in the population. As the expression of P fimbriae triggers pro-inflammatory responses in the host (Wullt *et al.*, 2001a, b; Samuelsson *et al.*, 2004), a relationship is likely to exist between the proportion of P-fimbriate bacteria and the severity of disease. Such a relationship would argue that P fimbrial phase transition frequencies are virulence factors in their own right and therefore mechanisms for modulating them might have evolved. Some supportive evidence comes from studies reporting repressed fimbrial expression in UPEC isolates from asymptomatic infections (Andersson *et al.*, 1991; Graham *et al.*, 2001; Hull *et al.*, 1998, 1999; Stenqvist *et al.*, 1987; Lim *et al.*, 1998; Klemm *et al.*, 2006; Roos and Klemm, 2006; Roos *et al.*, 2006a, b). Moreover, the ability of type 1 fimbriae to phase-vary was proposed to be advantageous to UPEC survival in the murine urinary tract (Snyder *et al.*, 2006). The aim of this study was to investigate the regulation of P fimbrial phase transition frequencies in multiple UPEC isolates of different clinical origin, and the investigation was focused on the regulatory sequence of *pap* operons.

In the present study, P fimbrial phase transition frequencies were measured in a clinical isolate genetic background. Previous studies have assessed *pap* phase variation in the *E. coli* K-12 genetic background (Blyn *et al.*, 1989, 1990; White-Ziegler *et al.*, 1990) that is non-native as it does not normally contain *pap* operons.

The isolate genetic background proved to be a more relevant regulatory context for assessing *pap* phase variation, as it allowed it to occur in high enough frequencies to account for the level of P fimbriation observed for UPEC isolates in this study.

Phase variation frequencies measured under identical growth conditions and in the same genetic background were found to be different between homologous *pap* operons from different UPEC isolates. Similar findings have been reported for homologous *pap* operons belonging to the same UPEC isolate (Blyn *et al.*, 1989). In the present study, differential expression of homologous *pap* operons from the same UPEC isolate was also observed under certain environmental conditions. These observations suggested that mechanisms for modulating *pap* phase variation frequencies exist in UPEC isolates. Prior to this work, DNA sequence information for only a few regulators and regulatory regions from *pap* and *pap*-related fimbrial operons was available, but contained sequence differences with a predicted potential to affect *pap* phase variation. An interesting example is the *pap* operon from *E. coli* isolate J96 that was shown in this study to phase-vary at higher off-to-on transition frequencies than P1 and P2 operons from *E. coli* CFT073. Its regulatory region is 95.8% and 94.9% identical to the P1 and P2 regulatory regions, respectively, and sequence differences exist in the *papI* gene and the *pap* DNA region containing PapB site 1, CAP and Lrp binding sites. The factors involved in the increased phase transition frequency of the Pap_{J96} operon are unknown, but sequence variation in the local regulatory proteins and/or the *pap* regulatory DNA region was proposed to play a role. The remainder of this chapter discusses the evidence provided by this study supporting the hypothesis that sequence variation in *pap* regulatory regions has an

impact on phase transition frequencies and the regulation of *pap* expression, and presents a hypothesis for the potential implications of this mechanism on fimbrial expression in UPEC isolates.

To investigate sequence variation to a greater extent, 82 *pap* regulatory regions belonging to 54 UPEC clinical isolates were sequenced in this study. Evidence of sequence differences that can affect *pap* phase variation and subsequently P fimbrial expression was presented for a protein binding site in *pap* regulatory DNA, as well as for the *papI* and *papB* genes encoding for the local transcriptional regulators PapI and PapB. Interestingly, the sequence variation identified in the *pap* transcriptional activator PapI was shown to be driven by positive selection, providing strong supportive evidence that mechanisms for modulating phase variation of homologous *pap* operons have evolved in UPEC isolates. The *pap* adhesin PapG, the type 1 fimbrial adhesin FimH, and the Dr adhesin DraE, have also been shown to accumulate functionally adaptive amino acid replacements under strong positive selection forces (Sokurenko *et al.*, 1998, 2004; Hommais *et al.*, 2003; Weissman *et al.*, 2006; Korotkova *et al.*, 2007), however this is the first demonstration of a fimbrial regulator evolving in a similar manner.

PapB binding site 1, located near the p_I promoter, was a highly variable region observed among the sequenced *pap* switches and contained insertions/deletions of 9 bp sequence repeats. Naturally-occurring site 1 variants with 4, 6, 7, 8, or 9 repeats were found to differ in their relative affinity for PapB binding. These results were in agreement with previous investigation of PapB binding to site 1 sequences with 4

and 6 repeats (Xia *et al.*, 1998). The present study extended the investigation of this finding to its impact on expression from both divergently transcribed *pap* promoters (p_I and p_{BA}). Overall, longer site 1 variants displayed higher off-to-on transition frequencies than short site 1 variants, and this may be partly attributed to higher levels of *papI* transcriptional activation, although differential *papI* activation was not solely dependent on the number of repeats present in the upstream PapB site 1. The differential in sensitivity of *pap* operon activation was subtle and appeared less prominent in culture conditions considered to promote P fimbrial expression. The reason for this remains unclear. The observation however that UPEC isolates with two *pap* operons tend to differ in their site 1 sequences suggests that differential sensitivity of activation generated by sequence change could be playing a role in coordinating expression of homologous *pap* operons present in the same isolate.

Sequence changes were also identified in the genes encoding for the local *pap* regulators. As both PapI and PapB are required for regulating P fimbrial expression, it was tempting to speculate that the observed variation in amino acid sequence alters protein regulatory function and represents adaptive changes that have evolved in an isolate background-dependent manner to modulate *pap* phase variation. The variation identified in PapB involved a single nucleotide deletion leading to a frameshift mutation, which is more likely to prevent phase-variable expression of the encoding *pap* operon, rather than modulate its phase transition frequencies. The resulting truncated polypeptides were predicted to be nonfunctional as they lacked both the DNA binding and oligomerisation regions required for PapB to function as a transcriptional regulator (Xia and Uhlin, 1999). In contrast, amino acid replacement

changes in PapI were demonstrated to alter the capacity of the regulator to activate expression of the same *pap* operon. The molecular mechanism by which PapI variants differentially affect *pap* operon expression was proposed to involve the interaction of PapI with Lrp on *pap* DNA. This interaction appeared to be influenced by changes in all three factors involved (see Fig. 6.10).

Previous work has established that PapI and PapB can work *in trans* to cross-regulate homologous fimbrial operons (Goransson *et al.*, 1988; van der Woude and Low, 1994; Morschhauser *et al.*, 1994). It was therefore appreciated that the observed sequence variation in *pap*-encoded regulators could not only modulate phase variation frequencies of the encoding *pap* operon, but could potentially affect phase variation frequencies of homologous *pap* operons present in the same isolate through positive cross-talk. Modulation of regulatory cross-talk potential was demonstrated for certain PapI variants, which were commonly found in combination with other PapI protein variants in UPEC isolates possessing multiple *pap* operons. The ability of PapI variants to cross-activate the same *pap* operon was found to differ and the extent of cross-activation also depended on the regulatory sequence of the operon. Similarly, the observed deletions in *papB* were predominantly found in isolates that contained homologous *pap* operons with intact *papB* copies. Full-length *papB* genes encoded for highly conserved regulators. In the context of these UPEC isolates, such mutations will reduce the *papB* copy number and potentially the level of the regulator in the cell, which can have an effect on PapB cross-talk between the *pap* operons of an isolate. Variation in PapB protein levels can prove more important in modulating phase transition frequencies than protein sequence variation, since

PapB's autoregulatory function is dependent on its expression level; PapB activates *pap* expression at low levels and represses it at increasing concentrations (Forsman *et al.*, 1989).

Phase variation is generally considered to be a stochastic event with a switching frequency that can in some cases be influenced by external inputs. This study provides evidence that the phase variation frequencies determining P fimbriae expression are not only modulated by external factors but are also influenced by DNA sequence variation in the regulatory region and the regulators controlling the *pap* switch. Variable regulatory DNA regions were shown to affect the sensitivity of *pap* operon activation, while regulatory protein variants differed in their capacity to activate expression of their own or homologous *pap* operons. Based on the findings of this study, it was hypothesised that modulation of *pap* phase variation frequencies via regulatory sequence changes has evolved as a mechanism for differential activation of homologous *pap* operons present in the same UPEC isolate and possibly for coordinating the timing of their expression. Therefore, a model can be drawn predicting different scenarios of sequential expression of homologous *pap* operons in a UPEC isolate (Fig. 7.1). A UPEC isolate with three homologous fimbrial operons is considered in this model: two encoding for P fimbriae and one encoding for F1C fimbriae. The two *pap* operons contain regulatory sequence variation that was commonly encountered in the clinical UPEC isolates examined in this study. Under certain environmental conditions and assuming a low level of PapB regulator present, operon P_2 will have a higher probability of switching on. This probability is amplified further, as P_2 will have a higher sensitivity of activation than

P_1 due to its longer PapB site 1 sequence. Expression of PapI variant I.3 will ensure P_2 remains on but eventually, high local levels of PapB will repress expression. Variant I.3 has the potential to cross-activate P_1, which results in a fraction of bacterial cells sequentially expressing P_1 fimbriae on their surface after P_2. The PapI variant encoded by P_1 (I.2) lacks the potential to re-activate P_2 but can cross-activate the homologous fimbrial operon F1C. Increasing local PapB levels will eventually repress P_1 operon expression and F1C fimbriae will be sequentially presented on the bacterial cell surface after P_1. A distinct pathway of sequential fimbrial expression is also presented in parallel, but this scenario is predicted to have a lower probability of occurring in the bacterial population, due to the regulatory sequence differences present in the P operons modelled.

Differential sensitivity to activation and differential ability of cross-activation by homologous fimbrial regulators achieves a programmed set of switching events at the single-cell level, and coordinates fimbrial expression in a hierarchical manner. Sequential expression of related fimbrial structures may be of benefit for several reasons. Firstly, it limits physical interference between surface structures. In addition, the existence of a 'master' fimbrial operon which when stimulated by appropriate environmental cues can initiate an activation relay by influencing the frequency of activation of related fimbrial operons lower in the regulatory hierarchy, provides the cell with the chance to display its entire repertoire of related fimbrial adhesins on its surface. This could perhaps aid sequential occupation of different urinary tract niches, due to slight differences in receptor specificity displayed by different adhesin alleles. Lastly, fimbrial operons containing inactivating mutations in certain genes, as seen

here for the PapB regulator could be complemented by positive regulatory cross-talk and thus expressed.

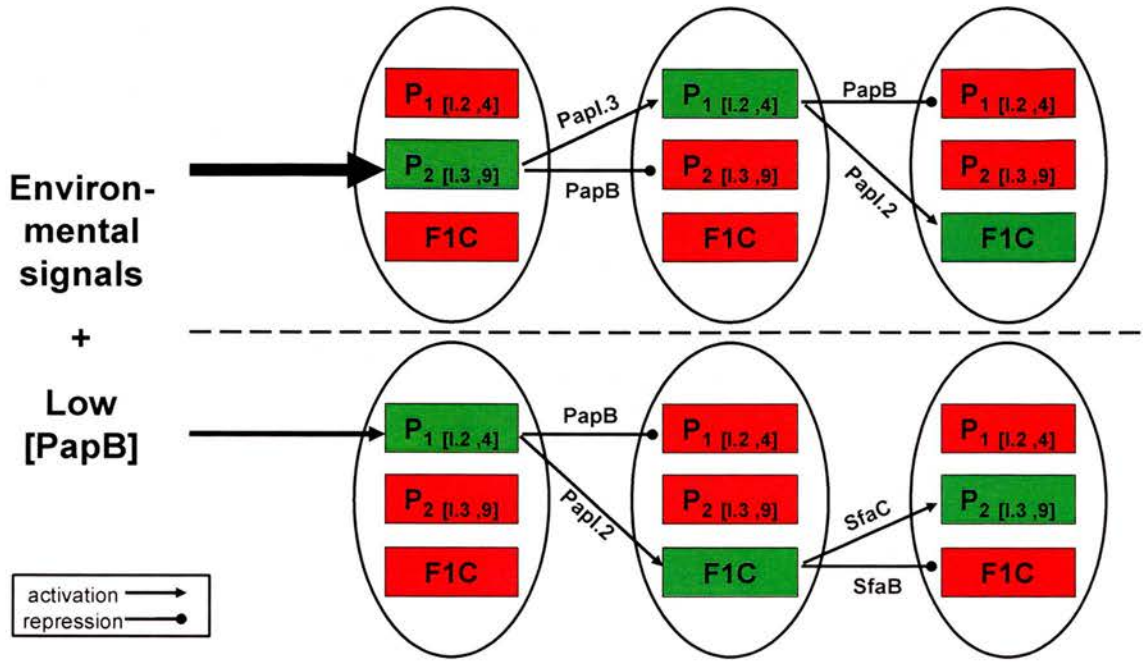


Fig. 7.1 Schematic diagram illustrating different scenarios of sequential expression of homologous fimbrial operons in a UPEC isolate. The model UPEC isolate contains three homologous fimbrial operons: P_1, P_2 encoding for P fimbriae and F1C encoding for F1C fimbriae. Regulatory sequence differences between P_1 and P_2 are shown in the encoded PapI variant (I.2 and I.3) and the number of 9 bp repeats present in the PapB binding site 1 sequence in each operon (4 and 9). Activated operons are shaded green, while red represents no expression. The two depicted scenarios have different probabilities (marked by arrow weight), due to (*i*) differential responsiveness of homologous P fimbrial operons to certain environmental conditions, and (*ii*) differential sensitivity of PapB site 1 variant operons to activation in the presence of low PapB levels. Once expression of the first operon is activated, the rules applied in the diagram are: *i*) the P_2 operon is more likely to cross-activate P_1, than vice versa, due to differences in the potential of encoded PapI variants to cross-talk, *ii*) increasing local PapB/SfaB levels will repress expression of the encoding P/F1C fimbrial operon, and *iii*) F1C operon activation is likely to follow P_1 or P_2 expression due to PapI's cross-activating ability, while SfaC is also capable of cross-activating P fimbrial operons (Goransson *et al.*, 1988).

This interplay between phase-varying *pap* operons becomes more important when integrated into the entire regulatory network operating in clinical *E. coli* isolates and controlling expression of multiple surface organelles, including fimbriae, autotransporter proteins, capsule and flagella. Its existence into such a regulatory context suggests a role in the dynamic bacterial adaptation and survival in changing environments and supports the view that phase variation is not a random uncontrolled process (Holden and Gally, 2004; van der Woude, 2006).

Surface presentation of UPEC is expected to affect the strain's pathogenic potential as most surface factors implicated in regulatory cross-talk are immunogenic. Testing this hypothesis *in vivo* is a challenge, as subtle changes in fimbrial phase variation resulting directly by a bacterial regulatory mechanism can be hard to distinguish from selection forces driven by the host's immune system. One of the original aims of this thesis was to test whether differences in P fimbrial phase transition frequencies impact on the strain's pathogenic potential *in vivo*. Preliminary work, which is not part of this thesis, tested the C3H/HeN mouse model of UTI for pro-inflammatory responses induced by isogenic P fimbrial mutants. The range of induced responses was unsuitable for differentiating the effects of isogenic isolate variants with subtle differences in P fimbrial phase transition frequencies (M. Totsika, unpublished). Future work needs to identify a suitable *in vivo* model or an appropriate *in vitro* cell system that will allow the study of fimbrial regulatory variants in the context of clinical isolates in order to understand the biological benefits of regulating the phase variation mechanism of P fimbrial expression. Understanding these regulatory pathways will provide further insights into bacterial pathogenesis.

References

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*. *PNAS* **82**, 5724-5727.

Abraham, S., Shin, J. & Malaviya, R. (2001). Type 1 fimbriated *Escherichia coli*-mast cell interactions in cystitis. *J Infect Dis* **183 Suppl 1**, S51-55.

Agace, W., Hedges, S., Andersson, U., Andersson, J., Ceska, M. & Svanborg, C. (1993a). Selective cytokine production by epithelial cells following exposure to *Escherichia coli*. *Infect Immun* **61**, 602-609.

Agace, W. W., Hedges, S. R., Ceska, M. & Svanborg, C. (1993b). Interleukin-8 and the neutrophil response to mucosal gram-negative infection. *J Clin Invest* **92**, 780-785.

Agace, W., Connell, H and C. Svanborg (1996). Host resistance to Urinary tract infection. In *Urinary Tract Infections*. Edited by L. T. Mobley and J. W. Warren. Washington, DC: American Society for Microbiology Press.

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.

Anderson, G. G., Martin, S. M. & Hultgren, S. J. (2004). Host subversion by formation of intracellular bacterial communities in the urinary tract. *Microbes Infect* **6**, 1094-1101.

Andersson, P., Engberg, I., Lidin-Janson, G., Lincoln, K., Hull, R., Hull, S. & Svanborg, C. (1991). Persistence of *Escherichia coli* bacteriuria is not determined by bacterial adherence. *Infect Immun* **59**, 2915-2921.

Aoki, S. K., Pamma, R., Hernday, A. D., Bickham, J. E., Braaten, B. A. & Low, D. A. (2005). Contact-dependent inhibition of growth in *Escherichia coli*. *Science* **309**, 1245-1248.

Asscher, A. W., Sussman, M., Waters, W. E., Davis, R. H. & Chick, S. (1966). Urine as a medium for bacterial growth. *Lancet* **2**, 1037-1041.

Backhed, F., Alsen, B., Roche, N., Angstrom, J., von Euler, A., Breimer, M. E., Westerlund-Wikstrom, B., Teneberg, S. & Richter-Dahlfors, A. (2002a). Identification of target tissue

glycosphingolipid receptors for uropathogenic, F1C-fimbriated *Escherichia coli* and its role in mucosal inflammation. *J Biol Chem* **277**, 18198-18205.

Backhed, F., Meijer, L., Normark, S. & Richter-Dahlfors, A. (2002b). TLR4-dependent recognition of lipopolysaccharide by epithelial cells requires sCD14. *Cell Microbiol* **4**, 493-501.

Baga, M., Goransson, M., Normark, S. & Uhlin, B. E. (1985). Transcriptional activation of a *pap* pilus virulence operon from uropathogenic *Escherichia coli*. *EMBO J* **4**, 3887-3893.

Bann, J. G., Pinkner, J. S., Frieden, C. & Hultgren, S. J. (2004). Catalysis of protein folding by chaperones in pathogenic bacteria. *PNAS* **101**, 17389-17393.

Barnich, N., Boudeau, J., Claret, L. & Darfeuille-Michaud, A. (2003). Regulatory and functional co-operation of flagella and type 1 pili in adhesive and invasive abilities of AIEC strain LF82 isolated from a patient with Crohn's disease. *Mol Microbiol* **48**, 781-794.

Berger, H., Hacker, J., Juarez, A., Hughes, C. & Goebel, W. (1982). Cloning of the chromosomal determinants encoding hemolysin production and mannose-resistant hemagglutination in *Escherichia coli*. *J Bacteriol* **152**, 1241-1247.

Bergsten, G., Samuelsson, M., Wullt, B., Leijonhufvud, I., Fischer, H. & Svanborg, C. (2004). PapG-dependent adherence breaks mucosal inertia and triggers the innate host response. *J Infect Dis* **189**, 1734-1742.

Bergsten, G., Wullt, B., Schembri, M. A., Leijonhufvud, I. & Svanborg, C. (2007). Do type 1 fimbriae promote inflammation in the human urinary tract? *Cell Microbiol* doi:10.1111/j.1462-5822.2007.00912.x

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242.

Berman, H., Henrick, K. & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980.

Bidet, P., Bonacorsi, S., Clermont, O., De Montille, C., Brahimi, N. & Bingen, E. (2005). Multiple insertional events, restricted by the genetic background, have led to acquisition of pathogenicity island IIJ96-like domains among *Escherichia coli* strains of different clinical origins. *Infect Immun* **73**, 4081-4087.

Blanco, J., Alonso, M. P., Gonzalez, E. A., Blanco, M. & Garabal, J. I. (1990). Virulence factors of bacteraemic *Escherichia coli* with particular reference to production of cytotoxic necrotising factor (CNF) by P-fimbriate strains. *J Med Microbiol* **31**, 175-183.

Blanco, M., Blanco, J. E., Alonso, M. P., Mora, A., Balsalobre, C., Munoa, F., Juarez, A. & Blanco, J. (1997). Detection of *pap*, *sfa* and *afa* adhesin-encoding operons in uropathogenic *Escherichia coli* strains: relationship with expression of adhesins and production of toxins. *Res Microbiol* **148**, 745-755.

Blattner, F. R., Plunkett, G., III, Bloch, C. A. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462.

Blomfield, I. C., Vaughn, V., Rest, R. F. & Eisenstein, B. I. (1991). 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., 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.

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. (2001). The regulation of pap and type 1 fimbriation in *Escherichia coli*. *Adv Microb Physiol* **45**, 1-49.

Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschape, H. & Hacker, J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* **62**, 606-614.

Blum, G., Falbo, V., Caprioli, A. & Hacker, J. (1995). Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and α-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. *FEMS Microbiol Lett* **126**, 189-195.

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.

Blyn, L. B., Braaten, B. A. & Low, D. A. (1990). Regulation of *pap* pilin phase variation by a mechanism involving differential Dam methylation states. *EMBO J* **9**, 4045-4054.

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.

Boscia, J. A., Kobasa, W. D., Abrutyn, E., Levison, M. E., Kaplan A. M., and D. Kaye (1986). Lack of association between bacteriuria and symptoms in the elderly. *Am J Med* **81**, 979-982.

Bouza, E., San Juan, R., Munoz, P., Voss, A. & Kluytmans, J. (2001). A European perspective on nosocomial urinary tract infections II. Report on incidence, clinical characteristics and outcome (ESGNI-004 study). *Clin Microbiol Infec* **7**, 532-542.

Braaten, B. A., Blyn, L. B., Skinner, B. S. & Low, D. A. (1991). Evidence for a methylation-blocking factor (mbf) locus involved in *pap* pilus expression and phase variation in *Escherichia coli*. *J Bacteriol* **173**, 1789-1800.

Braaten, B. A., Platko, J. V., van der Woude, M. W., Simons, B. H., de Graaf, F. K., Calvo, J. M. & Low, D. A. (1992). Leucine-responsive regulatory protein controls the expression of both the *pap* and *fan* pili operons in *Escherichia coli*. *PNAS* **89**, 4250-4254.

Braaten, B. A., Nou, X., Kaltenbach, L. S. & Low, D. A. (1994). Methylation patterns in *pap* regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. *Cell* **76**, 577-588.

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.

Brzuszkiewicz, E., Bruggemann, H., Liesegang, H. & other authors (2006). How to become a uropathogen: Comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *PNAS* **103**, 12879-12884.

Buchanan, K., Falkow, S., Hull, R. A. & Hull, S. I. (1985). Frequency among Enterobacteriaceae of the DNA sequences encoding type 1 pili. *J Bacteriol* **162**, 799-803.

Busby, S. & Ebright, R. H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**, 743-746.

Carbonetti, N. H., Boonchai, S., Parry, S. H., Vaisanen-Rhen, V., Korhonen, T. K. & Williams, P. H. (1986). Aerobactin-mediated iron uptake by *Escherichia coli* isolates from human extraintestinal infections. *Infect Immun* **51**, 966-968.

Cardenas, D. D. & Hooton, T. M. (1995). Urinary tract infection in persons with spinal cord injury. *Arch Phys Med Rehab* **76**, 272-280.

Casadesus, J. & Low, D. (2006). Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* **70**, 830-856.

Chassin, C., Goujon, J.-M., Darche, S. & other authors (2006). Renal collecting duct epithelial cells react to pyelonephritis-associated *Escherichia coli* by activating distinct TLR4-dependent and -independent inflammatory pathways. *J Immunol* **177**, 4773-4784.

Chen, S. L., Hung, C. S., Xu, J. & other authors (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: A comparative genomics approach. *PNAS* **103**, 5977-5982.

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.

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. *PNAS* **93**, 9827-9832.

Cross, A. S., Kim, K. S., Wright, D. C., Sadoff, J. C. & Gemski, P. (1986). Role of lipopolysaccharide and capsule in the serum resistance of bacteremic strains of *Escherichia coli*. *J Infect Dis* **154**, 497-503.

Cunningham, F. G. & Lucas, M. J. (1994). Urinary tract infections complicating pregnancy. *Baillieres Clin Obstet Gynaecol* **8**, 353-373.

Danese, P. N., Pratt, L. A., Dove, S. L. & Kolter, R. (2000). The outer membrane protein, Antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. *Mol Microbiol* **37**, 424-432.

Darouiche, R. O., Donovan, W. H., Del Terzo, M., Thornby, J. I., Rudy, D. C. & Hull, R. A. (2001). Pilot trial of bacterial interference for preventing urinary tract infection. *Urology* **58**, 339-344.

de Man, P., van Kooten, C., Aarden, L., Engberg, I., Linder, H. & Svanborg Eden, C. (1989). Interleukin-6 induced at mucosal surfaces by gram-negative bacterial infection. *Infect Immun* **57**, 3383-3388.

de Ree, J. M. & van den Bosch, J. F. (1987). Serological response to the P fimbriae of uropathogenic *Escherichia coli* in pyelonephritis. *Infect Immun* **55**, 2204-2207.

DiGiuseppe, P. A. & Silhavy, T. J. (2003). Signal detection and target gene induction by the CpxRA two-component system. *J Bacteriol* **185**, 2432-2440.

Dobrindt, U., Blum-Oehler, G., Hartsch, T., Gottschalk, G., Ron, E. Z., Funfstuck, R. & Hacker, J. (2001). S-fimbria-encoding determinant sfa_I is located on pathogenicity island III_{536} of uropathogenic *Escherichia coli* strain 536. *Infect Immun* **69**, 4248-4256.

Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414-424.

Dodson, K. W., Pinkner, J. S., Rose, T., Magnusson, G., Hultgren, S. J. & Waksman, G. (2001). Structural basis of the interaction of the pyelonephritic *E. coli* adhesin to its human kidney receptor. *Cell* **105**, 733-743.

Dorel, C., Lejeune, P. & Rodrigue, A. (2006). The Cpx system of *Escherichia coli*, a strategic signaling pathway for confronting adverse conditions and for settling biofilm communities? *Res Microbiol* **157**, 306-314.

Duguid, J. P. & Anderson, E. S. (1967). Terminology of bacterial fimbriae, or pili, and their types. *Nature* **215**, 89-90.

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984). Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. *Nature* **311**, 232-235.

Eisenstein, B. (1981). Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. *Science* **214**, 337-339.

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*. *PNAS* **84**, 6506-6510.

Emmerson, J. R., Gally, D. L. & Roe, A. J. (2006). Generation of gene deletions and gene replacements in *Escherichia coli* O157:H7 using a temperature sensitive allelic exchange system. *Biol Proced Online* **8**, 153-162.

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.

Evans, D. G., Evans, D. J., Jr. & Tjoa, W. (1977). Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor. *Infect Immun* **18**, 330-337.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* **22**, 521-565.

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.

Forsman, K., Sonden, B., Goransson, M. & Uhlin, B. E. (1992). Antirepression function in *Escherichia coli* for the cAMP-cAMP receptor protein transcriptional activator. *PNAS* **89**, 9880-9884.

Foxman, B. (1990). Recurring urinary tract infection: incidence and risk factors. *Am J Public Health* **80**, 331-333.

Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B. & Sobel, J. D. (2000). Urinary tract infection: self-reported incidence and associated costs. *Ann Epidemiol* **10**, 509-515.

Foxman, B. (2002). Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med* **113**, 5-13.

Foxman, B. (2007). Contributions of molecular epidemiology to the understanding of infectious disease transmission, pathogenesis, and evolution. *Ann Epidemiol* **17**, 148-156.

Frendeus, B., Wachtler, C., Hedlund, M., Fischer, H., Samuelsson, P., Svensson, M. & Svanborg, C. (2001). *Escherichia coli* P fimbriae utilize the Toll-like receptor 4 pathway for cell activation. *Mol Microbiol* **40**, 37-51.

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.

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.

Glasner, J. D., and N. T. Perna (2004). Comparative genomics of *E. coli*. *Microbiol Today* **31**, 124-125.

Godaly, G., Proudfoot, A. E., Offord, R. E., Svanborg, C. & Agace, W. W. (1997). Role of epithelial interleukin-8 (IL-8) and neutrophil IL-8 receptor A in *Escherichia coli*-induced transuroepithelial neutrophil migration. *Infect Immun* **65**, 3451-3456.

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.

Godaly, G., Hang, L., Frendeus, B. & Svanborg, C. (2000). Transepithelial neutrophil migration is CXCR1 dependent in vitro and is defective in IL-8 receptor knockout mice. *J Immunol* **165**, 5287-5294.

Goransson, M., Forsman, K. & Uhlin, B. E. (1988). Functional and structural homology among regulatory cistrons of pili-adhesin determinants in *Escherichia coli*. *Mol Gen Genet* **212**, 412-417.

Goransson, M., Forsman, K. & Uhlin, B. E. (1989a). Regulatory genes in the thermoregulation of *Escherichia coli* pili gene transcription. *Genes Dev* **3**, 123-130.

Goransson, M., Forsman, P., Nilsson, P. & Uhlin, B. E. (1989b). Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus-adhesin in *Escherichia coli*. *Mol Microbiol* **3**, 1557-1565.

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.

Graham, J. C., Leathart, J. B., Keegan, S. J., Pearson, J., Bint, A. & Gally, D. L. (2001). Analysis of *Escherichia coli* strains causing bacteriuria during pregnancy: selection for strains that do not express type 1 fimbriae. *Infect Immun* **69**, 794-799.

Gransden, W. R., Eykyn, S. J., Phillips, I., and Rowe, B. (1990). Bactreraemia due to *Escherichia coli*: a study of 861 episodes. *Rev Infect Dis* **12**, 1008-1018.

Guignot, J., Chaplais, C., Coconnier-Polter, M.-H. & Servin, A. L. (2007). The secreted autotransporter toxin, Sat, functions as a virulence factor in Afa/Dr diffusely adhering *Escherichia coli* by promoting lesions in tight junction of polarized epithelial cells. *Cell Microbiol* **9**, 204-221.

Guyer, D. M., Kao, J.-S. & Mobley, H. L. T. (1998). Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter associated bacteriuria and from fecal samples. *Infect Immun* **66**, 4411-4417.

Guyer, D. M., Henderson, I. R., Nataro, J. P. & Mobley, H. L. T. (2000). Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. *Mol Microbiol* **38**, 53-66.

Guyer, D. M., Radulovic, S., Jones, F.-E. & Mobley, H. L. T. (2002). Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. *Infect Immun* **70**, 4539-4546.

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 Pt 1**, 135-140.

Guzman, L., Belin, D., Carson, M. & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J Bacteriol* **177**, 4121-4130.

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.

Hacker, J., Knapp, S. & Goebel, W. (1983). Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* O6 strain. *J Bacteriol* **154**, 1145-1152.

Hacker, J., Schmidt, G., Hughes, C., Knapp, S., Marget, M. & Goebel, W. (1985). Cloning and characterization of genes involved in production of mannose-resistant, neuraminidase-susceptible (X) fimbriae from a uropathogenic O6:K15:H31 *Escherichia coli* strain. *Infect Immun* **47**, 434-440.

Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R. & Goebel, W. (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* **8**, 213-225.

Hacker, J., Kestler, H., Hoschutzky, H., Jann, K., Lottspeich, F. & Korhonen, T. K. (1993). Cloning and characterization of the S fimbrial adhesin II complex of an *Escherichia coli* O18:K1 meningitis isolate. *Infect Immun* **61**, 544-550.

Hacker, J., Blum-Oehler, G., Janke, B., Nagy, G., and W. Goebel (1999). Pathogenicity islands of extraintestinal *Escherichia coli*. In *Pathogenicity islands and other mobile virulence elements*. Edited by B. Kaper and J. Hacker. Washington, DC: American Society for Microbiology Press.

Hagberg, L., Jodal, U., Korhonen, T. K., Lidin-Janson, G., Lindberg, U. & Svanborg Eden, C. (1981). Adhesion, hemagglutination, and virulence of *Escherichia coli* causing urinary tract infections. *Infect Immun* **31**, 564-570.

Hagberg, L., Hull, R., Hull, S., Falkow, S., Freter, R. & Svanborg Eden, C. (1983). Contribution of adhesion to bacterial persistence in the mouse urinary tract. *Infect Immun* **40**, 265-272.

Hanson, L. A., Ahlstedt, S., Fasth, A. & other authors (1977). Antigens of *Escherichia coli*, human immune response, and the pathogenesis of urinary tract infections. *J Infect Dis* **136 Suppl**, S144-149.

Hasman, H., Chakraborty, T. & Klemm, P. (1999). Antigen-43-mediated autoaggregation of *Escherichia coli* is blocked by fimbriation. *J Bacteriol* **181**, 4834-4841.

Hayashi, T., Makino, K., Ohnishi, M. & other authors (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11-22.

Hedges, S., Anderson, P., Lidin-Janson, G., de Man, P. & Svanborg, C. (1991). Interleukin-6 response to deliberate colonization of the human urinary tract with gram-negative bacteria. *Infect Immun* **59**, 421-427.

Hedges, S., Svensson, M. & Svanborg, C. (1992). Interleukin-6 response of epithelial cell lines to bacterial stimulation *in vitro*. *Infect Immun* **60**, 1295-1301.

Hedges, S., Agace, W., Svensson, M., Sjogren, A. C., Ceska, M. & Svanborg, C. (1994). Uroepithelial cells are part of a mucosal cytokine network. *Infect Immun* **62**, 2315-2321.

Hedlund, M., Wachtler, C., Johansson, E., Hang, L., Somerville, J. E., Darveau, R. P. & Svanborg, C. (1999). P fimbriae-dependent, lipopolysaccharide-independent activation of epithelial cytokine responses. *Mol Microbiol* **33**, 693-703.

Hedlund, M., Duan, R. D., Nilsson, A., Svensson, M., Karpman, D. & Svanborg, C. (2001a). Fimbriae, transmembrane signaling, and cell activation. *J Infect Dis* **183 Suppl 1**, S47-50.

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

Henderson, I. R. & Owen, P. (1999). The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and OxyR. *J Bacteriol* **181**, 2132-2141.

Henderson, I. R., Owen, P. & Nataro, J. P. (1999). Molecular switches - the ON and OFF of bacterial phase variation. *Mol Microbiol* **33**, 919-932.

Hernday, A., Krabbe, M., Braaten, B. & Low, D. (2002). Self-perpetuating epigenetic pili switches in bacteria. *PNAS* **99**, 16470-16476.

Hernday, A. D., Braaten, B. A. & Low, D. A. (2003). The mechanism by which DNA adenine methylase and PapI activate the *pap* epigenetic switch. *Mol Cell* **12**, 947-957.

Hernday, A. D., Braaten, B. A., Broitman-Maduro, G., Engelberts, P. & Low, D. A. (2004). Regulation of the *pap* epigenetic switch by CpxAR: phosphorylated CpxR inhibits transition to the phase ON state by competition with Lrp. *Mol Cell* **16**, 537-547.

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 NJ & Gally, DL (2004). Switches, cross-talk and memory in *Escherichia coli* adherence. *J Med Microbiol* **53**, 585-593.

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.

Holden, N., Totsika, M., Dixon, L., Catherwood, K. & Gally, D. L. (2007). Regulation of P-fimbrial phase variation frequencies in *Escherichia coli* CFT073. *Infect Immun* doi:10.1128/IAI.01989-06.

Hommais, F., Gouriou, S., Amorin, C., Bui, H., Rahimy, M. C., Picard, B. & Denamur, E. (2003). The FimH A27V mutation is pathoadaptive for urovirulence in *Escherichia coli* B2 phylogenetic group isolates. *Infect Immun* **71**, 3619-3622.

Hooton, T. M., Scholes, D., Hughes, J. P., Winter, C., Roberts, P. L., Stapleton, A. E., Stergachis, A. & Stamm, W. E. (1996). A prospective study of risk factors for symptomatic urinary tract infection in young women. *N Engl J Med* **335**, 468-474.

Hooton, T. M., Scholes, D., Stapleton, A. E., Roberts, P. L., Winter, C., Gupta, K., Samadpour, M. & Stamm, W. E. (2000). A prospective study of asymptomatic bacteriuria in sexually active young women. *N Engl J Med* **343**, 992-997.

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.

Hull, R., Bieler, S., Falkow, S. & Hull, S. (1986). Chromosomal map position of genes encoding P adhesins in uropathogenic *Escherichia coli*. *Infect Immun* **51**, 693-695.

Hull, R. A., Nowicki, B., Kaul, A., Runyan, R., Svanborg, C. & Hull, S. I. (1994). Effect of *pap* copy number and receptor specificity on virulence of fimbriated *Escherichia coli* in a murine urinary tract colonization model. *Microb Pathog* **17**, 79-86.

Hull, R. A., Rudy, D. C., Wieser, I. E. & Donovan, W. H. (1998). Virulence factors of *Escherichia coli* isolates from patients with symptomatic and asymptomatic bacteriuria and neuropathic bladders due to spinal cord and brain injuries. *J Clin Microbiol* **36**, 115-117.

Hull, R. A., Rudy, D. C., Donovan, W. H., Wieser, I. E., Stewart, C. & Darouiche, R. O. (1999). Virulence properties of *Escherichia coli* 83972, a prototype strain associated with asymptomatic bacteriuria. *Infect Immun* **67**, 429-432.

Hull, R. A., Rudy, D. C., Donovan, W. H., Svanborg, C., Wieser, I. E., Stewart, C. & Darouiche R. O. (2000). Urinary tract infection prophylaxis using *Escherichia coli* 83972 in spinal cord injured patients. *J Urol* **163**, 872-877.

Hull, R. A., Donovan, W. H., Del Terzo, M., Stewart, C., Rogers, M. & Darouiche, R. O. (2002). Role of type 1 fimbria- and P fimbria-specific adherence in colonization of the neurogenic human bladder by *Escherichia coli*. *Infect Immun* **70**, 6481-6484.

Hultgren, S. J., Porter, T. N., Schaeffer, A. J. & Duncan, J. L. (1985). Role of type 1 pili and effects of phase variation on lower urinary tract infections produced by *Escherichia coli*. *Infect Immun* **50**, 370-377.

Hultgren, S. J., Normark, S. & Abraham, S. N. (1991). Chaperone-assisted assembly and molecular architecture of adhesive pili. *Annu Rev Microbiol* **45**, 383-415.

Hung, D. L., Pinkner, J. S., Knight, S. D. & Hultgren, S. J. (1999). Structural basis of chaperone self-capping in P pilus biogenesis. *PNAS* **96**, 8178-8183.

Hung, D. L., Raivio, T. L., Jones, C. H., Silhavy, T. J. & Hultgren, S. J. (2001). Cpx signaling pathway monitors biogenesis and affects assembly and expression of P pili. *EMBO J* **20**, 1508-1518.

Ikaheimo, R., Siitonen, A., Heiskanen, T., Karkkainen, U., Kuosmanen, P., Lipponen, P. & Makela, P. H. (1996). Recurrence of urinary tract infection in a primary care setting: analysis of a 1-year follow-up of 179 women. *Clin Infect Dis* **22**, 91-99.

Jakobsson, B., Berg, U., and Svensson, L. (1994). Renal scarring after acute pyelonephritis. *Arch Dis Child* **70**, 11-15.

Janke, B., Dobrindt, U., Hacker, J. & Blum-Oehler, G. (2001). A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. *FEMS Microbiol Lett* **199**, 61-66.

Jann, B., and K. Jann, M. (1997). Capsules of *Escherichia coli*. In *Mechanisms of virulence*, pp. 113–143. Edited by M. Sussman. Cambridge, England: Cambridge University Press.

Jodal, U. (1987). The natural history of bacteriuria in childhood. *Infect Dis Clin North Am* **1**, 713-729.

Johanson, I. M., Plos, K., Marklund, B. I. & Svanborg, C. (1993). Pap, *papG* and *prsG* DNA sequences in *Escherichia coli* from the fecal flora and the urinary tract. *Microb Pathog* **15**, 121-129.

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

Johnson, T. J., Kariyawasam, S., Wannemuehler, Y. & other authors (2007). The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* **189**, 3228-3236.

Jones, C. H., Dodson, K., and Hultgren, S. J. (1996). Structure, function and assembly of adhesive P pili. In *Urinary Tract Infections*. Edited by L. T. Mobley and J. W. Warren. Washington, DC: American Society for Microbiology Press.

Jordan, P. A., Iravani, A., Richard, G. A. & Baer, H. (1980). Urinary tract infection caused by *Staphylococcus saprophyticus*. *J Infect Dis* **142**, 510-515.

Kallenius, G. & Mollby, R. (1979). Adhesion of *Escherichia coli* to human periurethral cells correlated to mannose-resistant agglutination of human erythrocytes. *FEMS Microbiol Lett* **5**, 295-299.

Kallenius, G., Mollby, R. & Winberg, J. (1980). In vitro adhesion of uropathogenic *Escherichia coli* to human periurethral cells. *Infect Immun* **28**, 972-980.

Kaltenbach, L., Braaten, B. & Low, D. (1995). Specific binding of PapI to Lrp-*pap* DNA complexes. *J Bacteriol* **177**, 6449-6455.

Kaltenbach, L., Braaten, B., Tucker, J., Krabbe, M. & Low, D. (1998). Use of a two-color genetic screen to identify a domain of the global regulator Lrp that is specifically required for *pap* phase variation. *J Bacteriol* **180**, 1224-1231.

Kantele, A., Mottonen, T., Ala-Kaila, K. & Arvilommi, H. S. (2003). P fimbria-specific B cell responses in patients with urinary tract infection. *J Infect Dis* **188**, 1885-1891.

Kao, J., Stucker, D., Warren, J. & Mobley, H. (1997). Pathogenicity island sequences of pyelonephritogenic *Escherichia coli* CFT073 are associated with virulent uropathogenic strains. *Infect Immun* **65**, 2812-2820.

Kaper, J. B., Nataro, J. P. & Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**, 123-140.

Karr, J. F., Nowicki, B., Truong, L. D., Hull, R. A. & Hull, S. I. (1989). Purified P fimbriae from two cloned gene clusters of a single pyelonephritogenic strain adhere to unique structures in the human kidney. *Infect Immun* **57**, 3594-3600.

Karr JF, N. B., Truong LD, Hull RA, Moulds JJ, Hull SI (1990). pap-2-encoded fimbriae adhere to the P blood group-related glycosphingolipid stage-specific embryonic antigen 4 in the human kidney. *Infect Immun* **58**, 4055-4062.

Kawamura, T., Le, L. U., Zhou, H. & Dahlquist, F. W. (2007). Solution structure of *Escherichia coli* PapI, a key regulator of the *pap* pili phase variation. *J Mol Biol* **365**, 1130-1142.

Keegan, S. J. (2001). Characterisation and adherence mechanisms of *Escherichia coli* strains causing infections in patients with a reconstructed bladder. Ph.D. Thesis, University of Newcastle upon Tyne.

Keegan, S. J., Graham, C., Neal, D. E., Blum-Oehler, G., N'Dow, J., Pearson, J. P. & Gally, D. L. (2003). Characterization of *Escherichia coli* strains causing urinary tract infections in patients with transposed intestinal segments. *J Urol* **169**, 2382-2387.

Khan, A. S., Kniep, B., Oelschlaeger, T. A., Van Die, I., Korhonen, T. & Hacker, J. (2000). Receptor structure for F1C fimbriae of uropathogenic *Escherichia coli*. *Infect Immun* **68**, 3541-3547.

Kinane, D. F., Blackwell, C. C., Brettle, R. P., Weir, D. M., Winstanley, F. P. & Elton, R. A. (1982). ABO blood group, secretor state, and susceptibility to recurrent urinary tract infection in women. *Br Med J (Clin Res Ed)* **285**, 7-9.

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., Roos, V., Ulett, G. C., Svanborg, C. & Schembri, M. A. (2006). Molecular Characterization of the *Escherichia coli* asymptomatic bacteriuria strain 83972: the taming of a pathogen. *Infect Immun* **74**, 781-785.

Klumpp, D. J., Rycyk, M. T., Chen, M. C., Thumbikat, P., Sengupta, S. & Schaeffer, A. J. (2006). Uropathogenic *Escherichia coli* induces extrinsic and intrinsic cascades to initiate urothelial apoptosis. *Infect Immun* **74**, 5106-5113.

Knapp, S., Hacker, J., Jarchau, T. & Goebel, W. (1986). Large, unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* O6 strain 536. *J Bacteriol* **168**, 22-30.

Korhonen, T. K., Valtonen, M. V., Parkkinen, J., Vaisanen-Rhen, V., Finne, J., Orskov, F., Orskov, I., Svenson, S. B. & Makela, P. H. (1985). Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. *Infect Immun* **48**, 486-491.

Korhonen, T. K., Parkkinen, J., Hacker, J., Finne, J., Pere, A., Rhen, M. & H., H. (1986a). Binding of *Escherichia coli* S fimbriae to human kidney epithelium. *Infect Immun* **54**, 322-327.

Korhonen, T. K., Virkola, R. & Holthofer, H. (1986b). Localization of binding sites for purified *Escherichia coli* P fimbriae in the human kidney. *Infect Immun* **54**, 328-332.

Korotkova, N., Chattopadhyay, S., Tabata, T. A. & other authors (2007). Selection for functional diversity drives accumulation of point mutations in Dr adhesins of *Escherichia coli*. *Mol Microbiol* **64**, 180-194.

Krogfelt, K. A. (1991). Bacterial adhesion: Genetics, biogenesis, and role in pathogenesis of fimbrial adhesins of *Escherichia coli*. *Rev Infect Dis* **13**, 721-735.

Kucheria, R., Dasgupta, P., Sacks, S. H., Khan, M. S. & Sheerin, N. S. (2005). Urinary tract infections: new insights into a common problem. *Postgrad Med J* **81**, 83-86.

Kuehn, M. J., Heuser, J., Normark, S. & Hultgren, S. J. (1992). P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature* **356**, 252-255.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150-163.

Kunin, C. M. (1994). Urinary tract infections in females. *Clin Infect Dis* **18**, 1-10; quiz 11-12.

Lane, M. C. & Mobley, H. L. T. (2007). Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. *Kidney Int* doi: 10.1038/sj.ki.5002230

Latham, R. H., Running, K. & Stamm, W. E. (1983). Urinary tract infections in young adult women caused by *Staphylococcus saprophyticus*. *JAMA* **250**, 3063-3066.

Leffler, H. & Svanborg-Eden, C. (1980). Chemical identification of a glycosphingolipid receptor for *Escherichia coli* attaching to human urinary tract epithelial cells and agglutinating human erythrocytes. *FEMS Microbiol Lett* **8**, 127-134.

Leffler, H. & Svanborg-Eden, C. (1981). Glycolipid receptors for uropathogenic *Escherichia coli* on human erythrocytes and uroepithelial cells. *Infect Immun* **34**, 920-929.

Li, H., Qian, L., Chen, Z., Thibault, D., Liu, G., Liu, T. & Thanassi, D. G. (2004). The outer membrane usher forms a twin-pore secretion complex. *J Mol Biol* **344**, 1397-1407.

Li, X., Rasko, D. A., Lockatell, C. V., Johnson, D. E. & Mobley, H. L. (2001). Repression of bacterial motility by a novel fimbrial gene product. *EMBO J* **20**, 4854-4862.

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

Lindberg, F. P., Lund, B. & Normark, S. (1984). Genes of pyelonephritogenic *E. coli* required for digalactoside-specific agglutination of human cells. *EMBO J* **3**, 1167-1173.

Linder, H., Engberg, I., Baltzer, I. M., Jann, K. & Svanborg-Eden, C. (1988). Induction of inflammation by *Escherichia coli* on the mucosal level: requirement for adherence and endotoxin. *Infect Immun* **56**, 1309-1313.

Linder, H., Engberg, I., Hoschutzky, H., Mattsby-Baltzer, I. & Svanborg, C. (1991). Adhesion-dependent activation of mucosal interleukin-6 production. *Infect Immun* **59**, 4357-4362.

Lloyd, A. L., Rasko, D. A. & Mobley, H. L. T. (2007). Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. *J Bacteriol* **189**, 3532-3546.

Low, D., David, V., Lark, D., Schoolnik, G. & Falkow, S. (1984). Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infections. *Infect Immun* **43**, 353-358.

Low, D., Robinson, E. N., Jr., McGee, Z. A. & Falkow, S. (1987). The frequency of expression of pyelonephritis-associated pili is under regulatory control. *Mol Microbiol* **1**, 335-346.

Lund, B., Lindberg, F., Marklund, B. I. & Normark, S. (1987). The PapG protein is the α-D-galactopyranosyl-(1→4)-β-D-galactopyranose-binding adhesin of uropathogenic *Escherichia coli*. *PNAS* **84**, 5898-5902.

Lundstedt, A.-C., Leijonhufvud, I., Ragnarsdottir, B., Karpman, D., Andersson, B. & Svanborg, C. (2007). Inherited susceptibility to acute pyelonephritis: a family study of urinary tract infection. *J Infect Dis* **195**, 1227-1234.

Manning, S. D., Zhang, L., Foxman, B., Spindler, A., Tallman, P. & Marrs, C. F. (2001). Prevalence of known P-fimbrial G alleles in *Escherichia coli* and identification of a new adhesin class. *Clin Diagn Lab Immunol* **8**, 637-640.

Marklund, B. I., Tennent, J. M., Garcia, E. & other authors (1992). Horizontal gene transfer of the *Escherichia coli pap* and *prs* pili operons as a mechanism for the development of tissue-specific adhesive properties. *Mol Microbiol* **6**, 2225-2242.

Maroncle, N. M., Sivick, K. E., Brady, R., Stokes, F.-E. & Mobley, H. L. T. (2006). Protease activity, secretion, cell entry, cytotoxicity, and cellular targets of secreted autotransporter toxin of uropathogenic *Escherichia coli*. *Infect Immun* **74**, 6124-6134.

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.

Matsumoto, T., Takahashi, K., Manabe, N., Iwatsubo, E. & Kawakami, Y. (2001). Urinary tract infection in neurogenic bladder. *Int J Antimicrob Ag* **17**, 293-297.

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.

Melkerson-Watson, L. J., Rode, C. K., Zhang, L., Foxman, B. & Bloch, C. A. (2000). Integrated genomic map from uropathogenic *Escherichia coli* J96. *Infect Immun* **68**, 5933-5942.

Middendorf, B., Hochhut, B., Leipold, K., Dobrindt, U., Blum-Oehler, G. & Hacker, J. (2004). Instability of pathogenicity islands in uropathogenic *Escherichia coli* 536. *J Bacteriol* **186**, 3086-3096.

Mills, M., Meysick, K. C. & O'Brien, A. D. (2000). Cytotoxic necrotizing factor type 1 of uropathogenic *Escherichia coli* kills cultured human uroepithelial 5637 cells by an apoptotic mechanism. *Infect Immun* **68**, 5869-5880.

Minshew, B. H., Jorgensen, J., Counts, G. W. & Falkow, S. (1978). Association of hemolysin production, hemagglutination of human erythrocytes, and virulence for chicken embryos of extraintestinal *Escherichia coli* isolates. *Infect Immun* **20**, 50-54.

Mobley, H. L., Green, D. M., Trifillis, A. L., Johnson, D. E., Chippendale, G. R., Lockatell, C. V., Jones, B. D. & Warren, J. W. (1990). Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. *Infect Immun* **58**, 1281-1289.

Moch, T., Hoschutzky, H., Hacker, J., Kroncke, K.-D. & Jann, K. (1987). Isolation and characterization of the α-sialyl-β-2,3-galactosyl-specific adhesin from fimbriated *Escherichia coli*. *PNAS* **84**, 3462-3466.

Montgomerie, J. Z., Bindereif, A., Neilands, J. B., Kalmanson, G. M. & Guze, L. B. (1984). Association of hydroxamate siderophore (aerobactin) with *Escherichia coli* isolated from patients with bacteremia. *Infect Immun* **46**, 835-838.

Moreland, J., Gramada, A., Buzko, O., Zhang, Q. & Bourne, P. (2005). The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **6**, 21.

Morschhauser, J., Hoschutzky, H., Jann, K. & Hacker, J. (1990). Functional analysis of the sialic acid-binding adhesin SfaS of pathogenic *Escherichia coli* by site-specific mutagenesis. *Infect Immun* **58**, 2133-2138.

Morschhauser, J., Uhlin, B.-E. & Hacker, J. (1993). Transcriptional analysis and regulation of the *sfa* determinant coding for S fimbriae of pathogenic *Escherichia coli* strains. *Mol Gen Genet MGG* **238**, 97-105.

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.

Mu, X. Q., Jiang, Z. G. & Bullitt, E. (2005). Localization of a critical interface for helical rod formation of bacterial adhesion P-pili. *J Mol Biol* **346**, 13-20.

Mulvey, M. A., Lopez-Boado, Y. S., Wilson, C. L., Roth, R., Parks, W. C., Heuser, J. & Hultgren, S. J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* **282**, 1494-1497.

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.

Murray, M. G. & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**, 4321-4325.

Navas, E. L., Venegas, M. F., Duncan, J. L., Anderson, B. E., Kanerva, C., Chmiel, J. S. & Schaeffer, A. J. (1994). Blood group antigen expression on vaginal cells and mucus in women with and without a history of urinary tract infections. *J Urol* **152**, 345-349.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418-426.

Ng, T. W., Akman, L., Osisami, M. & Thanassi, D. G. (2004). The usher N Terminus is the initial targeting site for chaperone-subunit complexes and participates in subsequent pilus biogenesis events. *J Bacteriol* **186**, 5321-5331.

Nicholson, B. & Low, D. (2000). DNA methylation-dependent regulation of Pef expression in *Salmonella typhimurium*. *Mol Microbiol* **35**, 728-742.

Nilsson, L. M., Thomas, W. E., Trintchina, E., Vogel, V. & Sokurenko, E. V. (2006). Catch bond-mediated adhesion without a shear threshold: trimannose versus monomannose interactions with the FimH adhesin of *Escherichia coli*. *J Biol Chem*, M511496200.

Nilsson, P., Naureckiene, S. & Uhlin, B. E. (1996). Mutations affecting mRNA processing and fimbrial biogenesis in the *Escherichia coli pap* operon. *J Bacteriol* **178**, 683-690.

Nishiyama, M., Vetsch, M., Puorger, C., Jelesarov, I. & Glockshuber, R. (2003). Identification and characterization of the chaperone-subunit complex-binding domain from the type 1 pilus assembly platform FimD. *J Mol Biol* **330**, 513-525.

Nishiyama, M., Horst, R., Eidam, O. & other authors (2005). Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. *EMBO J* **24**, 2075-2086.

Norris, T. L. & Baumler, A. J. (1999). Phase variation of the *lpf* operon is a mechanism to evade cross-immunity between *Salmonella* serotypes. *PNAS* **96**, 13393-13398.

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.

Nou, X., Braaten, B., Kaltenbach, L. & Low, D. A. (1995). Differential binding of Lrp to two sets of *pap* DNA binding sites mediated by Pap I regulates Pap phase variation in *Escherichia coli*. *EMBO J* **14**, 5785-5797.

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.

Nowicki, B., Rhen, M., Vaisanen-Rhen, V., Pere, A. & Korhonen, T. K. (1985). Kinetics of phase variation between S and type-1 fimbriae of *Escherichia coli*. *FEMS Microbiol Lett* **28**, 237-242.

Oelschlaeger, T. A., Dobrindt, U., and J. Hacker (2002). Pathogenicity islands of uropathogenic *E. coli* and the evolution of virulence. *Int J Antimicrob Ag* **19**, 517-521.

Ofek, I., Mirelman, D. & Sharon, N. (1977). Adherence of *Escherichia coli* to human mucosal cells mediated by mannose receptors. *Nature* **265**, 623-625.

O'Hanley, P., Lark, D., Falkow, S. & Schoolnik, G. (1985). Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice. Gal-Gal pili immunization prevents *Escherichia coli* pyelonephritis in the BALB/c mouse model of human pyelonephritis. *J Clin Invest* **75**, 347-360.

Orskov, I., Orskov, F. (1990). Serological classification of fimbriae. *Curr Top Microbiol* **151**, 70-90.

Ott, M., Hacker, J., Schmoll, T., Jarchau, T., Korhonen, T. K. & Goebel, W. (1986). Analysis of the genetic determinants coding for the S-fimbrial adhesin (*sfa*) in different *Escherichia coli* strains causing meningitis or urinary tract infections. *Infect Immun* **54**, 646-653.

Ott, M., Schmoll, T., Goebel, W., Van Die, I. & Hacker, J. (1987). Comparison of the genetic determinant coding for the S-fimbrial adhesin (*sfa*) of *Escherichia coli* to other chromosomally encoded fimbrial determinants. *Infect Immun* **55**, 1940-1943.

Ott, M., Hoschutzky, H., Jann, K., Van Die, I. & Hacker, J. (1988). Gene clusters for S fimbrial adhesin (*sfa*) and F1C fimbriae (*foc*) of *Escherichia coli*: comparative aspects of structure and function. *J Bacteriol* **170**, 3983-3990.

Otto, K. & Silhavy, T. J. (2002). Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *PNAS* **99**, 2287-2292.

Parham, N. J., Pollard, S. J., Chaudhuri, R. R. & other authors (2005). Prevalence of pathogenicity island II_{CFT073} genes among extraintestinal clinical isolates of *Escherichia coli*. *J Clin Microbiol* **43**, 2425-2434.

Pecha, B., Low, D. & O'Hanley, P. (1989). Gal-Gal pili vaccines prevent pyelonephritis by piliated *Escherichia coli* in a murine model. Single-component Gal-Gal pili vaccines prevent pyelonephritis by homologous and heterologous piliated *E. coli* strains. *J Clin Invest* **83**, 2102-2108.

Pere, A., Leinonen, M., Vaisanen-Rhen, V., Rhen, M. & Korhonen, T. K. (1985). Occurrence of type-1C fimbriae on *Escherichia coli* strains isolated from human extraintestinal infections. *J Gen Microbiol* **131**, 1705-1711.

Pere, A., Nowicki, B., Saxen, H., Siitonen, A., Korhonen, T. K., Rhen, M. & Vaisanen-Rhen, V. (1987). Expression of P, type-1, and type-1C fimbriae of *Escherichia coli* in the urine of patients with acute urinary tract infection. *J Infect Dis* **156**, 567-574.

Pere, A. (1989). Serological study of *Escherichia coli* fimbriae associated with human extraintestinal infections. Ph.D. Thesis, Department of General Microbiology, University of Helsinki.

Perna, N. T., Plunkett, G., Burland, V. & other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529-533.

Peterson, S. N. & Reich, N. O. (2006). GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence *pap* expression. *J Mol Biol* **355**, 459-472.

Plos, K., Hull, S. I., Hull, R. A., Levin, B. R., Orskov, I., Orskov, F. & Svanborg-Eden, C. (1989). Distribution of the P-associated-pilus (*pap*) region among *Escherichia coli* from natural sources: evidence for horizontal gene transfer. *Infect Immun* **57**, 1604-1611.

Plos, K., Carter, T., Hull, S., Hull, R. & Svanborg Eden, C. (1990). Frequency and organization of *pap* homologous DNA in relation to clinical origin of uropathogenic *Escherichia coli*. *J Infect Dis* **161**, 518-524.

Plos, K., Lomberg, H., Hull, S., Johansson, I. & Svanborg, C. (1991). *Escherichia coli* in patients with renal scarring: genotype and phenotype of Gal α1-4Galβ-, Forssman- and mannose-specific adhesins. *Pediatr Infect Dis J* **10**, 15-19.

Porter, M. E., Mitchell, P., Roe, A. J., Free, A., Smith, D. G. E. & Gally, D. L. (2004). Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. *Mol Microbiol* **54**, 1117-1133.

Pouttu, R., Westerlund-Wikstrom, B., Lang, H. & other authors (2001). *matB*, a common fimbrillin gene of *Escherichia coli*, expressed in a genetically conserved, virulent clonal group. *J Bacteriol* **183**, 4727-4736.

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.

Rasko, D. A., Phillips, J. A., Li, X. & Mobley, H. L. (2001). Identification of DNA sequences from a second pathogenicity island of uropathogenic *Escherichia coli* CFT073: probes specific for uropathogenic populations. *J Infect Dis* **184**, 1041-1049.

Ritter, A., Blum, G., Emody, L., Kerenyi, M., Bock, A., Neuhieri, 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, J. A., Marklund, B. I., Ilver, D. & other authors (1994). The Gal(α1-4)Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *PNAS* **91**, 11889-11893.

Roberts, J. A., Kaack, M. B., Baskin, G., Chapman, M. R., Hunstad, D. A., Pinkner, J. S. & Hultgren, S. J. (2004). Antibody responses and protection from pyelonephritis following vaccination with purified *Escherichia coli* PapDG protein. *J Urol* **171**, 1682-1685.

Roe, A. J., Naylor, S. W., Spears, K. J. & other authors (2004). Co-ordinate single-cell expression of LEE4- and LEE5-encoded proteins of *Escherichia coli* O157:H7. *Mol Microbiol* **54**, 337-352.

Ronald, A. & Ludwig, E. (2001). Urinary tract infections in adults with diabetes. *Int J Antimicrob Ag* **17**, 287-292.

Roos, V. & Klemm, P. (2006). Global gene expression profiling of the asymptomatic bacteriuria *Escherichia coli* strain 83972 in the human urinary tract. *Infect Immun* **74**, 3565-3575.

Roos, V., Nielsen, E. M. & Klemm, P. (2006a). Asymptomatic bacteriuria *Escherichia coli* strains: adhesins, growth and competition. *FEMS Microbiol Lett* **262**, 22-30.

Roos, V., Schembri, M. A., Ulett, G. C. & Klemm, P. (2006b). Asymptomatic bacteriuria *Escherichia coli* strain 83972 carries mutations in the *foc* locus and is unable to express F1C fimbriae. *Microbiology* **152**, 1799-1806.

Roos, V., Ulett, G. C., Schembri, M. A. & Klemm, P. (2006c). The asymptomatic bacteriuria *Escherichia coli* strain 83972 outcompetes uropathogenic *E. coli* strains in human urine. *Infect Immun* **74**, 615-624.

Rose, R. E. (1988). The nucleotide sequence of pACYC184. *Nucl Acids Res* **16**, 355.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Rubin, R. H., Shapiro, E. D., Andriole, V. T., Davis, R. J. & Stamm, W. E. (1992). Evaluation of new anti-infective drugs for the treatment of urinary tract infection. Infectious Diseases Society of America and the Food and Drug Administration. *Clin Infect Dis* **15 Suppl 1**, S216-227.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Press.

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., 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.

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.

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.

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.

Schembri, M. A., Sokurenko, E. V. & Klemm, P. (2000). Functional flexibility of the FimH adhesin: insights from a random mutant library. *Infect Immun* **68**, 2638-2646.

Schembri, M. A. & Klemm, P. (2001). Coordinate gene regulation by fimbriae-induced signal transduction. *EMBO J* **20**, 3074-3081.

Schembri, M. A., Dalsgaard, D. & Klemm, P. (2004). Capsule shields the function of short bacterial adhesins. *J Bacteriol* **186**, 1249-1257.

Schembri, M. A., Blom, J., Krogfelt, K. A. & Klemm, P. (2005). Capsule and fimbria interaction in *Klebsiella pneumoniae*. *Infect Immun* **73**, 4626-4633.

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., Lorenz, R. G., Hultgren, S. J. (2002). Effect of trimethoprim-sulfamethoxazole on recurrent bacteriuria and bacterial persistence in mice infected with uropathogenic *Escherichia coli*. *Infect Immun* **70**, 7042-7049.

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.

Schmoll, T., Morschhauser, J., Ott, M., Ludwig, B., van Die, I. & Hacker, J. (1990). Complete genetic organization and functional aspects of the *Escherichia coli* S fimbrial adhesin determinant: nucleotide sequence of the genes *sfa B, C, D, E, F*. *Microb Pathogenesis* **9**, 331-343.

Schonwald, S., Begovac, J. & Skerk, V. (1999). Urinary tract infections in HIV disease. *Int J Antimicrob Ag* **11**, 309-311.

Schwan, W. R., Lee, J. L., Lenard, F. A., Matthews, B. T. & Beck, M. T. (2002). Osmolarity and pH growth conditions regulate *fim* gene transcription and type 1 pilus expression in uropathogenic *Escherichia coli*. *Infect Immun* **70**, 1391-1402.

Schwan, W. R., Beck, M. T., Hultgren, S. J., Pinkner, J., Woolever, N. L. & Larson, T. (2005). Down-regulation of the *kps* region 1 capsular assembly operon following attachment of *Escherichia coli* type 1 fimbriae to D-mannose receptors. *Infect Immun* **73**, 1226-1231.

Shahin, R. D., Engberg, I., Hagberg, L. & Svanborg Eden, C. (1987). Neutrophil recruitment and bacterial clearance correlated with LPS responsiveness in local gram-negative infection. *J Immunol* **138**, 3475-3480.

Shu Kin So, S. & Thanassi, D. G. (2006). Analysis of the requirements for pilus biogenesis at the outer membrane usher and the function of the usher C-terminus. *Mol Microbiol* **60**, 364-375.

Sims, J. B., de Lemos, J. A., Maewal, P., Warner, J. J., Peterson, G. E. & McGuire, D. K. (2005). Urinary tract infection in patients with acute coronary syndrome: A potential systemic inflammatory connection. *Am Heart J* **149**, 1062-1065.

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

Snyder, J. A., Lloyd, A. L., Lockatell, C. V., Johnson, D. E. & Mobley, H. L. T. (2006). Role of phase variation of type 1 fimbriae in a uropathogenic *Escherichia coli* cystitis isolate during urinary tract infection. *Infect Immun* **74**, 1387-1393.

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., Courtney, H., Maslow, J., Siitonen, A. & Hasty, D. (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., 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., 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. *PNAS* **95**, 8922-8926.

Sokurenko, E. V., Feldgarden, M., Trintchina, E., Weissman, S. J., Avagyan, S., Chattopadhyay, S., Johnson, J. R. & Dykhuizen, D. E. (2004). Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in *Escherichia coli*. *Mol Biol Evol* **21**, 1373-1383.

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

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.

Stenqvist, K., Sandberg, T., Lidin-Janson, G., Orskov, F., Orskov, I. & Svanborg-Eden, C. (1987). Virulence factors of *Escherichia coli* in urinary isolates from pregnant women. *J Infect Dis* **156**, 870-877.

Stromberg, N., Marklund, B. I., Lund, B., Ilver, D., Hamers, A., Gaastra, W., Karlsson, K. A. & Normark, S. (1990). Host-specificity of uropathogenic *Escherichia coli* depends on differences in binding specificity to Galα 1-4Gal-containing isoreceptors. *EMBO J* **9**, 2001-2010.

Stromberg, N., Nyholm, P., Pascher, I. & Normark, S. (1991). Saccharide orientation at the cell surface affects glycolipid receptor function. *PNAS* **88**, 9340-9344.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60-89.

Sung, M. A., Fleming, K., Chen, H. A. & Matthews, S. (2001). The solution structure of PapGII from uropathogenic *Escherichia coli* and its recognition of glycolipid receptors. *EMBO Rep* **2**, 621-627.

Sussman, M. (1997). *Escherichia coli: Mechanisms of virulence*. Cambridge, England: Cambridge University Press.

Svanborg, C., Frendeus, B., Godaly, G., Hang, L., Hedlund, M. & Wachtler, C. (2001). Toll-like receptor signaling and chemokine receptor expression influence the severity of urinary tract infection. *J Infect Dis* **183 Suppl 1**, S61-65.

Svanborg-Eden, C., Hanson, L. A., Jodal, U., Lindberg, U. & Akerlund, A. S. (1976). Variable adherence to normal human urinary-tract epithelial cells of *Escherichia coli* strains associated with various forms of urinary-tract infection. *Lancet* **1**, 490-492.

Svanborg-Eden, C. & Hansson, H. A. (1978). *Escherichia coli* pili as possible mediators of attachment to human urinary tract epithelial cells. *Infect Immun* **21**, 229-237.

Svanborg-Eden, C., Kulhavy, R., Marild, S., Prince, S. J. & Mestecky, J. (1985). Urinary immunoglobulins in healthy individuals and children with acute pyelonephritis. *Scand J Immunol* **21**, 305-313.

Svanborg-Eden, C., Hagberg, L., Hull, R., Hull, S., Magnusson, K. E. & Ohman, L. (1987). Bacterial virulence versus host resistance in the urinary tracts of mice. *Infect Immun* **55**, 1224-1232.

Svensson, M., Lindstedt, R., Radin, N. S. & Svanborg, C. (1994). Epithelial glucosphingolipid expression as a determinant of bacterial adherence and cytokine production. *Infect Immun* **62**, 4404-4410.

Swenson, D., Bukanov, N., Berg, D. & Welch, R. (1996). Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. *Infect Immun* **64**, 3736-3743.

Thanassi, D. G., Saulino, E. T. & Hultgren, S. J. (1998a). The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr Opin Microbiol* **1**, 223-231.

Thanassi, D. G., Saulino, E. T., Lombardo, M.-J., Roth, R., Heuser, J. & Hultgren, S. J. (1998b). The PapC usher forms an oligomeric channel: Implications for pilus biogenesis across the outer membrane. *PNAS* **95**, 3146-3151.

Thanassi, D. G. & Hultgren, S. J. (2000). Assembly of complex organelles: pilus biogenesis in Gram-negative bacteria as a model system. *Methods* **20**, 111-126.

Thanassi, D. G., Stathopoulos, C., Karkal, A. & Li, H. (2005). Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of gram-negative bacteria (review). *Mol Membr Biol* **22**, 63-72.

Thompson, J. D., Gibson,T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, **25,** 4876-4882.

Trautner, B., Darouiche, R., Hull, R., Hull, S. & Thornby J. (2002). Pre-inoculation of urinary catheters with *Escherichia coli* 83972 inhibits catheter colonization by *Enterococcus faecalis*. *J Urol* **167**, 375-379.

Trautner, B. W., Hull, R. A. & Darouiche, R. O. (2003). *Escherichia coli* 83972 inhibits catheter adherence by a broad spectrum of uropathogens. *Urology* **61**, 1059-1062.

Tseng, C. C., Huang, J. J., Wang, M. C., Wu, A. B., Ko, W. C., Chen, W. C. & Wu, J. J. (2007). PapG II adhesin in the establishment and persistence of *Escherichia coli* infection in mouse kidneys. *Kidney Int* doi:10.1038/sj.ki.5002111.

Tullus, K., Kuhn, I., Orskov, I., Orskov, F. & Mollby, R. (1992). The importance of P and type 1 fimbriae for the persistence of *Escherichia coli* in the human gut. *Epidemiol Infect* **108**, 415-421.

Ulett, G. C., Webb, R. I. & Schembri, M. A. (2006). Antigen-43-mediated autoaggregation impairs motility in *Escherichia coli*. *Microbiology* **152**, 2101-2110.

Ulett, G. C., Valle, J., Beloin, C., Sherlock, O., Ghigo, J.-M. & Schembri, M. A. (2007). Functional analysis of Antigen 43 in uropathogenic *Escherichia coli* reveals a role in long term persistence in the urinary tract. *Infect Immun* doi:10.1128/IAI.01952-06.

van der Woude, M. W., Braaten, B. A. & Low, D. A. (1992). Evidence for global regulatory control of pilus expression in *Escherichia coli* by Lrp and DNA methylation: model building based on analysis of *pap*. *Mol Microbiol* **6**, 2429-2435.

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., Kaltenbach, L. S. & Low, D. A. (1995). Leucine-responsive regulatory protein plays dual roles as both an activator and a repressor of the *Escherichia coli pap* fimbrial operon. *Mol Microbiol* **17**, 303-312.

van der Woude, M., Braaten, B. & Low, D. (1996). Epigenetic phase variation of the *pap* operon in *Escherichia coli*. *Trends Microbiol* **4**, 5-9.

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

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

Verger, D., Miller, E., Remaut, H., Waksman, G. & Hultgren, S. (2006). Molecular mechanism of P pilus termination in uropathogenic *Escherichia coli*. *EMBO Rep* **7**, 1228-1232.

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.

Vetsch, M., Erilov, D., Moliere, N., Nishiyama, M., Ignatov, O. & Glockshuber, R. (2006). Mechanism of fibre assembly through the chaperone-usher pathway. *EMBO Rep* **7**, 734-738.

Virkola, R., Westerlund, B., Holthofer, H., Parkkinen, J., Kekomaki, M. & Korhonen, T. K. (1988). Binding characteristics of *Escherichia coli* adhesins in human urinary bladder. *Infect Immun* **56**, 2615-2622.

Vitagliano, L., Ruggiero, A., Pedone, C. & Berisio, R. (2007). A Molecular dynamics study of pilus subunits: insights into pilus biogenesis. *J Mol Biol* **367**, 935-941.

Wallecha, A., Correnti, J., Munster, V. & van der Woude, M. (2003). Phase variation of Ag43 is independent of the oxidation state of OxyR. *J Bacteriol* **185**, 2203-2209.

Warren, J. W., Mobley, H. L. T., Hebel, J. R. & Trifillis, A. L. (1995). Cytolethality of hemolytic *Escherichia coli* to primary human renal proximal tubular cell cultures obtained from different donors. *Urology* **45**, 706-710.

Warren, J. W. (1996). Clinical presentations and epidemiology of urinary tract infections. In *Urinary Tract Infections*. Edited by H. L. T. Mobley and J. W. Warren. Washington, DC: American Society for Microbiology Press.

Warren, J. W. (2001). Catheter-associated urinary tract infections. *Int J Antimicrob Ag* **17**, 299-303.

Weiser, J. N. & Gotschlich, E. C. (1991). Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of *Escherichia coli* K-1. *Infect Immun* **59**, 2252-2258.

Weissman, S. J., Chattopadhyay, S., Aprikian, P. & other authors (2006). Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. *Mol Microbiol* **59**, 975-988.

Welch, R. A., Hull, R. & Falkow, S. (1983). Molecular cloning and physical characterization of a chromosomal hemolysin from *Escherichia coli*. *Infect Immun* **42**, 178-186.

Welch, R. A., Burland, V., Plunkett, G., 3rd & other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *PNAS* **99**, 17020-17024.

Weyand, N. J. & Low, D. A. (2000). Regulation of Pap phase variation. Lrp is sufficient for the establishment of the phase off *pap* DNA methylation pattern and repression of *pap* transcription *in vitro*. *J Biol Chem* **275**, 3192-3200.

Weyand, N. J., Braaten, B. A., van der Woude, M., Tucker, J. & Low, D. A. (2001). The essential role of the promoter-proximal subunit of CAP in *pap* phase variation: Lrp- and helical phase-dependent activation of *papBA* transcription by CAP from -215. *Mol Microbiol* **39**, 1504-1522.

White-Ziegler, C. A., Blyn, L. B., Braaten, B. A. & Low, D. A. (1990). Identification of an *Escherichia coli* genetic locus involved in thermoregulation of the *pap* operon. *J Bacteriol* **172**, 1775-1782.

White-Ziegler, C. A., Angus Hill, M. L., Braaten, B. A., van der Woude, M. W. & Low, D. A. (1998). Thermoregulation of *Escherichia coli pap* transcription: H-NS is a temperature-dependent DNA methylation blocking factor. *Mol Microbiol* **28**, 1121-1137.

White-Ziegler, C. A., Villapakkam, A., Ronaszeki, K. & Young, S. (2000). H-NS controls *pap* and *daa* fimbrial transcription in *Escherichia coli* in response to multiple environmental cues. *J Bacteriol* **182**, 6391-6400.

Willenbrock, H., Petersen, A., Sekse, C., Kiil, K., Wasteson, Y. & Ussery, D. W. (2006). Design of a seven-genome *Escherichia coli* microarray for comparative genomic profiling. *J Bacteriol* **188**, 7713-7721.

Williams, D. N., Lund, M. E. & Blazevic, D. J. (1976). Significance of urinary isolates of coagulase-negative Micrococcaceae. *J Clin Microbiol* **3**, 556-559.

Wullt, B., Connell, H., Rollano, P., Mansson, W., Colleen, S. & Svanborg, C. (1998). Urodynamic factors influence the duration of *Escherichia coli* bacteriuria in deliberately colonized cases. *J Urol* **159**, 2057-2062.

Wullt, B., Bergsten, G., Connell, H., Rollano, P., Gebretsadik, N., Hull, R. & Svanborg, C. (2000). P fimbriae enhance the early establishment of *Escherichia coli* in the human urinary tract. *Mol Microbiol* **38**, 456-464.

Wullt, B., Bergsten, G., Connell, H., Rollano, P., Gebratsedik, N., Hang, L. & Svanborg, C. (2001a). P-fimbriae trigger mucosal responses to *Escherichia coli* in the human urinary tract. *Cell Microbiol* **3**, 255-264.

Wullt, B., Bergsten, G., Samuelsson, M., Gebretsadik, N., Hull, R. & Svanborg, C. (2001b). The role of P fimbriae for colonization and host response induction in the human urinary tract. *J Infect Dis* **183**, S43-46.

Wullt, B., Bergsten, G., Samuelsson, M. & Svanborg, C. (2002). The role of P fimbriae for *Escherichia coli* establishment and mucosal inflammation in the human urinary tract. *Int J Antimicrob Agents* **19**, 522-538.

Wullt, B., Bergsten, G., Fischer, H. & other authors (2003). The host response to urinary tract infection. *Infect Dis Clin North Am* **17**, 279-301.

Xia, Y., Forsman, K., Jass, J. & Uhlin, B. E. (1998). Oligomeric interaction of the PapB transcriptional regulator with the upstream activating region of pili adhesin gene promoters in *Escherichia coli*. *Mol Microbiol* **30**, 513-523.

Xia, Y. & Uhlin, B. E. (1999). Mutational analysis of the PapB transcriptional regulator in *Escherichia coli*. Regions important for DNA binding and oligomerization. *J Biol Chem* **274**, 19723-19730.

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.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Zhang, D., Zhang, G., Hayden, M. S., Greenblatt, M. B., Bussey, C., Flavell, R. A. & Ghosh, S. (2004). A Toll-like receptor that prevents infection by uropathogenic bacteria. *Science* **303**, 1522-1526.

Zorc, J. J., Kiddoo, D. A. & Shaw, K. N. (2005). Diagnosis and management of pediatric urinary tract infections. *Clin Microbiol Rev* **18**, 417-422.