Molecular mechanisms underlying adhesion and regulation of virulence of human Enterotoxigenic *Escherichia coli*

by

Vipin Madhavan Thekke Palasseri
B.Tech, MSc

School of Medical Science, Griffith University, Gold Coast Campus, Queensland, Australia

Submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

14th Dec 2015

Abstract

Enterotoxigenic *Escherichia coli* (ETEC) is the key etiological agent of traveller's diarrhea. It accounts for 200 million cases of diarrhea and 380,000 deaths annually. ETEC uses proteinaceous appendages called pili in order to attach with host intestine, followed by secretion of enterotoxins leading to profuse diarrhea.

Among various ETEC pili that are characterised till date, CFA/I is the archetype of a class of pili called class 5 (α pili) pili. CFA/I is composed of two types of protein subunits. It consists of a cylindrical stalk of CfaB subunits and CfaE at the tip. Both intra and inter subunit bonding takes place via donor strand complementation. CfaE is widely accepted as an adhesin molecule that binds to human intestine and human erythrocytes. Also CfaE is regarded as nucleating molecule for pili synthesis (Li et al., 2007, Baker et al., 2009). An AraC family of transcriptional activator (AFTR), Rns is known to regulate CFA/I synthesis and another half of known ETEC pili. Present study focuses on three aspects of ETEC virulence.

First, Although CfaE is considered to be the key adhesin, an existing hypothesis claims that CfaB could independently bind to glycosphingolipid receptors, including asialo-GM1. Furthermore, it claims that CFA/I pili could be synthesised without CfaE (Jansson et al., 2006). Our studies indicate that, contrary to the existing model, CfaB does not bind asialo-GM1 independently of CfaE. Neither isolated CfaB subunits nor CfaB assembled into pili bind to asialo-GM1. Instead, we demonstrated that binding activity for asialo-GM1 resides in CfaE and is essential for pilus binding to intestinal epithelial cells. We concluded that the binding activities of CFA/I pili for asialo-GM1, erythrocytes and intestinal cells are

inseparable, depend on the same amino acid residues in CfaE and therefore represent the same or very similar binding mechanisms.

Secondly, role of Rns in regulating the expression of non-pilus genes in ETEC genome was studied. Rns, apart from regulating ETEC pili are known to regulate few genes in ETEC genome (Bodero et al., 2008). Our present study showed that Rns regulate a large number of genes belonging to numerous metabolic pathways. Rns also regulate some of the key virulence factors, *etp*B,*etp*A and *elt*A. Furthermore, we found that Rns binds to upstream of etpBAC operon and upregulates the expression of etpB and *etp*A.

Third, role of Rns regulated factors on host immune response was studied using Caco-2 as model. No significant role of any host related factors was found. At the same time, the study was able to show cytokine responses of IL-9, and IL-13 against ETEC infection which were not reported previously.

Statement of originality

This work has not been previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

(Signed)_____
Vipin Madhavan Thekke Palasseri

Table of Contents

Acknowledgements

First and foremost, I sincerely thank my principal supervisor Dr Harry Sakellaris for being a guiding light throughout this span of nearly four years. He gave me immense amount of freedom in adopting different experimental methods in spite of failures at times. He has been a great teacher and a sincere critic throughout my PhD life.

Secondly, I would like to thank my associate supervisor, Prof Glen Charles Ulett for providing much of the research amenities. He has always seen and treated me as part of his own group. His ideas and views were quite inspiring and productively useful during my work. I also thank him for being a perennial source of lollies, which he has generously supplied although these years.

I also thank my other associate supervisor Prof Victoria Korolik for providing me access to Glycomics and allowing me to use her lab facilities during protein purification.

My work involved collaboration with people belonging to other labs and institutions. In this regard, I show my sincere gratitude to Yun Kit, Jason Steen and Philip Hugenholtz, The University of Queensland, Dr Tim Perkins, University of Western Australia and Dr Steve Bottomely , Monash University.

I thank my present and past lab members, Ben, Matt, Deepak, Sophie, Josh, Michelle and Alison.

My word of gratitude to members of Prof Saso Ivanosky group, Saeed and Sudheesh.

I thank Griffith University for supporting my PhD research through GUIPRS, GUPRS and CAPRS scholarships.

Last but not least, I thank my parents and almighty for all the help and support till this day.

List of abbreviations

AFTR	Ara-C family of transcriptional activator
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
Caco-2	Colorectal adenocarcinoma
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
CDS	Coding sequences
CF	Colonisation factor
CFA	Colonisation factor antigen
cGMP	Cyclic guanosine monophosphate
ChIP	Chromatin Immunoprecipitation
CRISPR	Clustered regularly-interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
ERK	Extracellular signal-regulated kinase
FBS	Fetal bovine serum
HCl	Hydrochloric acid
IL	Interleukin
IP	Immunoprecipitation
Kb	Kilo bases
kDa	Kilo Dalton

MAPK	Mitogen-activated protein kinases
Mb	Mega bases
MBP	Maltose binding protein
MOI	Multiplicity of infection
MRHA	Mannose-resistant haemagglutination
mRNA	Messenger RNA
mTOR	Mechanistic target of rapamycin
NaCl	Sodium chloride
Ni-NTA	Nickel-Nitrilotriacetic acid
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PKA	Protein kinase A
qRT-PCR	Quantitative real-time PCR
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute medium
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
TBS	Tris-buffered saline
TGF	Transforming growth factor
TLR	Toll like receptor
TNF	Tumour necrosis factor
tRNA	Transfer RNA
VEGF	Vascular endothelial growth factor
XRE	Xenobiotic response element

List of figures

List of Tables

Acknowledgement of published papers included in this thesis

Section 9.1 of the Griffith University Code for the Responsible Conduct of Research ("Criteria for Authorship"), in accordance with Section 5 of the Australian Code for the Responsible Conduct of Research, states:

> To be named as an author, a researcher must have made a substantial scholarly contribution to the creative or scholarly work that constitutes the research output, and be able to take public responsibility for at least that part of the work they contributed. Attribution of authorship depends to some extent on the discipline and publisher policies, but in all cases, authorship must be based on substantial contributions in a combination of one or more of:
>
> - Conception and design of the research project
> - Analysis and interpretation of research data
> - Drafting or making significant parts of the creative or scholarly work or critically revising it so as to contribute significantly to the final output

Section 9.3 of the Griffith University Code ("Responsibilities of Researchers"), in accordance with Section 5 of the Australian Code, states:

> Researchers are expected to:
>
> - Offer authorship to all people, including research trainees, who meet the criteria for authorship listed above, but only those people.
> - Accept or decline offers of authorship promptly in writing.
> - Include in the list of authors only those who have accepted authorship
> - Appoint one author to be the executive author to record authorship and manage correspondence about the work with the publisher and other interested parties.
> - Acknowledge all those who have contributed to the research, facilities or materials but who do not qualify as authors, such as research assistants, technical staff, and advisors on cultural or community knowledge. Obtain written consent to name individuals.

Included in this thesis are papers in Chapters 1 and 5 which are co-authored with other researchers. My contribution to each co-authored paper is outlined at the front of the relevant chapter. The bibliographic details / status for these papers including all authors, are:

Appearing in Chapter 1:

T.P.Vipin Madhavan and Harry Sakellaris**.** Colonization Factors of Enterotoxigenic *Escherichia coli*. Advances in applied microbiology. 2015;90c:155-97. Epub 2015/01/18.

Appearing in Chapter 5:

T.P.Vipin Madhavan, Jason. A. Steen, Philip Hugenholtz and Harry Sakellaris. Genome Sequence of Enterotoxigenic *E. coli* (ETEC) strain 06:H16: CFA/II strain B2C. Genome Announc. 2(2):e00247-14.

Appearing in Abstract and Chapter 4:

Madhavan TPV, Riches JD, Scanlon MJ, Ulett GC, Sakellaris H. 2016.Binding of CFA/I pili of enterotoxigenic *Escherichia coli* to asialo-GM1 is mediated by the minor pilin CfaE. Infect Immun 84:1642–1649.

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in each published paper.

(Signed) _____

Vipin Madhavan Thekke Palasseri

(Countersigned) _____ ____

Supervisor: Dr Harry Sakellaris

Chapter 1 – Literature Review

ADHERENCE FACTORS OF ENTEROTOXIGENIC *E. coli*

1.1 Introduction

Escherichia coli is a widespread commensal inhabitant of the intestinal tract of animals, including humans, and is thought to play a beneficial role in the host. However, a variety of *E. coli* pathotypes capable of causing intestinal infections have also evolved. These include Enteropathogenic *E. coli* (EPEC), Enteroaggregative *E. coli* (EAEC), Enteroinvasive *E coli* (EIEC), Enterohaemorrhagic *E coli* (EHEC) , Enteroadherent *Escherichia coli* (EAEC) and Enterotoxigenic *E. coli* (ETEC) (Turner et al., 2006, Mathewson et al., 1983). Each pathotype causes diarrheal disease via distinct mechanisms and may induce different symptoms in the host. In the case of ETEC, infections are characterized by watery diarrhea, vomiting, stomach cramps and in some cases mild fever (Evans and Evans, 1996). Infections are self-limiting in otherwise healthy individuals, but can have devastating consequences in vulnerable individuals. ETEC is responsible for nearly 200 million cases of diarrhea and 380,000 deaths annually, mainly in children less than 5 years of age (Steffen et al., 2005, Wenneras and Erling, 2004). It is also the major cause of traveler's diarrhea in people visiting endemic areas (Northey et al., 2007) and a significant cause of infections in military personnel in the field (Nada et al., 2013, Monteville et al., 2006). ETEC is transmitted primarily via contaminated water and is therefore a major public health problem in developing countries that lack the infrastructure to provide reliably treated drinking water. ETEC also causes neonatal enteric infections in agricultural production animals including cattle, pigs and sheep, leading to financial losses worldwide (Levine, 1981). However, such strains appear to be specific for animal hosts as they are not encountered in human infections.

It has been estimated that the ETEC pathotype evolved around 25,000 to 30,000 years ago (Chattopadhyay et al., 2012). The bacterium has evolved a variety of virulence factors that enable it to colonize the intestine and induce diarrhea in the infected host. Two major types of toxin, heat-labile enterotoxin and heat-stable enterotoxin, are secreted by ETEC. These activate the host enzymes, adenylate cyclase and guanylate cyclase, respectively in intestinal epithelial cells (Sears and Kaper, 1996). This leads to the accumulation of cAMP or cGMP, which in turn activate the cystic fibrosis transmembrane conductance regulator (CFTR). CFTR is an ion channel that controls the flow of electrolytes and water across the membrane of enterocytes. Enterotoxin-mediated activation of the channel leads to a net flow of water from the cell into the intestinal lumen, resulting in profuse watery diarrhea (Sears and Kaper, 1996).

In addition to the well-studied enterotoxins, two other toxins potentially play a role in ETEC virulence. EatA, an autotransporter protein with serine protease activity is encoded by a gene which is carried on the ETEC virulence plasmid pCS1 (Patel et al., 2004). The *eatA* gene is found in a variety of clinical ETEC strains and it has been demonstrated that EatA damages the intestinal epithelium and contributes to fluid loss in a rabbit ileal loop model of infection (Patel et al., 2004). Another type of toxin, the cytolysin ClyA, is found in a variety of ETEC strains (Ludwig et al., 2004). Although ClyA is not expressed by ETEC under laboratory conditions, it may be expressed *in vivo*. Its lytic activity towards macrophages and HeLa cells suggest that it may have a role in virulence. However, this has not been proven in experimental infections (Oscarsson et al., 1999).

Before ETEC can initiate an infection, it must first colonize the host intestine. Unlike commensal strains of *E. coli* which are generally limited to the colon (Kalser et al., 1966, Swidsinski et al., 2005), ETEC colonizes the small intestine. The small intestine is sparsely inhabited by normal bacterial flora and ETEC appears to have adapted to this niche. In the

small intestine the bacterium is confined to the epithelial surface, unlike some other enteropathogens which invade into deeper tissues. Colonization of the small intestine is mediated by protein adhesins localized on the bacterial cell surface. Most of these adhesins are pili (fimbriae) or pilus-related molecules with polymeric structures. However, a few adhesins are not related to pili and are simple outer membrane proteins that do not form macromolecular structures.

ETEC adherence factors have been the subject of a great deal of research. This is because adherence to, and colonization of, the small intestine is an essential, first step in the pathogenesis of ETEC infections. ETEC adherence therefore represents a critical point at which infections may be prevented with vaccines that block the interaction of adhesins to their receptors. A great deal of work, which is beyond the scope of this review, has been done in the area of ETEC vaccine development. For comprehensive coverage of this field, the reader is referred to excellent reviews devoted to the area (Svennerholm and Tobias, 2008, Zhang and Sack, 2012). The current review focuses on the diversity, structure, assembly, function and regulation of adhesins in ETEC strains that infect humans.

1.2. Pilus and Pilus-Related Colonization Factors

1.2.1 General Characteristics

1.2.1.1 Morphology and Composition

Colonization of the intestinal epithelium by ETEC is mediated by a variety colonization factors (CFs) which are pilus (fimbrial) or pilus-related adhesins (Table 1.1). Pili are hair-like proteinaceous appendages that protrude from the bacterial cell surface and, in general, mediate the attachment of bacteria to surfaces, including the surfaces of cells and tissues of host animals and plants. They are composed of repeating protein subunits, otherwise known

as pilins, and consist of either a single structural subunit (homopolymeric) or more than one type of subunit (heteropolymeric). In general, these macromolecular, polymeric structures are readily observable by electron microscopy. However, a few such structures in ETEC are not readily observed on the bacterial surface. Despite their clear relationship to pili, they form structures that are been below the limit of resolution of electron microscopy and have usually been referred to as "afimbrial" CFs.

In addition to the "classical" CFs, a few adhesins that are unrelated to pili also mediate the attachment of ETEC to intestinal cells (Elsinghorst and Kopecko, 1992, Elsinghorst and Weitz, 1994, Fleckenstein et al., 1996).

CFs differ in morphology, serotype, amino acid sequence and receptor binding specificity. They are divided into four main morphological types; fimbrial (pilus), fibrillar, helical and afimbrial. ETEC pili are relatively rigid filamentous structures of 5-7 nm in diameter, while fibrillae are more flexible structures of 2-3 nm in diameter. Two ETEC CFs, CS5 and CS7 have distinctive structures 3.5-6.5 nm in diameter, which consist of two helically intertwined filaments. In contrast, four ETEC CFs, are designated as afimbrial (Table 1.1).

1.2.1.2 Adherence Function

The first CF of a human ETEC strain was defined by its ability to facilitate the colonization of the rabbit small intestine (Evans et al., 1975). CF expression and colonization of the rabbit intestine were later found to correlate with the ability of ETEC strains to agglutinate human erythrocytes (Evans et al., 1977b). The implication of this finding was that the receptors found on human erythrocytes were of similar structure to those engaged by ETEC to attach to the intestine. Thus, hemagglutination and intestinal attachment were thought to represent one and the same activity. Decades later, this presumption was proven correct (Baker et al., 2009, Sakellaris et al., 1999a). The initial discovery of hemagglutination by the CF of one ETEC

Table 1.1: Characteristics of ETEC Colonization Factors

Colonization Factor	Assembly Class	FUP Clade	Morph-ology	Sequence Information	Database Accession No.	Studies Identifying CF or Determining Sequence	Studies of Binding to Intestinal Cells
CFA/I	CU	α	Fimbrial	Operon	M55661	(Jordi et al., 1992)	(Viboud et al., 1996)
CS1	CU	α	Fimbrial	Operon	NC_007635	(Froehlich et al., 1994b) (Marron and Smyth, 1995)	(Fujiwara et al., 2001)*
CS2	CU	α	Fimbrial	Operon	AY515609	(Froehlich et al., 1995)	(Viboud et al., 1996)
CS4	CU	α	Fimbrial	Operon	AY288101	(Anantha et al., 2004)	(Viboud et al., 1996)
CS5	CU	α	Helical	Operon	AY513487	(Duthy et al., 1999)	(Viboud et al., 1996)
CS14 (PCF0166)	CU	α	Fimbrial	Operon	Z47800	(Anantha et al., 2004)	(Viboud et al., 1996)
CS17	CU	α	Fimbrial	Operon	AY281092	(Anantha et al., 2004)	(Anantha et al., 2004)
CS19	CU	α	Fimbrial	Operon	AY283611	(Anantha et al., 2004)	(Grewal et al., 1997a)
PCFO71	CU	α	Fimbrial	Operon	AJ224079	(Anantha et al., 2004)	(Viboud et al., 1996)
CS12 (PCF0159)	CU	γ2	Fimbrial	Operon	AY009096	(Tacket et al., 1987)	(Viboud et al., 1996)
CS18 (PCF020)	CU	γ2	Fimbrial	Operon	AF335469	(Honarvar et al., 2003)	(Viboud et al., 1996)
CS3	CU	γ2	Fibrillar	Operon	X16944	(Jalajakumari et al., 1989) (Yakhchali and Manning, 1997)	(Viboud et al., 1996)
CS6	CU	γ2	Afimbrial	Operon	U04844	(Wolf et al., 1997)	(Ghosal et al., 2009)
CS13 (PCF09)	CU	κ	Fibrillar	Operon	X71971	{Heuzenroeder, 1990 #465}	(Viboud et al., 1996)
CS23	CU	κ	Afimbrial	Operon	JQ434477	(Del Canto et al., 2012)	(Del Canto et al., 2012)
CS7	CU	U	Helical	Pilin gene	AY009095	(Hibberd et al., 1990)	(Viboud et al., 1996)
CS20	CU	U	Fimbrial	Pilin gene	AF438155	(Valvatne et al., 2004)	(Valvatne et al., 1996)
CS22	CU	U	Fibrillar	Pilin gene	AF145205	(Pichel et al., 2000)	(Pichel et al., 2000)
CS15 (Ag 8786)	CU	U	Afimbrial	Pilin N-terminus	None	(Aubel et al., 1991)	(Aubel et al., 1991)
CS26	CU	U	Fimbrial	Pilin gene	HQ203050	(Nada et al., 2011)	
CS8 (CFA/III)	Type IV	NA	Fimbrial	Operon	AB049751	(Taniguchi et al., 2001)	(Taniguchi et al., 2001)

CS21 (Longus)	Type IV	NA	Fimbrial	Operon	EF595770	(Gomez-Duarte et al., 2007)	(Mazariego-Espinosa et al., 2010)
CS11 (PCF0148)	U	NA	Fibrillar	None	None	(Knutton et al., 1987)	(Knutton et al., 1987)
CS10 (Antigen 2230)	U	NA	Afimbrial	None	None	(Darfeuille-Michaud et al., 1986)	(Darfeuille-Michaud et al., 1986)

* This study does not distinguish between CS1- and CS3-mediated binding to cultured intestinal cells
Abbreviations: CU - Chaperone-Usher; NA: Not Applicable; U: Unknown

strain was soon extended to other ETEC CFs and individual CFs were found to have distinct receptor specificities, as indicated by their differential agglutination of human, bovine, chicken, and guinea pig erythrocytes (Evans et al., 1979). Hemagglutination served as a model for binding to host tissues, but more direct evidence that ETEC CFs mediate colonization of the human intestine, came from studies of bacterial attachment to isolated human enterocytes and cultured intestinal epithelial cells, as listed in Table 1.1. In these studies, CF involvement in bacterial attachment was demonstrated in a number of ways, including the inhibition of bacterial attachment by competitive inhibition with purified CF proteins or by CF-specific antibodies. Alternatively, the attachment of recombinant CF-positive and CF-negative *E. coli* was compared. More definitive evidence for the role of CFs in intestinal colonization comes from studies of bacterial attachment in human intestinal organ cultures (Bakker et al., 1992) and from direct human challenge studies involving piliated and non-piliated ETEC variants (Satterwhite et al., 1978).

The natural intestinal receptor molecules for ETEC CFs are not known. However, CFs bind to glycoproteins extracted from cultured human intestinal epithelial cells (Wenneras et al., 1990) and to a wide variety of glycosphingolipids *in vitro* (Jansson et al., 2009, Jansson et al., 2006). The mechanism of receptor binding has been most extensively studied in CFA/I pili. In this pilus, the minor tip-associated pilin, CfaE, mediates binding to erythrocytes and intestinal cells (Baker et al., 2009, Sakellaris et al., 1999a). Several clustered amino acid residues on the surface of CfaE are required for binding to erythrocytes and at least one of these residues is involved in pilus binding to intestinal cells. It has also been reported that CFA/I has a second, distinct binding activity, mediated by the major pilin, CfaB, towards a wide variety of glycosphingolipids *vitro* (Jansson et al., 2009, Jansson et al., 2006). However, the mechanism of glycosphingolipid binding and the potential significance of CfaB in co-operative binding with CfaE or in determining tissue tropism has not been investigated.

1.2.1.3 Nomenclature

The naming of CFs in the early history of this field was confusing. CFs were often given the designation "CFA" followed by a roman numeral e.g. CFA/I or CFA/II. Alternatively, they were given the designation "PCF" (putative colonization factor) followed by the O-serotype of the ETEC strain e.g. PCFO166, or named "Antigen" followed by a numerical designation e.g. Antigen 8786. This was further complicated by the discovery that some of the original CFAs, were found to contain more than one serologically distinct pilus. For example, CFA/II was found to comprise three serologically distinct pili. The nomenclature of CFs has since been simplified and standardized (Gaastra and Svennerholm, 1996, Turner et al., 2006) so that ETEC CFs are now given a "CS" (Coli Surface antigen) designation, followed by an Arabic numeral. Thus, CFA/II which was found to consist of three distinct CFs, has been replaced by the individual components, CS1, CS2 and CS3 (Ericsson et al., 2008). This nomenclature has been applied to all ETEC pili except CFA/I and PCFO71. Despite the fact that pili belonging to the Type 4 pilus assembly class have been renamed CS8 and CS21 (Table 1.1), they are still widely referred to in the literature by their original names i.e. CFA/III and Longus, respectively.

Colonization factors have been classified according to a variety of characteristics including morphology, antigenic type and mode of assembly (Paranchych and Frost, 1988). The first two criteria for classification are not very useful as they may group evolutionarily unrelated pili together and separate related pili into different groups. The mechanism of CF assembly is a useful criterion for grouping pili into broad classes. The assembly classes for pili, in general, include (i) conjugative pili, (ii) pili assembled by the chaperone-usher (CU) pathway (iii) Type IV pili and (iv) pili assembled via the extracellular nucleation pathway (Nuccio

and Baumler, 2007). ETEC CFs fall within the CU and Type IV assembly classes. Although 50% of ETEC strains carry pili assembled via the extracellular nucleation pathway (Szabo et al., 2005), evidence for a role in colonizing the host is lacking. Therefore, this class of pili will not be discussed further.

1.2.1.4 Genetics

The genes encoding ETEC CFs are often located on plasmids (Aubel et al., 1991, Darfeuille-Michaud et al., 1986, Evans et al., 1975, Froehlich et al., 2005, Giron et al., 1994, Grewal et al., 1997a, Schifferli et al., 1990, Tacket et al., 1987, Taniguchi et al., 2001, Crossman et al., 2010, Smith et al., 1982, Wajima et al., 2013), suggesting that ETEC pilus genes may have been laterally acquired. This has been confirmed for the genes encoding the CFs, CFA/I and CS1, which are carried by the plasmids pCS1 and pCoo, respectively (Froehlich et al., 2004, Yamamoto et al., 1984). Although neither plasmid is self-transmissible, each plasmid can be mobilized by co-resident conjugative plasmids. When the conjugative R plasmid, R64, is introduced into ETEC carrying pCoo, the two plasmids form a co-integrate which is capable of conjugative transfer. In the case of the plasmid carrying the CFA/I genes, the native conjugative ETEC plasmid, pTra1 provides the necessary transfer functions, *in trans*, for the mobilization of pCS1.

In every case where complete nucleotide sequence data is available, it is clear that all of the genes required for the assembly of functional CFs are encoded by genes clustered together in operons (Fig. 1.1, Table 1.1). Sequence analysis shows that in some cases, these operons are flanked by remnants of insertion sequence (IS) elements, suggesting that the pilus operons may have at one time been mobile via transposition (Froehlich et al., 2005, Sakellaris and Scott, 1998).

1.2.2 Pili Assembled by the Chaperone-Usher Pathway

The CU pathway is the most extensively studied and best understood pilus assembly system. The defining features of the pathway include a periplasmic chaperone protein that stabilizes newly secreted pilins in the periplasm, prevents their premature polymerization and targets pilins to an assembly site in the outer membrane (Busch and Waksman, 2012). This assembly site contains an outer membrane "usher" protein which is required for the orderly secretion pilins and assembly of the pilus on the bacterial cell surface. Pili assembled by the CU pathway are found in a wide variety of Gram-negative bacteria and represent the largest group of pili found in ETEC. As well as sharing gene sequence similarity, these pilus systems often share similarity in the operonic organization of their genes. At a basic level, these operons consist of a chaperone gene, an usher gene and genes encoding pilins. The number of pilin genes within the operon determines the complexity of the pilus structure. Most of the pili assembled via this pathway are heteropolymeric. In ETEC, the heteropolymeric pili have two or three pilin genes; one usually encoding a major pilin comprising the bulk of the pilus shaft and others encoding minor pilins representing only a small fraction of the pilus structure. In some cases pili may consist of two pilins present in approximately equal proportions.

In order to classify pili at a finer level, it is necessary to determine phylogenetic relationships based on individual genes comprising the pilus operons. Classification based on pilus subunits is problematic due, in many cases, to the presence of multiple pilus subunits. Another approach involves determining the phylogenetic relationships between chaperone proteins, but again the presence of more than one chaperone in some pilus systems complicates the classification. The most suitable protein to compare for the determination of phylogeny is the usher protein, which is encoded by a single gene in all known CU pilus

systems. Thus CFs assembled by the CU pathway are classified according to sequence comparisons of their usher proteins. Fimbrial usher proteins (FUPs) have been grouped into six major phylogenetic clades designated α, β, γ, π, κ and σ. The pili of ETEC strains that infect humans fall within the α, γ and κ clades as indicated in Fig. 1.1.

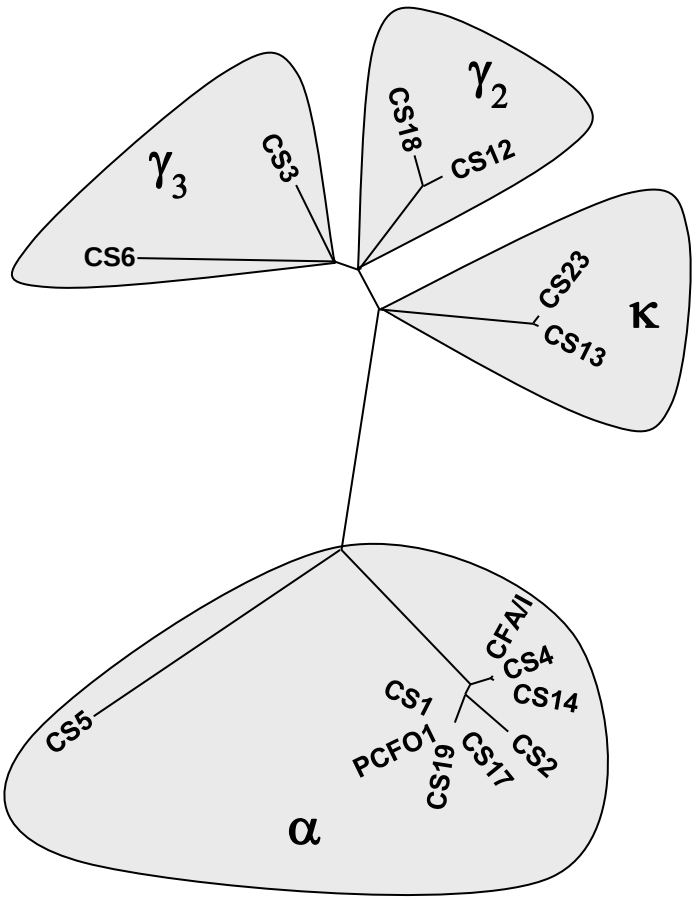

Figure 1.1. The relationship between CFs of human ETEC strains. The phylogenetic tree is based on fimbrial usher protein (FUP) sequences. CFs of ETEC isolated from humans are represented by the α, γ and κ FUP clades but not the β, π and σ FUP clades.

1.2.2.1 CFs of the α FUP Clade

ETEC pili belonging to the α clade of the FUP family include CFA/I, CS1, CS2, CS4, CS14, CS17, CS19 and CS5. This group of pili has also been described as belonging to the Class 5 pili according to an earlier classification scheme based on the sequence of the major pilus subunit (Chattopadhyay et al., 2012). It has been calculated that members of this class diverged from a common ancestor 0.7 to 1.6 million years ago, much earlier than the emergence of ETEC, and were acquired by ETEC no less than 700 years ago (Chattopadhyay et al., 2012)

The gene operons encoding the majority of α pili from ETEC, including CFA/I, CS1, CS2, CS4, CS17 and CS19, have identical organizations. They consist of four tandemly arranged genes encoding the chaperone, major pilin, usher and minor pilin in consecutive order (Fig 1.2). The CS14 operon has a similar genetic organization, but contains an additional, nearly identical, copy of the major pilin gene (Anantha et al., 2004, Gaastra et al., 2002). It is not known whether the two pilin genes have specialized functions that differentiate them.

Early characterizations of various ETEC pili, including α pili, were mainly restricted to the cloning and sequencing of the genes encoding pili. Subsequent studies were focused on the functions of genes involved in the assembly of α pili, as exemplified by CS1 pili which served as a model system for the α pili in ETEC (Nuccio and Baumler, 2007). CS1 pili are encoded by the four genes, *cooB*, *cooA*, *cooC* and *cooD*, arranged in an operon (Sakellaris and Scott, 1998). The major and minor pilins are encoded by *cooA* and *cooD*, respectively (Perez-Casal et al., 1990, Sakellaris et al., 1996). CooA comprises the great bulk of the pilus shaft while CooD is only found at the pilus tip and serves as the adhesin subunit. The stoichiometry of CooA:CooD is approximately 1000:1 (Sakellaris et al., 1996). The *cooB*

gene encodes a periplasmic chaperone (Voegele et al., 1997) and *cooC* encodes an outer membrane usher protein. CooB, C and D are indispensable for the assembly of major subunit CooA and are therefore regarded as assembly genes (Sakellaris and Scott, 1998).

The assembly of CS1 pili takes place via the CU pathway. However, it is worth noting that the assembly pathway has been extensively referred to in the literature as the "alternate chaperone-usher pathway." This terminology was first used because, although CS1 pili are assembled by a mechanism that was very similar to the classical CU pathway, the initial sequence analysis of the CS1 operon did not reveal an evolutionary relationship with the classical CU family (Soto and Hultgren, 1999). However, the subsequent expansion of sequence databases and the development of software capable of detecting distant evolutionary relationships, later allowed an evolutionary link to be made between pili assembled by the alternate and classical CU pathways. Therefore, the two terms are synonymous and the term "chaperone-usher pathway," will be used exclusively in this review.

Of the pilus systems belonging to the α clade, the assembly of only CS1 and CS5 has been studied. For CS1, the basic details of pilus assembly follow the paradigm that has emerged from the extensive studies of Pap and Type I pili of uropathogenic *E. coli*. All of the Coo proteins are secreted via the Sec pathway into the periplasm, where the chaperone, CooB, binds and stabilizes both the major and the minor pilins, CooA and CooD, respectively (Sakellaris et al., 1996, Sakellaris and Scott, 1998, Voegele et al., 1997). In the absence of CooB, both pilus subunits are degraded, as expected for a chaperone that facilitates the proper folding of pilus subunits and/or prevents their polymerization/aggregation in the periplasm. In addition, CooB interacts with CooC. However, unlike other chaperones of other CU systems,

CooB prevents the degradation of its usher protein. This would be in keeping with a role in maintaining the conformation of the usher.

The initiation of CS1 pilus assembly seems to differ from that of Type I and Pap pili. CS1 pilus assembly is initiated specifically by the minor tip pilin, CooD, complexed with its chaperone. The evidence for this comes from the demonstration that artificially modulating the level of CooD expression, modulates the number of pili assembled at the cell surface and the deletion of *cooD* abolishes pilus assembly (Sakellaris et al., 1999b). Since pili grow by the addition of pilins to the base of the structure, the assembly initiation function of CooD ensures that CooD will be positioned at the tip of the pilus (Sakellaris et al., 1999b), where it serves as an adhesin (Sakellaris et al., 1999a). In this way, CS1 differs from Type 1 and Pap pili, in which pilus assembly proceeds in the absence of the pilus-tip subunit. Instead, in these systems, the pilus-tip localization of the adhesin is determined by its affinity for the usher. Chaperone-adhesin complexes have much higher affinities for the usher than chaperone complexes with other pilins (Saulino et al., 1998). This ensures that the first stable interaction of the usher is with the adhesin rather than the other pilin subunits.

An important feature of all structures assembled by the CU pathway is the mechanism by which individual subunits interact with each other to produce a strong, stable polymeric structure. Subunits assembled by the CU pathway have incomplete Ig-like folds that lack an internal β-strand required to complete the fold (Poole et al., 2007). In a process termed "donor strand complementation,' an N-terminal β-strand extension of each subunit, termed the donor strand, interacts non-covalently with a hydrophobic groove at the C-terminus of the adjacent subunit to complete the Ig fold (Poole et al., 2007). This cis-complementation of the β-strand interlocks adjoining subunits and stabilizes the complemented subunit. A similar

process is involved in the binding of the chaperone to structural subunits in the periplasm. In this case, the chaperone binds to the structural subunit by donating a β-strand to the C-terminal groove of the subunit in a process termed "donor-strand exchange." Failure of a subunit to encounter and interact with the chaperone leads to subunit misfolding and degradation in the cell.

Due to the sequence similarity and identical organization of their operons, the mechanisms of assembly for CFA/I, CS2, CS4, CS17 and CS19 are expected to be identical to that for CS1. The presence of an extra pilin gene in the CS14 operon suggests that there will be a minor variation to the assembly of this pilus.

CS5 pilus assembly shares the basic features of CS1 assembly but also has some interesting variations. The operon encoding CS5 is unusual in that it contains two chaperone genes. CsfB is a periplasmic chaperone for the major pilin CsfA, while CsfF chaperones the minor pilins CsfD and CsfE. CsfD is essential for pilus assembly, while CsfE is a pilin-like protein that terminates pilus assembly and thus determines pilus length. (Duthy et al., 2002, Duthy et al., 1999). The adhesin subunit of CS5 pili has not been identified. However, sequence similarity between the CsfD and the CS1 adhesin, CooD, is high (53% identity; 67% similarity), suggesting that CsfD is the CS5 adhesin.

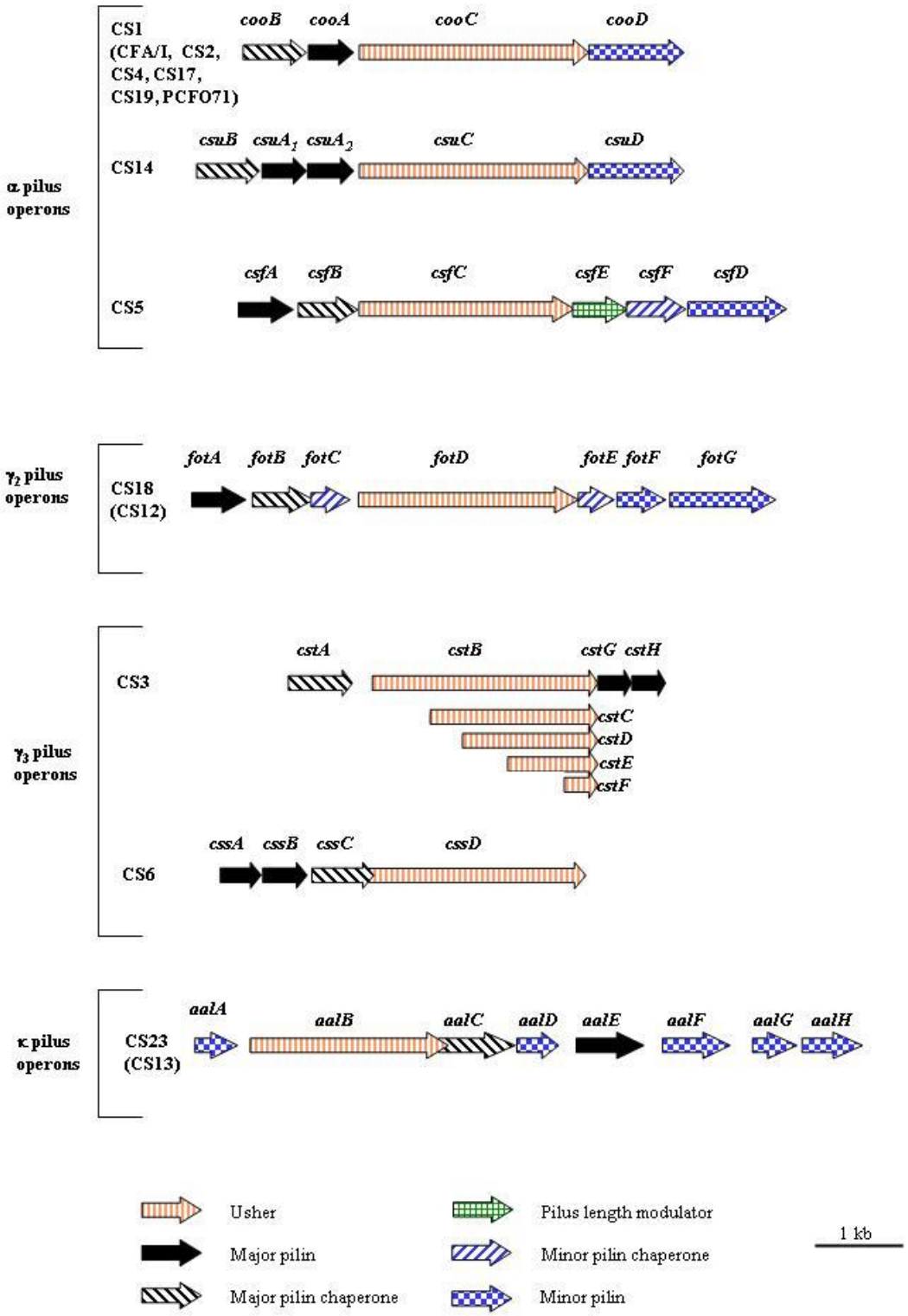

Figure 1.2. Genetic organization of the operons encoding ETEC CFs of the α FUP clade. Closely related CFs with identical operon organizations are listed in parenthesis.

1.2.2.2 CFs of the γ2 FUP Clade

The genes encoding CS18 were cloned, sequenced and analyzed to reveal a cluster of 7 genes, *fotA-G*, with sequence similarity and the same genetic organization as the *fasA-G* gene cluster encoding 987P pili from porcine ETEC (Honarvar et al., 2003, Viboud et al., 1993). The *fot* gene cluster includes an usher gene, three distinct pilin genes and three distinct chaperone-like genes. Although the functions of the *fot* genes have not been analysed, it has been inferred that the CS18 pilin and chaperone genes are likely to have the same functions as the corresponding *fas* genes encoding 987P pili.

In 987P, FasA is the major pilin while FasG, the adhesin, and the FasF minor pilin, are associated with the pilus tip and interspersed throughout the pilus shaft (Cao et al., 1995, Khan and Schifferli, 1994). FasC is a chaperone that binds to and stabilizes FasG, while FasE is a chaperone-like protein that is required for optimal extracellular transport of FasG (Edwards et al., 1996). The requirement for multiple pilin chaperones is an interesting characteristic which is also shared by members of the α family (CS5) and κ family (CS13 and CS23) of CFs (Table 1.1). In 987P, as in the CS5, the pilin chaperones show specificity for distinct pilins (Duthy et al., 1999, Edwards et al., 1996). The significance of such specificity is not clear, given that the pilin chaperones from other systems e.g. CS1, bind and stabilize both major and minor pilins (Voegele et al., 1997).

As for CS18, the genes encoding CS12 (*cswA-G*) have been sequenced and found to share sequence similarity and the same genetic organization as the 987P gene cluster. There has been no functional analysis of the genes encoding CS12, due to the high sequence and organizational similarity with the 987P gene cluster.

1.2.2.3 CFs of the γ3 FUP Clade

Two ETEC CFs, CS3 and CS6, belong to the γ_3- clade in the FUP classification scheme. CS3 is a narrow wire like appendage with a diameter of 2nM (Giron et al., 1995b). It has variously been described in the literature as a fibrillum, due to its small diameter, and a pilus. CS3 is encoded by the *cst*A-H gene cluster (Fig 1.2) (Altboum et al., 2001). *cst*A encodes a protein with sequence similarity to periplasmic pilin chaperones, while *cstH* encodes a CS3 pilin. The largest gene, *cstB* (Yakhchali and Manning, 1997), is predicted to encode a protein with sequence similarity to usher proteins. The description of *cstB* is unusual because of the claim that it contains an additional 4 overlapping internal genes, *cstC-F* (Fig. 1.2)." The conclusion that *cstB* contains internal overlapping genes was made based on the observed production of several smaller from the same DNA fragment comprising *cstB* (Jalajakumari et al., 1989). However, the biological significance of these smaller proteins has not been established and the existence of such genes in other CU pilus systems has not been described.

Until recently, it was thought that CS3 pili were composed of a single pilin encoded by *cstH* (Hall et al., 1989, Jalajakumari et al., 1989). However, further analysis of the *cst* gene cluster has revealed the existence of a previously unidentified gene, *cstG*, which encodes a second pilin (Savarino, 2006). Antibodies raised against CstH were capable of inhibiting pilus-mediated adherence to erythrocytes but antibodies to CstG were not, suggesting that CstH is the adhesin subunit. Other CFs e.g. CS1, contain two distinct subunits; a major subunit which comprises the great bulk of the pilus and a minor subunit which comprises only a small fraction of the pilus and is the adhesin subunit. However, unlike CS1 pili, in which the ratio of adhesin to non-adhesin subunits is approximately 1:1000 (Sakellaris et al., 1996), the ratio of CstH to CstG is 1:1.5 (Savarino, 2006).

CS6 is one of the most prevalent CFs in ETEC. Unlike most other CFs which are long filamentous structures protruding from the cell surface, CS6 has no readily identifiable macromolecular structure and is therefore classified as a non-pilus, or afimbrial colonisation factor (Ludi et al., 2008). However, the expression of CS6 on the bacterial cell surface is readily observed by immunogold electron microscopy. It has been suggested that CS6 may be a very short filament that does not protruded far from the cell surface, making it difficult to visualise by electron microscopy (Ludi et al., 2008).

In some respects, CS6 is similar to CS3, the only other CF representing the γ_3-fimbrial clade in ETEC. CS6 is encoded by a four gene operon consisting of *cssA, B, C* and *D* (Fig 1.2B) (Wolf et al., 1997). Based on sequence similarity, *cssA* and *cssB* encode two distinct structural subunits, while *cssC* encodes a periplasmic chaperone and *cssD* encodes an usher protein. As in CS3, the two structural subunits of CS6 are present in approximately equal proportion (Ghosal et al., 2009) and the subunits interact in an alternating fashion (Roy et al., 2012). CS6 is unusual in that one subunit, CssA, is substituted with a fatty acid, the function of which is unknown (Ghosal et al., 2009). The CS6 assembly system is also unusual in that the CssA subunit requires interaction with both the CssC chaperone and the CssB pilin, for maturation and function. In the absence of interaction with CssB, even CssC interaction is unable to prevent the degradation of CssA (Wajima et al., 2011).

It would be expected that CS6 follows the accepted paradigm for the assembly and surface expression of CU CFs. However, it has been reported that the usher, CssC, and the chaperone, CssD, are dispensable for surface localization and binding of CS6 to intestinal epithelial cells (Nicklasson et al., 2008). However, this unusual finding was contradicted in another study which demonstrated that all four genes are essential for CS6 surface expression and adherence (Wajima et al., 2011).

CS6 binding to intestinal cells is, at least in part, due to its ability to bind fibronectin (Fn), which is present in the extracellular matrix of the intestinal epithelium and on the surface of cultured intestinal epithelial cells (Ghosal et al., 2009). Binding to Fn occurs in a dose-dependent and saturable manner. The identities of the CS6 structural subunits which are responsible for cell-binding and Fn-binding are not known. Ghosal *et al* concluded that the CssA subunit, but not CssB, is responsible for binding to both Fn and intestinal cells (Ghosal et al., 2009). These conclusions were based largely on the observations that purified CssA bound Fn and intestinal cells and, in addition, inhibited the attachment of CS6-bearing ETEC to intestinal cells. Pre-incubation of ETEC with Fn blocked the adherence of ETEC to intestinal cells, while a synthetic peptide derived from the C-terminal sequence of CssA also inhibited ETEC adherence to intestinal cells (Ghosal et al., 2009). In contrast, others have concluded that both CssA and CssB mediate adherence of CS6 to intestinal epithelial cells via an unknown receptor and Fn, respectively (Roy et al., 2012). Interestingly, the study by Roy *et al* was confined to assays of subunit binding to Fn and intestinal cells and did not include binding assays with whole, piliated bacteria. The methods used to isolate CS6 subunits in each study was also quite different and may have accounted for the disparity in results and conclusions. Therefore, the mechanism of CS6-mediated binding to intestinal cells remains controversial and requires clarification. Adding to the uncertainty about the binding activities of CS6, Jansson *et al* have concluded that CS6 binds to sulfatide moieties of glycosphingolipids (Jansson et al., 2009), yet Roy *et al* found that CS6 does not bind to sulfatide (Roy et al., 2012). These contradictions are yet to be resolved.

1.2.2.4 CFs of the κ FUP Clade

Two ETEC CFs, CS13 and CS23, belong to the κ clade of the FUP family. The genes encoding these CFs have been cloned and sequenced (Del Canto et al., 2012, Heuzenroeder et

al., 1990). The gene cluster encoding CS23 consists of eight genes, *aalA-H*, with significant sequence similarity to, and the same genetic organization as, the *faeC-J* gene cluster encoding K88 fibrillae of porcine ETEC (Bakker et al., 1992). It has been reported that the CS13 gene cluster consists of only seven genes, *cshA-G*, (Genbank accession no. X71971), which are homologues of the *faeC-I* and *aalA-G* genes encoding K88 and CS23, respectively. However, the published sequence of the CS13 gene cluster terminates in a 108 nucleotide sequence with similarity to the 5' termini of *faeJ* and *aalG* (Sakellaris, unpublished). Thus it appears likely that the nucleotide sequence for CS13 is incomplete and the complete sequence may include an eighth structural gene.

Due to the high level of similarity between the κ CF and K88 gene clusters, there has been no functional analysis of the genes encoding CS13 and CS23. Instead, gene functions for CS13 and CS23 are inferred from the known functions of the genes in the well-studied K88 system (Verdonck et al., 2004). The K88 assembly system includes typical chaperone and usher genes, as do the CS13 and CS23 systems (Fig 1.2). The remaining genes encode a major structural subunit which is homologous to AalE in CS23 (Fig 1.2), and five distinct minor structural subunits that are homologous to AalA, D, F, G and H. Unlike many CFs which have tip-localized adhesin subunits, the major structural subunit of K88 is the adhesin responsible for binding intestinal receptors (Jacobs et al., 1987a, Jacobs et al., 1987b). The minor subunits include tip-localized subunits and subunits that are found throughout the K88 fibrillum (Bakker et al., 1992).

1.2.2.5 Structure of CFs Assembled by the Chaperone-Usher Family

Pilus subunits have a natural propensity to polymerize, making it difficult or impossible to prepare regular crystals suitable for structural determinations. For the determination of the

structure of subunits assembled by the CU pathway, this problem was overcome by manipulating donor strand complementation. Specifically, subunits were engineered to be self-complementing, soluble monomers. This was achieved by linking the short donor complementing strand to the C-terminus of the subunit via a flexible linker (Li et al., 2007). The donor strand in such constructs interacts with the C-terminal cleft to complete the Ig-fold of the subunit, thus stabilizing the protein and preventing the subunit from polymerizing. Using this approach to generate self-complemented soluble monomers, the tertiary structures of CFA/I (Li et al., 2009, Li et al., 2007), CS1 (Galkin et al., 2013) and CS6 subunits (Roy et al., 2012) have been determined.

In studies of CFA/I, the structures of both the CfaB major pilin and the CfaE minor pilin were determined. In addition, CfaE-CfaB and CfaB-CfaB hybrids were constructed by covalently linking subunits via the donor strand in order to analyze potential subunit-subunit interactions (Li et al., 2009). The donor strand consisted of the N-terminal amino acid sequence, VEKNITVTASVD, from CfaB (Poole et al., 2007). The features of both the major and minor subunits can be seen in the structure of the CfaE-CfaB hybrid, illustrated in Fig 1.3.

CfaB is a single domain protein with a 7-stranded β-sandwich fold belonging to the immunoglobulin-like family (Li et al., 2009). In contrast, CfaE folds into two domains, each adopting immunoglobulin-like β-sandwich folds which are very similar to that of CfaB (Li et al., 2007). In fact, the CfaB structure can be superimposed on each domain to yield root mean square deviations of 2.0 Å or 2.4 Å (Li et al., 2009). The N-terminal domain of CfaE is an adhesin domain, which binds to intestinal and erythrocytes receptors. The adhesin domain contains a concave surface which includes conserved, positively charged amino acid residues that are indispensable for adherence. The C-terminal domain of CfaE is a pilin domain which

connects the adhesin domain to the main pilus shaft (Li et al., 2007). The complementing donor strands interact with the C-terminal hydrophobic clefts of both CfaB and CfaE, as shown in Fig. 1.3. There are extensive interactions at the interface of the adhesin and pilin domains of CfaE, making CfaE conformationally rigid, unlike the structurally related adhesin, FimH from UPEC, which is more flexible at the interface between the pilin and adhesin domains. Interactions between CfaB domains in CfaB-CfaB molecules are minimal, allowing flexibility between subunits. Interactions between CfaE and CfaB also very limited.

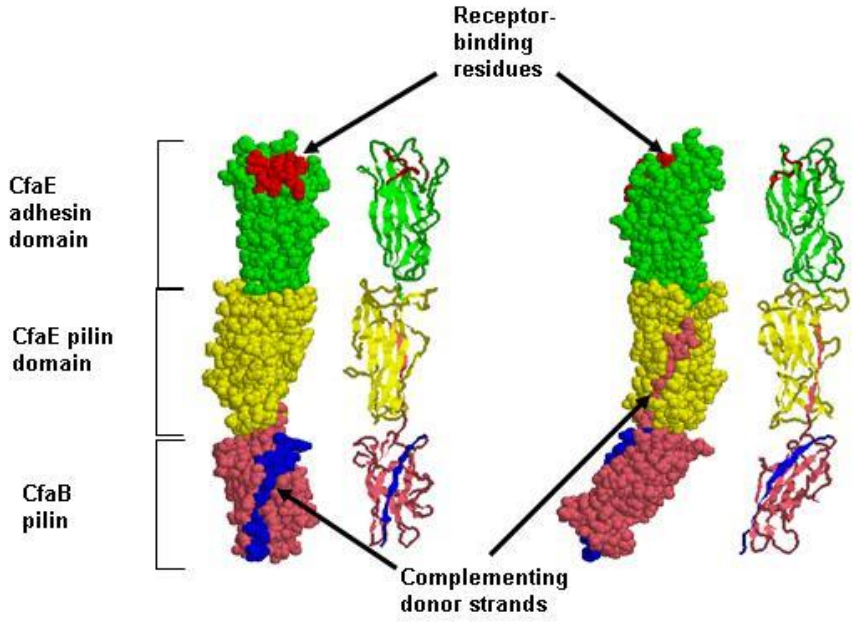

Figure 1.3. Structures of the major pilin, CfaB, and minor pilin, CfaE, of CFA/I pili. Space-fill and cartoon models of a CfaE-CfaB hybrid molecule (PDB 3F83) are represented from two different perspectives. Cartoon models illustrate the barrel-shaped β-sandwich folds of the major pilin, CfaB, and both the pilin and adhesin domains of the minor pilin, CfaE. Complementing donor strands interact with hydrophobic clefts on the surface of CfaB and the pilin domain of CfaE. The receptor binding pocket is located in a concave surface at the tip of the adhesin domain of CfaE. Amino acid residues involved in receptor binding are indicated in red.

The structure of the CFA/I pilus shaft was modelled by fitting the known tertiary structure of CfaB to 3D reconstructions of the pilus shaft based on electron micrographs of CFA/I pili (Andersson et al., 2012, Li et al., 2009). The modelling revealed that CFA/I is composed of CfaB subunits stacked in a helical conformation. The flexibility of CfaB-CfaB interactions gives the helical chains and cylindrical structure as a whole, more flexibility and resistance to shear stress. Furthermore, it was shown that proline^{13} in CfaB acts as a molecular hinge. The *cis-trans* summarization of proline^{13} switches the structure of the shaft between the normal helical conformation with a width of 74 Å to an extended or unwound fibrillar conformation with a width of 45 Å. The narrow fibrillar conformation may be essential to allow transit of the pilin through the pore formed by the usher during secretion from the cell. The wider helical conformation is the stable form maintained after transport through the usher (Li et al., 2009).

The structure of CS1 pili has also been determined (Galkin et al., 2013). The structure of the CS1 major pilin resembles that of CFA/I. Pilins interacts via donor strand complementation as in CFA/I and have the same barrel-shaped β-sandwich structure. Reconstructions of the pilus shaft suggest that CS1 adopts multiple structural states created by changes in the orientation and packing of pilins. It was suggested that this filament polymorphism confers greater flexibility on pilus fibres, thereby facilitating adherence to the intestinal epithelium under conditions of high shear. Both CS1 and CFA/I can adopt coiled and uncoiled conformations reminiscent of those observed in Type 1 pili and P pili. This allows such pili to act like springs that uncoil in response to shear, thereby allowing the bacterium to remain attached to the intestine during peristalsis (Andersson et al., 2012, Galkin et al., 2013, Mu et al., 2008).

Another adaptation to shear force was discovered in CFA/I pili. Receptor binding of CFA/I pili is greater at moderate hydrodynamic sheer stress than at low sheer (Tchesnokova et al., 2010). This is a function of the CfaE adhesin subunit as its binding to erythrocytes is also enhanced by sheer. This phenomenon appears to be similar to the shear-induced catch-bond formation in FimH, the adhesin subunit of Type I pili (Yakovenko et al., 2008). Although the mechanism in Type I pili is not fully understood, catch-bond formation depends on amino acid residues at the interface of the adhesin and pilin domains of FimH. Therefore a number of site directed mutations affecting the residues at the interdomain interface of CfaE were made to test whether they had a similar effect in CFA/I. A number of mutations were found to increase CFA/I binding to erythrocytes under low hydrodynamic shear (Tchesnokova et al., 2010). This was analogous to the findings in Type I pili. In order to determine how such mutations might affect sheer-induced receptor binding, the tertiary structure of one such mutant CfaE molecule was determined. Analysis of the structure showed that there were no changes in the conformation of the adhesin domain but the buried surface area at the interface between the pilin and adhesin domains decreased by 36% (Liu et al., 2013). It is not clear, however, how such an opening of the interdomain interface would change binding activity when the receptor binding pocket is at the pole of the CfaE molecule. Thus the structural basis of the catch-bond mechanism remains a mystery.

The structure of CS6 has also been elucidated (Roy et al., 2012). Most of the CU pathway pili, have a homopolymeric stalk composed of a major pilus subunit, with or without a minor pilus subunit at the distal tip. In contrast, CS6 is composed of two subunits, CssA and CssB, in equal proportion, interacting in an alternating pattern to form a heteropolymeric structure. Other features of CS6 are typical of CFs assembled by the CU pathway. Structure determinations of both CssA and CssB confirm that, like other CU systems, the subunits have

Ig-like β-barrel folds consisting of two β sheets. Despite considerable divergence in amino acid sequence (23% identity), CssA and CssB have only small differences in fold topology (Roy et al., 2012).

An important outcome of the structural studies on CS6, is an explanation of the mechanism by which CssA and CssB are incorporated into the CS6 structure in an alternating pattern. Structures of CssA complemented with donor strands derived from either CssA (homocomplemented) or CssB (heterocomplemented), showed that the fit between the donor strand and acceptor cleft was relatively poor in the homocomplemented subunit relative to the heterocomplemented subunit. Thus, linkages between unlike subunits are more stable and would be favored over linkages between like subunits. This is consistent with the finding that homocomplemented subunits of CS6 are less stable to temperature and chemical denaturants and expressed at lower levels in *E. coli* than heterocomplemented subunits (Roy et al., 2012).

1.2.3 Type IV pili

Type IV pili are very long pilus structures found in a wide variety of bacterial pathogens (Craig et al., 2004). Unlike pili of the CU family, whose main function is adherence, these pili have a variety of roles in bacterial pathogenesis. Apart from adherence to host cells, Type IV pili may confer twitching motility, cellular aggregation leading to microcolony formation, biofilm development and DNA uptake. The Type IV pili are classified into two groups, Type IVa and Type IVb, based on the length of the mature pilin sequence and the length of a leader sequence in the pilin which is cleaved during pilus assembly. In addition, the N-terminal residue of Type IVa pilins is phenylalanine, while for Type IVb, the N-terminal residues may be methionine, leucine or valine (Giron et al., 1997). Type IVa pili are expressed by a wide range of bacteria that infect several different tissues. In contrast, Type IVb pili are found only

in enteric bacteria. In ETEC, this includes Longus and CFA/III, otherwise known as CS21 and CS8, respectively. The naming of Longus is a reflection of the unusual length of the pilus. Most bacterial pili vary from approximately 0.5 - several μm in length. In contrast, Longus is up to 20μm in length. Although Longus and CFA/III are closely related, they are readily distinguished by their lengths and presentation on the bacterial surface. Longus is displayed at the cell pole while CFA//III is displayed peritrichously. With a length of 5-10μm, CFA/III is significantly shorter than Longus (Giron et al., 1997).

1.2.3.1 CFA/III and Longus

Of the ETEC Type IVb pili, CFA/III was the first to be characterized. As shown in Fig. 1.4, the CFA/III operon comprises a 14 gene operon, carried on a native plasmid in ETEC (Taniguchi et al., 2001). The CFA/III operon shows close sequence similarity to the *tcp* operon of *Vibrio cholerae*. The *cofA* gene encodes the major pilin, CofA, which is synthesized initially as a 26.5KDa prepilin protein. A prepilin peptidase, CofP, cleaves CofA prepilin into a mature protein of 20.5Kda (Taniguchi et al., 1999). The functions of other genes in the operon have been inferred based on their sequence similarity to other type IVb genes such as those of the *tcp* operon. *cofR* and *cofS* are predicted to be regulatory proteins, while *cofS* shows sequence similarity to the AraC family of transcriptional regulators, including *cfaD* and *Rns* from ETEC. *cofD* has the hallmarks of a membrane-bound lipoprotein, including an N-terminal lipoprotein recognition signal and an essential cysteine residue with a predicted role in lipid binding attachment. *CofH* is predicted to be an inner membrane nucleotide binding protein, due to characteristic sequence features, including conserved walker boxes A and B, Asp boxes and 2 pairs of CXXC cysteine residues. Due to its sequence similarity with genes like *tcpB* from *Vibrio cholerae, cofB* is predicted to encode a pilin-like protein. Three other proteins *cofE, cofF* and *cofI* have been assigned as genes

encoding inner membrane proteins without any detail of their exact function. No function has been predicted for the remaining genes in the operon.

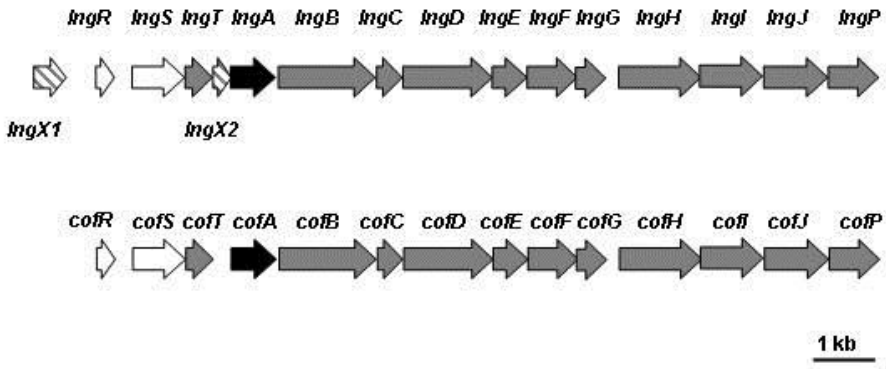

Figure 1.4 Genetic organization of the operons encoding CFA/III and Longus pili in ETEC. The 14 genes of the *cof* operon encode CFA/III pili. The major pilin is indicated by a black arrow. Regulatory genes are indicated as white arrows. Genes involved in assembly or of unknown function are indicated as gray arrows. The genes encoding Longus pili have the same organization and are designated by the prefix "*lng.*" The *lng* operon contains two additional ORFs, *lngX1* and *lngX2*, of unknown function, shown as hatched arrows.

The Longus operon contains all of the open reading frames corresponding to CFA/III and designated with the prefix *lng* (Fig. 1.4). In addition, the Longus operon contains two extra open reading frames, *lngX1* and *lngX2*, of unknown function (Gomez-Duarte et al., 2007). Sequence variation at the amino acid level is greatest within the structural pilus subunits, CofA and LngA of CFA/III and Longus, respectively (Gomez-Duarte et al., 2007). Although immune recognition is likely to be a selective pressure driving sequence variation, it is also possible that sequence variation reflects functional adaptation in receptor binding specificity (Gomez-Duarte et al., 2007). In contrast to the structural subunits, non-structural proteins of CFA/III and Longus are highly conserved.

Phylogenetic analysis of congruence among various *lng* loci has shown that there is no evidence of recombinational shuffling between the non-structural genes from different ETEC strains. However, phylogenetic trees of non-structural *lng* genes and *lngA* alleles are not congruent, suggesting that the structural subunit has moved laterally between *lng* gene clusters of different strains (Gomez-Duarte et al., 2007). This is consistent with the findings that trees of non-structural *lng* genes and housekeeping genes are congruent while those of *lngA* and housekeeping genes are incongruent (Gomez-Duarte et al., 2007). These findings support the hypothesis that horizontal transfer of *lngA* occurs via a single gene recombination rather than recombination of the entire operon. Based on the accumulation of mutations in non-structural genes, it has been estimated that CFA/III and Longus diverged 5-6 million years ago and that CFA/III was acquired by *E. coli* 0.73 to 3.64 million years ago, while Longus was acquired 0.97 to 4.83 million years ago (Gomez-Duarte et al., 2007).

Many of the properties conferred by Type IV pili on host bacteria are shown by Longus in ETEC, including self-aggregation and protection from antimicrobial agents (Clavijo et al., 2010), twitching motility, and adherence to cultured intestinal epithelial cells (Mazariego-

Espinosa et al., 2010). In contrast, CFA/III mediates only autoaggregation (Galkin et al., 2013, Kolappan et al., 2012).

1.2.3.2 Structure of Type IV Pili in ETEC

The purification of Type IV pilins for structure determinations has been problematic. As with pilins of the CU pathway, the pilus subunits have a propensity to polymerize, a property which interferes with formation of regular crystals suitable for structure determination. However, the mechanism for Type IV pilus polymerization is quite different to that of CU pili, thus requiring an alternative strategy for the prevention of protein aggregation. Type IV pilins polymerize via hydrophobic interactions between their N-terminal α-helices. Consequently, there are only two species, *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa*, from which full-length Type IV pilins have been successfully crystallized and their structures determined (Craig et al., 2003, Parge et al., 1995). For other Type IV pilins, it has been necessary to generate constructs lacking the N-terminal α-helix to prevent protein aggregation. Structural studies have revealed that a typical Type IV pilin consists of a conserved, hydrophobic, N terminal α-helix, which is involved in pilin polymerization (Craig et al., 2004, Craig et al., 2003, Parge et al., 1995). The amphipathic C-terminal portion of this structure is buried within a globular core domain consisting of a central, hydrophobic, antiparallel β-sheet. Another region of the protein, termed the αβ-loop, links an N-terminal α-helix with the central β-sheet. A fourth region of the protein, termed the D-region, is located between two disulphide-bonded cysteine residues. The αβ-loop and D-region form the outer edges of the globular head domains of Type IV pilins. They are solvent exposed, cover much of the protein surface, show a great deal of sequence variation and are responsible for functions such as interactions of the pilus with receptors and with other pili (Kirn et al., 2000, Xu et al., 2004, Fukakusa et al., 2012).

Determination of the tertiary structure of the CFA/III pilin, CofA, revealed an overall architecture that is similar to those of other Type IV pilins (Fukakusa et al., 2012, Kolappan

et al., 2012). Structures were determined independently by two research groups, using an N-terminal, 28-residue truncated derivative of CofA. The structures of truncated CofA, showed that the protein could be divided into three distinct regions similar to those of other Type IV pilins, as illustrated in Fig. 1.5 (Fukakusa et al., 2012, Kolappan et al., 2012). The structural core, comprising an α-helix and an antiparallel β-sheet was almost identical to that of Tcp pilin from *Vibrio cholerae*, the first Type IVb pilin for which the tertiary structure was determined. In contrast, the αβ-loop and D-region showed structural variation. Both the α-loop and D-region were highly solvent exposed and covered most of the surface of the protein, in keeping with the general features of other Type IV pilins.

CFA/III filament models were constructed by homology modeling and computational docking (Fukakusa et al., 2012, Kolappan et al., 2012). A full-length CofA model was generated by modeling CofA on the PAK pilin from *Pseudomonas aeruginosa* strain K, for which the tertiary structure, including the N-terminal α-helix, is known. The CofA structure was then superimposed onto pilin subunits in a Tcp filament model derived by electron microscopic reconstruction. The model CFA/III filament has a three start helical assembly with a diameter of 80 Å. The conserved N-terminal α-helices are themselves arranged helically over the pilus core and are stabilized by hydrophobic interactions. The variable globular domains are stacked around the surface of the pili while the αβ-loop and D-region interact to further stabilize the pilus structure. In comparison to Tcp and other Type IVb pili, the surface of the CFA/III pilus is relatively smooth and highly negatively charged (Kolappan et al., 2012).

Kolappan *et al* modeled the structure of the Longus major subunit LngA on the known tertiary structure for CofA and the Tcp filament model. The CFA/III and Longus filament

models were very similar and confirmed that the relatively smooth, highly negatively charged surface of these fibers is distinct from other known Type IV pili.

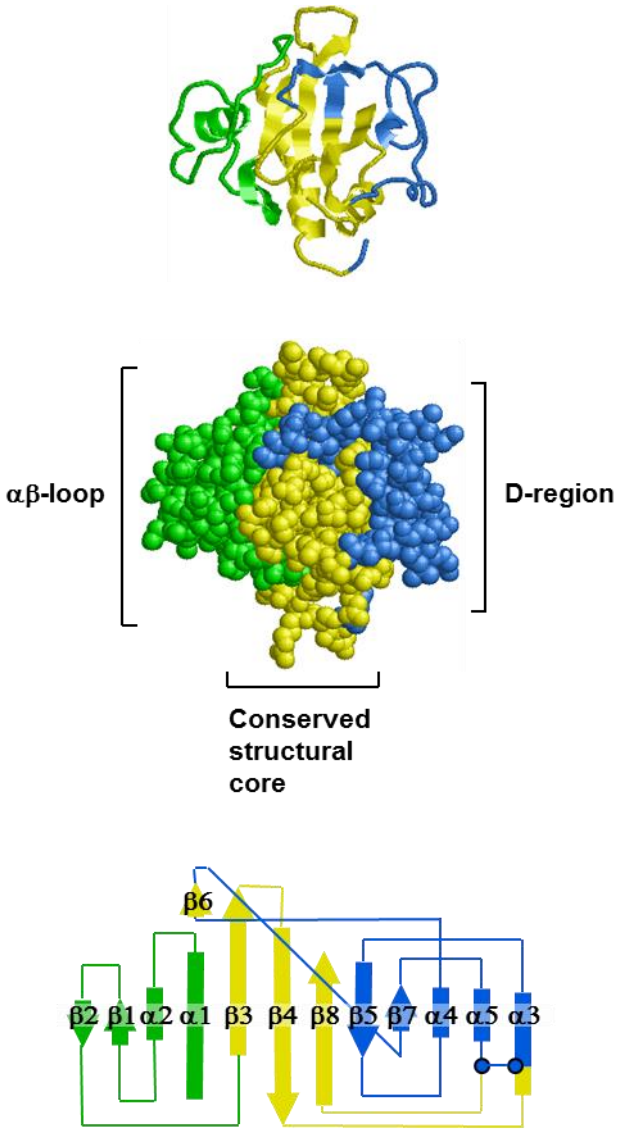

Figure 1.5. Structure of the CofA pilin from CFA/III pili of ETEC.

Cartoon (top) and space-fill (center) tertiary structure models of an N-terminal deletion derivative of CofA. The topological diagram (bottom) shows the relative positions of secondary structural elements that define the αβ-loop (green), conserved hydrophobic structural core (yellow) and D-region (blue). The disulfide-bonded cysteine residues that define the boundaries of the D-region are indicated as blue circles. Tertiary structure models were derived from PDB entry 3VOR. The topological diagram of CofA was adapted from Fukakusa et al (Fukakusa et al., 2012).

1.3. Non-Pilus Adhesins

1.3.1 Tia

Unlike some other enteric pathogens, ETEC does not invade the intestinal epithelium into deeper tissues. However, ETEC strain H10407 invades cultured ileocecal and colonic epithelial cells (Elsinghorst and Kopecko, 1992). Although it is able to invade intestinal cells, H10407 is unable to replicate inside the cell and is confined to endocytic vacuoles after invasion (Elsinghorst and Kopecko, 1992). The invasive properties of H10407 were attributed to two genes, *tia* and *tibA*. These two genes are chromosomally encoded and are believed to act independently (Elsinghorst and Kopecko, 1992).

The *tia* gene is borne on a 46kb pathogenicity island, along with a number of other ORFs believed to encode for a protein secretion apparatus (Fleckenstein et al., 2000). *tia* encodes a 25kDa outer membrane protein, Tia, which functions as both an adhesin and an invasin (Fleckenstein et al., 1996, Mammarappallil and Elsinghorst, 2000). Since cell invasion has not been described as a feature of ETEC infections, the functional significance of the cell invasion property observed *in vitro* is not clear. Although, the molecular structure of Tia has not been elucidated, secondary structure predictions suggest that it has eight transmembrane amphipathic β-sheets with four surface exposed loops on the outer membrane. Furthermore, it is believed that the interaction of Tia with the intestinal epithelium is partially or largely mediated via binding to heparin sulfate proteoglycans (Fleckenstein et al., 2002).

1.3.2 EtpA

ETEC strain H10407 also possesses a plasmid-borne two-partner secretion (TPS) locus consisting of three genes, *etpBAC*. EtpA is a secreted protein with sequence similarity to the high molecular weight (HMW) adhesins from *Haemophilus influenzae* and related adhesins from *Yersinia pestis, Y. pseudotuberculosis, Y. enterocolytica* and *Bordetella pertussis*

(Fleckenstein et al., 2006). EtpA mediates bacterial adherence to cultured intestinal cells *in vitro* and promotes bacterial colonization of the intestine in a murine model of infection (Roy et al., 2008). The binding activity of EtpA is dependent on its glycosylation, which in turn is catalyzed by EtpC. Although the nature of the glycosl group that modifies EtpA is unknown, EtpC has sequence similarity to a family of O-linked N-acetylglucosamine transferases that modify the serine/threonine residues of proteins (Fleckenstein et al., 2006). The effect of EtpA glycosylation on bacterial adherence is unusual in that the loss of glycosylation produces reduced bacterial adherence to colonic cells and hyperadherence to ileocecal cells. This suggest that glycosylation of EtpA modulates the attachment of ETEC to different cell types (Dorsey et al., 2006)

EtpB is a 65kDa pore-forming outer membrane protein with sequence similarity to known porins of other TPS systems, which are required for the secretion of their cognate partners. Furthermore, bioinformatic analysis suggests that it has domain features similar to those of many transporters (Meli et al., 2009).

As well as conferring adherence independently, EtpA also interacts with conserved residues in flagellin (FliC), in a serotype independent manner, to promote flagellum-mediated adherence to intestinal cells (Roy et al., 2009a). The implication of this may be significant in terms of vaccine development, since antibodies raised against EtpA and/or flagellin block the adherence of ETEC to intestinal cells and interfere with colonization of the host (Roy et al., 2009a).

1.3.3 TibA

TibA is encoded within the *tibDBCA* gene cluster (Lindenthal and Elsinghorst, 1999). *tibA* encodes a 104kDa outer membrane protein. *tibC* encodes a cytoplasmic heptosyltransferase

which glycosylates the TibA precursor by addition of a heptose moiety (Moormann et al., 2002, Elsinghorst and Weitz, 1994). The role of *tibB* has yet to be elucidated.

TibA is involved in adherence, invasion and autoaggregation (Elsinghorst and Weitz, 1994, Lindenthal and Elsinghorst, 2001). Although glycosylation of TibA is predicted to be essential for function, TibA mediates bacterial autoaggregation and biofilm formation in a glycosylation independent manner (Sherlock et al., 2005).

Structurally, TibA is a self-associating autotransporter (SAAT) (Klemm et al., 2006, Cote and Mourez, 2011). It is predicted to have a Sec complex homing N-terminal signal peptide and C-terminal membrane spanning domain (Cote and Mourez, 2011). Structural and functional studies employing transposon-based linker scanning mutagenesis and site-directed mutagenesis showed a modular organization of TibA function. Specifically, autoaggregation is mediated by the N-terminal half of the extracellular domain, while adherence to epithelial cells is specified by the C-terminal half of the extracellular domain (Cote and Mourez, 2011). However, mutations that affect invasion or biofilm formation, also affect either adherence or autoaggregation, indicating that these activities are related.

1.4 Regulation of Pilus Expression

1.4.1 AraC-family transcriptional regulators

The expression of pili of ETEC belonging to the α FUP clade is dependent on transcriptional activation. This was first demonstrated for CS1 and CS2 pili, which require for their expression a native plasmid that carries the *rns* gene (Caron et al., 1989). The *rns* gene encodes a 26kDa protein that belongs to the AraC family of transcriptional regulators (AFTRs) (Caron et al., 1989). Rns has significant sequence similarity to many regulators of pilus and virulence genes from a variety of pathogenic bacteria. These homologues include

CfaR/CfaD, CsvR, and FapR from ETEC as well as AggR, VirF and PerA from EAEC, *Shigella flexneri* and EPEC, respectively (Munson et al., 2002). Sequence similarity is greatest in the C-terminus due to the presence of two conserved helix-turn-helix (HTH) DNA binding motifs (Martin and Rosner, 2001, Munson and Scott, 2000). The sequence similarity between Rns and its homologues is reflected in function as well. Although it was first identified to activate CS1 and CS2 expression, Rns was shown to also activate the expression of other α pili including CS17, CS19, and PCFO71 (Basturea et al., 2008), the γ pilus, CS3, and is predicted to regulate the expression of CS20 (Valvatne et al., 1996). In fact, Rns can regulate the expression of almost half of the known pili in ETEC and is considered to be the prototype of transcriptional activators for ETEC pili (Basturea et al., 2008).

AraC transcriptional activators regulate a wide variety of genetic systems across different bacterial species. They are known to regulate multiple genes in the bacterial cell by binding to DNA sequences upstream and sometimes downstream of the transcription start site (Gallegos et al., 1997, Hart et al., 2008). Although Rns was first discovered as a regulator of pilus expression in ETEC, it also appears to regulate the expression of genes that do not encode pili. Munson *et al* identified a number of Rns-regulated genes by capturing ETEC DNA sequences that bound to a Rns fusion protein (Munson et al., 2002). The bound DNA fragments were then sequenced and analyzed. Sequencing to identify putative Rns binding sites, combined with *in vitro* transcription analysis showed that Rns influences the transcription of two genes, *yiiS* and *asl*A (Munson et al., 2002). Rns binds to sequences upstream of *yiiS* and activates its expression. To date, the function of *yiiS* is not known. The *yiiT* gene, which is situated downstream of *yiiS* is a paralogue of *uspA*, a universal stress response gene (Gustavsson et al., 2002) under the control of the extracytoplasmic stress response sigma factor, σ^E (Rezuchova et al., 2003). The *yiiS* and *yiiT* genes are co-transcribed. These facts suggest that *yiiS* may have a role in the bacterial stress response.

Many of the functionally characterized genes regulated by AFTRs have direct or indirect roles in bacterial virulence (Hart et al., 2008). Similarly, *yiiS* might have a role in ETEC virulence. Apart from *yiiS*, Rns binds upstream of *asl*A, which increases the ability of *E. coli* K1 to invade brain microvascular endothelial cells leading to meningitis in a rat model of infection (Hoffman et al., 2000). How such a gene might play a role in ETEC virulence is not clear, as ETEC is confined to the intestinal mucosa during an infection. Although Rns binds to the promoter regions of the genes described, no transcription analysis was conducted on *aslA* and therefore it is not possible to infer that *asl*A transcription is regulated by Rns (Munson et al., 2002). In contrast to the activation of genes, Rns also downregulates the expression of the inner membrane lipoprotein, NlpA. NlpA is involved in the synthesis of outer membrane vesicles (OMVs) in bacteria (Yu et al., 1986, McBroom et al., 2006). OMVs are known to carry even ETEC heat labile toxin (LT), suggesting that, *rns* may be indirectly influencing the release of toxin (Bodero et al., 2007).

The *cfaABCE* gene cluster encoding CFA/I is regulated by CfaD (CfaR) which shares 95% sequence identity with Rns. CfaD is encoded in region 2 of the plasmid encoding the CFA/I genes (Savelkoul et al., 1990). As well as binding to a sequence upstream of the *cfaABCE* gene cluster, CfaD also binds to a region upstream of another gene, *cex*E. The first 19 amino acids at the N-terminus of CexE show characteristics of a signal peptide and it was therefore predicted that CexE is transported via the general secretion pathway to outer membrane. Recent studies have substantiated these claims, as CexE was detected among the constituent proteins in outer membrane vesicles (Roy et al., 2010, Busch and Waksman, 2012). CexE has sequence similarity to the disperin protein Aap from enteroaggregative *E. coli* (Sheikh et al., 2002). This negatively charged protein coats the bacterial cell surface and prevents AAF/II fimbriae from collapsing onto the cell surface (Sheikh et al., 2002). By analogy, it is possible

that CexE may have a similar function in ETEC i.e. it may be required for the optimal display of pili to maximize adherence to intestinal cells.

The tertiary structures of Rns or other AFTRs regulating ETEC pili have not yet been elucidated. From the known structures of AFTRs like MarA from *E. coli* (Rhee et al., 1998) and ToxT from *Vibrio cholerae* (Lowden et al., 2010), a general pattern is observed. A typical AFTR consists of two conserved HTH DNA binding domains at the C-terminus and an N-terminal domain which is variable and may be disordered. The N terminal effector domain may be involved in dimerization and ligand binding for activation. In many AFTRs a flexible linker connects these two domains (Yang et al., 2010, Crossman et al., 2010). Sequence analysis and experimental data suggests that Rns is similar (Munson and Scott, 1999, Munson et al., 2001, Basturea et al., 2008, Crossman et al., 2010) to other members of AFTRs.

As with other AraC members, Rns binds to AT rich regions in the genome (Yang et al., 2010, Munson et al., 2001). DNase footprinting assays have revealed that Rns binds to a consensus sequence (Munson and Scott, 1999). Variation within this consensus sequence complicates genome-wide computational prediction of Rns binding sites (Munson et al., 2002). However, such predictions have revealed many potential Rns binding sites in ETEC isolates from Guinea-Bissau. Although not validated by experimental evidence, the results suggest that Rns may influence many genes, some of which may have roles in virulence. These include a Type VI secretion effector, an acetate operon repressor and the inner membrane protein YafU (Sahl et al., 2011).

In order to activate the transcription of genes encoding CS1, Rns binds to site I (-109.5) and site II (-37.5), upstream of P*coo*, the promoter for the *cooBACD* operon. Binding to both sites is important for maximal expression of CS1 (Munson et al., 2001), although Rns can bind to

these sites independently as well. Site II is particularly important, probably due to its proximity to the RNA polymerase binding site and because Rns probably interacts with the α subunit of RNA polymerase (Munson and Scott, 1999).

Rns regulates its own expression by binding to regions upstream and downstream of transcription start site for P_{rns}. A single upstream consensus binding sequence is located an unusually long distance, centred around 224.5bp region, from the transcriptional start site. In addition, Rns binds two sites within the Rns coding sequence, centred with in a range of +43.5 and +83.5. Each of these sites is essential for the autoregulation of Rns (Munson et al., 2001). This is unusual in bacterial transcriptional regulators and the mechanism by which Rns binding to coding sequences activates transcription is still not understood.

Unlike many AFTRs, an effector binding function in the N-terminal domain has not yet been found in Rns. Determining the three-dimensional structure of Rns or its N-terminal domain may facilitate the identification of putative ligand binding sites and effector molecules that modulate Rns function. Many AraC proteins bind to DNA as dimers. An early report that Rns does not dimerize (Basturea et al., 2008), has been contradicted by recent studies that have demonstrated dimerization both *in vitro* and *in vivo* (Mahon et al., 2012). The flexible linker connecting the N and C terminal domains of Rns has been found to be essential for Rns binding to DNA (Basturea et al., 2008, Crossman et al., 2010).

A significant amount of work has revealed greater detail about the mechanism by which Rns acts and has dispelled the original view that Rns is only involved in the transcriptional regulation of pili in ETEC. It appears that Rns plays a wider role in the regulation of genes in ETEC. More study in this area may reveal novel genes and transcriptional networks involved in ETEC virulence.

Studies of the regulation of CF expression by AraC homologues have increased our understanding of ETEC virulence. However, they have also opened up the possibility of the development of drugs for the treatment of ETEC infections. Since Rns regulates the expression of 50% of all ETEC CFs, and possibly controls the expression of other types of virulence gene, a small molecule inhibitor of Rns might be a widely effective drug against ETEC infections (Basturea et al., 2008). Such a drug-like molecule has been developed to inhibit the AraC family of transcriptional regulator, ToxT, which controls pilus and cholera toxin expression in *V.cholerae* (Hung et al., 2005). Orogastric administration of the inhibitor, virstatin, was successful in preventing bacterial colonization of the intestine in a mouse model of infection (Hung et al., 2005). Therefore, the development of drugs against ETEC infections seems like a realistic possibility.

1.4.2 Phase-variation

Regulation of the *fot* operon, which encodes CS18 pili, appears to differ from that of other ETEC pilus operons. There is no evidence for an AFTR that controls the transcription of the *fot* operon. Instead, CS18 expression is controlled by phase-variation, mediated by a 312 bp invertible DNA element (Honarvar et al., 2003). The invertible element is located between the first gene of the operon and two recombinase genes, *fotS* and *fotT*, located upstream of the operon. The invertible element is presumed to contain the promoter for the *fot* operon, such that the orientation of the element determines whether the operon is transcribed.

The mechanism by which CS18 phase-variation occurs is similar to that of the common Type I pilus in *E. coli* (Abraham et al., 1985, Olsen and Klemm, 1994). The *fotS* and *fotT* genes show significant sequence similarity with each other and with the site-specific recombinase genes *fimB* and *fimE*, which control phase-variable expression of the *fim* operon, encoding Type I pili. It was proposed that FotS and FotT interact with the 16 bp inverted repeats

flanking the invertible element to mediate a recombination that inverts the element and switches transcription of the operon on or off. FotS mediates recombination of the invertible element in both ON and OFF orientations, while FotT favors recombinations resulting in the OFF orientation. Despite the similarity between the phase variation mechanisms for CS18 and Type I pili, the accessory DNA binding proteins, leucine-responsive regulatory protein (Lrp) and integration host factor (IHF), which participate in *fim* switching do not appear to play a role in phase variation of CS18.

The operons encoding CS12 and 987P pili from ETEC strains that infect humans and pigs respectively, are closely related to that of CS18. In addition, their structural genes are organized identically. Despite this similarity, however, the nucleotide sequences of the CS12 and 987P do not reveal recombinase genes which might control their expression via phase variation. Instead, both CS12 and 987P operons are associated with upstream AFTR genes. Thus it appears that different regulatory elements can be adapted to control the expression of evolutionarily related pilus systems in ETEC.

1.5 Aims of this study

The aims of this study were to investigate three related aspects of ETEC virulence; pilus mediated adherence, global regulation of ETEC genes by the pilus gene regulator, Rns, and the role of Rns-regulated factors on the innate immune response of intestinal cells to ETEC.

Aim 1: The mechanism of pilus adherence to erythrocytes and intestinal cells has been studied only in CFA/I, the archetypal ETEC pilus. The mechanism by which the minor, tip-associated pilin, CfaE, mediates binding to erythrocytes and intestinal cells has been studied. However, nothing is known about the mechanism of binding to glycosphingolipids by the major pilin, CfaB, or its biological significance. The specific aim of this part of the study was to identify amino acid residues in CfaB, by site-directed mutagenesis, which participate in pilus binding to glycosphingolipids. This would allow the generation of ETEC strains expressing CFA/I variants that are unable to bind glycosphingolipids, thereby providing the means to test the biological significance of glycosphingolipid binding in an animal model of intestinal colonisation.

Aim 2: Rns and the closely related homologue CfaD/CfaR were originally identified as a AraC-familiy transcriptional activators (AFTRs) of ETEC pilus genes. Subsequent studies have revealed that they also regulate the expression of four non-pilus genes. In a variety of Gram-negative intestinal pathogens, AFTRs simultaneously regulate the expression pili and a variety of virulence factors suggesting that Rns may regulate a wider range of genes, including virulence genes, in ETEC. The specific aim of this part of the study was to determine the complete Rns regulon, in order to identify new genes with potential roles in virulence.

Aim 3: In general, intestinal pathogens interact with the host at the mucosal intestinal epithelium, thereby eliciting innate immune responses that deterrmine the outcome of the infection. ETEC strains affect the secretion of various cytokines and antimicrobial peptides *in vivo* and *in vitro*. Rns regulates the expression of proteins at the bacterial cell surface, where interactions with the host epithelium occur. Therefore the aim of this part of the study was to test the hypothesis that Rns influences the cytokine response to ETEC.

Chapter-2: Materials

2.1 Antibiotics

Antibiotics used in this study are listed in Table 2.1 below. All antibiotic solutions were filter-sterilised before use.

Table 2.1 Antibiotics used in this study

Antibiotic	**Working concentration (µg/ml)**	**Stock concentration (mg/ml)**	**Solvent**	**Storage temperature**
Ampicillin	100	100	dH_2O	4°C
Chloramphenicol	50	50	Ethanol	-20°C
Gentamicin	Varied	100	dH_2O	4°C
Kanamycin	50	50	dH_2O	4°C
Tetracycline	20	5	Ethanol	-20°C

2.2 Primers

Table 2.2: Oligonucleotide primers used in the study.

Primer name	**Sequence (5'→3')**	**Description**
HSP245	AAA AAA CAT ATG AAA TTT AAA AAA ACT ATT GGT GCA A	For cloning of a donor-strand-complemented form of *cfaB* designated dsc_{19}CfaB$(His)_6$. Anneals to 28 bases of the *cfaB* sequence beginning with the start codon. Includes 6 additional A residues and an *Nde*I site at 5' end. Used in conjunction with HSP246
HSP246	AAA AAA CTC GAG TTG CAA AAG ATC AAT TAC AGG ATC AAC ACT AGC TGT TAC AGT AAT ATT TTT CTC TAC TTG TTT ATT ATC GGA TCC CAA AGT CAT TAC AAG	For cloning of a donor-strand-complemented form of *cfaB* designated dsc_{19}CfaB$(His)_6$. Includes 6 additional A residues, an *Xho*I site, sequences encoding a tetrapeptide linker (DNKQ) and the first nineteen amino acid residues of CfaB at 5' end. Remaining sequence anneals to the 21 codons of *cfaB* upstream of the stop codon.
HSP257	AAA AAG GAT CCA AAG GAT AAA CGA TG	For cloning of *cfaE* into *Bam*HI/*EcoR*I sites of pHSG576. Contains *Bam*HI site and 6 additional A residues at 5' end. Used in conjunction with HSP258. Overlaps RBS and start codon.
HSP258	AAA AAG AAT TCC TAG AGT GTT TGA CTACTA C	For cloning of *cfaE* into *Bam*HI/*EcoR*I sites of pHSG576. Contains an *Eco*RI site and 5 additional A residues at 5' terminus. Overlaps with the stop codon. Used in conjunction with HSP257.

HSP259	TTA TTT CTA GAG GGG AAT TG	For inverse PCR to amplify a vector containing pET22b(+) and the sequences encoding the tetrapeptide linker (DNKQ) and the codons for the 19 N-terminal amino acid residues of mature *cfaB* from pGU6 . Vector used for cloning a self-complemented *cfaE* gene. Used in conjunction with HSP260.
HSP260	GAT AAT AAA CAA GTA GAG AAA A	For inverse PCR to amplify a vector containing pET22b(+) and the sequences encoding the tetrapeptide linker (DNKQ) and the codons for the 19 N-terminal amino acid residues of mature *cfaB* from pGU6 . Vector used for cloning a self-complemented *cfaE* gene. Used in conjunction with HSP259.
HSP261	AAA ATC TAG AGA ATG ATA AAG GAT AAA CGA TG	For amplifying and cloning a self-complemented *cfaE* gene. Contains *Xba*I site at 5' site. Overlaps start codon of *cfaE* and includes ribosome binding site. Used in conjunction with HSP306.
HSP277	ACA CCA AGT AGT CAA ACA CTC TAG	For deletion of *cfaE* gene from pEU2124. Homologous to 3' end of the *cfaE*. Used in conjunction with HSP278.
HSP278	CTA TGTA AAA ATA AAT AAA ATT TTA TTC AT	For deletion of *cfaE* gene from pEU2124. A stop codon is incorporated at the 5' end of the primer so as to abolish any further translation of the template after the primer. Used in conjunction with HSP277.
HSP281	GCA ATG AGT TTT TTA TGT TTG TCT TCT CAA	For introducing R67A mutation in pEU2124. Contains GCA codon for alanine at 5' end. Used in conjunction with HSP282.
HSP282	ATC ATA CAG ATG ATG GCT TCC	For introducing R67A mutation in pEU2124. Used in conjunction with HSP281.
HSP294	CGC TGG GTA TTA TTG CAG ATG	For quantifying expression of *rns* in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP295.
HSP295	CGCA GCC TTA CTC ATT CTC A	For quantifying expression of *rns* in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP294.

HSP298	ACG CTG CTG TTG ACC TTC TT	For quantifying expression of *gyr*B in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP299.
HSP299	TGT TCC TGC TTG CCT TTC TT	For quantifying expression of *gyrB* in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP298.
HSP306	GAG TGT TTG ACT ACT TGG TGT GAA	For amplifying and cloning a self-complemented *cfaE* gene. Used in conjunction with HSP261.
HSP307	GCA CGA TAT GAT ACA ACC TAT G	For introducing R181A mutation in *cfaE* cloned in pHSG576. Contains GCA codon for alanine. Used in conjunction with HSP308.
HSP308	TTT TAC ATT TAG CTT CAG AAC G	For introducing R181A mutation in *cfaE* cloned in pHSG576. Used in conjunction with HSP307.
HSP321	GTA AAA CCG GGA CAG CAA GA	For quantifying expression of *cot*B in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP322.
HSP322	TTC CTC TGT GGG TTC TCA GG	For quantifying expression of *cot*B in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP321.
HSP323	AGA ACG TTT CCC AAA TGA CG	For quantifying expression of Pyruvate dehydrogenase in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP324.
HSP324	TAC TGA GCA CGC TCA ACA CC	For quantifying expression of Pyruvate dehydrogenase in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP333.
HSP333	ATT CAG ACT CAT CCG CTC GT	For quantifying expression of biofilm formation regulatory protein (BssS) in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP334.
HSP334	TCG GAC TTA TTT GGG GTC TG	For quantifying expression of biofilm formation regulatory protein (BssS) in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP333.
HSP343	TCT GTG GGA AGT CGA AAA TG	For quantifying expression of XRE family of transcriptional regulator in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP344.

HSP344	TTC AGG TTC CGC TGA GAA AT	For quantifying expression of XRE family of transcriptional regulator in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP343.
HSP345	AGA GAC CCC CTT TGC AAG TC	For quantifying expression of ABC transporter permease in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP346.
HSP346	TTT CAC TCC AAA GTC TGA ACG A	For quantifying expression of ABC transporter permease in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP345.
HSP347	GGT TAA CGG CAC CAT CAG TT	For quantifying expression of Malate synthase in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP348.
HSP348	GAC ATG TTT TTC CGG CAA GT	For quantifying expression of Malate synthase in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP347.
HSP364	CGC GAA GAA GCA GAG TCT TT	For quantifying expression of *yii*S in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP365.
HSP365	GCC CCT CCA GAT CCT TAA TC	For quantifying expression of *yii*S in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP364.
HSP366	ACC TAC CGC GCT GGA TAT AA	For quantifying expression of *nlpA* in ETEC B2C using qRT-PCR. Forward primer used in conjunction with HSP367.
HSP367	CCC TCG AGT TCC ATA ATC TGC	For quantifying expression of *nlpA* in ETEC B2C using qRT-PCR. Reverse primer used in conjunction with HSP366.
HSP415	AGC CCC CCC TTC ATC	For generating DNA fragment for EMSA on upstream sequence of *etpBAC* operon. Forward primer. Used in conjunction with HSP416.
HSP416	CCC TAT ATT TCC AGA CAC CTG TTA T	For generating DNA fragment for EMSA on upstream sequence of *etpBAC* operon. Reverse primer. Used in conjunction with HSP415.

2.3 Plasmids

Table 2.3 List of plasmids used in the study

Plasmid	Backbone Vector	Relevant Characteristics	Reference/ Source
pHSG576		P_{lac}-*lacZ'*	(Takeshita et al., 1987)
pMAL-c2X		P_{tac}-*malE-lacZ'*	New England Biolabs
pET22b(+)		P_{T7}, *lacI*	Novagen
pJGX15W	pJRD184	P_{tet}-*cfaABCE*	(Giron et al., 1995a)
pEU2124	pJRD184	P_{tet}-*cfaABCE1**	(Sakellaris et al., 1999a)
pGU1	pJRD184	P_{tet}-*cfaABC*	This Study
pGU2	pHSG576	P_{lac}-*cfaE*	This Study
pGU3	pHSG576	P_{lac}-*cfaE1**	This Study
pGU4	pHSG576	Plac-*cfaE2*†	This Study
pGU5	pHSG576	P_{lac}-*cfaE3*‡	This Study
pGU6	pET22b(+)	P_{T7}- dsc_{19}*cfaB*$(His)_6$	This Study
pGU7	pET22b(+)	P_{T7}- dsc_{19}*cfaE*$(His)_6$	This Study
pMal-c2X		P_{tac} - *lacZ'*	New England Biolabs
pBAD30		P_{ara}	(Guzman et al., 1995)

2.4 Bacterial strains

Table 2.4: List of bacterial strains

Strain	**Genotype/Characteristics**	**Reference/ Source**
E. coli DH5α/F'-*lacI*^{q}	F'[*lacI*^{q}] *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17*(r_K^- m_K^+), λ^-	Novagen
E. coli BL21 DE3 Rosetta	F$^-$ *ompT gal dcm lon hsdS$_B$*(r_B^- m_B^-) λ(DE3 [*lacI lacUV5-T7* gene 1 *ind1 sam7 nin5*]) pSRARE (*cat*)	Novagen
E. coli MC4100	F$^-$ [*araD139*]$_{B/r}$ Δ(*argF-lac*)169* *lambda*$^-$ *e14-flhD5301* Δ(*fruK-yeiR*)725 (*fruA25*) *relA1 rpsL150(strR) rbsR22 Δ(fimB-fimE)632(::IS1) deoC1*	H. Sakellaris
E. coli B2C	O6:H16:CFA/II strain of enterotoxigenic *E. coli*	(DuPont et al., 1971)
E. coli B2C Rif^r	Spontaneous rifampicin resistant mutant of B2C	H. Sakellaris
E. coli B2C Rif^r *rns*	ETEC B2C Rif^r::pCactus-*mob-rns'*; single cross-over homologous recombination of pCactus-*mob-rns'* into chromosome inactivates *rns* gene.	H. Sakellaris

Chapter 3: Methods

3.1 General molecular biology techniques

All of the general molecular biology techniques used, including restriction digestion, ligation, phosphorylation of 5' ends and PCR were performed with enzymes and chemicals procured from New England Biolabs, or Thermofisher Scientific and by following manufacturer's instructions. DNA sequencing was done at Australian Genome Research Facility Ltd (AGRF), Brisbane.

3.2 Ammonium sulphate precipitation

Ammonium sulphate (Sigma-Aldrich) was added according to a standard ammonium sulphate saturation table (Sambrook and Russell, 2001) and kept in a tube roller for 2hours at 4°C. The precipitate was centrifuged at 8,200 x *g* for 10 min, suspended in 4 ml of PBS and dialyzed against 1 litre of PBS for 18 h, with changes of dialysis buffer every 6 h.

3.3 Bacterial strains, plasmids and culture conditions

Cultures were grown aerated in Luria-Bertani (LB) broth or agar (Scott, 1974), or 2YT broth (Sambrook and Russell, 2001), or CFA agar or broth (Evans et al., 1980) at 37^0 C with antibiotic selection where appropriate. Antibiotics used are given in Table 2.1. The *E. coli K*-12 strains DH5α (Hanahan, 1983) and DH5α/F'-*lacI*^{q} (New England Biolabs) were used as general hosts for cloning. *E. coli* strain BL21(DE3) Rosetta (Novagen) was used for high-level expression under the T7 promoter. MC4100 is a *lac*-deletion *E. coli* K-12 strain (Casadaban, 1976) used for the expression of the *cfa* genes under the control of the *lac*

promoter. Further details of the strains are given in Table 2.4. Plasmids used or constructed in this study are listed in Table 2.3.

3.4 Construction of a *cfaE* deletion of the CFA/I pilus operon

The *cfaE* gene was deleted from the CFA/I pilus operon by amplifying part of pEU2124, excluding *cfaE*, by inverse PCR using Phusion High Fidelity DNA polymerase (Thermofisher Scientific), with the primers HSP277 and HSP278. The amplicon was phosphorylated using T4 polynucleotide kinase (New England Biolabs) and ligated using T4 DNA ligase (New England Biolabs). DH5α/F'-*lacI*^{q} was transformed with the ligation reaction and presumptive clones were screened by PCR. The deletion, which was defined by the start and stop codons of *cfaE*, and the fidelity of the remaining sequence, was confirmed by DNA sequencing. The *cfaE* deletion construct was given the plasmid designation, pGU1 (Table 2.3).

3.5 Construction of *cfaE* complementation plasmid

The *cfaE* gene was PCR-amplified from pJGX15W using primers HSP257 and HSP258 and ligated into the *Bam*HI / *Eco*RI sites of pHSG576 Table 2.3. The construct was confirmed by DNA sequencing of the entire cloned insert and designated pGU2 (Table 2.3).

3.6 Site-directed mutagenesis of *cfaE*

To substitute arginine residues for alanine at positions 67 and 181 in the CfaE amino acid sequence, site directed mutations were introduced into pGU2. To introduce an R181A mutation into *cfaE*, pGU2 was PCR-amplified with primers HSP307 and HSP308. HSP307 contains a 5'-terminal alanine codon, GCA, which substituted the AGA arginine codon in the wild-type sequence. The amplicon was phosphorylated and ligated as described previously. Ligated samples were transformed into DH5α/F'-*lacI*^{q} and putative clones were confirmed by

DNA sequencing. The *cfaE* gene carrying the R181A mutation was designated *cfaE1* and the corresponding plasmid was designated pGU3. Similarly, an R67A mutation was generated by PCR-amplification of pGU2 with the primers HSP281 and HSP282. HSP281 contains a 5'-terminal alanine codon, GCA, which substituted the AGG arginine codon in the wild-type CfaE sequence. Clones carrying the *cfaE2* mutant allele were confirmed by DNA sequencing and designated pGU4. A double mutation (R67A, R181A) was generated by PCR amplification of pGU4 with primers HSP307 and HSP308.

3.7 Purification of pili

800ml LB broth cultures were incubated overnight at 37°C. Bacterial cells were removed by centrifugation at 8,000 x *g* for 10 min and cell pellets were washed in 60ml of phosphate buffered saline (PBS) before centrifugation again at 8,000 x g for 10 min. Cell pellets were suspended in 60 ml of PBS and subjected to mechanical shearing for 10 sec in a food blender. Bacterial cells were removed by centrifugation as above, and protein in the supernatant was precipitated at 10% (w/v) ammonium sulphate saturation. The precipitate was centrifuged at 8,200 x *g* for 10 min, suspended in 4 ml of PBS and dialyzed against 1 litre of PBS for 18 h; with changes of dialysis buffer every 6 h. Dialyzed pilus preparations were checked for purity by SDS PAGE and immunoblotting with anti-CFA/I antiserum and stored at 4°C.

3.8 Cloning, expression and purification of *in cis* donor strand-complemented pilins

In cis donor strand-complemented CfaB and CfaE, with the same amino acid sequence as previously described(Poole et al., 2007), were constructed by the addition of a C-terminal extension consisting of a 4 amino acid flexible hairpin loop, the 19 amino acid N-terminal sequence of mature CfaB, and a hexahistidine tag. The fusion proteins were given the same

designations, $dsc_{19}CfaB(His)_6$ and $dsc_{19}CfaB(His)_6$, used by Poole et al (Poole et al., 2007) To construct the dsc19CfaB$(His)_6$ fusion protein, *cfa*B was PCR-amplified from pJGX15W using primers HSP245 and HSP246 . HSP245 contains an introduced *Nde*I restriction site for cloning and anneals to 28 bases of the *cfaB* sequence beginning with the start codon. HSP246 contains an introduced *Xho*I restriction site for cloning, a sequence encoding a tetra-peptide linker (DNKQ), the first nineteen amino acid residues of CfaB and a sequence that anneals to the 21 codons of *cfaB* upstream of the stop codon. The PCR amplicon was digested with *Nde*I and *Xho*I and ligated into corresponding restriction sites of pET22b (+) (Novagen) downstream of the T7 promoter. This cloning resulted in an in-frame fusion of the extended *cfaE* insert with the hexahistidine coding sequence of the plasmid vector. The clone was confirmed by double restriction digestion and DNA sequencing and given the designation pGU6 (Table 2.3).

To construct the $dsc_{19}CfaE(His)_6$ fusion protein, the coding sequence for CfaB in pGU6 was replaced by the coding sequence for CfaE, while retaining the sequences encoding the tetra-peptide linker and donor strand-complementing, N-terminal sequence of CfaB. To achieve this, a section of pGU6 lacking the CfaB coding sequence was amplified using outward facing PCR primers HSP259 containing the vector-borne *Xba*I site and HSP260. The *cfa*E gene was PCR-amplified using primers HSP261 containing an *Xba*I site and HSP306. Both amplicons were phosphorylated with T4 polynucleotide kinase (New England Biolabs), digested with *Xba*I and ligated. Plasmid clones in which the CfaB coding sequence was replaced by the CfaE coding sequence were confirmed by DNA sequencing and given the designation pGU7 (Table 2.3)

To express and purify the fusion proteins, BL21(DE3) Rosetta/pGU6 and BL21(DE3) Rosetta/pGU7 were grown with aeration in 1 litre of 2YT broth supplemented with 0.4% (v/v) glycerol and appropriate antibiotics at 37°C. At OD_{600} = 0.4, IPTG was added to a

concentration of 1 mM and cultures were incubated at 16°C for 16 hours. Cells were harvested by centrifugation and resuspended in lysis buffer (50mM Tris-HCl pH 8, 300mM NaCl, 10% (v/v) glycerol, lysozyme 300µg/ml) supplemented with and protease inhibitor cocktail (Roche) as recommended by the manufacturer. The cell suspension was incubated for enzymatic lysis overnight at 4°C before centrifugation at 104,000 x *g* for 1 h (Beckman Optima L-100XP Ultracentrifuge). The presence of fusion protein was monitored in both the soluble and insoluble fractions by SDS PAGE and immunoblotting using anti-hexahistidine antibody (Sigma Aldrich). Proteins in the soluble fractions were precipitated by the addition of ammonium sulphate to 30% saturation and dialysed three times, for 6 h, against 1 litre changes of Ni-NTA binding buffer (sodium phosphate buffer pH 7.5, 300mM NaCl, 10% (v/v) glycerol, 0.02% (v/v) triton X-100, 25mM imidazole) supplemented with protease inhibitor cocktail (Roche). Protein solutions were loaded onto a Histrap column (Sigma) at a rate of 0.5 ml/min using a GE-AKTA Start System. Columns were washed with washing buffer (sodium phosphate buffer pH 7.5, 500mM NaCl, 50 mM imidazole, 10% (v/v) glycerol) and proteins were eluted with an imidazole concentration gradient of 100-500 mM in wash buffer. The purity of eluents was confirmed by SDS PAGE. Purified fractions were pooled and dialysed against PBS (pH7).

3.9 Transmission electron microscopy

Whole bacterial cells and purified pili were prepared, negatively stained with phosphotungstic acid and examined by transmission electron microscopy as previously described (Froehlich et al., 1995). Bacterial strains were grown overnight on CFA agar at 37° C with appropriate selection. The medium was LB agar if recombinant strains were being examined. In an Eppendorf tube, a small sample of each strain was suspended in 100 µl of buffer containing (1 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$). Bacteria were allowed to float

on the surface for 60 min at room temperature. 10 µl of the bacterial suspension was diluted into 15 µl of buffer on a parafilm strip. A formvar coated carbon grids(200 mesh) was floated on the diluted sample for 30 min. The grid was laid face-up on parafilm and excess bacterial suspension was removed with a paper wick. 10-15 µl of 0.5% (w/v) phosphotungstic acid (pH-7) was placed onto the grid surface and stained for 1min and finally wick dried.

3.10 Production of antiserum

A New Zealand White rabbit was inoculated subcutaneously with 200 µg of CFA/I pili in complete Freund's adjuvant. Repeat inoculations were made on days 30 and 54 after the first inoculation, with 200 µg of CFA/I pili in incomplete Freund's adjuvant. Serum was collected on day 76.

3.11 Pilus and pilus subunit binding assays

Binding of pili and pilus subunits to asialo-GM1 (Sigma Aldrich) was measured with an ELISA-based assay. Microtitre plates (Immulon 1B ELISA plate, Thermo Scientific) were coated with 100 µl of asialo-GM1 at 5µg/ml final concentration. Plates were air-dried at 4°C for 12 hours, after which they were washed five times with wash buffer consisting of PBS+0.5% (w/v) BSA. Wells were blocked with blocking buffer, consisting of PBS+5% (w/v) BSA, overnight at 4°C. Purified pili were subsequently added to asialo-GM1 coated wells and uncoated wells, which served as negative controls. No-pilus control solutions were also added to coated and uncoated wells. Microtitre plates were incubated for 1 hour at 37°C and washed five times with wash buffer before the addition of anti-CFA/I antiserum, diluted 1/2000 in blocking buffer, and incubation for 1 h at room temperature. After five washes to remove unbound antibody, anti-rabbit IgG-alkaline phospatase conjugate (Santacruz) was added at a dilution of 1/5,000 and incubated for 1 hour at room temperature. After five washes, 100µl of detection reagent (5mM diethanolamine, 2mM NaCl, 2mM $MgCl_2$, 15 mg

of *p*-nitrophenyl phosphate disodium hexahydrate, pH9.5) was added to each well and incubated at 37°C for 45mins. The absorbance of each well was measured at 405 nM. Absorbance values for no-glycolipid controls were subtracted from corresponding wells containing asialo-GM1. The values for no-pilus control wells were subtracted from the values for test wells.

For assays of pilus binding to erythrocytes, human type A blood was obtained from the Australian Red Cross blood service. Erythrocytes were washed by adding 200µl of blood to 10ml of TBS (50mM Tris-HCl, 150mM NaCL, pH7.5) and centrifuging at 3,000 x g for 3 minutes. The retained pellet was washed an additional three times and the final pellet was resuspended in 1ml of TBS before adjustment of the OD_{600} to a value of 1 (10^7 erythrocytes/ml). 60 µL of this suspension was added to the wells of an Immulon 1B ELISA plate (Thermo Fisher Scientific, Australia) and 60 µL of TBS was added to each erythrocyte-negative control well. Plates were dried overnight at 4°C. All subsequent steps involving, blocking, pilus binding, washing and detection were performed as described above for asialo-GM1 binding assays.

For assays of pilus binding to Caco-2 cells, cells were grown in 96-well tissue culture plates (Falcon) to 80-90% confluence in RPMI medium supplemented with 20% (v/v) fetal bovine serum. All subsequent steps in the Caco-2 binding assay were performed as described for asialo-GM1 binding assays.

Each binding experiment consisted of three technical replicates with the same pilus or protein preparation and three biological replicates with independently derived preparations. Differences in the mean values for biological replicates were analyzed for statistical significance with a one-tailed, equal variance Student's *t*-test.

3.12 Assay of bacterial adherence to Caco-2 cells

Caco-2 cells were grown to confluence in 24 well tissue culture plates and inoculated with suspensions of MC4100/pGU1/pGU2 (wild-type pili), MC4100/pGU1/ pGU5 (mutant pili with R67A, R181A substitutions in CfaE) and MC4100 (pilus-negative) at an MOI of 100. Bacteria and Caco-2 cells were co-incubated in RPMI medium for 2h at 37^0C, washed with PBS five times and then treated with 0.1% (v/v) Triton X-100. The Triton X-100 suspension was serially diluted and plated onto LB Ap Cm agar plates. After overnight incubation at 37°C, CFUs on each plate was calculated. Statistical analysis was performed as per pilus and subunit binding assays.

3.13 Whole genome sequencing and analysis of ETEC B2C strain

Genomic DNA was extracted from B2C with the Fermentas GeneJET genomic DNA purification kit (Thermo Fisher Scientific, Australia), according to the manufacturers' protocol. One hundred nanograms of genomic DNA was used to generate a library using the Nextera DNA sample prep kit (Illumina, CA). The libraries were sequenced at the Queensland Centre for Medical Genomics, University of Queensland, Australia, on an Illumina HiSeq 2000 sequencing system (Illumina, CA). The read pairs were overlapped where possible using the SeqPrep software (https://github.com/jstjohn/SeqPrep) and trimmed for quality using the Nesoni software(http://www.vicbioinformatics.com/software.nesoni.shtml). A *de novo* assembly of the overlapped and quality trimmed reads was generated using Velvet version 1.2.07 (Zerbino and Birney, 2008). The draft genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) and manually curated using Artemis (Rutherford et al., 2000). The draft genome sequence of ETEC strain B2C was deposited in Genbank with accession no. AUZS00000000 (Mdhavan et al., 2014).

3.14 PCR based analysis of ETEC B2C *rns* mutant

Genomic DNA of putative ETEC B2C *rns* mutants (ETEC B2C Rif/pCactus-*mob-rns'*) and ETEC B2C Rif/pBBR1MCS was extracted using Fermentas GeneJET genomic DNA purification kit (Thermo Fisher Scientific, Australia), according to the manufacturers' protocol. Primers HSP14 and HSP15 was used to amplify the plasmid pCactus insertion site. Amplicon size was analysed using agarose gel electrophoresis.

3.15 RNA extraction and purification

Bacterial strains ETEC B2C WT/pBBR1MCS and ETEC B2C Rif/pCactus-mob-rns' were grown in 3mL of colonisation factor antigen (CFA media) overnight at 37°C in a shaking incubator to make a primary inoculum. To 20ml of CFA broth, primary inoculum was inoculated at 1 % (v/v) concentration, incubated until an OD_{600} of 0.4 was reached. Two volumes of bacterial RNA protection reagent (Qiagen RNAprotect Bacteria Reagent) was added and the culture was centrifuged at 14,000 x *g*. The bacterial cell pellet was frozen at -80°C. Bacterial RNA was extracted from the frozen pellets using the Promega SV Total RNA Purification kit. Genomic DNA was removed using TURBO™ DNase (Thermo Fisher Scientific) following the manufacturers guidelines. Final RNA samples were quantified using. NanoDrop (NanoDrop products, USA)

3.16 RNA-seq

Ribosomal RNA(rRNA) was removed from purified RNA of both B2C WT and B2C *rns*/pBBR1MCS using Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina USA) following manufacturers guidelines. RNA-Seq libraries was generated using ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Illumina USA) following manufacturers guidelines. cDNA was purified using AMPure XP magnetic beads (Beckman Coulter USA) and quality check was

conducted using an Agilent BioAnalyzer (Agilent Technologies USA). Sequencing of cDNA libraries were done using MiSeq sequencer (Illumina USA). Quality check of reads were conducted FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and sequences were trimmed using Trimmomatic (Bolger et al., 2014) and rRNA sequences were removed using SortMeRNA (Kopylova et al., 2012). Reads were mapped to reference (ETEC B2C genome sequence) using bwa (http://bio-bwa.sourceforge.net/). Differential expression between B2C WT and B2C *rns*/pBBR1MCS was obtained using R (limma package) (Law et al., 2014).

3.17 qRT-PCR

cDNA was synthesised using M-MuLV Reverse Transcriptase (NEB, Australia) following the manufacturers guidelines. Reaction mix for qRT-PCR was prepared using SensiFAST SYBR No-ROX Kit (Bioline) and quantified using the LightCycler® 480 System (Roche) following the manufacture's guidelines. The internal control for all qRT-PCR reactions was *gyrB*. All of the primers used are listed in Table 2.2.

The relative fold difference in cDNA from each sample was calculated using the $2^{-\Delta\Delta C_T}$ method as previously described (Livak and Schmittgen, 2001). Graphs were plotted using GraphPad Prism (version 6).

3.18 Construction of ETEC B2C Rif^r*rns*/pBAD30-*flag-rns*:

The *rns* sequence containing an N terminal flag tag coding sequence (5'-GACTACAAAGACGATGACGACAAG-3') was synthesised (GenScript, USA) and inserted into the *Sac*I and *Hin*dIII sites of pBAD30. The clone was confirmed by DNA sequencing. ETEC B2C Rif^r*rns* was transformed with the clone using a Biorad electroporation system (Bio-Rad).

3.19 Construction of ETEC B2C Rif^r*rns*/pBAD30:

The *rns* mutant strain, B2C Rif^r *rns* was transformed with pBAD30 and transformants were selected with both chloramphenicol and ampicillin.

3.20 Chromatin Immunoprecipitation

ETEC B2C Rif^r/pCactus-*mob-rns'*/pBAD30-*flag-rns* and ETEC B2C Rif^r/pCactus-*mob-rns'*/pBAD 30 were grown in 50ml CFA broth supplemented with chloramphenicol and ampicillin to an OD600 of 0.4. At OD = 0.4 formaldehyde (Sigma-Aldrich) was added to a final concentration of 1% (w/v), and incubated for 20 min on ice. The reaction was quenched by adding glycine to a final concentration of 0.5M and incubating for 10 min on ice. The culture was centrifuged at 4,000 x *g* for 15 min. The cell pellet was washed in ice cold PBS (pH 7.5) thrice. The cell pellet was then resuspended in 1 ml of lysis buffer (20mM Tris-Cl [pH-8], 20% (w/v) sucrose, 200mM NaCl, 10 mM EDTA, 20 mg/ml lysozyme and 0.1mg/ml RNase A and Roche protease inhibitor), and incubated in ice for 30mins with shaking. 4 ml of IP buffer (50mM Hepes-KOH [pH-7.5], 200 mM NaCl, 10 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) sodium dodecyl sulphate and Roche protease inhibitor) was added. Samples were sonicated at 10% output for 20 sec, cooled on ice for 2–5 min, and repeated four to five times until the DNA fragments were between 100-500bp. Cell debris was removed by centrifugation at 17,000 x *g* for 15 min at 4°C. The cell debris was discarded and supernatant was stored. To 1ml of supernatant, 5µl of anti FLAG antibody was added (Cell Signaling Technology, USA), and incubated overnight at 4 °C in a rotator. 20µl of Protein A magnetic beads (New England Biolabs) was added to the solution and incubated for 90 min at room temperature in a rotator. Magnetic beads were collected using a magnetic stand (New England Biolabs). Beads were washed twice with IP buffer, once with IP buffer containing 500mM NaCl, once with wash buffer (10 mM Tris-HCl pH

8.0, 250 mM LiCl, 1 mM EDTA, 0.5% (v/v) Nonidet-P40, 0.5% (w/v) sodium deoxycholate), and finally with Tris-EDTA buffer (pH8). Protein-DNA complex was eluted in elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% (w/v) SDS) at 65°C for 30 min. Immunoprecipitated samples were decrosslinked by incubation for 2 h at 42°C and for 6 h - overnight at 65°C in 0.5 x elution buffer containing pronase (0.8 mg/ml). Decrosslinked samples were purified using a Qiagen PCR purification kit and resuspend in 100 µl.

3.21 Sequencing and analysis of Chip DNA samples

Sequencing and analysis of Chip DNA samples were conducted as described previously (Perkins et al., 2013).

3.22 Construction and purification of MBP-Rns fusion protein

The *rns* gene was PCR amplified from ETEC B2C Rif^r genomic DNA using primers HSP279 and HSP280. The amplicon was purified, restricted with *Xba*I and *Hin*dIII and ligated with similarly restricted pMAL-C2X. Confirmed clone was transformed into BL21(DE3) Rosetta. Expression and purification was conducted as described previously (Munson and Scott, 1999).

3.23 Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assays (EMSA) were conducted as described previously (Munson and Scott, 1999) using an EMSA kit (Life Technologies).

3.24 Heat extraction of pili

Crude pilus preparations were produced by heat treatment of piliated strains which results in the release of intact pili into the supernatant. Typically 100 ml bacteria were grown with aeration, overnight at 37^{o}C in CFA broth with appropriate antibiotic. Cultures were

centrifuged at 6000 x *g* for 10 min and the cell pellet was resuspended in 1ml of 1x PBS. Cell suspensions were heated at 65°C for 15 mins. Suspensions was centrifuged at 12800 x *g* for 30 min and supernatants were collected.

3.25 Mannose resistant haemagglutination assay (MRHA)

Bacterial strains were grown overnight at 37°C, on agar media with appropriate antibiotic selection. Bacterial growth was transferred to 400 µl of 1x TBS + 0.1M D-mannose, using a sterile swab. Erythrocytes were prepared by resuspending 200 µl of human type-A blood in 800µl of 1xTBS+ 0.1M D-mannose. Suspensions were centrifuged at 3000 x *g* for 1 min and the pellet was retained. This washing step was repeated 3 times and cells were finally suspended in 100 µl of 1xTBS+ 0.1M D-mannose. The MRHA assay was performed by mixing a 20 µl erythrocyte suspension with a 20 µl bacterial cell suspension on a glass slide and tilting the slide gently at different angles to ensure that the suspensions were mixed well for few minutes. Slides were examined for haemagglutination under light.

3.26 Caco-2 cell culture

Caco-2 cells (ATCC HTB-37) were obtained from ATCC and maintained using the recommended culture conditions and RPMI1640 culture media (Life Technology) with 20% (heat-inactivated) foetal bovine serum (FBS).

3.27 Adhesion assay

Caco-2 cells were grown and passaged as recommended by ATCC. After reaching a confluence of 80-90%, 1ml of 50 000 cells/ml were seeded into 24 well plates and cultured with appropriate antibiotic selection. After 24 hours, the plates were washed with PBS three times. To prepare the bacterial challenge inoculum 3ml of bacterial culture, grown in CFA media overnight at 37°C, were spun down and 7000RCF for 15 mins and resuspended in 3ml

of PBS. The procedure was repeated three times and resuspended to 1ml. The final suspension was diluted using Caco-2 growth medium containing appropriate antibiotics to achieve the target MOI. Infected Caco-2 cells were incubated for 2 hours at 37°C and then washed with PBS three times and lysed in 1ml of 0.01% triton X-100. The CFUs of bacteria were determined by serially diluting Caco-2 cell lysates on CFA agar. To measure invasion, the infected Caco-2 cells were treated with 200µg/ml of Gentamycin (Sigma-Aldrich) in RPMI supplemented with 20% heat inactivated FBS for 15mins following the infection period (2 hours). The supernatants were subsequently removed and the monolayers washed with PBS three times as described previously, prior to lysing the cells and plating serial dilutions for colony counts.

3.28 Bio-Plex assay

To measure cytokine expression of host cells following infection 400µl of supernatant was collected after the 2 hours of infection and stored for downstream assay. Cytokine expression was quantified the using Bio-Plex Pro™ Human Cytokine 27-plex Assay (M500KCAF0Y) in a Bio-Plex® 200 system (Bio-Rad). The results were analysed using Microsoft excel and GraphPad Prism Version 6.0.

Chapter 4: Roles of the Minor and Major Subunits of CFA/I Pili of ETEC in Glycosphingolipid Binding

T.P Vipin Madhavan, Glen C Ulett and Harry Sakellaris

This chapter contains unpublished paper.

My contribution in this work was design of experiment, conducting experiment, analysis and writing the manuscript.

(Signed) _____

Vipin Madhavan Thekke Palasseri

(Countersigned) _____

Supervisor: Dr Harry Sakellaris

4.1 Introduction

ETEC strains express a wide variety of colonization factors, most of which are pili (fimbriae) (Vipin Madhavan and Sakellaris, 2015). The CFA/I pilus was the first colonization factor to be identified in ETEC (Evans et al., 1977a). It belongs to a large group of related ETEC pili belonging to the α clade of the fimbrial usher protein family (FUP) (Vipin Madhavan and Sakellaris, 2015), which overlaps substantially with the Class 5 pili defined by an earlier pilus classification scheme (Anantha et al., 2004). This pilus family includes the colonization factors CFA/I, CS1, CS2, CS4, CS14, CS17, CS19 and CS5 (Vipin Madhavan and Sakellaris, 2015). The majority of what is known about pilus assembly, structure and function in this family comes from studies of CS1 and CFA/I pili (Bao et al., 2014, Chen et al., 2013, Chattopadhyay et al., 2012, Tchesnokova et al., 2010, 2008, Li et al., 2007, Anantha et al., 2004, Sakellaris et al., 1999a, Starks et al., 2006, Sakellaris et al., 1996, Froehlich et al., 1995). In both cases, pili are assembled via the alternate chaperone-usher pathway and consist of only two pilus subunits (pilins); a major pilin that comprises the pilus shaft and a single pilus tip-associated pilin. The other components of these pilus systems are a periplasmic pilin chaperone and an usher protein, which are essential for pilus assembly. The pilus tip subunits, CooD in the case of CS1 pili and CfaE in the case of CFA/I pili, are responsible for pilus binding to erythrocytes (Li et al., 2007, Sakellaris et al., 1999a). CfaE also mediates binding of CFA/I to Caco-2 intestinal cells (Anantha et al., 2004), and to cultured small intestinal biopsy tissue (Baker et al., 2009).

It is thought that the receptor for CFA/I pili may be a glycolipid. Studies of the binding activities of ETEC pili indicate that CFA/I, and a number of other ETEC pili, bind to the glycosphingolipid, asialo-GM1, and a variety of other non-acid glycosphingolipids (Jansson et al., 2006, Bachovchin et al., 1990). However, whereas the binding activity for erythrocytes

and intestinal epithelial cells is specified by CfaE, it has been reported that the binding activity for glycosphingolipids resides in the major pilin, CfaB (Jansson et al., 2006). Thus, the current model for CFA/I-mediated adherence proposes that the pilus has two unrelated binding activities; one for unidentified receptors on the surface of erythrocytes and intestinal epithelial cells, and another for asialo-GM1 and a variety of other glycosphingolipids (Jansson et al., 2006).

The initial aim of the current study was to elucidate the specific mechanism by which CfaB binds glycosphingolipids. During the course of this study, we were unable to demonstrate the direct binding of purified CfaB pilins to asialo-GM1, which served as a model glycosphingolipid. We were also unable to demonstrate CfaB-dependent binding of whole CFA/I pili to asialo-GM. Instead, we demonstrate that asialo-GM1 binding is associated with CfaE. Purified CfaE binds to asialo-GM1, while purified CfaB has no detectable binding activity. Consistent with these findings, site-directed mutations in CfaE abolished the binding of whole pili to asialo-GM1 and, in addition, abolished the binding of pili to Caco-2 intestinal cells. These same mutations also abolished CFA/I-mediated attachment of piliated bacteria to Caco-2 cells. We therefore conclude that CfaB is not able to bind to asialo-GM1 independently of CfaE and that asialo-GM1 binding is inseparable from CfaE-mediated binding of pili and piliated bacteria to erythrocytes and intestinal cells.

4.2 Results

4.2.1 Expression of CfaE is an absolute requirement for CFA/I pilus assembly

To study CfaB-mediated pilus binding to asialo-GM1, we attempted to purify CFA/I pili from an *E. coli* strain which expresses three of the four genes of the CFA/I operon, *cfaA, B and C.*

This approach was predicated on a prior report that the expression of *cfaA, B and* C is sufficient for the assembly of pili consisting of the major pilin, CfaB, but devoid of CfaE (Jansson et al., 2006). To construct such a mutant, a defined deletion of the *cfaE* gene was generated in pEU2124 (see Methods 3.4), yielding the plasmid pGU1 (Table 2.3). To develop a quantitative, ELISA-based assay for pilus binding to asialo-GM1, pili were extracted and purified from *E. coli* strain MC4100/pGU1/pGU2 (*cfaABCE*) and MC4100/pGU1/pHSG576 (*cfaABC*). Strain MC4100, which served as a negative control for pilus production, was subjected to the same process as pilus-producing strains. Pili prepared from each strain were examined by SDS PAGE and immunoblotting with anti-CFA/I antiserum (Fig. 4.1).

As expected, no pili were extracted from the negative control strain, MC4100 (Fig. 4.1, lane 1; panels A and B). A protein of approx. 14 kDa which reacted with anti-CFA/I antiserum (Fig. 4.1, lane 4; panels A and B), was consistent with CfaB, in the pilus preparation from MC4100/pGU1/GU2 (*cfaABCE*). Transmission electron microscopy (See methods 3.9) of the negatively stained samples showed the typical morphology expected of CFA/I pili (Fig. 4.1, panel C). In contrast to MC4100/pGU1/GU2, pili could not be purified from strain MC4100/pGU1/pHSG576 as judged by SDS PAGE, immunoblotting or electron microscopy (Fig. 4.1, lane 2, panels A and B; Fig. 4.1, panel D). The ability of MC4100/pGU1/pHSG576 to produce pili was investigated further by electron microscopic examination of negatively stained whole bacterial cells. As expected, the negative control strain, MC4100, was non-piliated (Fig. 4.2, panel A), while the positive control strain, MC4100/pGU1/GU2, was piliated (Fig. 4.1, panel B). However, cells of MC4100/ pGU1/pHSG576, which do not express CfaE, were non-piliated (Fig. 4.1, panel C). Examination of 500 individual cells by electron microscopy, failed to reveal any piliated cells. It was therefore concluded that expression of the minor pilin is essential for the assembly of pili.

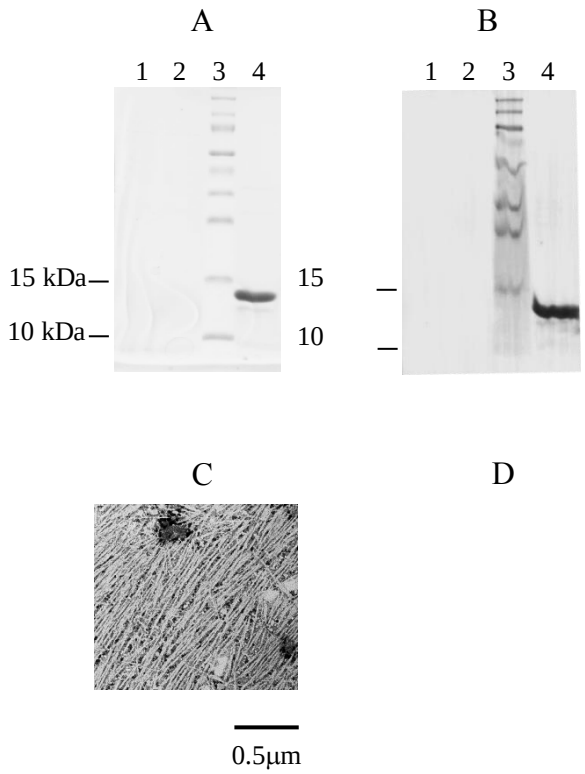

Fig. 4.1 Analysis of pili purified from strains of *E. coli* expressing *cfaABC* and *cfaABCE*.

SDS PAGE (panel A) and immunoblot (panel B) analysis of pilus preparations from: Lane 1, negative control strain MC4100; Lane 2, test strain MC4100/pGU1/pHSG576 (*cfaABC*); Lane 4, positive control strain MC4100/pGU1/pGU2 (*cfaABCE*). Lane 3 contains molecular weight markers. Transmission electron microscopy of negatively stained pilus preparations confirmed that although MC4100/pGU1/pGU2 produces CFA/I pili (panel C), test strain MC4100/pGU1/pHSG576 does not produce pili (panel D).

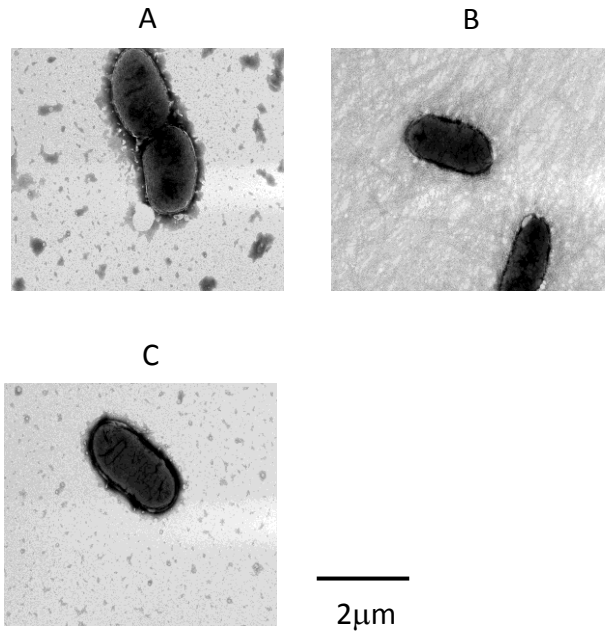

Fig. 4.2 Expression of all four *cfa* genes is essential for CFA/I pilus production.

Panel A: Negative control strain MC4100, which lacks the *cfa* genes, does not express pili on the cell surface. Panel B: Positive control strain MC4100/pGU1/pGU2 (*cfaABCE*) produces CFA/I pili. Panel C: Test strain MC4100/pGU1/pHSG576 (*cfaABC*) fails to produce pili.

4.2.2 Purified CfaE subunits, but not CfaB subunits, bind to asialo-GM1

As pili consisting of only the CfaB pilin could not be generated from strains expressing CfaA, B and C, an alternative approach involving the isolation of soluble CfaB pilins was employed to study the interaction of CfaB with asialo-GM1. Stable, natively folded CfaB protein was purified independently of CfaE by generating a donor strand-complemented, hexahistidine-tagged form of the pilin as previously described (Abedon, 2009, Li et al., 2007, Abedon and Culler, 2007). The fusion protein, termed dsc_{19}CfaB$(His)_6$, has an amino acid sequence which is identical to that of a CfaB fusion protein that was used to study donor strand complementation in CFA/I pili (Li et al., 2007). Similarly, a donor strand-complemented CfaE fusion protein, dsc_{19}CfaE$(His)_6$, was constructed and purified as a negative control for asialo-GM1 binding experiments. The CfaB and CfaE fusion proteins were purified and analysed by SDS PAGE and immunoblotting with anti-hexahistidine antibody (See methods 3.8). The purified CfaB fusion protein reacted with anti-hexahistidine antibody and had an apparent MW of 19-20 kDa (Fig. 4.3, panels A and B), as previously observed (Li et al., 2007), which is consistent with its predicted MW of 18.6kDa. The purified CfaE fusion protein also reacted with anti-hexahistidine antibody and had an apparent MW of 41 kDa, consistent with its predicted amino acid sequence (Fig. 4.3, panels C and D).

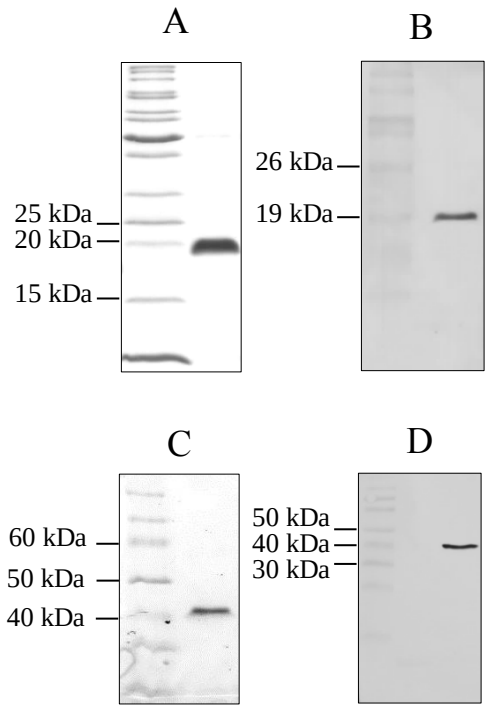

Fig. 4.3 Purification of donor strand-complemented CfaB and CfaE subunits.

The identity and purity of donor strand complemented CfaB (panels A and B) and CfaE (panels C and D) were confirmed by SDS PAGE (panels A and C) and immunoblotting with anti-hexahistidine antibody (panels B and D). The electrophoretic migrations of CfaB and CfaE fusion proteins were consistent with their predicted molecular weights of 18.6kDa and 41 kDa, respectively.

The binding of $dsc_{19}CfaB(His)_6$ to asialo-GM1 was tested in an ELISA-based assay (See methods 3.10 & 3.11) (Fig. 4.4). The $dsc_{19}CfaE(His)_6$ protein and heat-inactivated preparations of both dsc19CfaB$(His)_6$ and $dsc_{19}CfaE(His)_6$ served as negative controls for asialo-GM1 binding. Surprisingly, the negative control protein, $dsc_{19}CfaE(His)_{6,}$ bound strongly to asialo-GM1-coated ELISA plates while the test protein, $dsc_{19}CfaB(His)_6$, bound poorly. The level of $dsc_{19}CfaB(His)_6$ binding was identical to that of heat-inactivated dsc19CfaB$(His)_6$ and dsc19CfaE$(His)_6$, indicating that this low level of binding was non-specific.

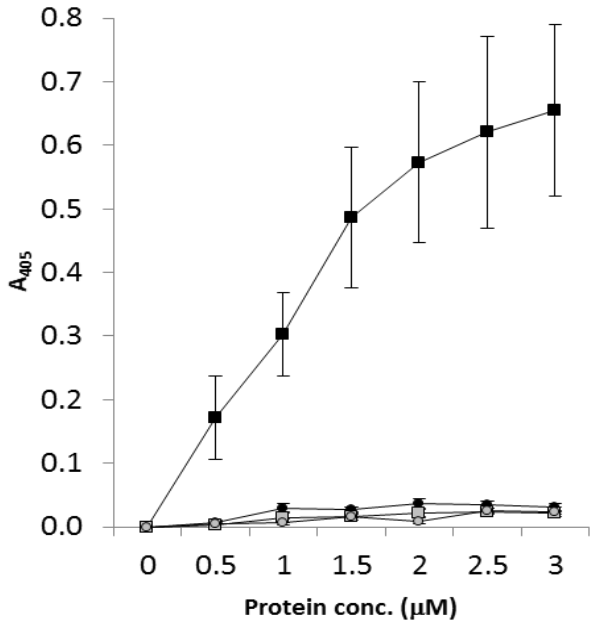

Fig. 4.4 Binding of donor strand-complemented CfaB and CfaE subunits to asialo-GM1.

Asialo-GM1 binding of donor strand-complemented subunits and negative-control, heat-inactivated subunits were compared in an ELISA-based assay. Purified CfaE subunits (black squares) bound to asialo-GM1, while binding of CfaB (black circles) was no greater than that of negative control, heat-inactivated CfaE and CfaB subunits (gray squares and circles, respectively). Data points are the means of three biological repeats, each consisting of 3 technical repeats, +/- SEM.

The ability of CfaE to bind asialo-GM1 has not been reported previously. However, the inability of $dsc_{19}CfaB(His)_6$ to bind asialo-GM1 is inconsistent with a previous report that pili consisting soley of CfaB bind to asialo-GM1 (Jansson et al., 2006). One potential explanation of these apparently conflicting data is that both CfaE and CfaB bind asialo-GM1, but that $dsc_{19}CfaB(His)_6$ does not fold into an active form with the normal activities of CfaB. To ensure that CfaB retained its native conformation for tests of its ability to bind asialo-GM1, we took the approach of producing pili in *E. coli* strains expressing CfaA, B and C and variant forms of CfaE which were unable to bind asialo-GM1. Pili consisting of mutant CfaE and wild-type CfaB could then be tested for any residual asialo-GM1 binding activity that might be associated with CfaB.

4.2.3 Binding of CFA/I pili to asialo-GM1 is mediated by CfaE

It appeared likely that amino acid residues within a pocket of CfaE, which are required for binding to erythrocyte receptors, might also be required for binding to asialo-GM1. It has been demonstrated that alanine substitutions of the arginine residues at positions 67 and 181 in CfaE abolish the hemagglutination activity of CfaE-coated polystyrene beads (Abedon and Culler, 2007). To test whether these residues were also involved in asialo-GM1 binding, and therefore whether pili containing a non-binding variant of CfaE could be generated, site directed mutations were introduced into the *cfaE* gene in pGU2 (Methods 3.6) (Table 2.3). The plasmids generated included pGU3 which encodes CfaE1 (R181A substitution), pGU4 which encodes CfaE2 (R67A substitution) and pGU5 which encodes CfaE3 (R181A and R67A substitutions). *E. coli* strain MC4100/pGU1 (*cfaA*, *B*, *C*) was complemented with pGU2, pGU3 or pGU4 to produce pili containing either the wild-type or mutant pili. Pili from each of these strains were extracted, purified and examined by SDS PAGE, immunoblotting and electron microscopy. Pilus extracts from all strains contained a 14 kDa

protein, consistent with the CfaB pilin (Fig. 4.5, panel A), which reacted with anti-CFA/I antiserum as expected (Fig. 4.5, panel B). Pilus extracts examined by electron microscopy showed typical morphology for CFA/I pili (data not shown).

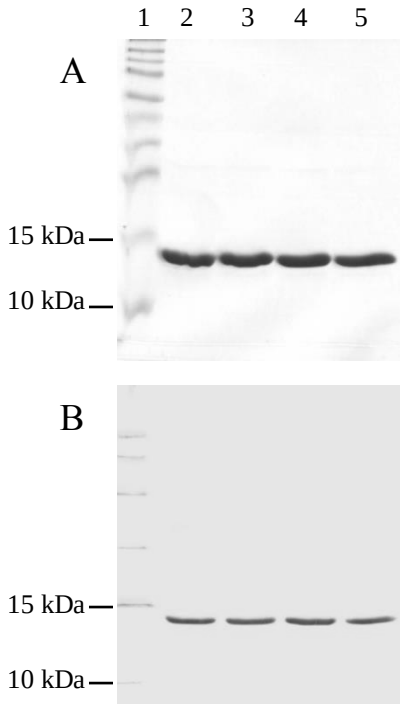

Fig. 4.5 Purification of wild-type and mutant CFA/I pili.

Purified pili were analysed by SDS PAGE (panel A) and immunoblotting with anti-CFA/I antiserum (panel B). Lane 1, protein molecular weight markers; Lane 2, wild-type pili; Lane 3, pili containing CfaE1 (R181A substitution); Lane 4, pili containing CfaE2 (R67A substitution) and Lane 5, pili containing CfaE3 (R67A, R181A substitutions).

Wild-type and mutant pili were then compared for their asialo-GM1 binding activity in an ELISA-based assay (Methods 3.11) (Fig. 4.6, panel A). Heat-inactivated pili containing wild-type CfaE or CfaE3 (R181A, R67A substitution), served as negative controls for asialo-GM1 binding activity. Pili containing wild-type CfaE, bound to asialo-GM1, as expected. However, binding was significantly reduced for pili containing CfaE1 and CfaE2 ($p < 0.05$ at pilus concentrations 10, 40, 60, 80 and 100 μg/ml). These data confirmed that CfaE residues from the same binding pocket interact with both erythrocyte receptors and asialo-GM1. Pili containing CfaE3 showed a significantly greater reduction in asialo-GM1 binding than pili containing CfaE1 ($p<0.05$ at all pilus concentrations). Similarly pili containing CfaE2 showed a greater reduction in asialo-GM1 binding, reaching statistical significance at pilus concentrations of 60, 80 and 100 μg/ml ($p<0.05$) but not at the lower concentrations of 10, 20 and 40 μg/ml where binding levels were low. However, the residual binding of pili containing CfaE3 was no greater than that of heat-inactivated wild-type pili ($p \geq 0.37$ at all pilus concentrations) or heat-inactivated mutant pili containing CfaE3 ($p > 0.31$ at 10 – 80 μg/ml; $p = 0.10$ at 100 μg/ml), indicating that it was non-specific. These data indicate that CfaE specifies CFA/I pilus binding to asialo-GM1, and that this occurs via the same residues required for binding to erythrocyte receptors. Since pilus binding to asialo-GM1 was abolished by mutations in CfaE, it was concluded that CfaB does not bind asialo-GM1 independently of CfaE.

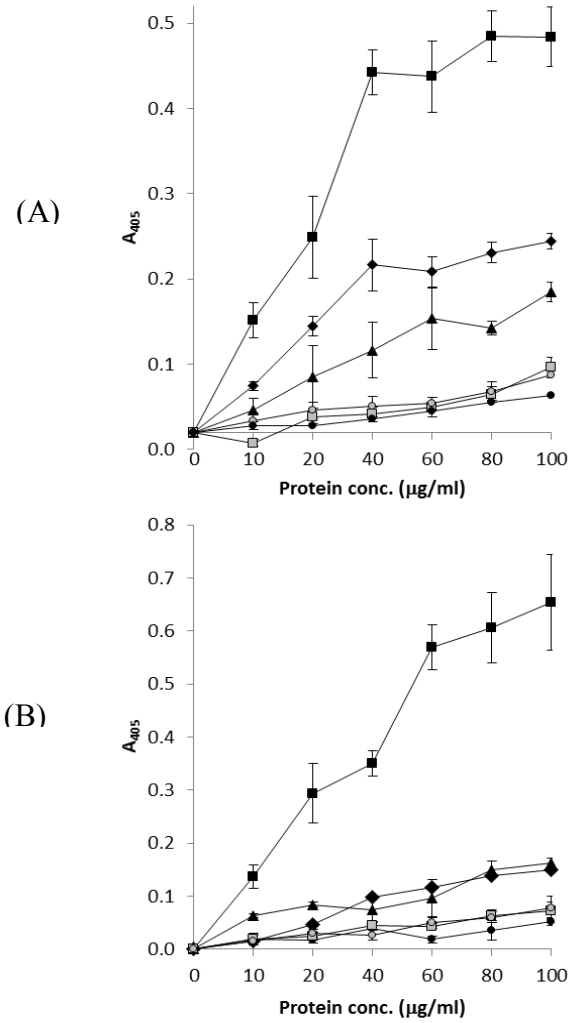

Fig. 4.6 Binding of mutant and wild-type CFA/I pili to asialo-GM1 and Caco-2 cells.

ELISA measurement of pilus binding to asialo-GM1 (panel A) and Caco-2 cells (panel B). Wild-type CFA/I pili (black squares) bound strongly to asialo-GM1 and Caco-2 cells, while binding was significantly reduced ($p < 0.05$ at all pilus concentrations) for pili containing CfaE1 (R181A substitution, black diamond) and CfaE2 (R67A, black triangle) and CfaE3 (R67A, R181A substitution, black circle). Binding of pili containing CfaE3 (R67A, R181A; black circle) was no greater than that of heat-inactivated wild-type pili (gray squares) or heat-inactivated CfaE3-containing pili (gray circles).

4.2.4 CfaE mutations that abolish asialo-GM1 binding, also abolish adherence to Caco-2 cells

To test the significance of CfaE residues required for asialo-GM1 binding in a model that has greater biological relevance, the binding of wild-type and mutant pili to cultured Caco-2 intestinal cells was determined (Fig. 4.6, panel B). The pattern of binding to Caco-2 cells was identical to that of pilus binding to asialo-GM1. In comparison to wild-type pili, mutant pili containing CfaE1 or CfaE2 had greatly reduced binding activity ($p < 0.05$ at all pilus concentrations). CfaE3 mutant pili did not have any greater binding activity than heat-inactivated wild type or CfaE3 pili, at any pilus concentration, indicating that low level binding to the cell culture was non-specific.

To determine whether the observations of pilus binding to Caco-2 cells also applied to whole piliated bacteria, the adherence of *E. coli* strain MC4100/pGU1/pGU2 (wild-type CFA/I pili), MC4100/pGU1/pGU5 (pili containing CfaE3) and the non-piliated, negative control strain, MC4100, were compared (Methods 3.12) (Fig. 4.7). The adherence of MC4100/pGU1/pGU5 was > 9-fold greater than both MC4100/pGU1/pGU2 ($p < 0.05$) and MC4100 ($p < 0.05$). However, the adherence of MC4100/pGU1/pGU5 and MC4100 to Caco-2 cells were not significantly different ($p = 0.37$).

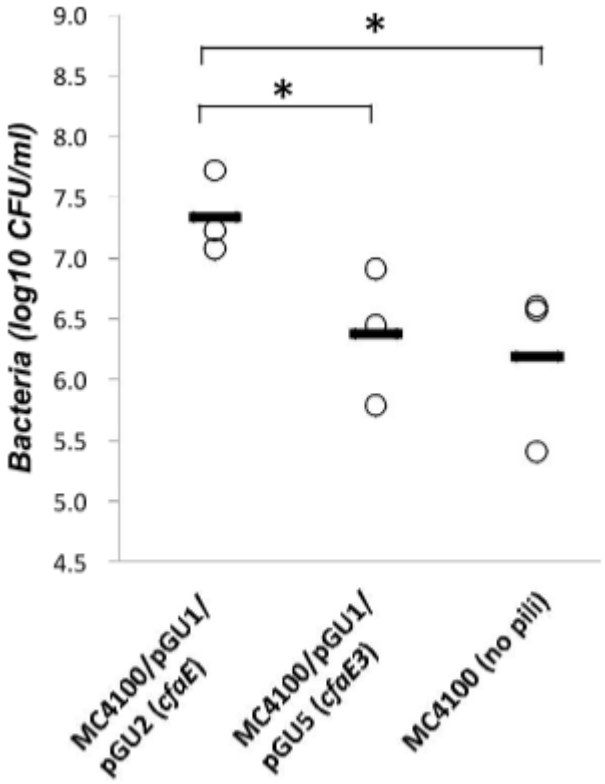

Fig. 4.7 Effect of *cfaE3* mutations on the adherence of CFA/I-piliated *E. coli* to Caco-2 intestinal cells

MC4100 strains expressing CFA/I pili containing CfaE3 (R67A, R181A substitutions) adhere significantly less to Caco-2 cells than strains expressing pili with the wild-type CfaE subunit. Significant differences in adherence ($p < 0.05$) are indicated by an asterisk.

4.3 Discussion

The initial aim of this study was to elucidate the mechanism by which CfaB binds to asialo-GM1, which would serve as a model for other glycosphingolipids that are bound by CFA/I pili. The first approach for developing a binding assay involved generating CFA/I pili consisting of only the major pilin, CfaB. This approach involved the expression of three of the four *cfa* genes, *cfaA, B* and *C,* based on a previous report that expression of these genes is sufficient to assemble pili devoid of CfaE (Jansson et al., 2006). However, in the current study, although CFA/I pili could be detected in extracts of an *E. coli* strain expressing all four *cfa* genes, *cfaA, B, C* and *D,* they could not be detected in extracts from a strain expressing *cfaA, B* and *C* by SDS PAGE, immunoblotting or electron microscopy. To examine the possibility that pilus assembly occurs at a very low frequency in the absence of CfaE, 500 independent $CfaE^-$ bacteria were examined for piliation. However, none of the cells showed any piliation. Therefore, it was concluded that the expression of CfaE is an absolute requirement for CFA/I pilus assembly. The reason for the discrepancy between our study and the study of Jansson *et al* (Jansson et al., 2006) is not clear. However, the results presented here are consistent with our understanding of the assembly of the closely related CS1 pilus, in which the minor tip-associated pilin, CooD, controls the initiation of, and is essential for, pilus assembly (Froehlich et al., 1994a, Sakellaris et al., 1999a). Our findings are also consistent with a prior report that insertional mutation of *cfaE* abolishes CFA/I assembly in the ETEC strain, H10407 (Baker et al., 2009).

Because, it was not possible to generate pili consisting solely of CfaB, an alternative approach involving the isolation of CfaB pilin for assays of asialo-GM1 binding was taken. This approach took advantage of prior work demonstrating that natively folded CFA/I pilins can be purified by generating *in cis,* donor strand-complemented pilins (Li et al., 2007).

Donor strand-complementation refers to the non-covalent interaction between the N-terminal β-strand of one subunit in an assembled pilus with a C-terminal acceptor cleft in the adjacent subunit. The inter-subunit donation of the N-terminal β-strand of CfaB simultaneously interlocks adjacent pilins and completes the immunoglobulin fold of adjacent pilins, whether they are CfaB or CfaE. When pilins are expressed in the absence of the pilus assembly machinery, they misfold and are subsequently degraded in the periplasm (Li et al., 2007). However, *in cis* donor strand-complementation can be engineered into pilins to produce self-complementing proteins that are stable, soluble and natively folded. These proteins are variants of the pilins which possess a C-terminal extension containing a flexible tetrapeptide linker, followed by a 19 amino acid sequence corresponding to the complementing N-terminal β-strand of mature CfaB, and a hexahistidine tag to facilitate purification by nickel affinity chromatography. The flexible linker allows the complementing β-strand to associate with the C-terminal cleft of the pilin to complete the Ig-fold of the protein. Binding assays with donor strand-complemented pilins gave the unexpected result that dsc_{19}CfaE$(His)_6$, which was included as a negative control protein, bound strongly to asialo-GM1 while dsc_{19}CfaB$(His)_6$ did not bind at all. Although the asialo-GM1 binding activity of CfaE had never been tested prior to this study, the negative result for CfaB contradicted the results of the study by Jansson *et al* (Jansson et al., 2006). Although dsc_{19}CfaB$(His)_6$ was expected to be natively folded and therefore to retain activity, it was possible that our specific expression and purification procedure resulted in an inactive protein. Therefore an alternative strategy for testing the binding activity of CfaB was devised.

To ensure the production of natively folded CfaB, we sought to produce CFA/I pili containing a mutant CfaE protein that was unable to bind asialo-GM1. By abolishing the asialo-GM1 binding activity of CfaE, the residual asialo-GM1 binding activity of CfaB could be measured. It seemed possible that amino acid residues in CfaE, which are involved in

binding erythrocytes, may also be required for asialo-GM1 binding. Therefore, alanine substitutions in CfaE were introduced to test the roles of R181 and R67 in asialo-GM1 binding. Substitution of either residue reduced the binding of whole pili to asialo-GM1, while substitution of both residues completely abolished pilus binding. This result showed that CfaB does not bind asialo-GM1 independently of CfaE and is consistent with the result of the binding experiments performed with donor strand-complemented CfaB and CfaE. In addition, we found that, although CFA/I-piliated *E. coli* bound to cultured intestinal cells, a piliated strain carrying the R181A and R67A substitutions in CfaE, did not show significantly greater adherence than non-piliated *E. coli*. Therefore we conclude that contrary to the existing model describing the adhesive properties of CFA/I pili, asialo-GM1 binding is not a distinct activity that resides in CfaB, independently of CfaE.

How can the discrepancy between the conclusions drawn from this study and a previous study by Jansson *et al* be explained? It was concluded that *E. coli* expressing *cfaA, B* and *C* assemble small amounts of CFA/I pili that are devoid of CfaE (Jansson et al., 2006). This result could not be replicated in our study and is not consistent with a previous study indicating that CfaE is required for pilus assembly (Baker et al., 2009). One potential explanation of how pili could be recovered from such a strain relies on the possibility that *cfaE* was indeed expressed in the strain generated by Jansson *et al*. Although the strain was described as a *cfaE* mutant, there is no description of how the *cfaE* mutation was made and there was no independent confirmation of the absence of CfaE in pili by immunoblot analysis of the pili. Regardless of the explanation, we have presented in this study, two independent lines of evidence that CFA/I binding to asialo-GM1 is not a distinct binding activity that resides in CfaB. Instead, the binding activities of CFA/I pili to asialo-GM1, erythrocytes and intestinal cells are inseparable and depend on common amino acid residues within CfaE.

Chapter 5: Identification of Rns-regulated genes in Enterotoxigenic *E. coli*

T.P Vipin Madhavan, Glen C Ulett and Harry Sakellaris

This chapter contains unpublished paper.

My contribution in this work was design of experiment, conducting experiment, analysis and writing the manuscript.

(Signed) _____

Vipin Madhavan Thekke Palasseri

(Countersigned) _____

Supervisor: Dr Harry Sakellaris

5.1 Introduction

Rns is an AFTR that was originally identified as a positively regulator of the expression of CS1 and CS2 pili in ETEC (Caron et al., 1989). Subsequent studies showed that Rns could regulate almost half of the known pilus systems in ETEC (Basturea et al., 2008), as well as four ETEC genes *cexE, yiiS, nlpA*, and *aslA* that do not encode pili (Pilonieta et al., 2007, Bodero et al., 2007, Munson et al., 2002).

In other Gram-negative intestinal pathogens of the intestine, AFTRs regulate the expression of large sets of genes. For example, the AFTR, PerA, regulates >40 genes required for Bfp pilus assembly, a Type III secretion system and secreted virulence proteins (Porter et al., 2004). In *V. cholerae*, ToxR controls the expression of virulence factors such as Tcp pili and cholera toxin as well as several other genes with potential or confirmed roles in virulence (Zhang et al., 2004, Zhang et al., 2003a, Zhang et al., 2003b). Similarly, the AFTR, RegA, from *Citrobacter rodentium* regulates the expression of >30 genes, including virulence genes encoding pilus and non-pilus adhesins, potential virulence factors and house-keeping genes (Yang et al., 2011, Yang et al., 2010, Hart et al., 2008). Therefore, in general, AFTRs controlling pilus expression in these pathogens also control the expression of large sets of non-pilus genes, including genes that encode other types of virulence factor. By analogy to EPEC, *V. cholerae* and *C. rodentium*, we hypothesized that Rns also controls the expression of a larger set of genes than has already been reported.

In order to test this hypothesis, we sequenced the genome of ETEC strain B2C, a $CS2^+$, $CS3^+$ strain isolated from a soldier in Vietnam (DuPont et al., 1971), and applied an RNA-seq strategy to determine the Rns regulon. It was discovered that a large set of genes was regulated by Rns. To distinguish genes that are directly regulated by Rns from those that are

regulated indirectly, a ChIP-seq approach was used to recover DNA sequences that bind to Rns. In addition, known virulence genes were analysed to determine whether they are regulated by Rns. Surprisingly, it was found that the well-studied virulence genes encoding heat-labile enterotoxin and a non-pilus adhesin, EtpA, are regulated by Rns.

5.2 Results

5.2.1 Genome sequencing, annotation and analysis of B2C genome:

In order to obtain a reference genome sequence for RNA-seq and ChIP-seq studies of Rns-mediated gene regulation in ETEC, the genome of ETEC strain B2C was sequenced, assembled and annotated (Methods 3.13). The draft genome sequence is available as 313 distinct contigs in the Genbank database under the accession number AUZS00000000 (Madhavan et al., 2014). Sequencing generated approximately 5.7 million read pairs for 200-fold coverage of the genome. The final assembly of 313 contigs contained 5,018,127 bp of sequence and the size of the B2C genome was estimated to be 5.2 Mb. An analysis of the genome revealed 4,804 coding sequences (CDS), 94 pseudogenes, including 64 containing frame-shift mutations, 2 clustered regularly interspaced short palindromic repeat (CRISPR) arrays, 9 rRNAs, and 83 tRNAs. Known ETEC virulence genes, including pilus and non-pilus colonization factors, enterotoxins and a secreted protease were found, as listed in Table 5.1.

The genes encoding colonization factors include two operons encoding CS2 and CS3 pili and the AFTR, Rns, which controls both CS2 and CS3 expression. The genome also carries *eltA* and *eltB*, which encode the two subunits of heat-labile enterotoxin, and *sta2* and *east1*, both of which encode heat-stable enterotoxins. In addition, the genome carries *etpA, etpB and etpC* which encode a non-pilus adhesin, its specific secretion protein and a glycosylase that modifies EtpA, respectively. Only a part of the sequence of *etpA* is represented in the

genome sequence at the ends of two contigs, 109 and 126 (Table 5.1). However, the gene is known to be intact in B2C since a high MW protein in culture supernatants has been identified as EtpA by mass spectrometry (Sakellaris, unpublished). Lastly, the genome carries a gene encoding the secreted, autotransporter protease EatA, which degrades the mucin layer of the intestinal epithelium (Kumar et al., 2013). Since these virulence genes are carried by plasmids in other ETEC strains, we presume that they are also carried on plasmids in B2C (Madhavan et al., 2014). A circular depiction of ETEC B2C genome is shown in Figure 5.1.

Table 5.1: List of known virulence genes in ETEC B2C genome.

Gene	**Function**	**Locus tag or position in contig**	**Extent of gene present in sequence**	**Contig number**
rns	AFTR of CS2 pilus expression	N444_16780	Complete	Contig000133
cotB	CS2 pilin chaperone	N444_16425	Complete	Contig000119
cotA	CS2 major pilin	N444_16430	Complete	Contig000119
cotC	CS2 usher	N444_16435	Complete	Contig000119
cotD	CS2 minor pilin	N444_16440	Complete	Contig000119
cstA	CS3 pilin chaperone	N444_14565	Complete	Contig000092
cstB	CS3 usher	N444_14560	Complete	Contig000092
cstG	CS3 major pilin	N444_14555	Complete	Contig000092
cstH	CS3 major pilin	N444_14550	Complete	Contig000092
eltA	Heat labile enterotoxin A subunit	Base pairs 1742 – 2518 of contig	Complete	Contig000092
eltB	Heat labile enterotoxin B subunit	N444_15170	Complete	Contig000100
sta2	Heat-stable enterotoxin A2	Base pairs 791 – 1008 of contig	Complete	Contig000095
east1	Enterotoxin	Base pairs 482 – 598 of contig	Complete	Contig000138
etpB	Outer membrane pore protein	N444_16715	Complete	Contig000126
etpA	Two-partner secreted adhesin EtpA	N444_16710	Base pairs 1-2002	Contig000126
etpA	Two-partner secreted adhesin EtpA	N444_16710	Base pairs 4277 - 4620	Contig000109
etpC	Glycosyltransferase	N444_15715	Complete	Contig000109
eatA	Proteolytic autotransporter	N444_06245	Complete	Contig000020

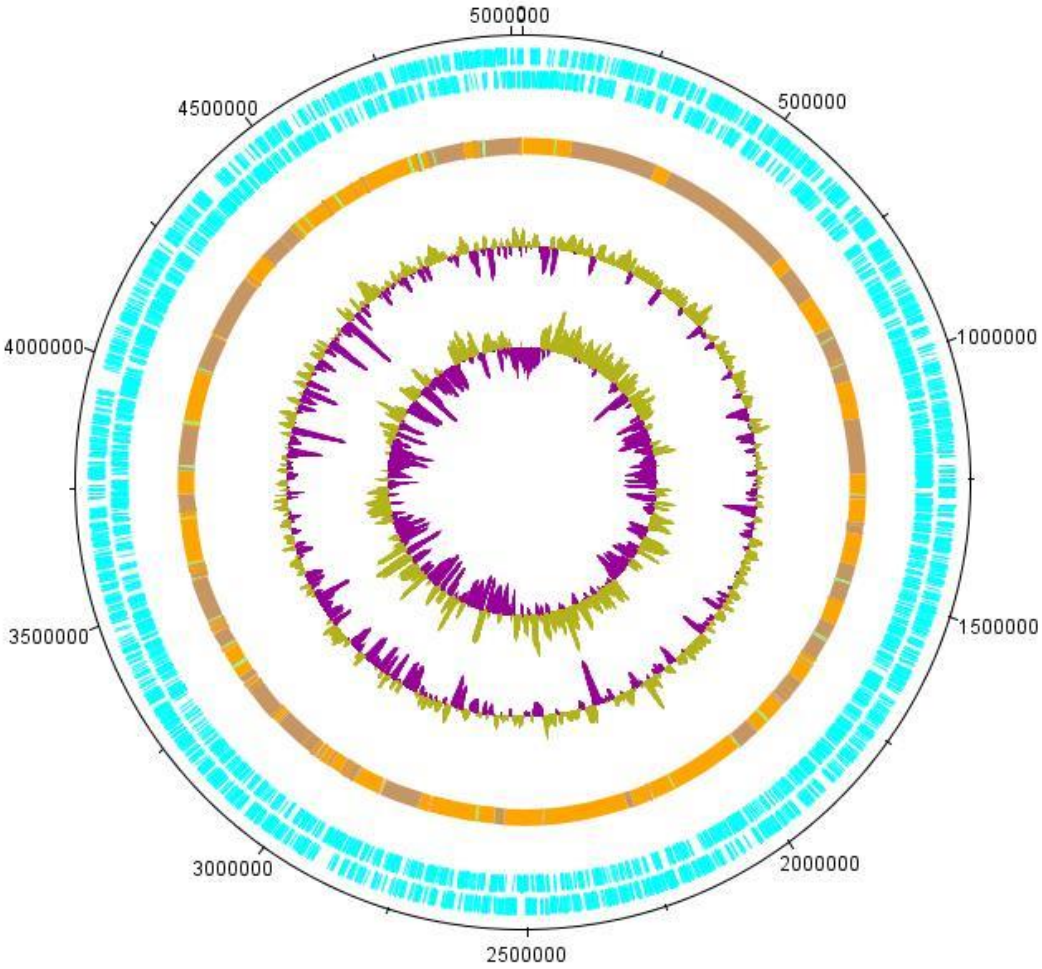

Fig 5.1: Circular depiction of ETEC B2C genome using DNA plotter (Carver et al., 2009). Each concentric ring show different features of the genome. Concatenated contigs are shown in light and dark yellow. Green lines shows separation of contigs. Coding sequences (CDS) in both DNA strands are shown in turquoise. G+C and GC skew plots are shown in inner and inner most rings respectively. Genome size is scaled in the outermost ring.

5.2.2 Confirmation of a *rns* mutation in ETEC strain B2C.

An unconfirmed insertional mutation of *rns* was constructed previously by selecting for a single cross-over homologous recombination of the suicide plasmid pCactus-*mob*-*rns'* into the chromosome of B2C Rif^r (Table 2.4). In order to confirm the mutation, a PCR analysis was conducted with the primers HSP14 and HSP15. A wild type *rns* locus amplified with HSP14 and HSP15 was predicted to yield a 0.8 kb locus, while insertional inactivation of *rns* was expected to yield a 9.4 kb fragment (Fig 5.2, panel A). Agarose gel electrophoresis of PCR fragments confirmed that the *rns* gene had been interrupted in the mutant strain B2C Rif^r *rns* (Fig 5.2, panel B).

Inactivation of *rns,* was tested phenotypically and at the level of transcription. While B2C Rif^r was able to cause mannose-resistant haemagglutination (MRHA) of bovine erythrocytes, the *rns* mutant was not able to cause haemagglutination. This is consistent with inactivation of *rns* because *rns* controls the expression of CS2 pili, which mediate MRHA. qRT-PCR was also conducted on genes that are known to be regulated by *rns*. These included *cot*B, *nlpA* and *yiiS*. The bacterial strains tested were B2C Rif^r/pBBR1MCS-2 (wild-type), B2C Rif^r *rns*/pBBR1MCS-2 (*rns* mutant) and B2C Rif^r *rns*/pBBR1MCS-2-*rns* (complemented mutant). The level of transcription in the wild-type and complemented mutant strains was compared to that of the *rns* mutant strain and expressed as a ratio (wild-type/*rns* mutant or complemented mutant/*rns* mutant). In both the wild-type and complemented mutant, the relative transcription of *cotB*, *yiiS* and *nlpA* was >1, indicating that Rns positively regulated their transcription. Transcription of *cotB* in the wild-type strain was 690-fold greater in the wild-type strain (p=0.054) and >3,000-fold greater in the complemented mutant strain (p = 0.014), than in the *rns* mutant (Fig. 5.3). The transcription of *yiiS* was 9.3-fold higher in the

wild-type (p<0.01) and 21-fold higher in the complemented mutant than in the *rns* mutant strain (p <0.05 in all cases). These data confirmed that *rns* was inactivated in B2C Rifr rns/pBBR1MCS-2. Surprisingly, the transcription of *nlpA* was upregulated by Rns, rather than repressed as previously described (Bodero et al., 2007).

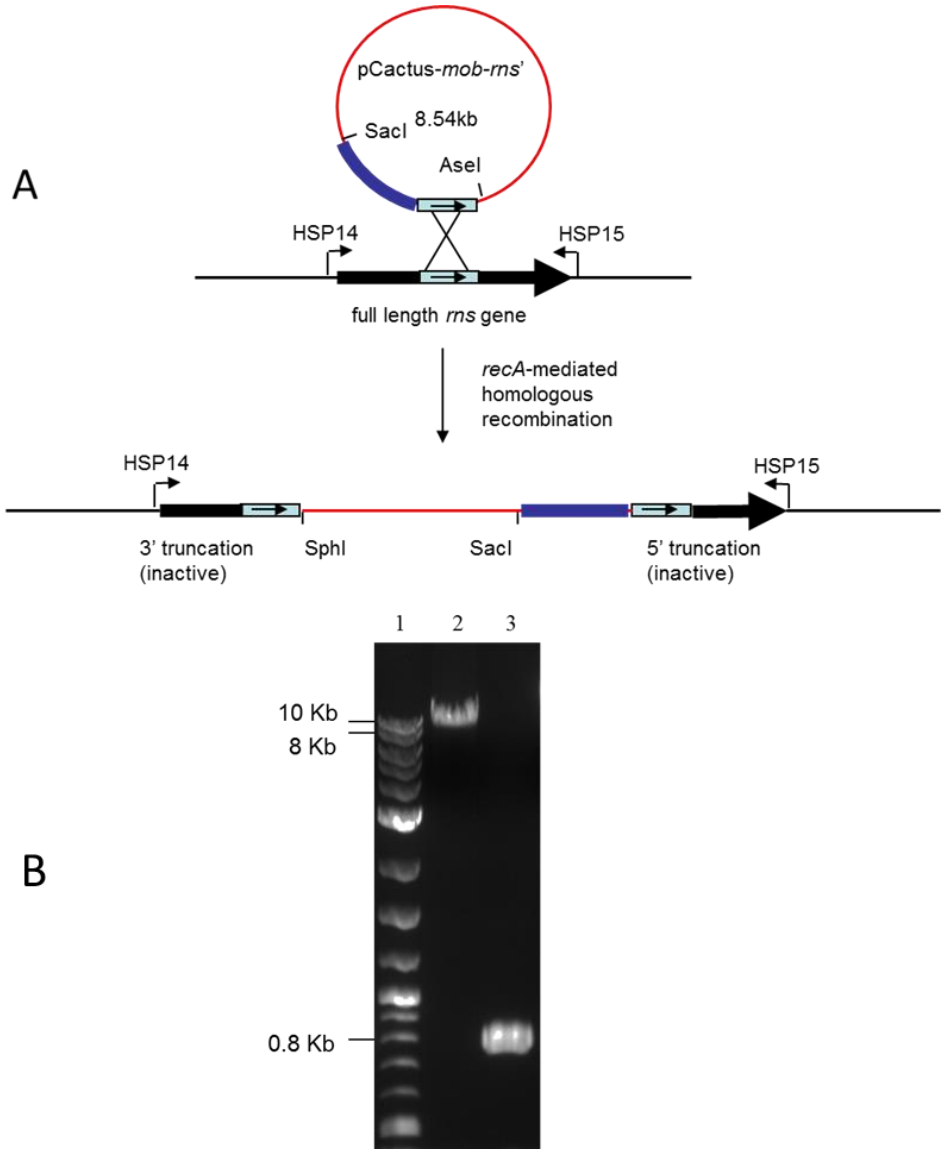

Fig 5.2: PCR confirmation of insertional mutation in *rns*.

Panel A: Organisation of *rns* locus in wild-type and *rns* mutant. Panel B: Agarose gel electrophoresis of PCR products amplified with primers HSP14 and HSP15, from genomic DNA of the presumptive B2C Rif^r *rns* mutant B2C Rif^r (lane 2) and ETEC B2C Rif^r (Lane 3). Lane 1 contains a molecular weight ladder.

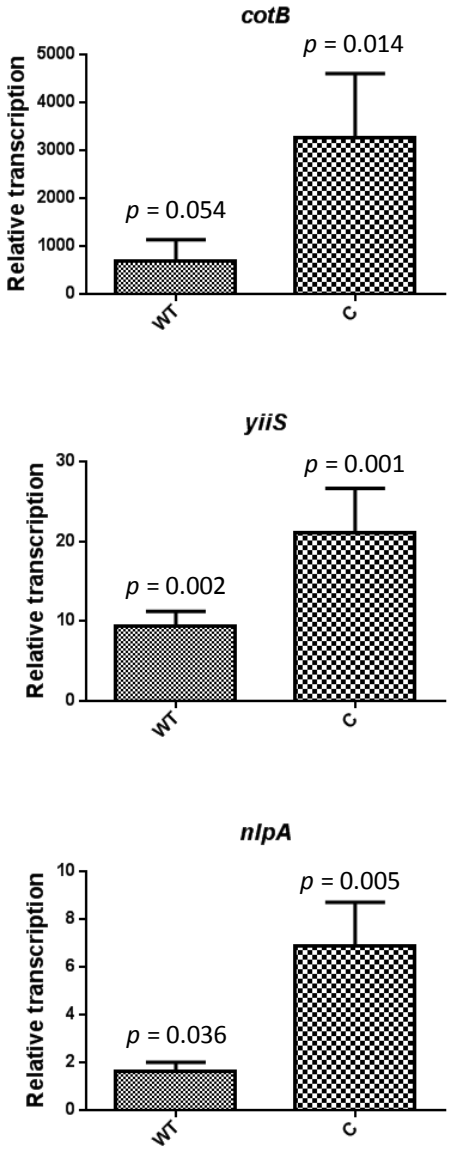

Fig 5.3: Transcription of *cot*B, *yii*S and *nlpA* in wild-type (WT) and complemented *rns* mutant (C) strains relative to the *rns* mutant strain. The results demonstrate that, relative to the *rns* mutant strain, transcription of *cot*B, *yii*S and *nlpA* is increased in the wild-type and complemented *rns* mutant strains, indicating that they are positively regulated by Rns. Statistical significance is indicated by *p* values.

5.2.3 RNA-seq demonstrates that Rns has a global influence on gene transcription.

In order to identify B2C genes that are regulated by Rns, RNA–seq was conducted on B2C Rif^r/pBBR1MCS (wild-type) and B2C Rif^r *rns* (*rns* mutant) (Methods 3.16). The results showed that Rns regulates a large number of genes which are listed in Table 5.2 and represented in the form of a heat map in Fig. 5.4. 100 genes were upregulated and 56 were downregulated by Rns. These genes are implicated in a wide variety of cellular functions including glycolysis, the TCA cycle, gluconeogenesis, extracellular stress signalling and DNA repair. A number of genes that are involved in phage assembly were also found to be upregulated.

The Rns-activated genes that were identified by RNA-seq included *cotA, B, C* and *D*, and *cstA, B, G* and *H*, encoding the CS2 and CS3 pili, respectively. These data validated the experiment since both CS2 and CS3 genes are transcriptionally activated by Rns. However, *nlpA* and *yiiS* which are also regulated by Rns, were not identified by RNA-seq in the current study.

Table 5.2: List of genes that are regulated by Rns. Fold change obtained using R (limma package) (Law et al., 2014)

Fold-regulation by Rns	Locus Tag	*p*-value	Gene	Product
819.4	N444_16925	4.89E-09		XRE family of transcriptional regulators
700.3	N444_16930	8.62E-11		hypothetical protein
341.2	N444_16940	5.39E-09		ATPase
318.1	N444_16890	1.76E-08		hypothetical protein
292.1	N444_16990	3.55E-07		hypothetical protein
270.1	N444_16965	4.26E-06		XRE family transcriptional regulator
174.1	N444_17575	1.65E-07		type II restriction endonuclease SsoII
143.2	N444_16895	9.29E-08		hypothetical protein
137.9	N444_16425	5.86E-15	*cotB*	hypothetical protein
118	N444_19305	7.92E-07		hypothetical protein
111.7	N444_16430	2.69E-13	*cotA*	CFA/I fimbrial subunit B
107.4	N444_16900	1.72E-07		hypothetical protein
106.9	N444_16935	1.54E-06		protein hokC
53.1	N444_19330	6.88E-06		terminase
49.1	N444_19325	9.83E-06		DNA packaging protein
48.6	N444_17570	8.45E-06		RNAI modulator protein
48.5	N444_16435	1.36E-10	*cotC*	hypothetical protein
47.3	N444_17015	1.45E-05		integrase
44.7	N444_14555	5.08E-15	*cstA*	hypothetical protein
41.6	N444_19310	2.70E-05		hypothetical protein
37.8	N444_10950	8.95E-10	*lacZ*	beta-D-galactosidase
28.5	N444_19345	6.96E-05		endopeptidase
26.7	N444_16985	5.28E-05		hypothetical protein
25.1	N444_16970	0.000181		hypothetical protein
25.1	N444_19350	0.000104		hypothetical protein
23.1	N444_19315	0.000109		portal protein
21.4	N444_16980	0.000157		hypothetical protein
21.3	N444_16975	0.000185		hypothetical protein
18.5	N444_17005	0.000254		hypothetical protein
14.2	N444_19340	0.000653		DNA-binding protein
13.2	N444_16945	0.001364		hypothetical protein
13	N444_16440	7.86E-08	*cotD*	hypothetical protein
11	N444_19985	1.16E-06		hypothetical protein
10	N444_24805	4.61E-05		ABC transporter permease
8.3	N444_10955	1.54E-09	*lacI*	lac repressor

Table 5.2 (continued): List of genes that are regulated by Rns. Fold change obtained using R (limma package) (Law et al., 2014)

Fold-regulation by Rns	Locus Tag	*p*-value	Gene	Product
5.3	N444_14565	3.84E-08		molecular chaperone
5.2	N444_10945	3.92E-06	*lacY*	galactoside permease
3.9	N444_20005	1.41E-08		integrase
3.7	N444_03600	3.35E-08		heat shock protein IbpA
3.6	N444_05155	1.01E-03		tartrate dehydrogenase
3.6	N444_10695	1.73E-06	*zntR*	zinc-responsive transcriptional regulator
3.3	N444_10700	2.73E-06		Alternative ribosome-rescue factor A
3.1	N444_10690	6.07E-06		hypothetical protein
2.9	N444_14560	3.96E-06	*cstB*	hypothetical protein
2.7	N444_18190	2.11E-04	*hspQ*	heat shock protein HspQ
2.6	N444_17340	3.24E-07	*bssS*	biofilm formation regulatory protein BssS
2.6	N444_18150	3.58E-05		SOS cell division inhibitor
2.6	N444_13170	2.01E-06		macrolide ABC transporter ATP-binding protein
2.5	N444_18365	1.71E-03		toxic peptide TisB
2.4	N444_15535	4.87E-04	*rpmE*	50S ribosomal protein L31
2.4	N444_02560	1.56E-03		LysR family transcriptional regulator
2.4	N444_14170	2.21E-07		protein disaggregation chaperone
2.4	N444_14550	5.01E-05	*cstH*	hypothetical protein
2.3	N444_20250	1.38E-05		5-hydroxymethyluracil DNA glycosylase
2.3	N444_04905	2.92E-06		glucose-1-phosphatase/inositol phosphatase
2.3	N444_16815	1.09E-06		hypothetical protein
2.3	N444_16820	4.46E-06		hypothetical protein
2.3	N444_16365	1.78E-06		XRE family transcriptional regulator
2.3	N444_17020	1.19E-04	*potC*	putrescine/spermidine ABC transporter permease
2.3	N444_17025	9.83E-06	*potD*	putrescine/spermidine ABC transporter substrate-binding protein
2.2	N444_11570	5.92E-07		heat shock protein 90
2.2	N444_09885	1.19E-05		alpha-galactosidase
2.2	N444_15510	1.76E-06	*hslU*	ATP-dependent protease
2.2	N444_15515	3.46E-06		ATP-dependent protease
2.1	N444_10005	4.75E-06	*groES*	co-chaperonin GroES
2.1	N444_07145	1.79E-06	*dnaK*	molecular chaperone DnaK
2.1	N444_11665	1.47E-05		hypothetical protein
2.1	N444_01975	4.12E-05		ribosome-associated heat shock protein Hsp15
2	N444_14865	1.11E-03		DNA-binding protein
2	N444_16810	1.33E-05		hypothetical protein
2	N444_12205	2.01E-04		hypothetical protein

Table 5.2 (continued): List of genes that are regulated by Rns. Fold change obtained using R (limma package) (Law et al., 2014)

Fold-regulation by Rns	Locus Tag	*p*-value	Gene	Product
2	N444_13660	3.52E-04		mobilization protein
2	N444_16370	4.31E-06		hypothetical protein
2	N444_22945	1.83E-04	*tolB*	translocation protein TolB
2	N444_07150	1.81E-05	*dnaJ*	molecular chaperone DnaJ
2	N444_08130	1.41E-04		magnesium-transporting ATPase
2	N444_08565	6.10E-05	*recN*	recombinase
2	N444_01970	3.29E-05	*hslO*	Hsp33 chaperonin
1.9	N444_16340	4.12E-05		XRE family transcriptional regulator
1.9	N444_04100	1.93E-05		RNA polymerase sigma factor RpoD
1.9	N444_21865	1.43E-05		TMAO/DMSO reductase
1.9	N444_09630	1.65E-03	*nrfA*	cytochrome C nitrite reductase subunit c552
1.9	N444_09330	1.83E-04		malate synthase
1.9	N444_11355	8.56E-05		cytochrome O ubiquinol oxidase
1.9	N444_01255	8.10E-04		hypothetical protein
1.8	N444_01980	1.08E-03		nucleotidase
1.8	N444_16335	2.88E-04		peptidase
1.8	N444_10010	2.80E-05	*groEL*	molecular chaperone GroEL
1.8	N444_11350	5.69E-05		cytochrome O ubiquinol oxidase
1.8	N444_15500	4.32E-04		ribonuclease activity regulator protein RraA
1.8	N444_04005	1.71E-03		glycogen synthesis protein GlgS
1.8	N444_19080	1.59E-03		alanine racemase
1.7	N444_01005	1.02E-04	*aceE*	pyruvate dehydrogenase
1.7	N444_21870	2.44E-04		hydroxyisourate hydrolase
1.7	N444_10140	5.72E-04	*miaA*	tRNA delta(2)-isopentenylpyrophosphate transferase
1.7	N444_00745	1.06E-03		ATP-dependent helicase
1.7	N444_05295	2.61E-04		heat shock protein HtpX
1.6	N444_04580	1.72E-03		acetyl-CoA acetyltransferase
1.6	N444_11985	1.68E-03	*yciS*	membrane protein
1.6	N444_05050	3.40E-04		glyceraldehyde-3-phosphate dehydrogenase
1.6	N444_11395	4.26E-04	*lon*	DNA-binding ATP-dependent protease La
-1.5	N444_04340	1.50E-03		serine/threonine transporter
-1.5	N444_07155	1.67E-03		transposase
-1.6	N444_07160	1.15E-03		transposase IS2
-1.6	N444_19515	7.80E-04		hypothetical protein
-1.6	N444_18335	8.58E-04		transposase IS401

Table 5.2 (continued): List of genes that are regulated by Rns. Fold change obtained using R (limma package) (Law et al., 2014)

Fold-regulation by Rns	Locus Tag	*p*-value	Gene	Product
-1.6	N444_21685	1.75E-03	*kpsS*	capsule polysaccharide transporter
-1.6	N444_14570	1.37E-03		ATPase AAA
-1.6	N444_09675	1.72E-03		hypothetical protein
-1.7	N444_22155	1.55E-03		pirin
-1.7	N444_06235	6.11E-04		EAL protein (cyclic di-GMP esterase)
-1.7	N444_06230	1.28E-03		resolvase
-1.8	N444_14575	9.00E-04		hypothetical protein
-1.8	N444_12915	1.75E-03		PTS N'-diacetylchitobiose transporter subunit IIC
-1.8	N444_04310	4.05E-04		membrane protein
-1.8	N444_11805	1.79E-03	*oppB*	peptide ABC transporter permease
-1.8	N444_08035	1.45E-03		transposase
-1.8	N444_21505	4.52E-04		dimethyl sulfoxide reductase subunit A
-1.9	N444_17650	1.41E-03		transposase
-1.9	N444_04910	1.02E-03		hypothetical protein
-1.9	N444_04335	4.10E-05		propionate/acetate kinase
-2	N444_01555	2.89E-05	*zraP*	zinc resistance protein
-2	N444_17595	4.96E-04		hypothetical protein
-2	N444_04915	1.01E-04		NAD(P)H:quinone oxidoreductase
-2	N444_15780	1.32E-03		hypothetical protein
-2	N444_12305	2.84E-04		gamma-aminobutyraldehyde dehydrogenase
-2.1	N444_04315	1.09E-04		membrane protein
-2.1	N444_05070	1.50E-05		PrkA family serine protein kinase
-2.1	N444_06240	1.59E-03		AraC family transcriptional regulator
-2.2	N444_21540	8.75E-06		bifunctional D-altronate/D-mannonate dehydratase
-2.3	N444_19090	9.60E-04		SpoVR family protein
-2.4	N444_25355	2.67E-05		Lipid A biosynthesis protein
-2.4	N444_14685	4.96E-04		hypothetical protein
-2.4	N444_10880	1.12E-04		hypothetical protein
-2.4	N444_15000	2.90E-06		tail sheath protein
-2.4	N444_06940	1.31E-03		hypothetical protein
-2.5	N444_21545	2.07E-05		oxidoreductase
-2.5	N444_05075	1.55E-05		hypothetical protein
-2.6	N444_12325	2.20E-04		spermidine/putrescine ABC transporter substrate-binding protein
-2.6	N444_14995	8.59E-07		major tail tube protein

Table 5.2 (continued): List of genes that are regulated by Rns. Fold change obtained using R (limma package) (Law et al., 2014)

Fold-regulation by Rns	Locus Tag	*p*-value	Gene	Product
-2.7	N444_12310	8.21E-05		spermidine/putrescine ABC transporter permease
-2.8	N444_15005	5.85E-05		late control protein D
-2.8	N444_14985	1.61E-05		tail protein
-2.8	N444_14990	8.87E-07		tail protein
-3.2	N444_16765	7.00E-05		hypothetical protein
-3.7	N444_14540	0.000601	*yeiA*	hypothetical protein
-3.7	N444_05965	0.001283		conjugal transfer protein TraA
-6.7	N444_19880	6.64E-09		D-arabinitol 4-dehydrogenase
-6.8	N444_22705	0.000881		hypothetical protein
-8.3	N444_18500	8.45E-12		Antigen 43
-8.7	N444_19870	2.97E-06		cytochrome C biogenesis protein CcdA
-12.1	N444_19875	1.74E-08		xylulokinase
-15.3	N444_14025	0.000573		single-stranded DNA-binding protein

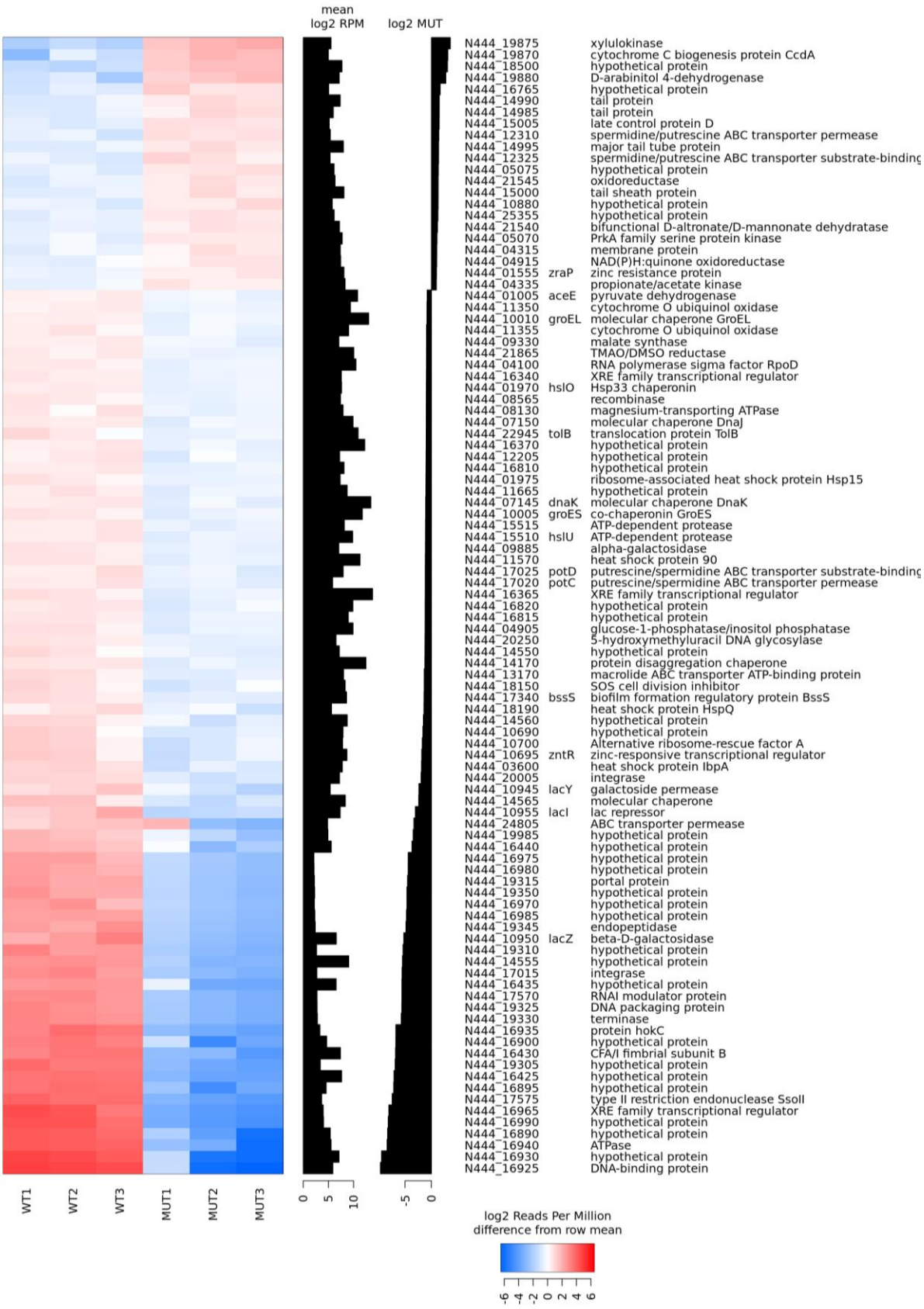

Fig 5.4: Heat map of *rns* regulated genes obtained after RNA-Seq.

Table 5.3: Distribution of Rns DNA binding sites in the ETEC B2C genome. The coordinate position of each statistically significant peak after sequence analysis obtained using MACS that are distributed across various contigs are listed below.

Contig	Start	End	Length	Fold enrichment	Name
contig000001	96,218	96,402	185	2.49	ETEC_B2C_rns_peak_1
contig000001	112,119	112,318	200	1.87	ETEC_B2C_rns_peak_2
contig000002	125,071	125,280	210	1.81	ETEC_B2C_rns_peak_3
contig000005	1,480	1,824	345	1.77	ETEC_B2C_rns_peak_4
contig000006	82,649	82,930	282	1.96	ETEC_B2C_rns_peak_5
contig000008	41,127	41,439	313	1.76	ETEC_B2C_rns_peak_6
contig000009	26,961	27,126	166	1.62	ETEC_B2C_rns_peak_7
contig000013	8,413	9,005	593	2.24	ETEC_B2C_rns_peak_8
contig000019	3,291	3,566	276	1.87	ETEC_B2C_rns_peak_9
contig000019	15,860	16,059	200	1.98	ETEC_B2C_rns_peak_10
contig000019	29,943	30,165	223	1.69	ETEC_B2C_rns_peak_11
contig000020	110	256	147	2.22	ETEC_B2C_rns_peak_12
contig000020	3,519	3,817	299	1.62	ETEC_B2C_rns_peak_13
contig000023	1,536	1,721	186	2.10	ETEC_B2C_rns_peak_14
contig000029	49,443	49,667	225	1.89	ETEC_B2C_rns_peak_15
contig000041	95	1,228	1134	1.27	ETEC_B2C_rns_peak_16
contig000043	30,470	30,618	149	1.93	ETEC_B2C_rns_peak_17
contig000043	34,268	34,421	154	2.31	ETEC_B2C_rns_peak_18
contig000043	76,881	77,165	285	1.67	ETEC_B2C_rns_peak_19
contig000047	35,239	35,449	211	1.66	ETEC_B2C_rns_peak_20
contig000047	44,554	44,837	284	1.92	ETEC_B2C_rns_peak_21
contig000047	108,240	108,428	189	1.86	ETEC_B2C_rns_peak_22
contig000054	8,063	8,592	530	1.74	ETEC_B2C_rns_peak_23
contig000054	14,835	15,292	458	1.94	ETEC_B2C_rns_peak_24
contig000054	18,013	18,203	191	1.62	ETEC_B2C_rns_peak_25
contig000055	82,121	82,314	194	1.69	ETEC_B2C_rns_peak_26
contig000055	114,708	115,006	299	2.00	ETEC_B2C_rns_peak_27
contig000066	263	485	223	1.27	ETEC_B2C_rns_peak_28

Table 5.3 (continued): Distribution of Rns DNA binding sites in the ETEC B2C genome. The coordinate position of each statistically significant peak after sequence analysis obtained using MACS that are distributed across various contigs are listed below.

Contig	Start	End	Length	Fold enrichment	Name
contig000067	125	303	179	1.52	ETEC_B2C_rns_peak_29
contig000075	126	306	181	1.35	ETEC_B2C_rns_peak_30
contig000085	230	401	172	1.81	ETEC_B2C_rns_peak_31
contig000089	309	583	275	1.73	ETEC_B2C_rns_peak_32
contig000105	99	326	228	1.53	ETEC_B2C_rns_peak_33
contig000106	145	419	275	1.91	ETEC_B2C_rns_peak_34
contig000108	45,527	45,898	372	1.92	ETEC_B2C_rns_peak_35
contig000111	1,918	2,205	288	1.56	ETEC_B2C_rns_peak_36
contig000117	70,932	71,139	208	1.69	ETEC_B2C_rns_peak_37
contig000131	899	1,201	303	1.60	ETEC_B2C_rns_peak_38
contig000155	134	355	222	1.36	ETEC_B2C_rns_peak_39
contig000161	111	345	235	1.43	ETEC_B2C_rns_peak_40
contig000163	135	334	200	1.70	ETEC_B2C_rns_peak_41
contig000167	3,261	3,502	242	1.70	ETEC_B2C_rns_peak_42
contig000167	9,972	10,156	185	1.69	ETEC_B2C_rns_peak_43
contig000176	27,642	27,788	147	2.03	ETEC_B2C_rns_peak_44
contig000181	21,777	22,054	278	1.85	ETEC_B2C_rns_peak_45
contig000182	8,032	8,184	153	1.97	ETEC_B2C_rns_peak_46
contig000186	496	697	202	1.49	ETEC_B2C_rns_peak_47
contig000195	72	292	221	1.98	ETEC_B2C_rns_peak_48
contig000205	265	458	194	1.78	ETEC_B2C_rns_peak_49
contig000223	10,396	10,679	284	1.87	ETEC_B2C_rns_peak_50
contig000255	29,401	29,625	225	2.04	ETEC_B2C_rns_peak_51
contig000274	66,941	67,148	208	1.81	ETEC_B2C_rns_peak_52
contig000274	72,153	72,328	176	1.97	ETEC_B2C_rns_peak_53
contig000288	2,429	2,679	251	1.53	ETEC_B2C_rns_peak_54
contig000289	100	302	203	1.82	ETEC_B2C_rns_peak_55

As *nlpA* and *yiiS* were not identified in the RNA-seq analysis, we sought to further substantiate the RNA-seq results by performing qRT-PCR on a panel of randomly selected genes that appeared to be regulated by Rns from the RNA-Seq analysis. These data are expressed as transcription levels in the wild-type and complemented *rns* mutant strains, relative to the *rns* mutant strain (Table 5.4).

Table 5.4: Comparison of relative transcription of Rns-regulated genes analysed by RNA-seq and qRT-PCR.

Gene or function	Locus Tag	Relative transcription in wild-type by RNA-seq (*p* value)	Relative transcription in wild-type by qRT-PCR (*p* value)	Relative transcription in complemented mutant by qRT-PCR (*p* value)
Xylulokinase	N444_19875	0.08 (*p*=0.01)	0.60 (*p*=0.01)	0.01 (*p*<0.01)
ABC transporter permease	N444_24805	10.00 (*p*=0.01)	18.34 (*p*=0.18)	337.47 (*p*=0.02)
Malate synthase	N444_09330	1.90 (*p*=0.01)	37.34 (*p*=0.01)	143.30 (*p*=0.01)
Acetyl-CoA acetyltransferase	N444_04580	1.60 (*p*=0.01)	8.98 (*p*=0.02)	2.28 (*p*=0.06)
groES (chaperonin)	N444_10005	2.10 (*p*=0.01)	2.67 (*p*=0.01)	1.76 (*p*=0.11)
bssS (biofilm formation regulatory protein)	N444_17340	2.60 (*p*=0.01)	10.63 (*p*<0.01)	82.98 (*p*<0.01)
Pyruvate dehydrogenase	N444_01005	1.70 (*p*=0.01)	0.28 (*p*<0.01)	0.54 (*p*=0.02)
XRE family of transcriptional regulators	N444_16925	819.4 (*p*=0.01)	1.60 (*p*=0.35)	4.75 (*p*<0.01)

In general, there was agreement between RNA-seq and qRT-PCR data on the whether genes were negatively or positively regulated. One exception was pyruvate dehydrogenase (N444_01005) which appeared, from the RNA-seq data, to be slightly activated but appeared

to be repressed from the qRT-PCR data. The level of regulation varied significantly with the measurement method. An extreme example of this was the positive regulation of the gene encoding an XRE family transcriptional regulator, which at its greatest was positively regulated 4.75-fold when measured by qRT-PCR but was positively regulated 819-fold when measured by RNA-seq. To investigate whether Rns regulates any of the genes listed in Table 5.4 directly or indirectly, a bioinformatic analysis of the DNA sequences upstream of each genes was conducted to identify potential Rns-binding sites. A previously determined Rns binding motif (Munson and Scott, 2000), [TA][CTG]A[TA][TA][AT][AT][ACT][AT]TAT[CT][TAG][TACG][TAC], was used to probe for putative binding sites with the searching algorithm, Fuzznuc (Rice et al., 2000). However, Rns binding motifs were not found upstream of any of the genes analysed.

5.2.4 Rns regulates the transcription of key virulence genes in ETEC.

RNA-seq analysis confirmed that the genes encoding CS2 and CS3 pili are regulated by *rns* but did not identify the Rns-regulated genes *nlpA* and *yiiS*. This suggested that potentially, other Rns-regulated genes may not be identified by RNA-seq. As AFTRs that regulate pilus expression in a number of Gram-negative pathogens also regulate the transcription of non-pilus virulence factors, we tested the hypothesis that Rns regulates the transcription of a variety of known, non-pilus virulence factors in ETEC, using a directed approach with qRT-PCR. In this regard four key virulence genes of ETEC, e*tpA*, *etpB, elt*A and *eatA*, were tested for Rns-regulation by qRT-PCR. The *etpA* and *etpB* genes encode a non-pilus adhesin and its transporter, respectively. The *eltA* and *eatA* genes encode the catalytic subunit of heat-labile enterotoxin and a proteolytic autotransporter, respectively. A comparison of the transcription of each gene was made in the *rns* mutant (B2C Rif^r *rns*/pBBR1MCS-2) and wild-type strain(B2C Rif^r/pBBR1MCS-2), as well as in the *rns* mutant and complemented mutant

strain(B2C Rif^r *rns*/pBBR1MCS-2-*rns*). Transcription of *etpA*, *etpB* and *eltA* was higher in the wild type and complemented mutant strains than it was in the *rns* mutant (Fig 5.5), indicating that Rns positively regulates the expression of these virulence genes. This was surprising as there have been no prior reports that these well-studied virulence genes are regulated by Rns. In contrast, Rns had a small but significant effect on the transcription (activation) of *eatA*.

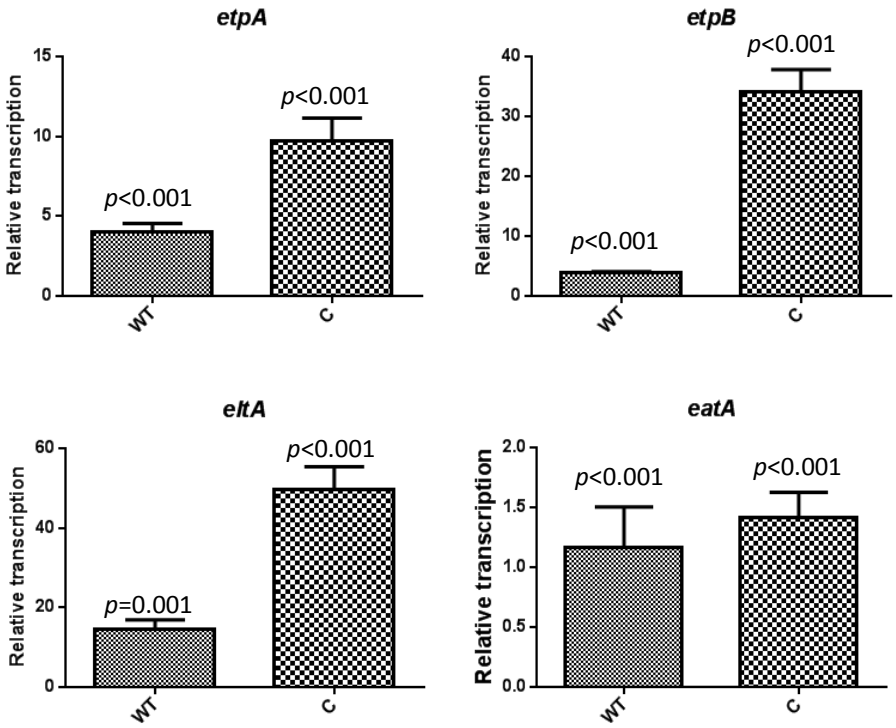

Fig 5.5: Rns-regulation of known virulence genes in B2C.

qRT-PCR measurement of the transcription of *etpA, etpB, eltA* and *eatA* in wild-type (WT) and complemented *rns* mutant (C) strains, relative to the *rns* mutant of B2C. The results indicate *that etpA, etpB* and *eltA* are positively regulated by Rns.

5.2.5 Rns binds to a DNA sequence upstream of the *etpBAC* operon.

To investigate whether Rns regulates *etpA*, *etpB* and *eltA* directly or indirectly, the DNA sequences upstream of the *etpBAC* and the *eltAB* operons was analysed with searching algorithm, Fuzznuc (Rice et al., 2000), for the presence the Rns binding motif (Munson and Scott, 2000). The analysis revealed a putative Rns binding site 148bps upstream of *etpBAC* operon. However, there was no readily identifiable Rns binding sequence upstream of the *eltAB* operon. An alignment of the putative Rns-binding sequence with other Rns-binding sequences is shown in Fig 5.6. To test whether the putative Rns binding site was biologically significant, an electrophoretic mobility shift assay (Methods 3.22) was conducted on a DNA fragment encompassing the sequence. A *malE-rns* fusion construct was cloned and MBP-Rns fusion protein was expressed and purified (Methods 3.22) for use in EMSA. A test DNA fragment containing the putative Rns-binding site was PCR-amplified with the primers HSP415 and HSP416, and a negative control DNA sequence was generated by PCR-amplifying an internal fragment of *gyrB* with the primers HSP298 and HSP299. Reactions of MBP-Rns and DNA fragments were subjected to electrophoresis and DNA was visualised with SYBR Green (Fig 5.7). Reaction of the test DNA with MBP-Rns led to the formation of two high molecular weight complexes (Fig 5.7, lanes 2-4), that were absent from the negative control reaction containing only DNA (lane 1). As expected, increasing concentrations of MBP-Rns were accompanied by increasing concentrations of the higher molecular weight complexes and decreasing concentrations of uncomplexed DNA. The inability of MBP-Rns to cause a shift in the mobility of the negative control *gyrB* DNA fragment (Fig 5.7, lanes 5 and 6) indicated that the binding of MBP-Rns to the sequence upstream of the *etpBAC* operon was specific.

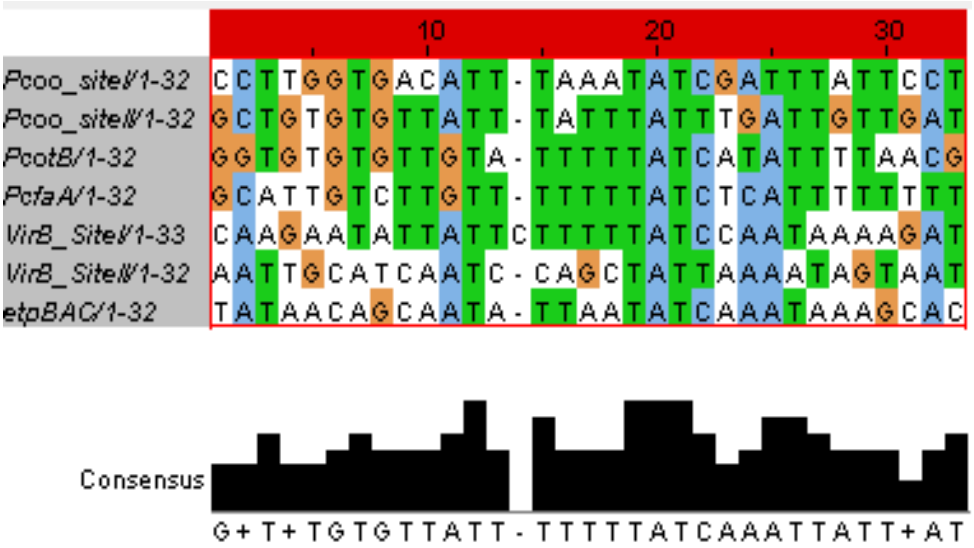

Fig 5.6: Multiple sequence alignment of Rns binding sites upstream of the *etpBAC* operon and other known binding sites. The Jalview image was generated using Kalign (Lassmann and Sonnhammer, 2005). A consensus sequence is shown at the bottom.

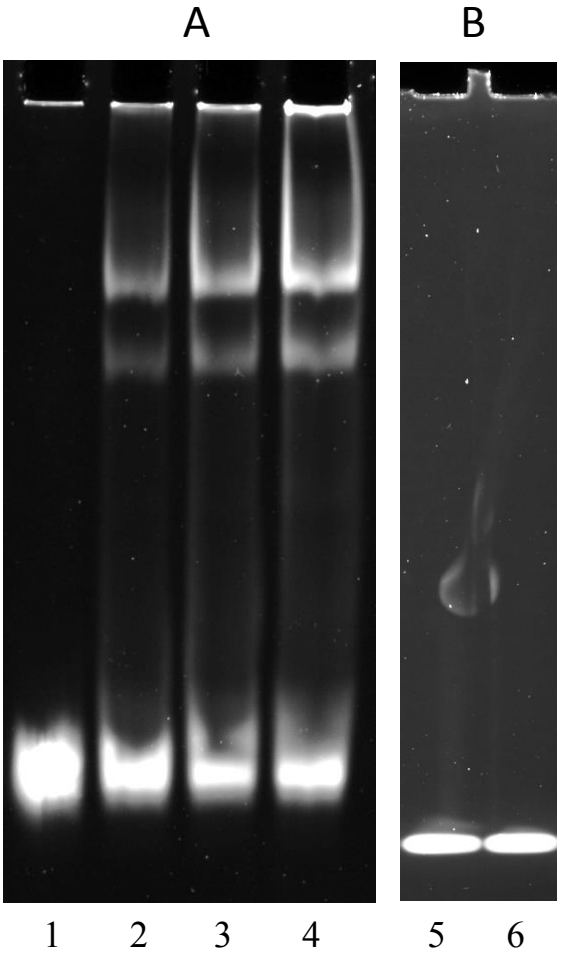

Fig 5.7: Electrophoretic mobility shift assay (EMSA) of MBP::Rns binding to DNA sequence upstream of *etpBAC* operon (Panel A) or a negative control sequence internal to *gyrB* (Panel B). Lane 1: no MBP::Rns reacted with DNA. Lane 2, 3 and 4: DNA reacted with MBP-Rns at 1 μM, 2 μM and 4 μM, respectively. Lane 5: no MBP-Rns reacted with *gyrA* DNA. Lane 6: *gyrA* DNA reacted with MBP-Rns at a concentration of 4 μM.

5.2.6 Chip-seq

In order to determine which of the genes identified by RNA-seq were directly regulated by Rns, and to potentially identify Rns-regulated genes that were not identified by RNA-seq, a Chip-seq analysis of B2C was conducted. This would allow the identification of Rns-binding sequences in the B2C genome. A plasmid expressing a Rns-Flag fusion protein was constructed to allow for expression and immunoprecipitation of Rns with a commercially available antibody to the Flag epitope (Methods 3.18). The functionality of Rns-Flag was confirmed by testing B2C Rif^r *rns*/pBAD30 and B2C Rif^r *rns*/pBAD30-*rns-flag* for the expression of CS2 pili. The two strains were grown overnight in CFA broth containing ampicillin (100μg/ml) and L-arabinose (2 mM). Bacterial cells were harvested and subjected to heat-extraction for the release of pili from the cell surface (Methods, 3.24). Heat extracts analyzed by SDS PAGE showed that Rns-Flag restored the expression of an approximately 14 kDa protein corresponding with the major CS2 pilin, confirming that Rns-Flag was functional as a transcriptional activator.

In order to find the genome wide binding sites of Rns in ETEC genome, Chip-seq was conducted (Methods 3.19). The sequencing data showed a large number of putative binding sites (Table 5.3). However, there was no enrichment of DNA fragments that are known to contain Rns-binding sites. Therefore the Chip-seq data were considered to be unreliable.

5.3 Discussion

Previous studies with AFTRs have revealed their global influence in bacterial gene expression (Hart et al., 2008, Yang et al., 2010). The hypothesis of this project was that Rns exerts a similar influence in the ETEC genome. In this regard, a combination of different techniques incorporating transcriptomics and DNA binding assays were employed to investigate this hypothesis.

As an initial step, a single crossover *rns* mutant was created by homologous recombination of a suicide vector into the *rns* locus in B2C Rif^r. A putative *rns* mutant was confirmed using a PCR analysis for insertional disruption of *rns*, as well as phenotypic and qRT-PCR analysis to confirm the abrogation of *rns* expression. The abolition of haemagglutination and the reduction of *cotB* and *yiiS* transcription were consistent with the insertional disruption of *rns*. However, an unexpected outcome of this analysis was that Rns positively regulates *nlpA* transcription. It has been previously reported that that transcription of *nlpA*, from the ETEC strain H10407, is repressed by Rns (Bodero et al., 2007). These apparently conflicting data may be due to the different ways in which the experiments were conducted in each study. In the current study, a *rns* mutation was made in the ETEC strain B2C and expression of *nlpA* was compared with the wild type and the complemented *rns* mutant. However, in the study reported by Bodero et al, an *nlpA-lacZ* fusion was integrated into the chromosome of an *E. coli* K12 strain, MC4100, for β-galactosidase assays in the presence and absence of Rns. Therefore, the hosts and the assays for *nlpA* expression were different. In addition, the culture conditions were different. In the current study, bacteria were grown in CFA broth, a medium that promotes pilus expression in ETEC, while in the study by Bodero et al, the bacteria were grown in LB broth, a medium that inhibits the expression of pili in ETEC.

In order to investigate the global influence of Rns on the ETEC transcriptome, RNA-seq was employed to measure differences in gene transcription in wild-type vs *rns* mutant strains of B2C. The results showed that 156 genes were regulated, of which 100 were upregulated and 56 downregulated. Many of these genes are clustered together, suggesting that they comprise operons.

To confirm the validity of the RNA-seq data, six randomly chosen genes were tested for Rns-regulation using qRT-PCR. Four of the genes that were positively regulated according to the RNA-seq data, were also positively regulated according to the qRT-PCR data. Similarly, one gene was negatively regulated when measured by both techniques. In some cases, the levels of gene regulation as determined by RNA-seq and qRT-PCR were similar eg. *groES* was upregulated by a factor of 2.1 as measured by RNA-seq and 2.67 as measured by qRT-PCR. In other cases there were significant discrepancies eg. The malate synthase gene was upregulated by a factor of 1.9 when measured by RNA-seq and by a factor of 37 when measured by qRT-PCR. Similar discrepancies between the two techniques have been described in the past (Trost et al., 2015). However, the large discrepancy observed with the transcriptional regulator of the XRE family (N444_16925) is not consistent with previous experience of RNA-seq and qRT-PCR and requires further investigation. In only one case did the RNA-seq and qRT-PCR data conflict. The gene encoding pyruvate dehydrogenase appeared to by moderately activated according to the RNA-seq data but repressed according to the qRT-PCR data.

To assess whether Rns might regulate the genes listed in Table 5.4 directly, a bioinformatic analysis for putative Rns binding sites upstream of the genes was conducted. None of the sequences contained putative Rns-bindings sites. There are two potential explanations of this finding. The first hypothesis is that the genes are regulated indirectly by Rns. In support of

this hypothesis a number of genes involved in responses to stress are upregulated eg. a number of genes encoding chaperonins or proteases involved in the refolding or degradation of misfolded proteins is positively regulated. These include *groEL, groES, hspQ, dnaJ, dnaK* and *lon*. It is possible that the high level of pilus subunit expression in the wild-type strain leads to the activation of these proteins, which are required for protein quality control in the cell. The Lon protease is also involved in regulating gene expression in wide variety of cellular process (Tsilibaris et al., 2006, Takaya et al., 2002). In addition genes involved in DNA repair (*recN*) and phage assembly were also upregulated, suggesting cellular stress.

An alternative possibility is that direct regulation of ETEC genes occurs via binding of Rns to operators that have diverged sufficiently from the consensus binding motif that they are no longer recognisable. One such sequence has been found in the past. The *nlpA-yicS* region of ETEC strain H10407 contains two distinct Rns-binding sequences, one of which was predicted by bioinformatic analysis, while the other was not predicted due to sequence divergence (Bodero et al., 2007).

In order to potentially discover new Rns-binding sites in the ETEC genome, that are not necessarily identifiable by bioinformatic analysis, and to identify genes that are directly regulated by Rns, a Chip-seq experiment was attempted. Although the analysis identified 56 regions in the ETEC genome, none of them covered known binding sites for Rns, including the operators for the CS2 and CS3 pilus operons and for *rns*. Therefore the data were not validated and were considered to be unreliable.

Because the RNA-seq analysis failed to identify two Rns-regulated genes, *nlpA* and *yiiS*, it appeared possible that other genes, including known virulence genes, may not have been identified as being regulated by Rns. Therefore we tested the hypothesis that the known virulence genes *etpB, etpA, eltA* and *eatA* are regulated by Rns. qRT-PCR analysis of

transcription showed that *etpB, etpA and eltA* are indeed regulated by Rns. This was a novel finding and is consistent with examples of master regulators of virulence in other bacterial pathogens. The analogy between *V. cholerae* and ETEC is particularly interesting since both organisms express pili under the control of an AFTR and the same AFTRs control the expression of cholera toxin and heat-labile toxin, which are functionally identical and evolutionarily closely related enterotoxins.

Bioinformatic analysis revealed a potential Rns-binding site sequences upstream of the *etpBAC* operon but not upstream of the *eltAB* operon. The biological significance of this Rns-binding sequence was confirmed by conducting an EMSA on DNA fragment containing the putative binding. Although the role of the *etpBAC* operon in ETEC virulence (Roy et al., 2009b, Meli et al., 2009) has been demonstrated, no regulator of this operon has been elucidated until now. EMSA showed two shifted bands. One possible explanation of this observation is the existence of another binding site in the DNA fragment (Munson et al., 2002). But no such binding site was seen in bioinformatic analysis. A potential alternative explanation could well be found in a recent observation that Rns forms dimers (Mahon et al., 2012). This dual oligomerisation state may lead to two different bands in EMSA.

In summary, an important aspect of this research has been the novel finding that Rns positively regulates the expression of both pilus (CS2 and CS3) and non-pilus (EtpA) adhesins. In addition, the Rns-regulation of heat-labile enterotoxin is also a novel discovery. Future work will be aimed at testing whether an operator for Rns exists upstream of the *eltAB* operon, using EMSA and DNase I footprinting, and whether Rns also regulates the expression of the two heat-stable enterotoxin genes in strain B2C.

Chapter 6: Understanding *rns* regulated virulence factors of ETEC and their effect towards the host immune response.

T.P Vipin Madhavan, Benjamin Luke Duell, Glen C Ulett and Harry Sakellaris

This chapter contains unpublished paper.

My contribution in this work was design of experiment, conducting experiment, analysis and writing the manuscript.

(Signed) _____

Vipin Madhavan Thekke Palasseri

(Countersigned) _____

Supervisor: Dr Harry Sakellaris

6.1 Introduction

Bacterial infection leads to different types of immune responses in the host that are geared towards eliminating the pathogen and preventing further infection. In order to survive in and invade the host, bacteria use multiple strategies for evasion of host immune responses which include attachment and colonization of body surfaces, evasion of immune recognition at mucosal surfaces, resistance to antibacterial effectors on epithelial surfaces, mechanisms of invasion and crossing of epithelia, evasion of phagocyte responses and interfering with cytokine production (Hornef et al., 2002). Many studies have helped to unravel the immune response against ETEC, as well as the immune evasion mechanisms of ETEC. These studies have employed different models such as porcine and murine animal infection models, and human, murine and porcine intestinal epithelial cell cultures *in vitro*.

Understanding the mechanism underlying host immune responses and evasion strategies used by bacteria is vital for the development of better treatment approaches including vaccines (Byrd et al., 2003). Earlier studies by Levine et al showed that development of immunity in human volunteers can occur after diarrheal episodes due to ETEC H10407 or ETEC B7A strains (Levine et al., 1980, Levine et al., 1979). Thus, challenging the host with colonisation factor antigens (CFAs) could aid in the development of immunity against ETEC infection (Freedman et al., 1998). Genetically modified polyvalent vaccines expressing different types of ETEC pili have shown to induce mucosal immune responses without eliciting inflammatory response (Daley et al., 2007) and offer additional insight into immune protection mechanisms operating against ETEC.

ETEC infection has a profound influence on host immune responses. Studies using mouse models have shown activation of MAPK pathways and down regulation of epidermal growth factor receptor (EGFR). In addition, Lyzs (lysozyme) is downregulated and this may affect intestinal immunity. ETEC is also believed to activate caspase 1, there by leading to activation of inflammasome (Ren et al., 2014).

An important aspect of the host response to ETEC is the secretion of cytokines. The type of cytokine expression that is induced following exposure to ETEC depends on the host environment and the given model system that is used to study it. ETEC infection of HCT-8 cell lines showed profound secretion of the pro inflammatory cytokine IL-8 involving the activation of TLR5 signalling (Huang et al., 2004). While infection of HT29 cells leads to secretion of pro inflammatory cytokines including TGF-β1, TGF-β2, TGF-β3, IL-1α, IL-6, IL-18 and TNFα, the response does not encompass IL-4 and IL-10 (Bahrami et al., 2011). Interestingly, however, studies on human volunteers showed a higher prevalence of symptomatic ETEC Traveller's Diarrhea (TD) following infection with LT expressing strains among people with high expressing alleles of IL-10 promoters (Flores et al., 2008). Thus, multiple different cytokines with opposing functions can be induced by ETEC and are likely to have a role in disease pathogenesis and progression following infection.

Multiple surface and secretory proteins of ETEC have been shown to trigger immune responses. These include FliC, EatA, Antigen 43, EtpA, CfaA, FimD2, CfaB (CFA/I), CexE, TibA, GroeL, DnaK, OmpA, NmpC, Bfr, OmpW, and YghJ (Roy et al., 2010). High concentrations of these virulence proteins were also observed in outer membrane vesicles (OMVs) suggesting a potential to interact with the immune system in the absence of whole bacterial cells (Roy et al., 2010). Further, EAST1 triggers IL-8 responses (Jiang et al., 2002). ETEC LT upregulates NF-κb and MAPK pathway activity; the latter is dependent on PKA while NF-κb activates through Rap1. MAPK activation leads to increased adherence (Wang et al., 2012). CFA/I expressed in a Salmonella vaccine vector illicit elevated serum IgG1 and mucosal IgA (Pascual et al., 1999).

ETEC is known to exhibit several immune evasion strategies that may counteract the host response during infection. For example, ETEC can downregulate antimicrobial peptides (AMPs) that are important in killing bacteria. CT and LT downregulate LL-37 and HBD-1;

CT transcriptionally downregulates AMPs involved in signal transduction and can effect protein kinase A (PKA), ERK MAPK and Cox-2 (Chakraborty et al., 2008). An unknown heat stable protein secreted by ETEC was shown to inhibit 1κBα degradation, thereby preventing NF-κB regulation of pro inflammatory cytokine responses. This process is believed to be mediated through clathrin mediated endocytosis (Wang and Hardwidge, 2012). ETEC infection induces autophagy through a pathway involving mTOR, ERK1/2 and AMPK, which is believed to prevent spread of infection (Tang et al., 2014). ETEC was also shown to induce cytotoxicity and apoptosis in J774 macrophage cells (Lai et al., 1999) suggesting a mechanism whereby the bacteria can proactively kill host cells.

Coculture of ETEC with non-pathogenic microorganisms has also shown a possible role in influencing host responses through microbe-microbe interactions; particularly in relation to the suppression of inflammatory responses. *Saccharomyces cerevisiae* downregulated ERK1/2 and p38 MAPK phosphorylation in differentiated IPEC-1 cells, thereby supressing the secretion of IL-6 and IL-8 (Zanello et al., 2011). *Lactobacillus amylovorus* inhibits the phosphorylation of IKKa, IKKb, IkBa and NF-κB, IL-8 and IL-1b, which are upregulated following ETEC infection. Furthermore, the negative regulators Tollip and IRAK-M, HSP72 and HSP 90 are also modulated (Finamore et al., 2014).

The AraC family of transcriptional regulators (AFTRs) regulate of variety of genes in bacterial genomes, including different metabolic pathways. Our findings suggest that Rns, an AFTRs, regulates genes in similar fashion (Madhavan TP et al unpublished). In this regard we sought to understand the effect of *rns* regulated factors on the host immune response. Using human epithelial colorectal adenocarcinoma-2 (Caco-2) cells as an *in vitro* model, we performed adhesion and invasion assay using ETEC B2C wild type (WT) and its *rns* mutant strain (ETEC B2C *rns* mutant) to gain insight into this novel host-pathogen interaction.

The hypothesis of this Chapter was that ETEC *rns* regulated factors would have a significant effect on the response of epithelial cells manifested as altered production of one or more cytokines.

6.2 Results:

6.2.1 Adhesion of ETEC WT and *rns:*

ETEC WT expresses pili while the *rns* mutant does not. To test the hypothesis that the WT binds more strongly than the *rns* mutant, adhesion assays were conducted and the results are shown in Fig 6.1. There was significantly fewer CFU obtained post infection of the *rns* mutant compared to WT at MOI 100 ($p<0.05$), indicating that WT bound more strongly than the *rns* mutant. There was more binding of the *rns* mutant with increasing MOI until value reached 300. Surprisingly, binding dropped with further increases in MOI for the *rns* mutant. An MOI of 250 for *rns* mutant showed equal binding to WT with 100 MOI.

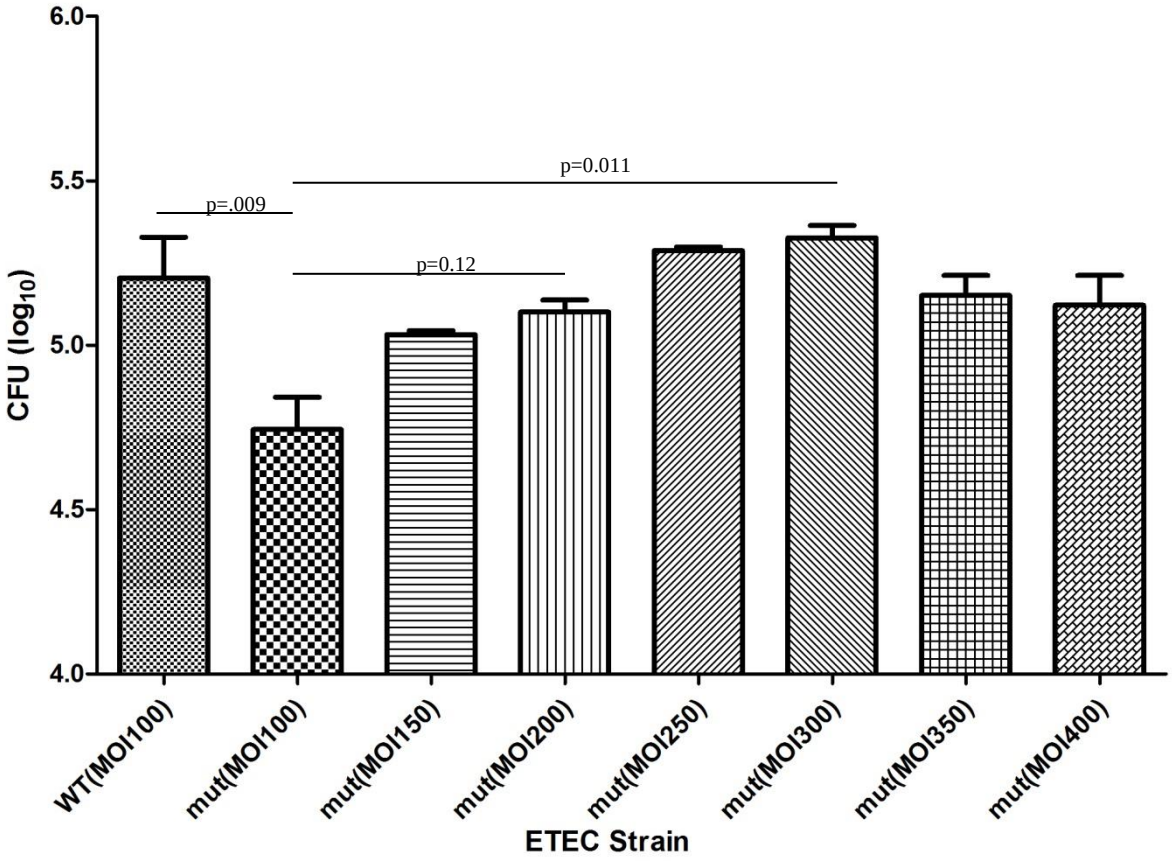

Fig 6.1: Adhesion assay of ETEC strains comparing one ETEC B2C WT MOI (100) and varying ETEC B2C *rns* mutant MOI (100-400). ETEC B2C WT is represented as WT and ETEC B2C *rns* mutant is represented as mut. Each strain sample is represented in X-axis while CFU obtained post infection is represented along Y-axis in log_{10} scale. Data were normalised against the challenge inoculum. Data represent the mean $\pm$ S.D. of technical triplicates from one independent experiment, representative of three. For statistical comparison, One-way ANOVA test was conducted to compare each group.

6.2.2 Invasion of Caco-2 cells by ETEC WT and *rns* mutant.

The following adhesion and invasion assay was conducted independently of Fig 6.1; using ETEC strains on Caco-2 cells at a fixed MOI (100) to compare the adhesion and invasion between WT and *rns* mutant (see Fig 6.2). Invasion assay showed that ETEC does invade host cells to a lesser extent compared to the total bacteria that binds to cells. There was no significant difference in invasion between WT and *rns* mutant upon treatment with gentamycin (200µg/ml).

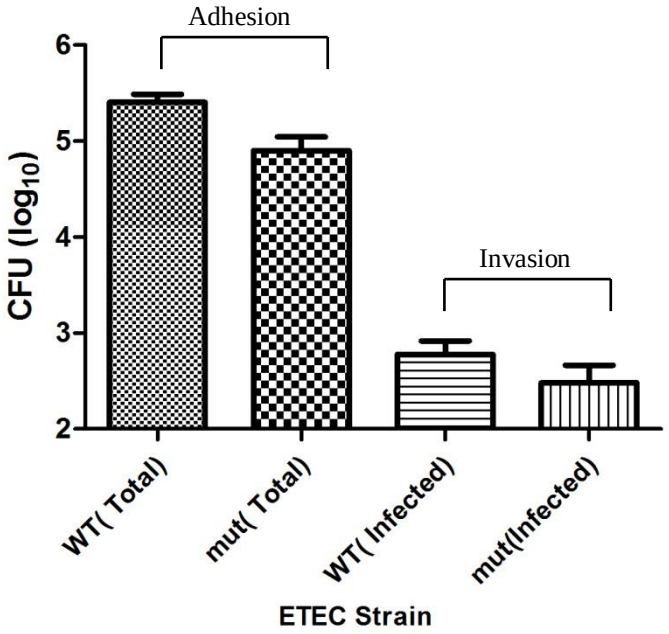

Fig 6.2: Adhesion and invasion assay conducted using ETEC B2C WT and ETEC *rns* mutant on Caco-2 cells at MOI 100. ETEC B2C WT is represented as WT and ETEC B2C *rns* mutant is represented as mut. Adhesion was calculated as previously. For invasion assay replica plates for samples used in adhesion assay was used in triplicates.

6.2.3 ETEC infection triggers cytokine expression in Caco-2 cells:

Caco-2 cells were infected with ETEC B2C WT and *rns* mutant. Uninfected (UIF) samples served as negative control. Bioplex analysis for of supernatants collected from infected and uninfected samples was conducted for 27 human cytokines. Only four cytokines showed significant expression due to infection. These were IL-8, IL-9, IL-13 and VEGF (Fig 6.3). Although there was statistically significant difference between uninfected and all infected samples for most of the expressed cytokines (with an exception of VEGF), there was no significant difference between WT and *rns* mutant for any of the cytokines.

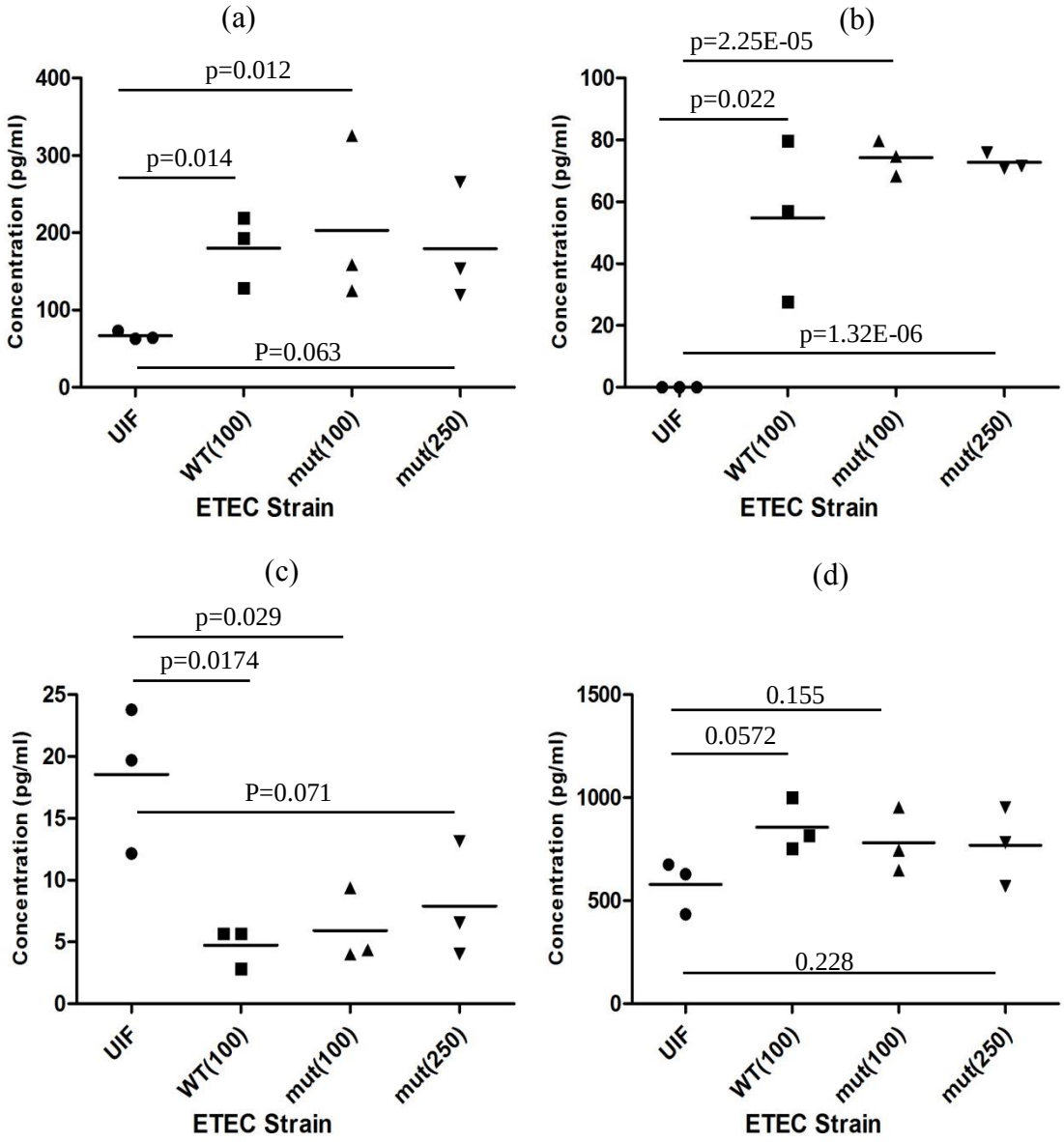

Fig 6.3: Bioplex analysis of human cytokine expression of Caco-2 cells against ETEC infection; IL-8 (a), IL-9 (b), IL-13 (c) and VEGF (d). Uninfected (UIF) and infected samples are shown along X-axis. MOI of each strain is shown in brackets. Each horizontal bar represents mean. For statistical comparison paired two tailed t-test was conducted between two groups.

6.3 Discussion:

The primary aims of this chapter were to compare the adhesion and invasion of WT and *rns* mutant towards Caco-2 cells and investigate the host response to infection. The major finding was a significant difference in binding between WT and *rns* mutant towards the host cells and significant changes in the expression of three cytokines.

The mechanism of infection is largely adhesive in nature (Madhavan and Sakellaris, 2015). Many previous studies have shown the mode of adhesion of ETEC on Caco-2 cells. Darfeuille-Michaud et al showed that ETEC strains expressing CFA/I, CFA/II and CFA/III could bind to Caco-2 cells. They further hypothesized that there may exist different receptor for each CFA (Darfeuille-Michaud et al., 1990). CS19 and CS20 was first characterised for their functions by their ability to bind Caco-2 cells, using strains that express them alone (Valvatne et al., 1996, Grewal et al., 1997b). Tobias et al showed that purified CS6 pili could bind to Caco-2 cells. Furthermore it was shown that the combined effect of CS6 subunits, CssC and CssD enable CS6 to bind to Caco-2 receptor (Tobias et al., 2008). In all these studies polyclonal or monoclonal antibodies generated against respective colonisation factors could abolish binding to Caco-2 cells which were preincubated with corresponding antibodies, showing the specificity of binding. Hence these studies showed that these colonisation factors could be promising vaccine targets (Byrd and Cassels, 2003). However, a study of Rns regulated factors have not been undertaken.

The present study showed that under the same MOI, there is a significant difference in adhesion between WT and *rns* mutant. It was show that in order to normalise adhesion of WT at MOI of 100, a *rns* mutant MOI of 250 is required. When we increased the MOI of *rns* mutant beyond 300 we found a drop in the CFU for reasons that could not be explained.

ETEC is not known to be an invasive pathogen. We searched for an invasive phenotype on ETEC using both WT and *rns* mutant. In our studies we found a few cells were seen after

gentamycin treatment, with no statistically significant difference in invasion of ETEC with either of strains. Similar observation of few cells invading human epithelial cells, were observed previously (Boudeau et al., 1999). These observations were not significant to conclude ETEC as an invasive pathogen, in the perspective of a bacterial infection. Similarly, in our studies, the number of cells obtained after the invasion assay was very low (2500 to 2800 CFU). This shows ETEC does not have an invasive phenotype. Few cells remaining after invasion assay could be a possible artefact. This could possibly due to following reasons; Gentamicin concentration used (200µg/ml) is low for complete inhibition of ETEC B2C strain, or Gentamicin stock used was slightly degraded, or few cells were trapped within the well and not exposed to Gentamicin treatment.

Our previous studies showed that Rns regulates a large number of genes in ETEC genome. Thus we hypothesized that Rns regulated factors would have a profound role in the immune response from the host. In order to test this hypothesis, we analyzed the cytokine response of Caco-2 cells against ETEC B2C WT and *rns* mutant. Among the 27 cytokines that were analyzed only four, IL-8, IL-9, IL-13 and VEGF showed significant expression. IL-8 expression has been observed previously (Huang et al., 2004). But expression of IL-9, IL-13 and VEGF has not been shown before. We saw an upregulation of IL-9 and down regulation of IL-13. IL-9 is known to have anti-apoptotic role (Demoulin et al., 2001, Singhera et al., 2008). ETEC infection is known to induce autophagy and apoptosis (Tamayo et al., 2005, Tang et al., 2014). The IL-9 response during ETEC infection may mediate an anti-apoptotic action to protect mucosal epithelial cells of intestine. Another interesting observation has been that ETEC infection leads to down regulation of the IL-13 response. IL-13 is known to have a role in triggering physiological changes in the gut, such as hyper secretion of glycoproteins, and to help in the expulsion of gut parasites and bacteria (Ochoa-Reparaz et al., 2008, Wynn, 2003). No such role has been reported in ETEC so far. It could be argued

that ETEC downregulates the IL-13 response to evade host defences and to achieve greater attachment to the intestine. Similar to IL-9, IL-13 is also known to trigger anti-apoptotic response. In this regard further studies are required to understand the *in vivo* significance of apoptosis in ETEC- host interactions. A significant expression of VEGF was also observed in infected samples. Although VEGF is known to trigger angiogenesis, the implication of this cytokine in bacterial infection is unknown. Contradicting our original hypothesis, there was no statistically significant difference in the expression of any cytokine between the WT and the *rns* mutant strains. This may be either due to the fact that Rns regulated factors having no role in immune response, or may reflect limitations in the experimental design or procedure. In summary this chapter showed that abolition of *rns* expression led to decreased bacterial attachment to Caco-2 cells. Rns regulated factors have no effect on host immune response. To conclude, Rns influences bacterial adhesion to intestinal cells but not the innate immune response.

Conclusion

Present study was aimed at investigating three aspects of ETEC virulence.

Role of minor pilin subunit of CFA/I, CfaE in pili synthesis, and binding to glycosphingolipids was investigated using whole pili level as well as individual subunits. CFA/I binds strongly asialo-GM1 and Caco-2 cells, while CFA/I with CfaE mutations failed to show significant binding. Similarly dsc_{19}CfaE$(His)_6$ bound to asialoGM1 significantly while dsc_{19}CfaE$(His)_6$ not. Although both MC4100 strains expressing CFA/I and CFA/I with mutated CfaE were able to generate pili, deletion of *cfa*E led to abrogation of pilli synthesis. To conclude, expression of CfaE is indispensable for pilus synthesis and glycosphingolipid binding. Also we propose, glycosphingolipid binding and human epithelial cell binding are not distinct but similar activity involving similar mechanisms. Present study concludes that CfaE is the sole adhesin in CFA/I pili, and role of CfaB is confined to structural support.

Sequencing of ETEC B2C genome gave deeper insights into the genome features of this ETEC strain. ETEC B2C encodes key virulence factors similar to other sequenced ETEC genomes of H10407 (Crossman et al., 2010) and E24377A (Sahl and Rasko, 2012). Contradicting previous observations, studies conducted to confirm the abrogation of *rns* expression showed *nlpA* is upregulated by Rns. Rns regulated non-pilus genes were investigated in ETEC B2C genome using transcriptomic and gel shift assays. RNA-seq analysis showed, 156 Rns regulated genes, of which 100 were up regulated and 56 downregulated. These genes were widely distributed across different metabolic pathways and included few transcriptional activators. Confirmation of a few genes predicted by RNA-seq using qRT-PCR showed discrepancy in fold change between RNA-seq data and qRT-PCR. qRT-PCR on key virulence genes, *etp*B, *etp*A , *elt*A. Furthermore we found Rns binds to upstream of etpBAC operon. Putting together, Rns directly or indirectly controls multiple

genes in ETEC genome. Similar to many other AFTRs, Rns is a global regulator of virulence in ETEC.

Since Rns is found to regulate a large number of genes in ETEC genome, study was conducted to find the role of Rns related factors on host immune response. Using Caco-2 cell line as a model of human gut epithelium, adhesion and invasion assay performed using ETEC B2C WT and ETEC *rns* mutant. There was a significant drop in adhesion with ETEC *rns* mutant relative to ETEC B2C WT. Cytokine analysis was conducted on supernatant of culture using Bioplex analysis. Analysis revealed that three cytokines, IL-8, IL-9, and IL-13 showed significant expression upon infection with ETEC. No significant change was observed between WT and *rns* mutant. Hence to conclude, expression of Rns influence adhesion but no significant effect on host immune response. No invasive phenotype was observed with or without the expression of Rns.

Apart from the aims of study, present study on ETEC gave few interesting observations. Similarity in sequence and organisation of genes constituting α pili operon reflects as symmetry in mode of pili synthesis and adhesion. Rns too display symmetry with other AFTRs in exerting global influence in bacterial genome.

Appendix 1

(700927401) Phlebotomy umbrella cover for school of medical science (MSC), Griffith University.

HUMAN RESEARCH ETHICS COMMITTEE

ETHICAL CLEARANCE CERTIFICATE

This certificate generated on 27-09-2011.

This certificate confirms that protocol 'approved procedure: Venepuncture, umbrella ethical cover for venous blood collection from human volunteers to maintain, develop and standardise the techniques and procedures in Haematology research laboratory.' (GU Protocol Number MSC/13/11/HREC) has ethical clearance from the Griffith University Human Research Ethics Committee (HREC) and has been issued with authorisation to be commenced.

The ethical clearance for this protocol runs from 05-09-2011 to 26-09-2016. The named members of the research team for this protocol are:

Dr Indu Singh

The research team has been sent correspondence that lists the standard conditions of ethical clearance that apply to Griffith University protocols.

The HREC is established in accordance with the *National Statement on Ethical Conduct on Research Involving Humans.* The operation of this Committee is outlined in the HREC Standard Operating Procedure, which is available from

www.gu.edu.au/or/ethics.

Please do not hesitate to contact me if you have any further queries about this matter.

Gary Allen
Manager, Research Ethics
Office for Research
G39 room 3.55 Gold Coast Campus
Griffith University
Phone: 3735 5585
Facsimile: 5552 9058
Email: g.allen@griffith.edu.au

Bibliography:

2008. Enterotoxigenic *Escherichia coli*: advances in technical and laboratory aspects of research and development of vaccines. *Wkly Epidemiol Rec,* 83**,** 92-5.

ABEDON, S. T. 2009. Kinetics of phage-mediated biocontrol of bacteria. *Foodborne Pathog Dis,* 6**,** 807-15.

ABEDON, S. T. & CULLER, R. R. 2007. Optimizing bacteriophage plaque fecundity. *J Theor Biol,* 249**,** 582-92.

ABRAHAM, J. M., FREITAG, C. S., CLEMENTS, J. R. & EISENSTEIN, B. I. 1985. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc Natl Acad Sci U S A,* 82**,** 5724-7.

ALTBOUM, Z., BARRY, E. M., LOSONSKY, G., GALEN, J. E. & LEVINE, M. M. 2001. Attenuated Shigella flexneri 2a Delta guaBA strain CVD 1204 expressing enterotoxigenic *Escherichia coli* (ETEC) CS2 and CS3 fimbriae as a live mucosal vaccine against Shigella and ETEC infection. *Infect Immun,* 69**,** 3150-8.

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

ANDERSSON, M., BJORNHAM, O., SVANTESSON, M., BADAHDAH, A., UHLIN, B. E. & BULLITT, E. 2012. A structural basis for sustained bacterial adhesion: biomechanical properties of CFA/I pili. *J Mol Biol,* 415**,** 918-28.

AUBEL, D., DARFEUILLE-MICHAUD, A. & JOLY, B. 1991. New adhesive factor (antigen 8786) on a human enterotoxigenic *Escherichia coli* O117:H4 strain isolated in Africa. *Infect Immun,* 59**,** 1290-9.

BACHOVCHIN, W. W., PLAUT, A. G., FLENTKE, G. R., LYNCH, M. & KETTNER, C. A. 1990. Inhibition of IgA1 proteinases from *Neisseria gonorrhoeae* and *Hemophilus influenzae* by peptide prolyl boronic acids. *J Biol Chem,* 265**,** 3738-43.

BAHRAMI, B., MACFARLANE, S. & MACFARLANE, G. T. 2011. Induction of cytokine formation by human intestinal bacteria in gut epithelial cell lines. *J Appl Microbiol,* 110**,** 353-363.

BAKER, K. K., LEVINE, M. M., MORISON, J., PHILLIPS, A. & BARRY, E. M. 2009. CfaE tip mutations in enterotoxigenic *Escherichia coli* CFA/I fimbriae define critical human intestinal binding sites. *Cell Microbiol,* 11**,** 742-54.

BAKKER, D., WILLEMSEN, P. T., WILLEMS, R. H., HUISMAN, T. T., MOOI, F. R., OUDEGA, B., STEGEHUIS, F. & DE GRAAF, F. K. 1992. Identification of minor fimbrial subunits involved in biosynthesis of K88 fimbriae. *J Bacteriol,* 174**,** 6350-8.

BAO, R., ESSER, L., POOLE, S., MCVEIGH, A., CHEN, Y. X., SAVARINO, S. J. & XIA, D. 2014. Preliminary X-ray diffraction analysis of CfaA, a molecular chaperone essential for the assembly of CFA/I fimbriae of human enterotoxigenic *Escherichia coli*. *Acta Crystallogr F Struct Biol Commun,* 70**,** 196-9.

BASTUREA, G. N., BODERO, M. D., MORENO, M. E. & MUNSON, G. P. 2008. Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding. *J Bacteriol,* 190**,** 2279-85.

BODERO, M. D., HARDEN, E. A. & MUNSON, G. P. 2008. Transcriptional regulation of subclass 5b fimbriae. *BMC Microbiol,* 8**,** 180.

BODERO, M. D., PILONIETA, M. C. & MUNSON, G. P. 2007. Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic *Escherichia coli*. *J Bacteriol,* 189**,** 1627-32.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BOUDEAU, J., GLASSER, A. L., MASSERET, E., JOLY, B. & DARFEUILLE-MICHAUD, A. 1999. Invasive ability of an *Escherichia coli* strain isolated from the ileal mucosa of a patient with Crohn's disease. *Infect Immun,* 67**,** 4499-509.

BUSCH, A. & WAKSMAN, G. 2012. Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1112-22.

BYRD, W. & CASSELS, F. J. 2003. Mucosal immunization of BALB/c mice using enterotoxigenic *Escherichia coli* colonization factors CFA/I and CS6 administered with and without a mutant heat-labile enterotoxin. *Vaccine,* 21**,** 1884-93.

BYRD, W., MOG, S. R. & CASSELS, F. J. 2003. Pathogenicity and immune response measured in mice following intranasal challenge with enterotoxigenic *Escherichia coli* strains H10407 and B7A. *Infect Immun,* 71**,** 13-21.

CAO, J., KHAN, A. S., BAYER, M. E. & SCHIFFERLI, D. M. 1995. Ordered translocation of 987P fimbrial subunits through the outer membrane of *Escherichia coli*. *J Bacteriol,* 177**,** 3704-13.

CARON, J., COFFIELD, L. M. & SCOTT, J. R. 1989. A plasmid-encoded regulatory gene, rns, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. *Proc Natl Acad Sci U S A,* 86**,** 963-7.

CARVER, T., THOMSON, N., BLEASBY, A., BERRIMAN, M. & PARKHILL, J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics,* 25**,** 119-20.

CASADABAN, M. J. 1976. Regulation of the regulatory gene for the arabinose pathway, araC. *J Mol Biol,* 104**,** 557-66.

CHAKRABORTY, K., GHOSH, S., KOLEY, H., MUKHOPADHYAY, A. K., RAMAMURTHY, T., SAHA, D. R., MUKHOPADHYAY, D., ROYCHOWDHURY, S., HAMABATA, T., TAKEDA, Y. & DAS, S. 2008. Bacterial exotoxins downregulate cathelicidin (hCAP-18/LL-37) and human beta-defensin 1 (HBD-1) expression in the intestinal epithelial cells. *Cell Microbiol,* 10**,** 2520-37.

CHATTOPADHYAY, S., TCHESNOKOVA, V., MCVEIGH, A., KISIELA, D. I., DORI, K., NAVARRO, A., SOKURENKO, E. V. & SAVARINO, S. J. 2012. Adaptive evolution of class 5 fimbrial genes in enterotoxigenic *Escherichia coli* and its functional consequences. *J Biol Chem,* 287**,** 6150-8.

CHEN, L. K., LIU, Y. L., HU, A., CHANG, K. C., LIN, N. T., LAI, M. J. & TSENG, C. C. 2013. Potential of bacteriophage PhiAB2 as an environmental biocontrol agent for the control of multidrug-resistant Acinetobacter baumannii. *BMC Microbiol,* 13**,** 154.

CLAVIJO, A. P., BAI, J. & GOMEZ-DUARTE, O. G. 2010. The Longus type IV pilus of enterotoxigenic *Escherichia coli* (ETEC) mediates bacterial self-aggregation and protection from antimicrobial agents. *Microb Pathog,* 48**,** 230-8.

COTE, J. P. & MOUREZ, M. 2011. Structure-function analysis of the TibA self-associating autotransporter reveals a modular organization. *Infect Immun,* 79**,** 1826-32.

CRAIG, L., PIQUE, M. E. & TAINER, J. A. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol,* 2**,** 363-78.

CRAIG, L., TAYLOR, R. K., PIQUE, M. E., ADAIR, B. D., ARVAI, A. S., SINGH, M., LLOYD, S. J., SHIN, D. S., GETZOFF, E. D., YEAGER, M., FOREST, K. T. & TAINER, J. A. 2003. Type IV pilin structure and assembly: X-ray and EM analyses of *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol Cell,* 11**,** 1139-50.

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

DALEY, A., RANDALL, R., DARSLEY, M., CHOUDHRY, N., THOMAS, N., SANDERSON, I. R., CROFT, N. M. & KELLY, P. 2007. Genetically modified enterotoxigenic *Escherichia coli* vaccines induce mucosal immune responses without inflammation. *Gut,* 56**,** 1550-6.

DARFEUILLE-MICHAUD, A., AUBEL, D., CHAUVIERE, G., RICH, C., BOURGES, M., SERVIN, A. & JOLY, B. 1990. Adhesion of enterotoxigenic *Escherichia coli* to the human colon carcinoma cell line Caco-2 in culture. *Infect Immun,* 58**,** 893-902.

DARFEUILLE-MICHAUD, A., FORESTIER, C., JOLY, B. & CLUZEL, R. 1986. Identification of a nonfimbrial adhesive factor of an enterotoxigenic *Escherichia coli* strain. *Infect Immun,* 52**,** 468-75.

DEL CANTO, F., BOTKIN, D. J., VALENZUELA, P., POPOV, V., RUIZ-PEREZ, F., NATARO, J. P., LEVINE, M. M., STINE, O. C., POP, M., TORRES, A. G. & VIDAL, R. 2012. Identification of Coli Surface Antigen 23, a novel adhesin of enterotoxigenic *Escherichia coli*. *Infect Immun,* 80**,** 2791-801.

DEMOULIN, J. B., VAN SNICK, J. & RENAULD, J. C. 2001. Interleukin-9 (IL-9) induces cell growth arrest associated with sustained signal transducer and activator of transcription activation in lymphoma cells overexpressing the IL-9 receptor. *Cell Growth Differ,* 12**,** 169-74.

DORSEY, F. C., FISCHER, J. F. & FLECKENSTEIN, J. M. 2006. Directed delivery of heat-labile enterotoxin by enterotoxigenic *Escherichia coli*. *Cell Microbiol,* 8**,** 1516-27.

DUPONT, H. L., FORMAL, S. B., HORNICK, R. B., SNYDER, M. J., LIBONATI, J. P., SHEAHAN, D. G., LABREC, E. H. & KALAS, J. P. 1971. Pathogenesis of *Escherichia coli* diarrhea. *N Engl J Med,* 285**,** 1-9.

DUTHY, T. G., MANNING, P. A. & HEUZENROEDER, M. W. 2002. Identification and characterization of assembly proteins of CS5 pili from enterotoxigenic *Escherichia coli*. *J Bacteriol,* 184**,** 1065-77.

DUTHY, T. G., STAENDNER, L. H., MANNING, P. A. & HEUZENROEDER, M. W. 1999. CS5 pilus biosynthesis genes from enterotoxigenic *Escherichia coli* O115:H40. *J Bacteriol,* 181**,** 5847-51.

EDWARDS, R. A., CAO, J. & SCHIFFERLI, D. M. 1996. Identification of major and minor chaperone proteins involved in the export of 987P fimbriae. *J Bacteriol,* 178**,** 3426-33.

ELSINGHORST, E. A. & KOPECKO, D. J. 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli*. *Infect Immun,* 60**,** 2409-17.

ELSINGHORST, E. A. & WEITZ, J. A. 1994. Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli* tib locus is associated with a 104-kilodalton outer membrane protein. *Infect Immun,* 62**,** 3463-71.

ERICSSON, C. D., DUPONT, H. L. & STEFFEN, R. 2008. *Travelers' diarrhea,* Hamilton, Ont. ; Lewiston, NY, BC Decker.

EVANS, D. G., EVANS, D. J., JR. & CLEGG, S. 1980. Detection of enterotoxigenic *Escherichia coli* colonization factor antigen I in stool specimens by an enzyme-linked immunosorbent assay. *J Clin Microbiol,* 12**,** 738-43.

EVANS, D. G., EVANS, D. J., JR. & DUPONT, H. L. 1977a. Virulence factors of enterotoxigenic *Escherichia coli*. *J Infect Dis,* 136 Suppl**,** S118-23.

EVANS, D. G., EVANS, D. J., JR. & TJOA, W. 1977b. Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor. *Infect Immun,* 18**,** 330-7.

EVANS, D. G., SILVER, R. P., EVANS, D. J., JR., CHASE, D. G. & GORBACH, S. L. 1975. Plasmid-controlled colonization factor associated with virulence in *Esherichia coli* enterotoxigenic for humans. *Infect Immun,* 12**,** 656-67.

EVANS, D. J. & EVANS, D. G. 1996. *Escherichia coli* in Diarrheal Disease. *In:* BARON, S. (ed.) *Medical Microbiology.* 4th ed. Galveston (TX) Chapter 25.

EVANS, D. J., JR., EVANS, D. G. & DUPONT, H. L. 1979. Hemagglutination patterns of enterotoxigenic and enteropathogenic *Escherichia coli* determined with human, bovine, chicken, and guinea pig erythrocytes in the presence and absence of mannose. *Infect Immun,* 23**,** 336-46.

FINAMORE, A., ROSELLI, M., IMBINTO, A., SEEBOTH, J., OSWALD, I. P. & MENGHERI, E. 2014. *Lactobacillus amylovorus* inhibits the TLR4 Inflammatory signaling triggered by enterotoxigenic *Escherichia coli* via modulation of the negative regulators and involvement of TLR2 in intestinal Caco-2 Cells and Pig Explants. *PLoS One,* 9**,** e94891,1-10.

FLECKENSTEIN, J. M., HOLLAND, J. T. & HASTY, D. L. 2002. Interaction of an outer membrane protein of enterotoxigenic *Escherichia coli* with cell surface heparan sulfate proteoglycans. *Infect Immun,* 70**,** 1530-7.

FLECKENSTEIN, J. M., KOPECKO, D. J., WARREN, R. L. & ELSINGHORST, E. A. 1996. Molecular characterization of the tia invasion locus from enterotoxigenic *Escherichia coli*. *Infect Immun,* 64**,** 2256-65.

FLECKENSTEIN, J. M., LINDLER, L. E., ELSINGHORST, E. A. & DALE, J. B. 2000. Identification of a gene within a pathogenicity island of enterotoxigenic *Escherichia coli* H10407 required for maximal secretion of the heat-labile enterotoxin. *Infect Immun,* 68**,** 2766-74.

FLECKENSTEIN, J. M., ROY, K., FISCHER, J. F. & BURKITT, M. 2006. Identification of a two-partner secretion locus of enterotoxigenic *Escherichia coli*. *Infect Immun,* 74**,** 2245-58.

FLORES, J., DUPONT, H. L., LEE, S. A., BELKIND-GERSON, J., PAREDES, M., MOHAMED, J. A., ARMITIGE, L. Y., GUO, D. C. & OKHUYSEN, P. C. 2008. Influence of host interleukin-10 polymorphisms on development of traveler's diarrhea due to heat-labile enterotoxin-producing *Escherichia coli* in travelers from the United States who are visiting Mexico. *Clin Vaccine Immunol,* 15**,** 1194-8.

FREEDMAN, D. J., TACKET, C. O., DELEHANTY, A., MANEVAL, D. R., NATARO, J. & CRABB, J. H. 1998. Milk immunoglobulin with specific activity against purified colonization factor antigens can protect against oral challenge with enterotoxigenic *Escherichia coli*. *J Infect Dis,* 177**,** 662-7.

FROEHLICH, B., HOLTZAPPLE, E., READ, T. D. & SCOTT, J. R. 2004. Horizontal transfer of CS1 pilin genes of enterotoxigenic *Escherichia coli*. *J Bacteriol,* 186**,** 3230-7.

FROEHLICH, B., HUSMANN, L., CARON, J. & SCOTT, J. R. 1994a. Regulation of rns, a positive regulatory factor for pili of enterotoxigenic *Escherichia coli*. *J Bacteriol,* 176**,** 5385-92.

FROEHLICH, B., PARKHILL, J., SANDERS, M., QUAIL, M. A. & SCOTT, J. R. 2005. The pCoo plasmid of enterotoxigenic *Escherichia coli* is a mosaic cointegrate. *J Bacteriol,* 187**,** 6509-16.

FROEHLICH, B. J., KARAKASHIAN, A., MELSEN, L. R., WAKEFIELD, J. C. & SCOTT, J. R. 1994b. CooC and CooD are required for assembly of CS1 pili. *Mol Microbiol,* 12**,** 387-401.

FROEHLICH, B. J., KARAKASHIAN, A., SAKELLARIS, H. & SCOTT, J. R. 1995. Genes for CS2 pili of enterotoxigenic *Escherichia coli* and their interchangeability with those for CS1 pili. *Infect Immun,* 63**,** 4849-56.

FUJIWARA, S., HASHIBA, H., HIROTA, T. & FORSTNER, J. F. 2001. Inhibition of the binding of enterotoxigenic *Escherichia coli* Pb176 to human intestinal epithelial cell line HCT-8 by an extracellular protein fraction containing BIF of Bifidobacterium longum SBT2928: suggestive evidence of blocking of the binding receptor gangliotetraosylceramide on the cell surface. *Int J Food Microbiol,* 67**,** 97-106.

FUKAKUSA, S., KAWAHARA, K., NAKAMURA, S., IWASHITA, T., BABA, S., NISHIMURA, M., KOBAYASHI, Y., HONDA, T., IIDA, T., TANIGUCHI, T. & OHKUBO, T. 2012. Structure of the CFA/III major pilin subunit CofA from human enterotoxigenic *Escherichia coli* determined at 0.90 A resolution by sulfur-SAD phasing. *Acta Crystallogr D Biol Crystallogr,* 68**,** 1418-29.

GAASTRA, W., SOMMERFELT, H., VAN DIJK, L., KUSTERS, J. G., SVENNERHOLM, A. M. & GREWAL, H. M. 2002. Antigenic variation within the subunit protein of members of the colonization factor antigen I group of fimbrial proteins in human enterotoxigenic *Escherichia coli*. *Int J Med Microbiol,* 292**,** 43-50.

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

GALKIN, V. E., KOLAPPAN, S., NG, D., ZONG, Z., LI, J., YU, X., EGELMAN, E. H. & CRAIG, L. 2013. The structure of the CS1 pilus of enterotoxigenic *Escherichia coli* reveals structural polymorphism. *J Bacteriol,* 195**,** 1360-70.

GALLEGOS, M. T., SCHLEIF, R., BAIROCH, A., HOFMANN, K. & RAMOS, J. L. 1997. Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev,* 61**,** 393-410.

GHOSAL, A., BHOWMICK, R., BANERJEE, R., GANGULY, S., YAMASAKI, S., RAMAMURTHY, T., HAMABATA, T. & CHATTERJEE, N. S. 2009. Characterization and studies of the cellular interaction of native colonization factor CS6 purified from a clinical isolate of enterotoxigenic *Escherichia coli*. *Infect Immun,* 77**,** 2125-35.

GIRON, J. A., GOMEZ-DUARTE, O. G., JARVIS, K. G. & KAPER, J. B. 1997. Longus pilus of enterotoxigenic *Escherichia coli* and its relatedness to other type-4 pili--a minireview. *Gene,* 192**,** 39-43.

GIRON, J. A., LEVINE, M. M. & KAPER, J. B. 1994. Longus: a long pilus ultrastructure produced by human enterotoxigenic *Escherichia coli*. *Mol Microbiol,* 12**,** 71-82.

GIRON, J. A., VIBOUD, G. I., SPERANDIO, V., GOMEZ-DUARTE, O. G., MANEVAL, D. R., ALBERT, M. J., LEVINE, M. M. & KAPER, J. B. 1995a. Prevalence and association of the longus pilus structural gene (lngA) with colonization factor antigens, enterotoxin types, and serotypes of enterotoxigenic *Escherichia coli*. *Infect Immun,* 63**,** 4195-8.

GIRON, J. A., XU, J. G., GONZALEZ, C. R., HONE, D., KAPER, J. B. & LEVINE, M. M. 1995b. Simultaneous expression of CFA/I and CS3 colonization factor antigens of enterotoxigenic *Escherichia coli* by delta aroC, delta aroD Salmonella typhi vaccine strain CVD 908. *Vaccine,* 13**,** 939-46.

GOMEZ-DUARTE, O. G., CHATTOPADHYAY, S., WEISSMAN, S. J., GIRON, J. A., KAPER, J. B. & SOKURENKO, E. V. 2007. Genetic diversity of the gene cluster encoding longus, a type IV pilus of enterotoxigenic *Escherichia coli*. *J Bacteriol,* 189**,** 9145-9.

GREWAL, H. M., VALVATNE, H., BHAN, M. K., VAN DIJK, L., GAASTRA, W. & SOMMERFELT, H. 1997a. A new putative fimbrial colonization factor, CS19, of human enterotoxigenic *Escherichia coli*. *Infect Immun,* 65**,** 507-13.

GREWAL, H. M. S., VALVATNE, H., BHAN, M. K., VANDIJK, L., GAASTRA, W. & SOMMERFELT, H. 1997b. A new putative fimbrial colonization factor, CS19, of human enterotoxigenic *Escherichia coli*. *Infection and Immunity,* 65**,** 507-513.

GUSTAVSSON, N., DIEZ, A. & NYSTROM, T. 2002. The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Mol Microbiol,* 43**,** 107-17.

GUZMAN, L. M., BELIN, D., CARSON, M. J. & BECKWITH, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol,* 177**,** 4121-30.

HALL, R. H., MANEVAL, D. R., JR., COLLINS, J. H., THEIBERT, J. L. & LEVINE, M. M. 1989. Purification and analysis of colonization factor antigen I, coli surface antigen 1, and coli surface antigen 3 fimbriae from enterotoxigenic *Escherichia coli*. *J Bacteriol,* 171**,** 6372-4.

HANAHAN, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol,* 166**,** 557-80.

HART, E., YANG, J., TAUSCHEK, M., KELLY, M., WAKEFIELD, M. J., FRANKEL, G., HARTLAND, E. L. & ROBINS-BROWNE, R. M. 2008. RegA, an AraC-like protein, is a global transcriptional regulator that controls virulence gene expression in *Citrobacter rodentium*. *Infect Immun,* 76**,** 5247-56.

HEUZENROEDER, M. W., ELLIOT, T. R., THOMAS, C. J., HALTER, R. & MANNING, P. A. 1990. A new fimbrial type (PCFO9) on enterotoxigenic *Escherichia coli* 09:H-LT+ isolated from a case of infant diarrhea in central Australia. *FEMS Microbiol Lett,* 54**,** 55-60.

HIBBERD, M. L., MCCONNELL, M. M., FIELD, A. M. & ROWE, B. 1990. The fimbriae of human enterotoxigenic *Escherichia coli* strain 334 are related to CS5 fimbriae. *J Gen Microbiol,* 136**,** 2449-56.

HOFFMAN, J. A., BADGER, J. L., ZHANG, Y., HUANG, S. H. & KIM, K. S. 2000. *Escherichia coli* K1 aslA contributes to invasion of brain microvascular endothelial cells in vitro and in vivo. *Infect Immun,* 68**,** 5062-7.

HONARVAR, S., CHOI, B. K. & SCHIFFERLI, D. M. 2003. Phase variation of the 987P-like CS18 fimbriae of human enterotoxigenic *Escherichia coli* is regulated by site-specific recombinases. *Mol Microbiol,* 48**,** 157-71.

HORNEF, M. W., WICK, M. J., RHEN, M. & NORMARK, S. 2002. Bacterial strategies for overcoming host innate and adaptive immune responses. *Nat Immunol,* 3**,** 1033-40.

HUANG, D. B., DUPONT, H. L., JIANG, Z. D., CARLIN, L. & OKHUYSEN, P. C. 2004. Interleukin-8 response in an intestinal HCT-8 cell line infected with enteroaggregative and enterotoxigenic *Escherichia coli*. *Clin Diagn Lab Immunol,* 11**,** 548-51.

HUNG, D. T., SHAKHNOVICH, E. A., PIERSON, E. & MEKALANOS, J. J. 2005. Small-molecule inhibitor of *Vibrio cholerae* virulence and intestinal colonization. *Science,* 310**,** 670-4.

JACOBS, A. A., ROOSENDAAL, B., VAN BREEMEN, J. F. & DE GRAAF, F. K. 1987a. Role of phenylalanine 150 in the receptor-binding domain of the K88 fibrillar subunit. *J Bacteriol,* 169**,** 4907-11.

JACOBS, A. A., VENEMA, J., LEEVEN, R., VAN PELT-HEERSCHAP, H. & DE GRAAF, F. K. 1987b. Inhibition of adhesive activity of K88 fibrillae by peptides derived from the K88 adhesin. *J Bacteriol,* 169**,** 735-41.

JALAJAKUMARI, M. B., THOMAS, C. J., HALTER, R. & MANNING, P. A. 1989. Genes for biosynthesis and assembly of CS3 pili of CFA/II enterotoxigenic *Escherichia coli*: novel regulation of pilus production by bypassing an amber codon. *Mol Microbiol,* 3**,** 1685-95.

JANSSON, L., TOBIAS, J., JAREFJALL, C., LEBENS, M., SVENNERHOLM, A. M. & TENEBERG, S. 2009. Sulfatide recognition by colonization factor antigen CS6 from enterotoxigenic *Escherichia coli*. *PLoS One,* 4**,** e4487,1-10.

JANSSON, L., TOBIAS, J., LEBENS, M., SVENNERHOLM, A. M. & TENEBERG, S. 2006. The major subunit, CfaB, of colonization factor antigen i from enterotoxigenic *Escherichia coli* is a glycosphingolipid binding protein. *Infect Immun,* 74**,** 3488-97.

JIANG, Z. D., GREENBERG, D., NATARO, J. P., STEFFEN, R. & DUPONT, H. L. 2002. Rate of occurrence and pathogenic effect of enteroaggregative *Escherichia coli* virulence factors in international travelers. *J Clin Microbiol,* 40**,** 4185-90.

JORDI, B. J., WILLSHAW, G. A., VAN DER ZEIJST, B. A. & GAASTRA, W. 1992. The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic *Escherichia coli*. *DNA Seq,* 2**,** 257-63.

KALSER, M. H., COHEN, R., ARTEAGA, I., YAWN, E., MAYORAL, L., HOFFERT, W. & FRAZIER, D. 1966. Normal viral and bacterial flora of the human small and large intestine. *N Engl J Med,* 274**,** 558-63 contd.

KHAN, A. S. & SCHIFFERLI, D. M. 1994. A minor 987P protein different from the structural fimbrial subunit is the adhesin. *Infect Immun,* 62**,** 4233-43.

KIRN, T. J., LAFFERTY, M. J., SANDOE, C. M. & TAYLOR, R. K. 2000. Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by *Vibrio cholerae*. *Mol Microbiol,* 35**,** 896-910.

KLEMM, P., VEJBORG, R. M. & SHERLOCK, O. 2006. Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol,* 296**,** 187-95.

KNUTTON, S., LLOYD, D. R. & MCNEISH, A. S. 1987. Identification of a new fimbrial structure in enterotoxigenic *Escherichia coli* (ETEC) serotype O148:H28 which adheres to human intestinal mucosa: a potentially new human ETEC colonization factor. *Infect Immun,* 55**,** 86-92.

KOLAPPAN, S., ROOS, J., YUEN, A. S., PIERCE, O. M. & CRAIG, L. 2012. Structural characterization of CFA/III and Longus type IVb pili from enterotoxigenic *Escherichia coli*. *J Bacteriol,* 194**,** 2725-35.

KOPYLOVA, E., NOE, L. & TOUZET, H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics,* 28**,** 3211-7.

KUMAR, P., LUO, Q., VICKERS, T. J., SHEIKH, A., LEWIS, W. G. & FLECKENSTEIN, J. M. 2013. EatA, an Immununogenic Protective Antigen of Enterotoxigenic *Escherichia coli* Degrades Intestinal Mucin. *Infect Immun*.

LAI, X. H., XU, J. G., MELGAR, S. & UHLIN, B. E. 1999. An apoptotic response by J774 macrophage cells is common upon infection with diarrheagenic *Escherichia coli*. *FEMS Microbiol Lett,* 172**,** 29-34.

LASSMANN, T. & SONNHAMMER, E. L. 2005. Kalign--an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics,* 6**,** 298.

LAW, C. W., CHEN, Y., SHI, W. & SMYTH, G. K. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol,* 15**,** R29.

LEVINE, M. M. 1981. Adhesion of enterotoxigenic *Escherichia coli* in humans and animals. *Ciba Found Symp,* 80**,** 142-60.

LEVINE, M. M., NALIN, D. R., HOOVER, D. L., BERGQUIST, E. J., HORNICK, R. B. & YOUNG, C. R. 1979. Immunity to enterotoxigenic *Escherichia coli*. *Infect Immun,* 23**,** 729-36.

LEVINE, M. M., RENNELS, M. B., CISNEROS, L., HUGHES, T. P., NALIN, D. R. & YOUNG, C. R. 1980. Lack of person-to-person transmission of enterotoxigenic *Escherichia coli* despite close contact. *Am J Epidemiol,* 111**,** 347-55.

LI, Y. F., POOLE, S., NISHIO, K., JANG, K., RASULOVA, F., MCVEIGH, A., SAVARINO, S. J., XIA, D. & BULLITT, E. 2009. Structure of CFA/I fimbriae from enterotoxigenic *Escherichia coli*. *Proc Natl Acad Sci U S A,* 106**,** 10793-8.

LI, Y. F., POOLE, S., RASULOVA, F., MCVEIGH, A. L., SAVARINO, S. J. & XIA, D. 2007. A receptor-binding site as revealed by the crystal structure of CfaE, the colonization factor antigen I fimbrial adhesin of enterotoxigenic *Escherichia coli*. *J Biol Chem,* 282**,** 23970-80.

LINDENTHAL, C. & ELSINGHORST, E. A. 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect Immun,* 67**,** 4084-91.

LINDENTHAL, C. & ELSINGHORST, E. A. 2001. Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestine epithelial cells. *Infect Immun,* 69**,** 52-7.

LIU, Y., ESSER, L., INTERLANDI, G., KISIELA, D. I., TCHESNOKOVA, V., THOMAS, W. E., SOKURENKO, E., XIA, D. & SAVARINO, S. J. 2013. Tight conformational coupling between the domains of the enterotoxigenic *Escherichia coli* fimbrial adhesin CfaE regulates binding state transition. *J Biol Chem,* 288**,** 9993-10001.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LOWDEN, M. J., SKORUPSKI, K., PELLEGRINI, M., CHIORAZZO, M. G., TAYLOR, R. K. & KULL, F. J. 2010. Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes. *Proc Natl Acad Sci U S A,* 107**,** 2860-5.

LUDI, S., FREY, J., FAVRE, D., VIRET, J. F., ZIETHLOW, V., BOGHENBOR, K. K. & STOFFEL, M. H. 2008. Morphological and immunocytochemical analysis of *Escherichia coli*-specific surface antigens in wildtype strains and in recombinant Vibrio cholerae. *Antonie Van Leeuwenhoek,* 93**,** 185-92.

LUDWIG, A., VON RHEIN, C., BAUER, S., HUTTINGER, C. & GOEBEL, W. 2004. Molecular analysis of cytolysin A (ClyA) in pathogenic *Escherichia coli* strains. *J Bacteriol,* 186**,** 5311-20.

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

MADHAVAN, T. P., STEEN, J. A., HUGENHOLTZ, P. & SAKELLARIS, H. 2014. Genome Sequence of Enterotoxigenic *Escherichia coli* Strain B2C. *Genome Announc,* 2.

MAHON, V., FAGAN, R. P. & SMITH, S. G. 2012. Snap denaturation reveals dimerization by AraC-like protein Rns. *Biochimie,* 94**,** 2058-61.

MAMMARAPPALLIL, J. G. & ELSINGHORST, E. A. 2000. Epithelial cell adherence mediated by the enterotoxigenic *Escherichia coli* tia protein. *Infect Immun,* 68**,** 6595-601.

MARRON, M. B. & SMYTH, C. J. 1995. Molecular analysis of the cso operon of Enterotoxigenic *Escherichia coli* reveals that CsoA is the adhesin of CS1 fimbriae and that the accessory genes are interchangeable with those of the cfa operon. *Microbiology-Uk,* 141**,** 2849-2859.

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

MATHEWSON, J. J., DUPONT, H. L., MORGAN, D. R., THORNTON, S. A. & ERICSSON, C. D. 1983. Enteroadherent *Escherichia coli* associated with travellers' diarrhoea. *Lancet,* 1**,** 1048.

MAZARIEGO-ESPINOSA, K., CRUZ, A., LEDESMA, M. A., OCHOA, S. A. & XICOHTENCATL-CORTES, J. 2010. Longus, a type IV pilus of enterotoxigenic *Escherichia coli*, is involved in adherence to intestinal epithelial cells. *J Bacteriol,* 192**,** 2791-800.

MCBROOM, A. J., JOHNSON, A. P., VEMULAPALLI, S. & KUEHN, M. J. 2006. Outer membrane vesicle production by *Escherichia coli* is independent of membrane instability. *J Bacteriol,* 188**,** 5385-92.

MELI, A. C., KONDRATOVA, M., MOLLE, V., COQUET, L., KAJAVA, A. V. & SAINT, N. 2009. EtpB is a pore-forming outer membrane protein showing TpsB protein features involved in the two-partner secretion system. *J Membr Biol,* 230**,** 143-54.

MONTEVILLE, M. R., RIDDLE, M. S., BAHT, U., PUTNAM, S. D., FRENCK, R. W., BROOKS, K., MOUSTAFA, M., BLAND, J. & SANDERS, J. W. 2006. Incidence, etiology, and impact of diarrhea among deployed US military personnel in support of Operation Iraqi Freedom and Operation Enduring Freedom. *Am J Trop Med Hyg,* 75**,** 762-7.

MOORMANN, C., BENZ, I. & SCHMIDT, M. A. 2002. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect Immun,* 70**,** 2264-70.

MU, X. Q., SAVARINO, S. J. & BULLITT, E. 2008. The three-dimensional structure of CFA/I adhesion pili: traveler's diarrhea bacteria hang on by a spring. *J Mol Biol,* 376**,** 614-20.

MUNSON, G. P., HOLCOMB, L. G., ALEXANDER, H. L. & SCOTT, J. R. 2002. In vitro identification of Rns-regulated genes. *J Bacteriol,* 184**,** 1196-9.

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

MUNSON, G. P. & SCOTT, J. R. 1999. Binding site recognition by Rns, a virulence regulator in the AraC family. *J Bacteriol,* 181**,** 2110-7.

MUNSON, G. P. & SCOTT, J. R. 2000. Rns, a virulence regulator within the AraC family, requires binding sites upstream and downstream of its own promoter to function as an activator. *Mol Microbiol,* 36**,** 1391-402.

NADA, R. A., ARMSTRONG, A., SHAHEEN, H. I., NAKHLA, I., SANDERS, J. W., RIDDLE, M. S., YOUNG, S. & SEBENY, P. 2013. Phenotypic and genotypic characterization of enterotoxigenic *Escherichia coli* isolated from U.S. military personnel participating in Operation Bright Star, Egypt, from 2005 to 2009. *Diagn Microbiol Infect Dis,* 76**,** 272-7.

NADA, R. A., SHAHEEN, H. I., KHALIL, S. B., MANSOUR, A., EL-SAYED, N., TOUNI, I., WEINER, M., ARMSTRONG, A. W. & KLENA, J. D. 2011. Discovery and phylogenetic analysis of novel members of class b enterotoxigenic *Escherichia coli* adhesive fimbriae. *J Clin Microbiol,* 49**,** 1403-10.

NICKLASSON, M., SJOLING, A., LEBENS, M., TOBIAS, J., JANZON, A., BRIVE, L. & SVENNERHOLM, A. M. 2008. Mutations in the periplasmic chaperone leading to loss of surface expression of the colonization factor CS6 in enterotoxigenic *Escherichia coli* (ETEC) clinical isolates. *Microb Pathog,* 44**,** 246-54.

NORTHEY, G., EVANS, M. R., SARVOTHAM, T. S., THOMAS, D. R. & HOWARD, T. J. 2007. Sentinel surveillance for travellers' diarrhoea in primary care. *BMC Infect Dis,* 7**,** 126.

NUCCIO, S. P. & BAUMLER, A. J. 2007. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev,* 71**,** 551-75.

OCHOA-REPARAZ, J., RYNDA, A., ASCON, M. A., YANG, X., KOCHETKOVA, I., RICCARDI, C., CALLIS, G., TRUNKLE, T. & PASCUAL, D. W. 2008. IL-13 production by regulatory T cells protects against experimental autoimmune encephalomyelitis independently of autoantigen. *J Immunol,* 181**,** 954-68.

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

OSCARSSON, J., MIZUNOE, Y., LI, L., LAI, X. H., WIESLANDER, A. & UHLIN, B. E. 1999. Molecular analysis of the cytolytic protein ClyA (SheA) from *Escherichia coli*. *Mol Microbiol,* 32**,** 1226-38.

PARANCHYCH, W. & FROST, L. S. 1988. The physiology and biochemistry of pili. *Adv Microb Physiol,* 29**,** 53-114.

PARGE, H. E., FOREST, K. T., HICKEY, M. J., CHRISTENSEN, D. A., GETZOFF, E. D. & TAINER, J. A. 1995. Structure of the fibre-forming protein pilin at 2.6 A resolution. *Nature,* 378**,** 32-8.

PASCUAL, D. W., HONE, D. M., HALL, S., VAN GINKEL, F. W., YAMAMOTO, M., WALTERS, N., FUJIHASHI, K., POWELL, R. J., WU, S., VANCOTT, J. L., KIYONO, H. & MCGHEE, J. R. 1999. Expression of recombinant enterotoxigenic *Escherichia coli* colonization factor antigen I by Salmonella typhimurium elicits a biphasic T helper cell response. *Infect Immun,* 67**,** 6249-56.

PATEL, S. K., DOTSON, J., ALLEN, K. P. & FLECKENSTEIN, J. M. 2004. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. *Infect Immun,* 72**,** 1786-94.

PEREZ-CASAL, J., SWARTLEY, J. S. & SCOTT, J. R. 1990. Gene encoding the major subunit of CS1 pili of human enterotoxigenic *Escherichia coli*. *Infect Immun,* 58**,** 3594-600.

PERKINS, T. T., DAVIES, M. R., KLEMM, E. J., ROWLEY, G., WILEMAN, T., JAMES, K., KEANE, T., MASKELL, D., HINTON, J. C., DOUGAN, G. & KINGSLEY, R. A. 2013. ChIP-seq and transcriptome analysis of the OmpR regulon of Salmonella enterica serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization. *Mol Microbiol,* 87**,** 526-38.

PICHEL, M., BINSZTEIN, N. & VIBOUD, G. 2000. CS22, a novel human enterotoxigenic *Escherichia coli* adhesin, is related to CS15. *Infect Immun,* 68**,** 3280-5.

PILONIETA, M. C., BODERO, M. D. & MUNSON, G. P. 2007. CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic *Escherichia coli*. *J Bacteriol,* 189**,** 5060-7.

POOLE, S. T., MCVEIGH, A. L., ANANTHA, R. P., LEE, L. H., AKAY, Y. M., PONTZER, E. A., SCOTT, D. A., BULLITT, E. & SAVARINO, S. J. 2007. Donor strand complementation governs intersubunit interaction of fimbriae of the alternate chaperone pathway. *Mol Microbiol,* 63**,** 1372-84.

PORTER, M. E., MITCHELL, P., ROE, A. J., FREE, A., SMITH, D. G. & 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**,** 1117-33.

REN, W., YIN, J., DUAN, J., LIU, G., ZHU, X., CHEN, S., LI, T., WANG, S., TANG, Y. & HARDWIDGE, P. R. 2014. Mouse intestinal innate immune responses altered by enterotoxigenic *Escherichia coli* (ETEC) infection. *Microbes Infect,* 16**,** 954-61.

REZUCHOVA, B., MITICKA, H., HOMEROVA, D., ROBERTS, M. & KORMANEC, J. 2003. New members of the *Escherichia coli* sigmaE regulon identified by a two-plasmid system. *FEMS Microbiol Lett,* 225**,** 1-7.

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

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.

ROY, K., BARTELS, S., QADRI, F. & FLECKENSTEIN, J. M. 2010. Enterotoxigenic *Escherichia coli* elicits immune responses to multiple surface proteins. *Infect Immun,* 78**,** 3027-35.

ROY, K., HAMILTON, D., ALLEN, K. P., RANDOLPH, M. P. & FLECKENSTEIN, J. M. 2008. The EtpA exoprotein of enterotoxigenic *Escherichia coli* promotes intestinal colonization and is a protective antigen in an experimental model of murine infection. *Infect Immun,* 76**,** 2106-12.

ROY, K., HAMILTON, D., OSTMANN, M. M. & FLECKENSTEIN, J. M. 2009a. Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic *Escherichia coli* in a murine model. *Vaccine,* 27**,** 4601-8.

ROY, K., HILLIARD, G. M., HAMILTON, D. J., LUO, J., OSTMANN, M. M. & FLECKENSTEIN, J. M. 2009b. Enterotoxigenic *Escherichia coli* EtpA mediates adhesion between flagella and host cells. *Nature,* 457**,** 594-8.

ROY, S. P., RAHMAN, M. M., YU, X. D., TUITTILA, M., KNIGHT, S. D. & ZAVIALOV, A. V. 2012. Crystal structure of enterotoxigenic *Escherichia coli* colonization factor CS6 reveals a novel type of functional assembly. *Mol Microbiol,* 86**,** 1100-15.

SAHL, J. W. & RASKO, D. A. 2012. Analysis of Global Transcriptional Profiles of Enterotoxigenic *Escherichia coli* Isolate E24377A. *Infection and Immunity,* 80**,** 1232-1242.

SAHL, J. W., STEINSLAND, H., REDMAN, J. C., ANGIUOLI, S. V., NATARO, J. P., SOMMERFELT, H. & RASKO, D. A. 2011. A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect Immun,* 79**,** 950-60.

SAKELLARIS, H., BALDING, D. P. & SCOTT, J. R. 1996. Assembly proteins of CS1 pili of enterotoxigenic *Escherichia coli*. *Mol Microbiol,* 21**,** 529-41.

SAKELLARIS, H., MUNSON, G. P. & SCOTT, J. R. 1999a. A conserved residue in the tip proteins of CS1 and CFA/I pili of enterotoxigenic *Escherichia coli* that is essential for adherence. *Proc Natl Acad Sci U S A,* 96**,** 12828-32.

SAKELLARIS, H., PENUMALLI, V. R. & SCOTT, J. R. 1999b. The level of expression of the minor pilin subunit, CooD, determines the number of CS1 pili assembled on the cell surface of *Escherichia coli*. *J Bacteriol,* 181**,** 1694-7.

SAKELLARIS, H. & SCOTT, J. R. 1998. New tools in an old trade: CS1 pilus morphogenesis. *Mol Microbiol,* 30**,** 681-7.

SAMBROOK, J. & RUSSELL, D. W. 2001. *Molecular Cloning: a laboratory manual,* Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

SATTERWHITE, T. K., EVANS, D. G., DUPONT, H. L. & EVANS, D. J., JR. 1978. Role of *Escherichia coli* colonisation factor antigen in acute diarrhoea. *Lancet,* 2**,** 181-4.

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**,** 2177-85.

SAVARINO, S. J. 2006. Adhesin as immunogen against enterotoxigenic *Escherichia coli*. USA patent application 20060153878.

SAVELKOUL, P. H., WILLSHAW, G. A., MCCONNELL, M. M., SMITH, H. R., HAMERS, A. M., VAN DER ZEIJST, B. A. & GAASTRA, W. 1990. Expression of CFA/I fimbriae is positively regulated. *Microb Pathog,* 8**,** 91-9.

SCHIFFERLI, D. M., BEACHEY, E. H. & TAYLOR, R. K. 1990. The 987P fimbrial gene cluster of enterotoxigenic *Escherichia coli* is plasmid encoded. *Infect Immun,* 58**,** 149-56.

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

SHEIKH, J., CZECZULIN, J. R., HARRINGTON, S., HICKS, S., HENDERSON, I. R., LE BOUGUENEC, C., GOUNON, P., PHILLIPS, A. & NATARO, J. P. 2002. A novel dispersin protein in enteroaggregative *Escherichia coli*. *J Clin Invest,* 110**,** 1329-37.

SHERLOCK, O., VEJBORG, R. M. & KLEMM, P. 2005. The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. *Infect Immun,* 73**,** 1954-63.

SINGHERA, G. K., MACREDMOND, R. & DORSCHEID, D. R. 2008. Interleukin-9 and -13 inhibit spontaneous and corticosteroid induced apoptosis of normal airway epithelial cells. *Exp Lung Res,* 34**,** 579-98.

SMITH, H. R., WILLSHAW, G. A. & ROWE, B. 1982. Mapping of a plasmid, coding for colonization, factor antigen I and heat-stable enterotoxin production, isolated from an enterotoxigenic strain of *Escherichia coli*. *J Bacteriol,* 149**,** 264-75.

SOTO, G. E. & HULTGREN, S. J. 1999. Bacterial adhesins: common themes and variations in architecture and assembly. *J Bacteriol,* 181**,** 1059-71.

STARKS, A. M., FROEHLICH, B. J., JONES, T. N. & SCOTT, J. R. 2006. Assembly of CS1 pili: the role of specific residues of the major pilin, CooA. *J Bacteriol,* 188**,** 231-9.

STEFFEN, R., CASTELLI, F., DIETER NOTHDURFT, H., ROMBO, L. & JANE ZUCKERMAN, N. 2005. Vaccination against enterotoxigenic *Escherichia coli*, a cause of travelers' diarrhea. *J Travel Med,* 12**,** 102-7.

SVENNERHOLM, A. M. & TOBIAS, J. 2008. Vaccines against enterotoxigenic *Escherichia coli*. *Expert Rev Vaccines,* 7**,** 795-804.

SWIDSINSKI, A., LOENING-BAUCKE, V., LOCHS, H. & HALE, L. P. 2005. Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. *World J Gastroenterol,* 11**,** 1131-40.

SZABO, E., SKEDSMO, A., SONNEVEND, A., AL-DHAHERI, K., EMODY, L., USMANI, A. & PAL, T. 2005. Curli expression of enterotoxigenic *Escherichia coli*. *Folia Microbiol (Praha),* 50**,** 40-6.

TACKET, C. O., MANEVAL, D. R. & LEVINE, M. M. 1987. Purification, morphology, and genetics of a new fimbrial putative colonization factor of enterotoxigenic *Escherichia coli* O159:H4. *Infect Immun,* 55**,** 1063-9.

TAKAYA, A., TOMOYASU, T., TOKUMITSU, A., MORIOKA, M. & YAMAMOTO, T. 2002. The ATP-Dependent Lon Protease of Salmonella enterica Serovar Typhimurium Regulates Invasion and Expression of Genes Carried on Salmonella Pathogenicity Island 1. *Journal of Bacteriology,* 184**,** 224-232.

TAKESHITA, S., SATO, M., TOBA, M., MASAHASHI, W. & HASHIMOTO-GOTOH, T. 1987. High-copy-number and low-copy-number plasmid vectors for lacZ alpha-complementation and chloramphenicol- or kanamycin-resistance selection. *Gene,* 61**,** 63-74.

TAMAYO, E., MERINO, R., GONZALEZ-ROJAS, J., MARQUINA, R., SANTIUSTE, I., AMADO, J. A., RAPPUOLI, R., DEL GIUDICE, G. & MERINO, J. 2005. The *Escherichia coli* heat-labile enterotoxin induces apoptosis of immature lymphocytes in vivo via a glucocorticoid-dependent pathway. *Eur J Immunol,* 35**,** 3505-15.

TANG, Y., LI, F., TAN, B., LIU, G., KONG, X., HARDWIDGE, P. R. & YIN, Y. 2014. Enterotoxigenic *Escherichia coli* infection induces intestinal epithelial cell autophagy. *Vet Microbiol,* 171**,** 160-4.

TANIGUCHI, T., AKEDA, Y., HABA, A., YASUDA, Y., YAMAMOTO, K., HONDA, T. & TOCHIKUBO, K. 2001. Gene cluster for assembly of pilus colonization factor antigen III of enterotoxigenic *Escherichia coli*. *Infect Immun,* 69**,** 5864-73.

TANIGUCHI, T., YASUDA, Y., TOCHIKUBO, K., YAMAMOTO, K. & HONDA, T. 1999. The gene encoding the prepilin peptidase involved in biosynthesis of pilus colonization factor antigen III (CFA/III) of human enterotoxigenic *Escherichia coli*. *Microbiol Immunol,* 43**,** 853-61.

TCHESNOKOVA, V., MCVEIGH, A. L., KIDD, B., YAKOVENKO, O., THOMAS, W. E., SOKURENKO, E. V. & SAVARINO, S. J. 2010. Shear-enhanced binding of intestinal colonization factor antigen I of enterotoxigenic *Escherichia coli*. *Mol Microbiol,* 76**,** 489-502.

TOBIAS, J., LEBENS, M., KALLGARD, S., NICKLASSON, M. & SVENNERHOLM, A. M. 2008. Role of different genes in the CS6 operon for surface expression of Enterotoxigenic *Escherichia coli* colonization factor CS6. *Vaccine,* 26**,** 5373-80.

TROST, B., MOIR, C. A., GILLESPIE, Z. E., KUSALIK, A., MITCHELL, J. A. & ESKIW, C. H. 2015. Concordance between RNA-sequencing data and DNA microarray data in transcriptome analysis of proliferative and quiescent fibroblasts. *R Soc Open Sci,* 2**,** 150402.

TSILIBARIS, V., MAENHAUT-MICHEL, G. & VAN MELDEREN, L. 2006. Biological roles of the Lon ATP-dependent protease. *Res Microbiol,* 157**,** 701-13.

TURNER, S. M., SCOTT-TUCKER, A., COOPER, L. M. & HENDERSON, I. R. 2006. Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*. *FEMS Microbiol Lett,* 263**,** 10-20.

VALVATNE, H., SOMMERFELT, H., GAASTRA, W., BHAN, M. K. & GREWAL, H. M. 1996. Identification and characterization of CS20, a new putative colonization factor of enterotoxigenic *Escherichia coli*. *Infect Immun,* 64**,** 2635-42.

VALVATNE, H., STEINSLAND, H., GREWAL, H. M., MOLBAK, K., VUUST, J. & SOMMERFELT, H. 2004. Identification and molecular characterization of the gene encoding coli surface antigen 20 of enterotoxigenic *Escherichia coli*. *FEMS Microbiol Lett,* 239**,** 131-8.

VERDONCK, F., COX, E. & GODDEERIS, B. M. 2004. F4 fimbriae expressed by porcine enterotoxigenic *Escherichia coli*, an example of an eccentric fimbrial system? *J Mol Microbiol Biotechnol,* 7**,** 155-69.

VIBOUD, G. I., BINSZTEIN, N. & SVENNERHOLM, A. M. 1993. A new fimbrial putative colonization factor, PCFO20, in human enterotoxigenic *Escherichia coli*. *Infect Immun,* 61**,** 5190-7.

VIBOUD, G. I., MCCONNELL, M. M., HELANDER, A. & SVENNERHOLM, A. M. 1996. Binding of enterotoxigenic *Escherichia coli* expressing different colonization factors to tissue-cultured Caco-2 cells and to isolated human enterocytes. *Microb Pathog,* 21**,** 139-47.

VIPIN MADHAVAN, T. P. & SAKELLARIS, H. 2015. Colonization Factors of Enterotoxigenic *Escherichia coli*. *Adv Appl Microbiol,* 90c**,** 155-197.

VOEGELE, K., SAKELLARIS, H. & SCOTT, J. R. 1997. CooB plays a chaperone-like role for the proteins involved in formation of CS1 pili of enterotoxigenic *Escherichia coli*. *Proc Natl Acad Sci U S A,* 94**,** 13257-61.

WAJIMA, T., SABUI, S., FUKUMOTO, M., KANO, S., RAMAMURTHY, T., CHATTERJEE, N. S. & HAMABATA, T. 2011. Enterotoxigenic *Escherichia coli*

CS6 gene products and their roles in CS6 structural protein assembly and cellular adherence. *Microb Pathog,* 51**,** 243-9.

WAJIMA, T., SABUI, S., KANO, S., RAMAMURTHY, T., CHATTERJEE, N. S. & HAMABATA, T. 2013. Entire sequence of the colonization factor coli surface antigen 6-encoding plasmid pCss165 from an enterotoxigenic *Escherichia coli* clinical isolate. *Plasmid,* 70, 343-352.

WANG, X., GAO, X. & HARDWIDGE, P. R. 2012. Heat-labile enterotoxin-induced activation of NF- κB and MAPK pathways in intestinal epithelial cells impacts enterotoxigenic *Escherichia coli* (ETEC) adherence. *Cell Microbiol,* 14**,** 1231-41.

WANG, X. & HARDWIDGE, P. R. 2012. Enterotoxigenic *Escherichia coli* prevents host NF-κB activation by targeting IκBα polyubiquitination. *Infect Immun,* 80**,** 4417-25.

WENNERAS, C. & ERLING, V. 2004. Prevalence of enterotoxigenic *Escherichia coli*-associated diarrhoea and carrier state in the developing world. *J Health Popul Nutr,* 22**,** 370-82.

WENNERAS, C., HOLMGREN, J. & SVENNERHOLM, A. M. 1990. The binding of colonization factor antigens of enterotoxigenic *Escherichia coli* to intestinal cell membrane proteins. *FEMS Microbiol Lett,* 54**,** 107-12.

WOLF, M. K., DE HAAN, L. A., CASSELS, F. J., WILLSHAW, G. A., WARREN, R., BOEDEKER, E. C. & GAASTRA, W. 1997. The CS6 colonization factor of human enterotoxigenic *Escherichia coli* contains two heterologous major subunits. *FEMS Microbiol Lett,* 148**,** 35-42.

WYNN, T. A. 2003. IL-13 effector functions. *Annu Rev Immunol,* 21**,** 425-56.

XU, X. F., TAN, Y. W., LAM, L., HACKETT, J., ZHANG, M. & MOK, Y. K. 2004. NMR structure of a type IVb pilin from Salmonella typhi and its assembly into pilus. *J Biol Chem,* 279**,** 31599-605.

YAKHCHALI, B. & MANNING, P. A. 1997. Epitope analysis of the CS3 fimbrial subunit of human enterotoxigenic *Escherichia coli* and the construction of novel CS3::ST and CS3::LT-B immunogens. *Behring Inst Mitt***,** 124-34.

YAKOVENKO, O., SHARMA, S., FORERO, M., TCHESNOKOVA, V., APRIKIAN, P., KIDD, B., MACH, A., VOGEL, V., SOKURENKO, E. & THOMAS, W. E. 2008. FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *J Biol Chem,* 283**,** 11596-605.

YAMAMOTO, T., HONDA, T., MIWATANI, T. & YOKOTA, T. 1984. A virulence plasmid in *Escherichia coli* enterotoxigenic for humans: intergenetic transfer and expression. *J Infect Dis,* 150**,** 688-98.

YANG, J., TAUSCHEK, M., HART, E., HARTLAND, E. L. & ROBINS-BROWNE, R. M. 2010. Virulence regulation in *Citrobacter rodentium*: the art of timing. *Microb Biotechnol,* 3**,** 259-68.

YANG, J., TAUSCHEK, M. & ROBINS-BROWNE, R. M. 2011. Control of bacterial virulence by AraC-like regulators that respond to chemical signals. *Trends Microbiol,* 19**,** 128-35.

YU, F., INOUYE, S. & INOUYE, M. 1986. Lipoprotein-28, a cytoplasmic membrane lipoprotein from *Escherichia coli*. Cloning, DNA sequence, and expression of its gene. *J Biol Chem,* 261**,** 2284-8.

ZANELLO, G., BERRI, M., DUPONT, J., SIZARET, P. Y., D'INCA, R., SALMON, H. & MEURENS, F. 2011. Saccharomyces cerevisiae modulates immune gene expressions and inhibits ETEC-mediated ERK1/2 and p38 signaling pathways in intestinal epithelial cells. *PLoS One,* 6**,** e18573,1-13.

ZHANG, D., MANOS, J., MA, X., BELAS, R. & KARAOLIS, D. K. 2004. Transcriptional analysis and operon structure of the tagA-orf2-orf3-mop-tagD region on the Vibrio pathogenicity island in epidemic V. cholerae. *FEMS Microbiol Lett,* 235**,** 199-207.

ZHANG, D., RAJANNA, C., SUN, W. & KARAOLIS, D. K. 2003a. Analysis of the Vibrio pathogenicity island-encoded Mop protein suggests a pleiotropic role in the virulence of epidemic *Vibrio cholerae*. *FEMS Microbiol Lett,* 225**,** 311-8.

ZHANG, D., XU, Z., SUN, W. & KARAOLIS, D. K. 2003b. The vibrio pathogenicity island-encoded mop protein modulates the pathogenesis and reactogenicity of epidemic *vibrio cholerae*. *Infect Immun,* 71**,** 510-5.

ZHANG, W. & SACK, D. A. 2012. Progress and hurdles in the development of vaccines against enterotoxigenic *Escherichia coli* in humans. *Expert Rev Vaccines,* 11**,** 677-94.