University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Hodson, Carla

Title:
Control of virulence expression in enterotoxigenic Escherichia coli by the master regulator Rns, and targeting Rns for inhibition of bacterial virulence

Date:
2016

Persistent Link:
https://hdl.handle.net/11343/123851

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

Control of virulence expression in enterotoxigenic *Escherichia coli* by the master regulator Rns, and targeting Rns for inhibition of bacterial virulence

Carla Hodson
orcid.org/0000-0003-4259-2254

Submitted for the degree of Master of Philosophy

July, 2016

Department of Microbiology and Immunology

The University of Melbourne

This thesis is being submitted in total fulfilment of the degree

Abstract

Enterotoxigenic Escherichia coli (ETEC) is a major cause of diarrhoea, particularly for children in developing countries, and is the most common cause of travellers' diarrhoea. The ETEC pathotype is defined by the presence of at least one colonisation factor (CF) and either or both of two toxins, the heat-labile (LT) and the heat-stable (ST) enterotoxins. In addition, a variety of accessory virulence factors have been identified that aid in pathogenesis, but are not present in every strain of ETEC. CFs are antigenically diverse and immunologically distinct, but many are transcriptionally controlled by a conserved AraC-like regulator, Rns.

Rns activates transcription of its target genes by binding to curved, AT-rich regions of DNA in the target promoter, releasing the global gene silencer H-NS, and allowing RNA-polymerase to transcribe the target gene. Rns is part of a family of AraC-like virulence regulators which also includes the master virulence regulators ToxT and RegA, from *Vibrio cholerae* and *Citrobacter rodentium* respectively. These are master virulence regulators that control the expression of an assortment of virulence genes. To date the complete Rns regulon has not been identified, and more gene targets with an impact of ETEC virulence may yet be unidentified.

In order to identify all of the gene targets of Rns, mRNA-sequencing was performed on ETEC strain H10407 and its isogenic *rns* negative mutant. This experiment identified four potential new virulence targets that required the presence of Rns for full expression. These are the *agn43*, *etpB*, *yghJ* and *rtr* genes which encode two autotransporter adhesins, a mucinase and the AraC-family negative regulator (ANR) respectively. Results from further transcriptional assays suggest that Rns directly activates the expression of *etpB*, *yghJ* and *rtr* but indirectly activates the expression of *agn43*.

With antimicrobial resistance an increasing problem, alternative methods of bacterial control are urgently required. Rns is absolutely required for the expression of CFs, an essential virulence factor of ETEC, and this makes it an attractive drug target. With the aim of identifying an inhibitor of Rns activity, and therefore of ETEC-induced diarrhoea, chemical compounds from two sources (those from a commercial small molecule library as well as those from structure based and computer-aided drug design) were screened for their ability to inhibit Rns activity. Several hit compounds were identified and the IC_{50} determined.

Declaration

This is to certify that:

i) This thesis comprises only my original work towards the Master of Philosophy degree except where otherwise indicated.

ii) Due acknowledgement has been made in the text to all other material used.

iii) This thesis is less than 50,000 words in length, exclusive of tables, bibliographies and appendices.

_____

Carla Hodson

Preface

Certain work described in chapter four of this thesis was performed by Dr. Jessica Holien at the St. Vincent's Institute of Medical Research, and is described as such within the text.

Acknowledgements

I would like to thank my supervisors Roy Robins-Browne and Ji Yang for all their hard work and guidance. To all other members of the Robins-Browne lab, thank you for your assistance and advice. Thank you also to my committee members, your regular input was invaluable and greatly appreciated. To the wonderful student support team of our department, especially Rebecca, thank you for helping me navigate the administrative labyrinth that is a research studentship. Finally, thank you to the Microbiology and Immunology teaching staff, who gave me a second home within the department and the opportunity to learn through teaching.

I could not have completed this thesis without your support.

Table of Contents

List of tables

List of figures

1 Introduction

1.1 Disease caused by ETEC

1.1.1 Diarrhoeal disease burden worldwide

Diarrhoea is a leading cause of death for young children in developing countries, second only to pneumonia in the infectious disease category (1, 2). The Global Burden of Disease (GBD) study estimated that in 2013, diarrhoeal disease was responsible for over half a million deaths in children under the age of five (1, 3).

Diarrhoea is usually defined as the passage of three or more loose or liquid stools per day. Death from diarrhoea is primarily due to dehydration and electrolyte imbalance, and the significant mortality of diarrhoea is compounded by secondary effects such as malnutrition and the related developmental defects caused by repeated episodes of diarrhoea during early life (4).

One recent large-scale study, the Global Enteric Multicentre Study (GEMS), which collected stool samples from children with and without diarrhoea from seven sites in Africa and South America (The Gambia, Mali, Mozambique, Kenya, India, Bangladesh, and Pakistan), found that the primary causes of moderate to severe diarrhoea in young children include rotavirus, *Cryptosporidium*, *Shigella*, and enterotoxigenic *Escherichia coli* (ETEC) (5).

1.1.2 Pathogenic *E. coli*

E. coli is a well-known member of the commensal microbiota of humans and other animals, as well as being a favourite 'laboratory workhorse' for use in research. But *E. coli* is also capable of causing disease, including urinary tract infections, sepsis, and diarrhoea (6). *E. coli* strains become pathogenic through the acquisition of virulence factors, often encoded on plasmids or other mobile genetic elements.

Pathotypes of diarrhoeagenic of *E. coli* include enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC) enteroaggregative *E. coli* (EAEC), enterohaemorrhagic *E. coli* (EHEC), and

enterotoxigenic *E. coli* (ETEC). Each pathotype causes disease through a different set of virulence factors, and the aetiology, symptoms, and severity of the disease they cause differs between pathotypes. EPEC is a major cause of illness in infants in developing countries, as is ETEC, which is also the most common cause of traveller's diarrhoea (5, 7). EAEC is being increasingly detected as an important pathogen worldwide (8). EHEC can affect people of all ages, and is the major cause of haemorrhagic colitis and haemolytic uremic syndrome (HUS) (9). Each of these diarrhoeagenic pathotypes is an important cause of disease within their own pathogenic niche.

1.1.3 ETEC as a cause of diarrhoea

ETEC is defined by its primary virulence factors, that is, the expression of at least one colonisation factor (CF) and one or both of two enterotoxins, heat-stable toxin (ST) and heat-labile toxin (LT). Accessory virulence factors have also been identified, which appear to play a role in pathogenesis, but are not present in all ETEC strains. ETEC disease typically lasts 3-5 days, and the severity ranges from mild diarrhoea to a severe cholera-like illness (10). ETEC primarily affects young children, as natural immunity to specific ETEC strains develops from repeated exposure. Accordingly, adults who live in regions where ETEC is endemic are generally not susceptible to the strains that are circulating in the area (11, 12).

ETEC is also the leading cause of travellers' diarrhoea (TD). In recent studies, incidence of TD in Europeans visiting Africa, South-East Asia, or Latin America ranged from as low as 8%, to as high as 50% (7, 13-16). TD is rarely life-threatening in healthy adults, but may develop into long-term gastrointestinal illness, and can trigger the onset of irritable bowel syndrome (17, 18).

As well as humans, ETEC naturally infects pigs and cows (19-22). Neonatal diarrhoea in pigs is a significant source of economic loss for the industry. In some piggeries, routine antimicrobial prophylaxis is used to protect piglets from diarrhoea, leading to high levels of antimicrobial resistance (23). One vaccine is commercially available and others are being trialled in an attempt to

control diarrhoea in piglets without the use of antimicrobials (24, 25). In these trials, piglets are protected by suckling milk containing antibodies from vaccinated sows (26).

1.2 The evolution of ETEC

The ETEC pathotype comprises a diverse group of *E. coli* strains, with a large range of O- and H-serogroups, and sequence types (27, 28). Nonetheless, most ETEC belong to *E. coli* phylogenetic groups A, B1 or C, and correlations between genetic background, serogroups, CFs and toxins do exist (27-31). Escobar-Paramo et. al. originally suggested that the ETEC pathotype had arisen on many occasions, as individual *E. coli* strains acquired virulence plasmids, and that a certain genomic background was required for effective use of the ETEC virulence genes (29). A more recent study, however, analysed whole genome sequences of 362 ETEC isolates, chosen specifically for their diversity, and identified 21 relatively stable ETEC lineages, including five particularly successful clades that have emerged in the past 200 years (28). Despite the diversity within the ETEC pathotype, population structure clearly exists, and there is a robust relationship between chromosomal background and precise CFs and toxin profiles.

Two strains of ETEC are commonly described as archetypal examples of the pathotype. These are H10407, the first ETEC strain to be fully sequenced, and E24377/A, which has also been fully sequenced, and which produces a different set of CFs and accessory virulence factors to H10407 (32, 33).

1.3 ETEC pathogenesis

In its simplest form, the pathogenesis of ETEC involves adherence to the intestinal epithelium mediated by CFs, followed by the delivery of ST and/or LT to host cells. Accessory virulence factors have also been identified that assist pathogenesis by improving adhesion, increasing the efficiency of toxin delivery, and compounding the toxic effect of ST and LT.

1.3.1 Colonisation factors

Over 25 different colonisation factors (CFs) that mediate adherence to human enterocytes have been discovered. The prevalence of each varies in different geographical regions (31). Current nomenclature designates CFs as CS1 through CS28 (for coli surface antigen), with the exception of the prototypical ETEC CF, which is known as CFA/I and is distinct from CS1. The structure of several CFs has been elucidated. They are generally fimbrial in nature and are comprised of a major subunit that polymerises in a helix to form the length of a rod-like structure, capped with a receptor binding domain, or with receptor binding ability throughout the structure (34).

The most common CFs seen in ETEC that infects humans are CFA/I, CS1-6, and CS21 (35). CFA/I is the archetypal CF of class 5 fimbriae, a class of ETEC CFs made up of some of the most common CFs, including CFA/I, CS1, CS2, CS4, CS14, CS17 and CS19 (36-38). These fimbriae are synthesised via the alternate chaperone pathway, and their operons consist of four genes: encoding a chaperone, a major fimbrial subunit, an usher pore protein, and a minor fimbrial subunit (38). In most cases, the genes for these proteins are carried on a plasmid which also encodes a gene regulator for the CF operon. The fimbriae themselves are made up of repeating major subunits arranged in a tight, rigid helix approximately 7 nm thick, and are capped with a minor subunit, the adhesin of the fimbriae (39). This rigid helix is able to unwind in response to shear forces, increasing the ability of class 5 CFs to remain attached to host cells (40). The minor subunit is required for initiation of fimbrial biosynthesis. All proteins involved in the synthesis of these fimbriae possess an amino-terminal signal sequence for Sec-dependent translocation into the periplasm and, once there, the chaperone protein ensures that the structural subunits do not aggregate or degrade within the periplasm, and are instead guided to the usher for correct assembly (41). In general, it appears that although the major subunit tends to be the primary antigenic determinant, it is the minor subunit that actually mediates adhesion to host cells. Importantly, these class 5 fimbriae, though closely related genetically, are antigenically distinct, and the regions of the proteins that interact with the host are under positive selection for mutation (42).

Unlike CFA/I, which is often expressed alone, CS1-6 are usually seen expressed in specific combinations. CS1 and CS2 are usually produced along with the fibrillae CS3, in either a CS1+3, or CS2+3 pattern (43). CS4, 5, and 6 exist in a similar pattern, with CS4 and CS5 usually accompanied by CS6. Unlike CS3, CS6 is also often found expressed alone, with neither CS4 nor CS5. Structurally, CS3 is a thin and flexible fimbriae (43). CS6 is a 'unique' class of CF, with no significant homology to any other known CF. CS6 has two major fimbrial subunits, present in approximately equal proportions in the final structure, which is a flexible fibre lacking the rod-like stability of most other ETEC fimbriae (44, 45).

CS21, along with CS8, CS9 and CS16, are type IV pili, and are synthesised in a similar way to the type II secretion system (46). These CFs have a more complex structure and biosynthetic pathway, with multiple accessory proteins involved in assembly, secretion, and anchoring of the pilus to the outer membrane. CS21 pili are over 20 μm long, much longer than most ETEC fimbriae, and CS21 is also known as 'longus' for 'long pilus'. CS21 is commonly produced by a variety of ETEC serogroups, often in combination with other CFs, especially CFA/I and CS6 (47).

Immunity to each of these CFs develops naturally, and by adulthood most people who live in regions where ETEC is endemic are protected from ETEC utilising the CFs that they have regularly encountered throughout their lives. However, this immunity is rarely cross-protective, and this has been a major stumbling block in the development of a comprehensive vaccine against ETEC. Additionally, more and more CFs are being discovered, and in some epidemiological studies up to half of the diarrhoeagenic ETEC strains analysed express an unknown CF (27, 35). This means that any vaccine developed against the currently described CFs will still fail to protect against many strains of ETEC.

1.3.2 Enterotoxins

Clinical isolates of ETEC may produce ST-only, LT-only, or both ST and LT. The relative frequencies of different toxin profiles varies worldwide and in different studies, but both toxins are important virulence factors in human disease (31, 48).

Two forms of ST exist. These are STh, a 19 amino-acid peptide first isolated in humans, and STp, an 18 amino-acid peptide first isolated in pigs, but also pathogenic for humans (49, 50). 14 amino-acids are identical between STp and STh, and the toxins are antigenically and functionally similar (51). Both mimic the endogenous hormone guanylin to activate guanylate cyclase in intestinal epithelial cells (52). This leads to an increase in cyclic GMP in the enterocytes, which activates ion channels, most importantly the cystic fibrosis transmembrane conductance regulator, causing a net loss of chloride ions from the cells (53). Water osmotically follows this ion gradient into the lumen of the gut, leading to watery diarrhoea. One important difference between ST and guanylin is the presence of an extra cysteine pair in ST; guanylin has only two disulphide bridges while ST has three (54). This third bond appears to lock the ST peptide into a conformation that activates guanylate cyclase irreversibly. By contrast, guanylin allows guanylate cyclase to switch between active and inactive conformations (55).

LT is very similar to cholera toxin in structure, function and antigenicity (56). It is an AB_5 toxin, made up of one active subunit (LT-A) residing atop a ring of 5 identical binding subunits (LT-B) (figure 1.1) (57, 58). The B subunits bind GM1 ganglioside and allow the A subunit entry into the cell, where it activates adenylate cyclase to induce cyclic AMP (cAMP) production (60). As with ST, this causes a net secretion of ions from intestinal epithelial cells. This is followed by water loss, in the form of diarrhoea.

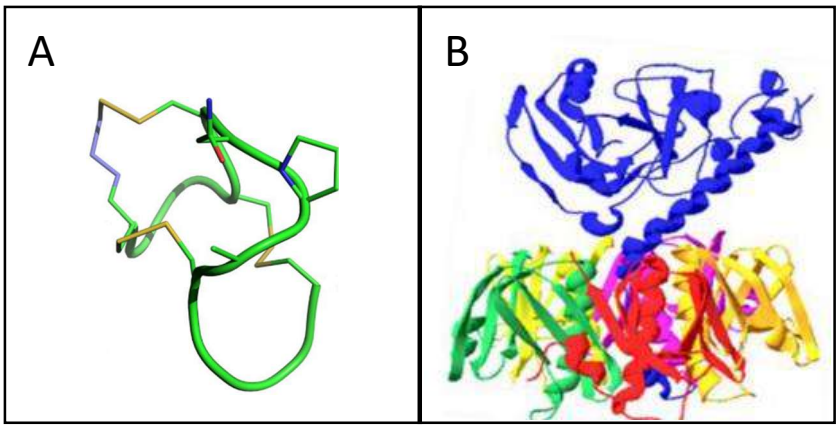


Figure 1.1. The structures of the heat-labile and heat-stable toxins

A) Structure of the heat-stable toxin. Figure from Taxt *et. al.* 2010 (58). B) A ribbon diagram of heat-labile toxin, with the A subunit (depicted in blue), and five identical B subunits. Figure from Mudrak and Kuehn, 2010 (59).

1.3.3 Accessory virulence factors

Despite the clear evidence of the requirement for CFs and enterotoxin for ETEC virulence, other virulence factors have also been proposed. These are proteins that contribute to the virulence of some strains of ETEC, but are not encoded by every member of the pathotype.

1.3.3.1 Adhesins

Tia is a 25 kDa non-fimbrial adhesin discovered in ETEC strain H10407 and shown in vitro to allow *E. coli* to adhere to and invade some strains of cultured human intestinal epithelial cells (61-63). The importance of Tia in ETEC pathogenesis has not been fully investigated, and traditionally ETEC has never been seen to invade cells in vivo. However, Tia may contribute to virulence as a non-fimbrial adhesin.

The *tibCADB* gene cluster was also discovered in H10407, and is also able to promote invasion of intestinal epithelial cells by ETEC (61, 64). TibA is a 104 kDa auto-transporter which is glycosylated by TibC, and which, aside from invasion, also mediates adhesion, and can self-associate to mediate cell aggregation and biofilm formation (64-67). TibB contains a LuxR-like DNA-binding domain, and, along with TibD, is expected to be involved in the transcriptional regulation of the *tibCA* genes (68). Transcription of the *tibDB* genes themselves is stimulated by the cAMP receptor protein (CRP) (68).

The *etpBAC* operon encodes a secreted adhesin, EtpA, which is glycosylated by EtpC and leaves the cell via the membrane pore EtpB (69). After secretion, EtpA binds to intestinal epithelial cells, and anchors ETEC to the host cell surface through interaction with the flagella (70). EtpA binds to a conserved region on FliC, which has been suggested as a vaccine target (71).

EaeH is a chromosomally encoded protein, homologous to intimin of EPEC, which is upregulated upon contact with intestinal epithelial cells and aids in the adhesion of ETEC to enterocytes (72, 73).

Finally, TleA is recently described as a serine protease autotransporter of *Enterobacteriaceae* (SPATE)-family auto-transporter, with homology to the temperature-sensitive haemagglutinin (Tsh) found in avian *E. coli* (74). TleA is able to confer upon previously non-adherent *E. coli* the ability to adhere to Caco-2 cells, (although a *tleA* deletion mutant of ETEC strain 1766a did not lose this ability), and may also have some effect on the host immune response (74).

1.3.3.2 The EAST-1 toxin

Enteroaggregative *E. coli* heat-stable toxin 1 (EAST-1), encoded by the *astA* gene, is a toxin of EAEC that has also been found in ETEC and other enteric pathogens (75, 76). It displays a high degree of homology to ST, and is thought to work through the same guanylate-cyclase C activation pathway to increase intracellular cGMP and induce the secretion of ions and water into the intestinal lumen (75).

1.3.3.3 Other accessory virulence factors

The small intestine secretes mucins to form a physical barrier that limits the access of bacteria to the intestinal epithelium (77). ETEC encodes proteins that allow it to penetrate this protective mucous layer. The most common mucin secreted in the human intestine is MUC2. ETEC accessory virulence factors EatA and YghJ are mucinases, both of which can break down MUC2. YghJ is also able to break down MUC3, a mucin which remains bound to intestinal cells (78-81). This allows for more efficient delivery of LT by exposing receptors on the intestinal epithelium. YghJ is found in commensal as well as pathogenic *E. coli*, but is more highly expressed in pathogens (80).

CexE is a homologue of dispersin, a virulence factor of EAEC which is involved in assisting adhesion by the AAF/I fimbriae (32). Dispersin is secreted by EAEC through the Aat transporter, and remains on the cell surface, reducing the aggregation of bacteria and enhancing the efficiency of binding to host cells (82, 83). A homologue of the Aat transporter is also encoded by *E. coli* H10407, immediately downstream of *cexE* (32).

Finally, bacterial dynamins, LeoA and LeoBC, have been identified in *E. coli* H10407 (84, 85). These proteins are involved in membrane remodelling, and in H10407 they improve the secretion of LT, most likely through the creation of outer membrane vesicles (84, 85).

1.4 Regulation of ETEC virulence genes

Intestinal pathogens such as ETEC can sense changes to their environment, and regulate the expression of their genes accordingly. Virulence mechanisms are energetically expensive, and are only useful to ETEC when the bacterium is in the small intestine, the site where the enterotoxins have the greatest effect (86). As well as the wildly changing environment encountered as they move through the gastrointestinal tract, ETEC also have to contend with the host immune system, share resources with other bacteria, and survive in the external environment between human hosts. Global gene regulators are those that enact broad changes of gene expression levels across the genome, to help *E. coli* respond to these challenges. Many of these also affect the expression of virulence genes. Global regulators which have been investigated in the context of ETEC virulence are described below.

1.4.1 H-NS: A global repressor

H-NS is a global gene silencer which binds AT-rich regions of DNA and polymerises to form long filaments along the DNA helix, preventing the formation of RNA-polymerase and inhibiting gene transcription (87). Genes that have been introduced to the *E. coli* genome by horizontal transmission often have a lower GC-percentage than the rest of the genome, and these tend to include virulence genes (88). Newly acquired genetic elements may be detrimental to the cell, and so H-NS and other global repressors have evolved to control the expression of these 'foreign' DNA sequences (89). H-NS silencing can be displaced by the preferential binding to DNA by gene activators (90). With H-NS no longer occluding the RNA-polymerase binding site, transcription of the target gene can occur. The binding of H-NS to target DNA can also be relaxed by environmental conditions, such as high temperatures and a high salt content (91).

H-NS repressed genes include the toxins encoded by ETEC (table 1.1). H-NS represses the transcription of *eltAB* (encoding LT) and both STs encoded by ETEC strain H10407, *hstA* (encoding STh) and *hstB* (encoding STp) (92, 93). Upon entering a human host, ETEC encounters high temperatures (37°C) and high salt content. Both of these signals reduce the ability of H-NS to bind DNA, resulting in increased expression of toxins once ETEC enters a human host (94). In addition, the transcription of *hns* is repressed upon contact with the intestinal epithelium, thus increasing the expression of toxins at their site of activity (95).

1.4.2 Glucose starvation

A number of studies have investigated the effects of cAMP and its receptor protein, CRP, on ETEC virulence regulation. Glucose starvation leads to an increase of cAMP, which in turn activates CRP, a global transcriptional messenger (96). CRP-cAMP controls the regulation of a large number of metabolic processes within *E. coli*, by binding to promoter DNA to activate or repress transcription (97).

In H10407, CRP represses the transcription of *eltAB*. One study has shown that CRP may be able to directly bind the *eltAB* promoter and inhibit the formation of RNA-polymerase, thus inhibiting transcription (93), while another study suggests that CRP does not repress *eltAB* directly (98). In ETEC strain E24377/A, however, CRP has been shown to either activate P*eltA* transcription, or have no effect on it at all (table 1.1) (93, 95). Another study suggests that, while CRP represses the transcription and translation of *eltAB*, it actually increases the secretion of functional LT, presumably through regulation of a secretion system (99). In both ETEC strains, CRP activates transcription of *hstA*, but, in H10407, CRP represses transcription of *hstB* (table 1.1) (93). While CRP does affect the expression of ETEC toxin genes, this effect is inconsistent between strains. This reflects the fact that ETEC are of a diverse genetic background, and the regulatory systems that interact with virulence may not be identical in every ETEC isolate.

Table 1.1. The effect of various environmental conditions and regulators on the transcription of virulence genes of ETEC

Arrows in black indicate data from ETEC strain H10407, arrows in blue indicate data from ETEC strain E24377/A. Up arrows indicate upregulation of the virulence gene, down arrows indicate downregulation of the virulence gene, and a dash indicates no change.

Environmental condition	Regulator	*cfaABCE*	*cooBACD*	*cstABCD*	*cexE*	*etpBAC*	*hstA*	*hstB*	*hltAB*	*tib*	*leoA*	*astA*
Low temperature Low salt concentrations	H-NS						↓(93) ↓(93)	↓(93)	↓(92) ↓(93)			
Glucose starvation	CRP						↑1,(95) ↑(95),(93)	↓(93)	↓(98, 93) ↑(95), – (93)	↑(68)		
Iron starvation	IscR	↑(103)			↑(103)	↑(103)	↓(103)	↓(103)	– (103)		↓(103)	
Bile salts	Unknown	– (108)	↓(108)	↓(108)		– (108) ↑(108)	– (108) – (108)		– (108) ↑(108)			↑(108) – (108)
Preconditioned media	Unknown		↑(108)[a]	↑(108)[a]			↑(108)		↑(108)			

[a] Preconditioned media of EPEC or *S. boydii* only, not of commensal *E. coli*

One accessory virulence factor, the *tib* operon, is positively regulated by CRP binding to the promoter sequence (68).

1.4.3 Iron starvation and IscR

IscR is a regulator that responds to iron starvation, and activates virulence processes in a range of pathogens, including *E. coli* H10407 (100-103). Iron starvation is a universal challenge for bacteria that live inside humans, and is commonly utilised as an environmental signal for pathogens to activate virulence processes (104). Through the activity of IscR, iron starvation leads to increased transcription of *cfaABCE* and *etpBAC*, and decreased transcription of *hstA, hstB*, and *leoA* in *E. coli* H10407 (table 1.1) (103). Interestingly, while the transcription of genes encoding LT is not directly affected, LT secretion does decrease. This is likely to be an effect of the repression of *leoA*, which is known to be involved in the secretion of LT. Based on this transcription profile, IscR may be active during the early stages of ETEC pathogenesis, to increase colonisation before the toxins are required.

1.4.4 Bile salts

Bile acids are secreted by the small intestine, and are a known environmental signal used by intestinal pathogens (105-107). Detection of bile salts results in an upregulation of *hltAB* and *astA* in H10407, but has no effect on the transcription of *cfaABCE, etpBAC*, or *hstA* (table 1.1) (108). In E24377/A, however, the addition of bile salts to growth media results in increased transcription of *etpBAC*, but has no effect on the transcription of *hstA, hlt*, or *astA* (table 1.1). Bile salts also downregulate the transcription of genes encoding CS1 (*cooBACD*) and CS3 (*cstABCD*), the CFs carried by E24377/A.

1.4.5 Preconditioned media

The intestine is home to a wide variety of microorganisms. In an attempt to simulate some of the inter-bacterial interactions that take place in that rich environment, the effects of growing ETEC in media that had been preconditioned by the growth of other bacteria has also been investigated. Media that have been preconditioned with a commensal strain of *E.coli*, an EPEC strain,

or *Shigella boydii* all increase toxin expression by ETEC strain E24377/A (table 1.1) (108). Interestingly, preconditioned media from pathogens, but not commensal *E. coli*, increased CS1 and CS3 expression in E24377/A. Quorum sensing has long been associated with bacterial virulence, and the interruption of quorum-sensing pathways has been shown to inhibit virulence regulation in *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Vibrio cholerae*, among others (109-111). It is not surprising, therefore, that ETEC may also respond to the presence of other intestinal bacteria.

1.5 Rns, a virulence activator of ETEC

The global regulators described above influence the expression of virulence genes as part of a broader influence on the transcriptome of *E. coli*. ETEC strains also possess a regulator dedicated to the control of virulence mechanisms; the 'master' virulence regulator, Rns.

Rns, encoded by the 798 bp gene *rns*, was first discovered as a regulator of the CFs CS1 and CS2 (112). Since then, homologous regulators have been found to control the expression of many CFs across the ETEC pathotype (113-118). Rns-like regulators within the ETEC pathotype include CfaD, the activator of CFA/I, and CsvR, the activator of CS4 and CS5. These three proteins share approximately 80% pairwise identity. CfaD and Rns are particularly similar, with only 14 out of 266 amino-acid differences between them, giving a value of 95% pairwise identity. CfaD and Rns are functionally interchangeable, and, for clarity, I will refer to both as 'Rns' within this thesis (119).

1.5.1 Structure

Rns is an AraC-like regulator (112). These regulators generally coordinate one of three types of environmental responses: changes to carbon utilisation, a stress response, or virulence mechanisms (120). AraC-like regulators are generally around 250 to 300 amino-acids in length, and are defined by a conserved DNA binding domain of 100 amino-acid residues, usually in the C-terminal domain of the protein (121). This DNA binding domain contains two helix-turn-helix (HTH) motifs, at least one of which contacts the target gene DNA. This domain also mediates the activation of gene transcription (122). AraC-like regulators also usually contain an N-terminal domain, joined to

the C-terminal domain by a flexible linker sequence (121). In Rns, the linker sequence is made up of residues 100 to 104, and this flexibility is essential, most likely to allow the N- and C-terminal domains to move relative to each-other (123, 124). Most AraC-like regulators function as dimers, and Rns has also been reported to dimerise (125). The N-terminal domain is often involved in dimerisation and/or effector binding, and is absent from some family members. In the case of Rns, a putative dimerisation domain, identified on the basis of on homology to the related regulator, RegA from *Citrobacter rodentium*, is on the C-terminal side of the flexible linker sequence (figure 1.2) (126). The function of the N-terminal domain is yet to be characterised.

AraC-like virulence regulators activate the expression of their target genes at the transcriptional level, most commonly by counteracting the negative regulation by H-NS (127). Figure 1.3 is a ribbon-diagram of the Rns protein, with the DNA binding helices highlighted in red, the putative dimerisation domain in blue, and the flexible linker sequence in orange.

1.5.2 Mechanism of action of Rns

The primary mechanism of action of Rns and its homologues is through anti-repression of H-NS. Rns binds preferentially to specific sites within the promoter regions of its target genes, displacing H-NS, and allowing RNA polymerase to bind and begin transcription (128). The orientation of the Rns binding sites are highly variable, and in some cases Rns binds downstream of the transcription start site (TSS) (115).

Rns binds to DNA by inserting its two HTH motifs into two adjacent segments of the major groove of the DNA strand, along one face of the helix (figure 1.4) (129). Direct contact by Rns to the DNA has been characterised at three thymine C5-methyl groups, where one thymine is on the reverse strand of DNA, and so is annotated as an adenine in the binding consensus sequence $\underline{A}N_6\underline{T}A\underline{T}$ (129). At some target promoters, Rns is able to bind to the sequence $\underline{G}N_6\underline{T}A\underline{T}$. The binding site of Rns is highly variable, which has made the identification of novel Rns targets difficult. Figure 1.5 shows

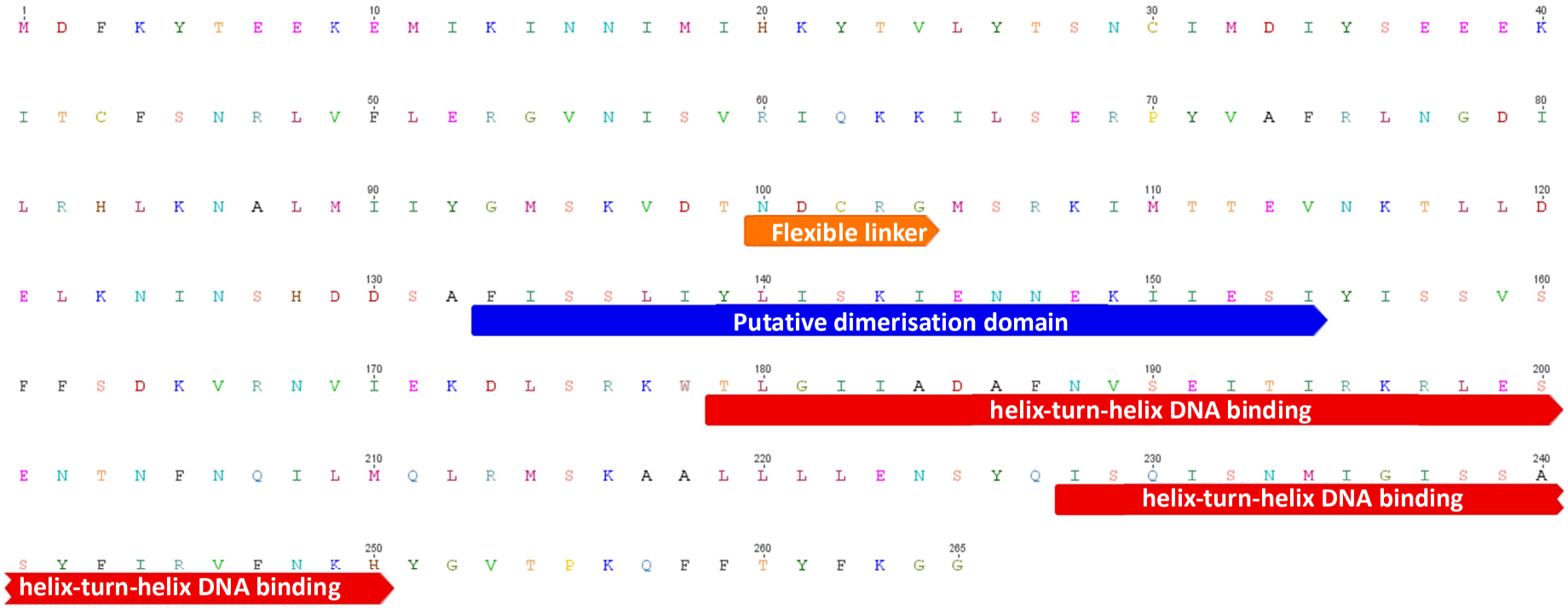


Figure 1.2. Amino acid sequence of Rns

Protein sequence of Rns with functional regions labelled.

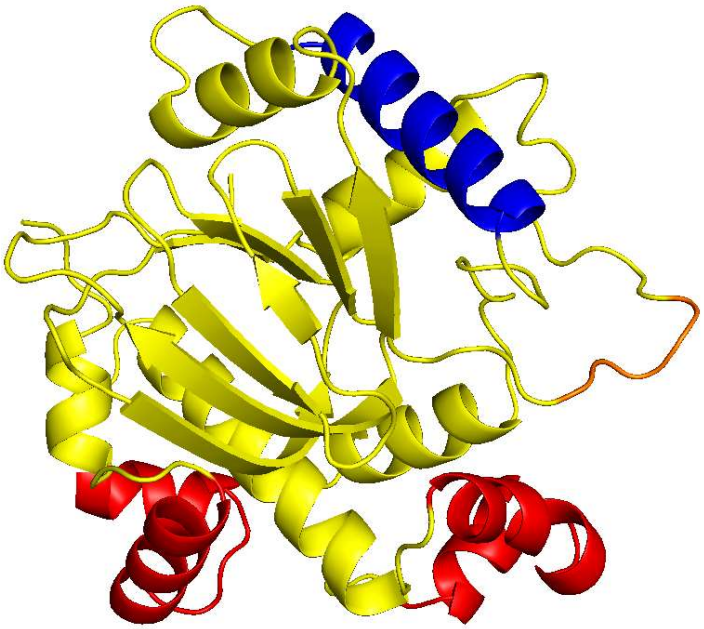


Figure 1.3. The structure of Rns

A ribbon diagram of Rns in yellow with the helix-turn-helix DNA binding motifs in red, the putative dimerization domain in blue, and the flexible linker in orange. This diagram was kindly provided by Dr. Jessica Holien at the St. Vincent's Institute of Medical Research.

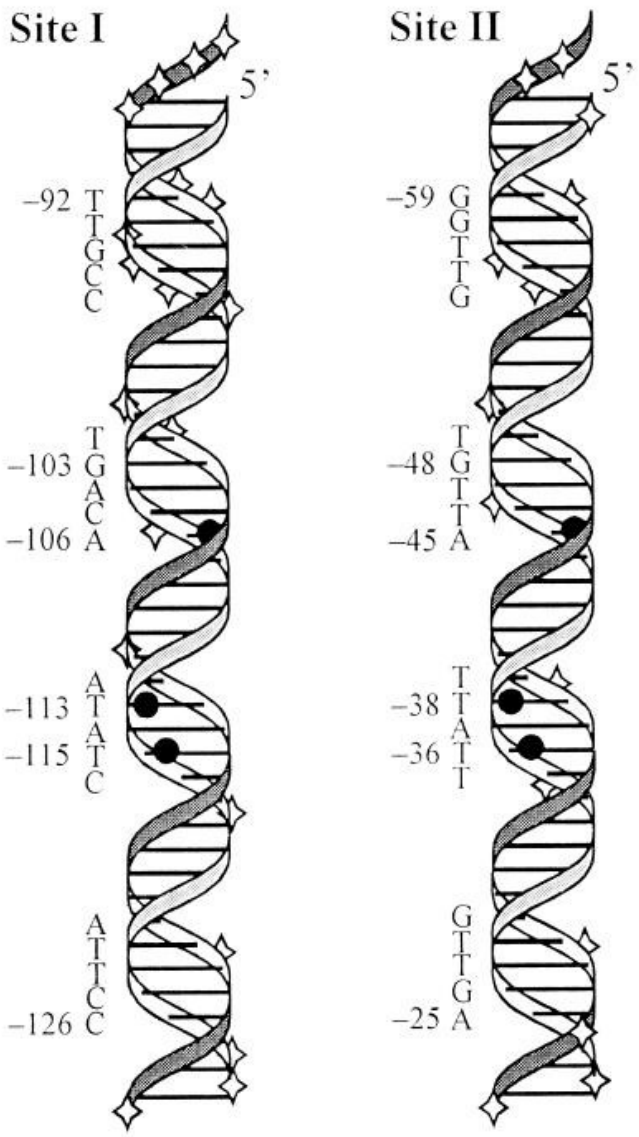


Figure 1.4 DNA binding by Rns

Diagram depicting the binding sites of Rns to P*coo*, promoter region of the operon encoding CS1. Thymine residues which make contact with the Rns protein are indicated with black circles. Bases which are protected from Dnase I activity by Rns binding are specified using standard abbreviations, and bases which are degraded are marked with diamonds. Numbering is relative to the transcription start site of *cooB*. Figure from Munson et al, 1999 (129).

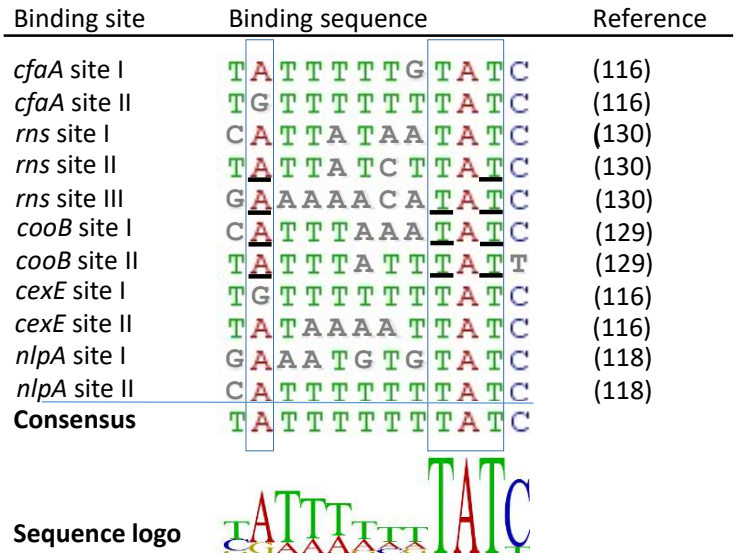


Figure 1.5. Sequence alignment of experimentally determined Rns binding sites

Alignment, consensus sequence and sequence logo of Rns binding sites which have been determined by DNase I footprinting. The AN_6TAT sequence which makes contact with the Rns protein is boxed. Sites of Rns contact which have been confirmed experimentally using uracil interference assays are underlined. Nucleotides which match the consensus sequence are shown in colour; while mismatches are shown in grey. The consensus sequence and sequence logo were generated using Geneious® V7.1.7 (Biomatters).

the experimentally determined Rns binding sites, with a consensus sequence and sequence logo as determined by Geneious 7.1.7 (Biomatters Limited).

The environmental factors that Rns detects as a signal to initiate virulence mechanisms are not known, although elevated temperatures relieve H-NS repression of Rns target genes (131). The master virulence regulator of *C. rodentium*, RegA, is a homologue of Rns, and it is activated by bicarbonate ions (132). Rns may interact with bicarbonate in the same way, or may respond to the presence of another factor in the small intestine.

1.5.3 Rns target genes

Rns was originally identified as an activator of CS1 and CS2, but since its discovery in 1989, more Rns target genes with virulence functions have been discovered. As shown in table 1.2, Rns-like proteins appear to control the expression of nearly 50% of all known ETEC CFs.

Rns binding sites have been experimentally confirmed in the promoter regions of operons encoding CFA/I, CS1, CS17, and CS19. In each case, Rns binds to a region overlapping the -35 sequence, as well as a second site approximately 100 bp upstream of the TSS (36, 116, 128, 129). In the case of CS4 and CS5, gene activation is by a related Rns-like regulator, CsvR. As mentioned, the pairwise identity between Rns and CsvR is approximately 80%, and they are able to substitute for each-other with some loss of activity (133).

Searching the genome sequence of ETEC strain H10407 for possible Rns-binding sequences revealed the Rns activation target *cexE* (116). CexE is a homologue of an EAEC protein, dispersin, which assists in fimbrial adhesion (83). In both ETEC and EAEC, this protein is encoded on the same plasmid as the regulator, along with a secretion system, *aatPABC*, used for export. Rns binds two

TABLE 1.2 Known CFs of ETEC and their method of genetic regulation

CF	Activated by an Rns-like regulator	Reference
CFA/I	Yes	(119)
CS1	Yes	(112)
CS2	Yes	(112)
CS3	Yes	(114)
CS4	Yes	(133)
CS5	Unknown	(133)
CS6	No	(114)
CS7	Unknown	
CS8	Unknown	
CS10	Unknown	
CS11	Unknown	
CS12	Probable[1]	Genbank Accession no. AY009096.1
CS13	Unknown	
CS14	Yes	(34)
CS15	Unknown	
CS17	Yes	(36)
CS18	No	(134)
CS19	Yes	(36)
CS20	Yes	(135)
CS21	No	(136)
CS22	Unknown	
CS23	No	(137)
CS26	Unknown	
CS27	Unknown	
CS28	Unknown	

[1] An Rns-like gene is present downstream of the operon encoding CS12

sites upstream of *cexE*, one between -33 and -44 (relative to the TSS), overlapping the -35 sequence, and another between -479 and -490 (116).

Rns and related regulators also auto-activate, increasing transcription from their own promoter (138). This mechanism results in a rapid increase of Rns concentrations in the cell as soon as the Rns response is triggered by an environmental cofactor. One Rns binding site is present in *rns* upstream (from -220 to -233, relative to the TSS) and two downstream (+36 to +49) and (+75 to +89) of the TSS.

An Rns-activated gene with no clear link to virulence is *yiiS*, which is a part of the *E. coli* stress response (117, 139). YiiS upregulation by Rns is less strong than what is seen with other Rns-regulated promotors. Potential Rns binding sites have been identified directly upstream (from -166 to -177, -194 to -205, and -264 to -276, relative to the ATG translation start site) of *yiiS*, suggesting that the gene is under direct regulation.

Only one gene has been reported as being repressed by Rns. This is *nlpA*, the product of which is involved in the formation of outer membrane vesicles (118, 140). One Rns binding site overlaps the start codon of *nlpA*, restricting transcription by RNA polymerase, and another is between 152 and 195 bp upstream of ATG. This second binding site is not required for repression (118).

1.5.4 AraC-like negative regulators (ANRs)

Auto-regulation allows Rns and similar regulators to quickly increase the production of virulence factors in response to an environmental stimulus. However, this establishes a positive-feedback loop which must eventually be curtailed. A recent study investigating AggR, an Rns homologue of EAEC, reported the protein Aar, whose synthesis is upregulated by AggR. Aar in turn inhibits the transcription of *aggR* itself (141). In this way, AggR production begins to decrease steadily after an initial period of rapid increase. A gene with significant homology to *aar* is present immediately upstream of *rns* in ETEC strain H10407, previously annotated as ETEC_p948_1070, and

referred to throughout this thesis as *rtr*, for Rns transcriptional repressor. Rtr is able to functionally substitute for Aar to repress *aggR* (141). This family of proteins is named ANRs, for AraC-like negative regulators, and has been identified in many pathogens which carry Rns homologues.

1.5.5 Master regulators of virulence

Homologues of Rns are found in many intestinal pathogens, including AggR of enteroaggregative *E. coli* (EAEC), RegA of *C. rodentium*, ToxT of *V. cholerae*, and VirF of S*higella flexneri* (142-145). Like Rns, the major mechanism of action of these regulators is through de-repression of H-NS (90, 120). Importantly, Rns homologues are 'master' regulators of virulence, and coordinate the activation of multiple virulence determinants. For example, AggR activates fimbriae and secretion systems in EAEC, ToxT of *V. cholerae* activates transcription of both cholera toxin and the toxin co-regulated pilus (TCP), and RegA of *C. rodentium* regulates the transcription of multiple fimbriae, secretion systems, and other putative virulence factors (143, 146, 147).

No in-depth analysis of the Rns regulon has been performed to identify its effect on the entire ETEC genome, or to identify other virulence genes which may be under its control. It is likely that other Rns target genes exist, and that a thorough analysis of the Rns regulon may identify accessory virulence factors, whether novel or previously characterised, that form part of the Rns response.

1.6 ETEC prevention and treatment

1.6.1 Treatment of ETEC

1.6.1.1 Treatment for childhood diarrhoea

In combating ETEC disease, treatment and prevention options are limited. The primary recommended treatment for childhood diarrhoea is rehydration with reduced osmolarity oral rehydration solution (ORS). This therapy focuses on limiting the effect of dehydration, rather than targeting the cause of the infection (148). ETEC causes diarrhoea by inducing the secretion of sodium

and chloride ions into the intestinal lumen and reducing their absorption, resulting in a loss of water down the resultant osmotic gradient. ORS is designed to replace both the water and the sodium chloride. The third core ingredient of ORS is glucose, which stimulates the sodium-glucose co-transporter to increase the absorption of sodium ions (148-150). Finally, citrate is added to buffer the acidosis that often accompanies dehydration (151). The low osmolarity of ORS is intended to counteract the osmotic pressure building in the lumen of the intestine in response to the action of the enterotoxins (152).

Zinc is another additive to ORT that is recommended by the World Health Organisation (WHO). 10 to 20 mg of zinc per day, for 10 to 14 days, has been consistently shown to improve disease outcomes in diarrhoea (153-156). Zinc is important in many physiological processes, including the maintenance of the intestinal mucosa and fluid transport, the regulation of immunity, and the reduction of oxidative stress (157). Zinc has also been shown to inhibit ion secretion in response to cAMP induction by LT and cholera toxin (158, 159).

1.6.1.2 The use of antimicrobial agents to treat ETEC-associated diarrhoea

While antimicrobials may in some cases be effective in reducing the duration and severity of infectious diarrhoea, there have been calls to restrict the antimicrobial treatment of ETEC infections, particularly for TD, over fears that this practice contributes significantly to the development and spread of antimicrobial resistant strains, while having only minor effects on symptoms (15, 160, 161). Ciprofloxacin and rifaximin can be effective, and are popular treatment options (162-166). ETEC resistance to ampicillin, tetracycline, and erythromycin has been high in recent studies worldwide, but resistance to ciprofloxacin and rifaximin is currently low according to most reports (167-175). However, some studies have detected high levels of ciprofloxacin resistance. In 2011, 16 out of 17 ETEC strains isolated from an outbreak in India were resistant to ciprofloxacin, and in an outbreak in Iran in 2013, the rate was 37.1% of 73 *E. coli* isolates (170, 171). In a study comparing isolates collected in the early 2000s to those collected in the early 1990s, there was a significant

increase in the proportion of ETEC strains that were resistant to nalidixic acid, ciprofloxacin, trimethoprim/sulphamethoxazole, and amoxicillin/clavulanate (176). Antimicrobial-resistant bacteria are becoming such a threat that we are rapidly moving toward a post-antimicrobial age, and as such, alternate methods of pathogen control are urgently required (177).

1.6.2 Vaccines targeting ETEC

One strategy to reduce the impact of ETEC disease is the development of a vaccine. Natural immunity to ETEC develops against the CFs and O-antigen on the bacterial surface. However, these are less than ideal vaccine targets, as ETEC belong to a wide variety of O-antigen types, and may possess only one of an ever-increasing number of known CFs.

Vaccine development strategies generally aim to include as many different CFs as possible. The non-toxic subunit LT-B is often used as an adjuvant, and to induce protection against LT (178). Protection against ST does not develop naturally, and has proven difficult to induce (59).

The most extensively studied potential vaccine against ETEC is the recombinant cholera-toxin B subunit/inactivated whole-cell (WC) vaccine (rCTB-WC). This is made up of inactivated *E. coli* expressing the major CFs CFA/I and CS1-5, and recombinant cholera-toxin B (CT-B) (non-toxic) subunit (179). This vaccine is based on the cholera vaccine, Dukoral, as CT-B may offer limited cross-protection against LT-B of ETEC (180). Human trials found that while the rCTB-WC vaccine was safe and immunogenic, it offered little protection against diarrhoea for travellers to regions where ETEC is endemic (181, 182). Work is ongoing to improve this vaccine through overexpression of the colonisation factors on vaccine strains, and the inclusion of mutated LT-B in place of CT-B (183, 184).

Live attenuated strains of ETEC, other *E. coli*, or other bacteria displaying CFs have also been investigated as possible vaccines. ACE527 is a combination of three attenuated strains of ETEC, between them expressing CFA/I, CS1 - 6, and LT-B (185, 186). This vaccine is also immunogenic, but, as with rCTB-WC, researchers found no significant protection against ETEC-induced diarrhoea in general (187).

A third vaccine candidate is ST. Ruan *et. al.* developed, in stages, a multi-epitope fusion, carrying antigens from CFA/I and CS1-6, and a second fusion of non-toxic mutants of both ST and LT (188, 189). These were then combined to form a complex multi-epitope fusion antigen that elicited antibodies able to both inhibit adherence of ETEC to Caco-2 cells, and to neutralise the effect of ST and LT on T-84 cells (190). Despite promising results in vitro, the authors of these studies acknowledged the limitations of this kind of vaccine. Certainly there are virulent strains of ETEC that produce CFs other than the seven included in the fusion antigen. Additionally, it has been suggested that anti-toxin antibodies have little effect *in vivo* due to the direct nature of delivery of toxins from ETEC to the host cells, which gives little opportunity for antibodies to neutralise them (191).

Clearly, there are many challenges facing the development of an effective vaccine targeting ETEC, and success so far is extremely limited. Alternate strategies are required to help control ETEC disease.

1.7 Inhibition of virulence gene activation

1.7.1 Virulence inhibitors as a novel form of treatment

While the CFs of ETEC are antigenically distinct on a protein level, many share a common mechanism of transcriptional regulation: Rns. This makes Rns an attractive single drug target that could have a broad-range effect on virulence. Chemical inhibition of the transcriptional activation of virulence genes has been proposed as a novel method to reduce the virulence of certain pathogens, and has several advantages over the traditional antimicrobial treatment of bacterial infections. As a novel intervention, resistance in the current population will be low or non-existent. Moreover, a drug that specifically inhibits virulence gene activation would be expected to exert a lower selection pressure on the target organism when compared to traditional antimicrobials, which target essential cell processes. Bacteria affected by a virulence inhibitor may suffer a fitness cost when inside a host, but are not killed outright, and those living in the external environment should remain unaffected. As such there is low pressure for resistance to become widespread outside the host. Additionally,

virulence inhibitors will be pathogen-specific, and so any resistance that may develop would also be pathogen-specific, without inter-species transfer leading to widespread resistance.

Another benefit of virulence inhibitors over traditional antimicrobials is the effect that they can have on the intestinal microbiota. The reduction in the number and variety of microorganisms in the gut following antimicrobial exposure can lead to intestinal disease, most notably a predisposition to *Clostridium difficile* infection. In addition, the repeated use of antimicrobials within an individual can increase the incidence of resistance within their own microbiota, paving the way for highly resistant opportunistic infections at a later date (192).

AraC-like regulators make ideal targets for inhibitors of virulence gene activation, for two reasons. First, when involved in virulence, they tend to be global virulence gene regulators, which activate multiple virulence genes while repressing house-keeping genes. By inhibiting the action of the regulator, it is possible to shut down multiple virulence determinants at once, without affecting essential cell functions. Second, AraC-like regulators are strong activators which, in many cases, are essential for the expression of their target genes, indicating that functional inhibition of these activators should be effective in preventing infection by bacterial pathogens (127).

However, there are also concerns which have been raised regarding the use of virulence inhibitors. One possible unintended consequence, for example, is that, in cases where virulence gives the bacteria a strong evolutionary advantage, virulence processes may be strengthened to overcome the effects of an inhibitor, creating hyper-virulent strains of the pathogen (193). This risk must be considered during the development process of any virulence inhibitor.

An additional complication of the use of virulence inhibitors is that, as a narrow-spectrum therapy, they rely on an accurate diagnosis of the causative agent. This is not always practical, especially in the economically disadvantaged regions where ETEC is endemic. Moreover, many children in these regions carry multiple infectious agents (5). This means that even if accurate

diagnostics are available, determining which pathogen or pathogens are contributing to disease is no simple task.

The fact that virulence inhibitors do not actually kill the target pathogen also has its drawbacks. If a virulence inhibitor reduces, for instance, toxin production, but does nothing to stop the colonisation of the bacteria, the infection may not be cleared. In this case, pathogenesis could resume as soon as treatment with an inhibitor is ceased. Fortunately, this is not a problem in the case of ETEC, as Rns does indeed control expression of the CFs essential for colonisation of the small intestine. If ETEC is successfully flushed through to the large intestine, it is no longer a functional pathogen, as only the small intestine is susceptible to the effects of the enterotoxins (59). This makes ETEC a particularly good candidate for treatment with a virulence inhibitor that targets Rns.

1.7.2 Existing inhibitors of Rns homologues

1.7.2.1 Inhibition of RegA by Regacin

A virulence inhibitor has been identified that targets RegA, the AraC-like global virulence gene regulator of *C. rodentium* (194). *C. rodentium* is a natural mouse pathogen, and is commonly used as a model organism for attaching and effacing pathogens such as EPEC (195). In *C. rodentium*, RegA activates the expression of a collection of genes required for virulence in mice (143). RegA is activated in the presence of bicarbonate ions, and controls the transcription of a variety of virulence genes, including the LEE pathogenicity island. The RegA inhibitor, Regacin, works by blocking the DNA binding domain of RegA, and is an effective treatment and prevention agent of *C. rodentium* infection in mice when administered before, concurrently, or up to 12 hours after inoculation (194).

1.7.2.2 ToxT inhibitors

An inhibitor of virulence gene regulation has also been developed against *ToxT*. ToxT is an AraC-like virulence gene regulator in *V. cholerae* that activates transcription of the cholera toxin and the toxin co-regulated pilus, an adhesin. Virstatin is a small molecule which interferes with the ability of ToxT to dimerise, negating its gene-activating ability (196, 197). It is an effective preventative in a

mouse model of *V. cholerae*, and a successful treatment when administered up to 12 hours post-infection (198). Virstatin is effective against the two most common serogroups of *V cholerae* that cause the majority of disease, but there are rare strains that carry a variant ToxT that is not affected by the inhibitor (198).

Interestingly, Virstatin also has an effect on *Acinetobacter baumannii*, a nosocomial pathogen which commonly causes pneumonia and/or sepsis. *A. baumannii* was chosen as a potential target of Virstatin because the type IV pilus it uses in the formation of biofilms is similar to the TCP of *V. cholerae* (199). In the presence of Virstatin, *A. baumannii* is less motile, and shows a reduced ability to form biofilms (200). This cross-reactivity was unexpected, as virulence inhibitors are generally highly specific for a certain regulator. Further study is required to investigate the mechanism of phenotype inhibition by Virstatin on *A. baumannii*, and to investigate the broader significance of this cross-species reactivity.

Conjugated linoleic acid (CLA) is another inhibitor of ToxT that reduces cholera toxin production (201). Whereas Virstatin targets the dimerisation of ToxT, an effect which leads to a stronger repression of some promoters than others (197), CLA works through a different mechanism, and is more broadly effective than Virstatin (202). Additionally, CLA has already been approved as safe for use as a dietary supplement by the US Food and Drug Administration, making it an attractive drug candidate.

1.8 Aims of this project

Rns-mediated activation of transcription has been well defined for a number of virulence genes, but more potential targets may remain to be identified. AraC-like regulators often activate the transcription of a wide range of gene targets, but to date, no thorough investigations into the genome-wide effects of Rns have been performed. The first aim of this project was to use mRNA-sequencing to reveal the entire Rns regulon in ETEC strain H10407. Further, we wished to identify and describe an environmental signal that activates Rns. AraC-like regulators are used by bacteria to

adapt to a changing environment, but the precise signal used by Rns to trigger the transcription of the Rns regulon has not previously been investigated.

The second aim of this project was to identify a chemical inhibitor of Rns-mediated gene activation. Small-molecule inhibitors are a novel type of antimicrobial which reduce the virulence of a pathogen, rather than having a bactericidal or bacteriostatic effect. In this case, we were searching for a chemical that would inhibit the gene-activating activity of Rns. To this end, we used two different screening methods; a high-throughput chemical screen, and a virtual screen using computer modelling.

The work described in this thesis improves our understanding of the Rns regulon, and provides proof-of-concept experiments characterising an inhibitor of virulence gene activation in ETEC.

2 Methods

2.1 Chemical reagents

All chemical reagents were purchased from Ajax Chemicals, Sigma-Aldrich or Chem-Supply unless otherwise specified.

2.2 Bacterial strains, media and growth conditions

2.2.1 Strains and plasmids

Bacterial strains and plasmids used in this study are listed in tables 2.1 and 2.2 respectively. Periodically, PCR was performed on *E. coli* H10407 to ensure that no plasmids had been lost, using primers p52.F and p52.R, p58.F and p58.R, hlt.F and hlt.R, and hstB.F and hstU.R, to detect plasmids p52, p58, p666, and p948, respectively (table 2.3).

2.2.2 Growth media

Strains were routinely grown in Luria-Bertani (LB) broth or agar plates. LB contains 1% (w/v) Tryptone, 0.5% (w/v) yeast extract, and 1% (w/v) NaCl. When the expression of CFA/I by *E. coli* H10407 was desired, CFA medium was used. CFA broth contains 1% (w/v) Casamino acids, 0.15% (w/v) yeast extract, 0.005% (w/v) $MgSO_4$, and 0.0005% (w/v) $MnCl_2$ (210). For the selection of *lacZ* reporter transformants, minimal medium was used. Minimal medium contains 0.2 % (w/v) D-glucose, trace amounts of thiamine [approximately 0.000001 % (w/v)], and 0.1 % (w/v) Casamino acids. SOB and SOC media were used to grow *E. coli* for electroporation. SOB contains 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM $MgCl_2$, and 10 mM $MgSO_4$. SOC is SOB containing 20 mM glucose. To make solid media, 1.2% (w/v) agar was added.

2.2.3 Culture conditions

Strains were generally grown overnight in LB broth at 37°C, with shaking aeration, or on LB agar plates at 37°C. Where relevant, antibiotics were used at the following concentrations: ampicillin 100 µg/ml, chloramphenicol 10 µg/ml, kanamycin 20 µg/ml. Trimethoprim was used at 40 µg/ml

Table 2.1. *E. coli* strains used in this study

Strain	Relevant characteristics[a]	Reference
H10407	ETEC isolated from a patient with diarrhoea. STp, STh, LT, CFA/I	(203)
H10407Δ*rns*	H10407Δ*rns*::*kan*^{R}	This study
H10407Δ*cfaABCE*	H10407Δ*cfaABCE*::*kan*^{R}	This study
H10407Δ*rns*(pRns)	H10407Δ*rns*::*kan*^{R} (pACYC184::*rns*)	This study
H10407Δ*rns* (pACYC184)	H10407Δ*rns*::*kan*^{R} (pACYC184)	This study
MC4100	Chemically competent *E. coli* strain for cloning	(204)
Top10	Chemically competent *E. coli* strain for cloning	Life Technologies
BL31 DE3	Chemically competent *E. coli* strain for protein expression	(205)
JW3933-3	*E. coli* K12 strain lacking *oxyR*	(206)

[a] STp, heat-stable enterotoxin (porcine subtype); STh, heat-stable enterotoxin (human subtype); LT, heat-labile enterotoxin; *kan*^{R}, kanamycin resistance

Table 2.2. Plasmids used in this study

Plasmid	Characteristic or purpose [a]	Reference
pKD46	λ-red recombinase system, Amp^R	(207)
pKD4	Kanamycin resistance cassette	(207)
pACYC184	Medium copy number cloning vector, Tet^R Chl^R	New England Biolabs
pRns	*rns* cloned into pACYC184, Chl^R	(208)
TOPO-TA	High copy number cloning vector, Amp^R, Kan^R	Life technologies
pGEM-T-Easy	High copy number cloning vector, Amp^R	Promega
pMAL-c2x	Production of MBP-Rns fusion protein, Amp^R	New England Biolabs
pMAL-c2x::*rns*	*rns* cloned into pMAL-c2x, Amp^R	This study
pMU2385	Low copy number plasmid for promoter-lacZ fusions, Tp^R	(209)
pMU2385::P*cfa*	Promoter of *cfa* cloned into pMU2385, Tp^R	This study
pMU2385::P*etpB*	Promoter of *etpB* cloned into pMU2385, Tp^R	This study
pMU2385::P*yghJ*	Promoter of *yghJ* cloned into pMU2385, Tp^R	This study
pMU2385::P*agn43*	Promoter of *agn43* cloned into pMU2385, Tp^R	This study
pMU2385::P*mtr*	Promoter of *mtr* cloned into pMU2385, Tp^R	(194)
pTyrR	*tyrR* cloned into pACYC184, Chl^R	(194)

[a] Abbreviations: Amp^R, ampicillin resistant; Tet^R, tetracycline resistant; Chl^R, chloramphenicol resistant; Kan^R, kanamycin resistant; Tp^R, trimethoprim resistant.

Table 2.3. Oligonucleotides used in this study

Oligonucleotide	Sequence[a]	Purpose
hstA.F	TTTCCCCTCTTTTAGTCAGTCAA	Detection of STp
hstB.F	TGCTAAACCAGTAGAGTCTTCAAAA	Detection of STh
hstU.R[b]	GCAGGATTACAACAMARTTCACAGCAG	Detection of STp or STh
hlt.F	ACGGCGTTACTATCCTCTC	Detection of LT
hlt.R	TGGTCTCGGTCAGATATGTG	Detection of LT
p52.F	GAGAGCTGAGGGAGGTTCAGG	Detection of plasmid p52
p52.R	CTGCAGAGGGGATTGGCTGAG	Detection of plasmid p52
p58.F	CGCTTTTTCCCTGGTTGCCCG	Detection of plasmid p58
p58.R	GGAGACAGCCGTGAGTCACG	Detection of plasmid p58
rnskoF	ATCAGAAAGGCCATATGTTGCATTCAGATTGAACGGAGATATA CTAAGGC*TGTGTAGGCTGGAGCTGCTTC*	Deletion of *rns*
rnskoR	GAATTTTCAAGTAGTAATAACGCAGCCTTGCTCATTCTTAATTG CATTAAAATC*GGTCCATATGAATATCCTCCTTAG*	Deletion of *rns*
cfaAkoF	GATGGAAGCTCAGGAGGAAATATGCATAAATTATTCTATTTAC TAAG*TGTGTAGGCTGGAGCTGCTTC*	Deletion of *cfaABCE*
cfaEkoR	CGCAATAGCGCCAATATTGTTGTTATCTAGAGTGTTTGACTACT TGGTGT*GGTCCATATGAATATCCTCCTTAG*	Deletion of *cfaABCE*
pKD4Fs	TGACGAGTTCTTCTGAGCGGGAC	Confirmation of gene deletions
pKD4Rs	TCTAGCTATCGCCATGTAAGCC	Confirmation of gene deletions
cfaAseqF	CAGCATTAGTGAGTGCCACAAG	Confirmation of *cfaABCE* gene deletions
cfaEseqR	AGTCGACGTAGAGATTTATTGCGACGCAAACTG	Confirmation of *cfaABCE* gene deletions
cfaDseqF	CAGACTTGTATTTCTTGAAAGAGGAG	Confirmation of *rns* gene deletions
cfaDseqR	AGCTGGTCCTGATGCATG	Confirmation of *rns* gene deletions

[a] Restriction digest sequences that were used in this study are underlined. Regions complementary to pKD4 are in italics.

Table 2.3 (cont.) Oligonucleotides used in this study

Oligonucleotide	Sequence[a]	Purpose
Pagn43.F	GGATCCGAATCTGTCCGGTCTCAG	Creation of P*agn43-lacZ* transcriptional fusion
Pagn43.R	AAGCTTAGTGGTATTCACTGCCTGC	Creation of P*agn43-lacZ* transcriptional fusion
PetpB.G	GGATCCGGCAACAGTCCGGTGACG	Creation of P*etpB-lacZ* transcriptional fusion
PetpB.R	AAGCTTGCTCCAGCATCTGGAGAC	Creation of P*etpB-lacZ* transcriptional fusion
Prtr.F	GGATCCTTCAATCTGAATGCAACATATGGCC	Creation of P*rtr-lacZ* transcriptional fusion
Prtr.R	ATCTAGACTAAATCCATGTTTTTTTCTGCCAGC	Creation of P*rtr-lacZ* transcriptional fusion
MBPRns.F	GGATCCATGGATTTTAAATACACTGAAGAAAAAGAAATGATAAAAA	Generation of MBP::Rns fusion protein
MBPRns.R	AAGCTTCAATTCAGTTTGCATCGCAATAAATCTC	Generation of MBP::Rns fusion protein
agn43.EMSA.F	AGGATCCGGAACTGGATTTTCTTCTGAACAG	Amplification of P*agn43* for EMSA
agn43.EMSA.R	GAAGCTTCCGGCAGGTACAAAACAGCATGG	Amplification of P*agn43* for EMSA
agn43EMSAControl.F	CACACATGTCACCCTGAAACAG	Amplification of *agn43* coding region for EMSA control
agn43EMSAControl.R	GCCATCAATACTGAACTTACCGC	Amplification of *agn43* coding region for EMSA control
etpB.EMSA.F	GGATCCGGCAGCAAGCCAGTAATG	Amplification of P*etpB* for EMSA
etpB.EMSA.R	GGATCCAGGCGAAAGAACGGTTCTG	Amplification P*etpB* for EMSA
yghJ.EMSA.F	GGATCCGTGAGAACATCTATTTTCTTAC	Amplification of P*yghJ* for EMSA
yghJ.EMSA.R	GGATCCCATTAATAACGCAAGTGAC	Amplification of P*yghJ* for EMSA

[a] Restriction digest sequences that were used in this study are underlined. Regions complementary to pKD4 are in italics.

when supplementing LB, and at 12 μg/ml when added to minimal medium. 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) was used at 25 μg/ml. Long-term storage of strains was in 67% (v/v) glycerol, 33% (v/v) LB at −80°C.

2.3 Software

2.3.1 Sequence viewing and analysis

Geneious version 7.1.7 (Biomatters) was used for all viewing and analysis of DNA and protein sequence data, including primer design and DNA sequence data.

2.3.2 Gene database searches

Gene sequences were from databases at the National Centre for Biotechnology Information (NCBI). These databases, and the BLAST search function used to identify homologous sequences, are available at www.ncbi.nlm.nih.gov.

2.3.3 mRNA sequencing analysis

RNA sequencing reads were aligned to the *E. coli* H10407 genome using Nesoni (Victorian Bioinformatics Consortium), and the data were analysed using the Voom/Limma software (Victorian Bioinformatics Consortium) to detect differential gene expression (211, 212). Results were visualised using Degust (Victorian Bioinformatics Consortium).

2.3.4 Statistical analysis

Prism Version 6.01 (Graphpad) was used for all graph plotting and statistical analysis. Results are expressed as the mean ± standard error of the mean (SEM). Student's t-test was used to determine the statistical significance of differences between results, and a p value < 0.05 was taken to indicate statistical significance, unless otherwise stated.

2.4 Polymerase chain reaction (PCR)

2.4.1 Oligonucleotides

Oligonucleotides used in this study are listed in table 2.3. All oligonucleotides were ordered from GeneWorks Pty Ltd.

2.4.2 PCR conditions

PCR amplifications were performed using a BioRAD C1000 Thermal Cycler. For routine detection of genes and gene deletions, GoTaq® Green Master Mix (Promega) or MyTaq™ Red Mix (Bioline) were used. To amplify DNA for cloning, Platinum® Taq HiFi (Life Technologies) or, Phusion® High-Fidelity PCR Master Mix (New England Biolabs) were used. To amplify DNA for sequencing, GoTaq® Green Master Mix (Promega) or Phusion® High-Fidelity PCR Master Mix (New England Biolabs) were used. Primers were used at 0.4 to 2 µM in general, or at 0.1 to 0.2 µM with Phusion® High-Fidelity PCR Master Mix (New England Biolabs). Template DNA was plasmid DNA, H10407 gDNA, or, for colony PCR, single colonies suspended in water. In general, thermocycling conditions were as follows: An initial denaturing step at 95°C for at least three minutes, followed by 30 or 35 cycles at 95°C for 30 seconds, an appropriate annealing temperature (specific to each primer set) for 30 - 45 seconds, and an extension period at 72°C for at least 15 seconds per 500 bp expected product size. When using Phusion® High-Fidelity polymerase (New England Biolabs), denaturing was performed at 98°C instead of 95°C, and the annealing time was reduced to 20 seconds. When using Platinum® Taq HiFi (Life Technologies), the extension temperature was reduced to 68°C, and the extension period was extended to 1 minute per kbp expected product size.

2.4.3 Addition of adenine overhangs to blunt-ended PCR products

When PCR products that were generated with blunt-ends required adenine overhangs for cloning into some vectors, A-tailing was performed using TaqPolymerase (New England Biolabs) in Standard Taq Buffer with 4 mM dATP (New England Biolabs). Reaction tubes were held at 70°C for 30 minutes.

2.4.4 Purification of PCR products

DNA resulting from PCR reactions were generally purified using the MinElute PCR purification kit (QIAGEN), the Wizard SV PCR reaction clean-up kit (Promega), or the UltraClean PCR clean-up kit (MoBio), according to the manufacturers' instructions. PCR products that were used for sequencing were cleaned using exoSAP-IT (USB) or the MinElute PCR purification kit (QIAGEN), according to the manufacturers' instructions.

2.5 DNA isolation and purification

2.5.1 Ethanol precipitation

Ethanol precipitation was used to purify and concentrate DNA. 1 volume of DNA to be precipitated was added to 2.5 volumes of ice-cold 100% ethanol and 0.1 volumes of 3 M sodium acetate (pH 5.2). Reactions were incubated on ice for 20 minutes, then centrifuged at room temperature for 15 minutes. The supernatant was discarded, and the remaining pellet resuspended in 100 µl of ice-cold 100% ethanol, incubated on ice for two minutes, and again centrifuged at room temperature for three minutes. The supernatant was discarded, leaving a pellet of DNA.

2.5.2 Plasmid DNA extraction

Plasmid DNA was extracted and purified using the Wizard plus SV Miniprep DNA purification kit (Promega), according to the manufacturer's instructions.

2.5.3 Purification of DNA from agarose gel

DNA was purified from agarose gels after electrophoresis using the QIAquick gel purification kit (QIAGEN) according to the manufacturer's instructions.

2.6 Enzymatic manipulations of DNA

2.6.1 Restriction enzyme digestion

Restriction enzymes (New England Biolabs or Promega) were used at 1.5 µl per 30 µl reaction volume, to cut between 0.1 and 5 µg of DNA. The recommended buffer (New England Biolabs or Promega) was used at 1 x concentration. Reaction tubes were held at 37°C for 2.5 hours.

2.6.2 Ligation reactions

Ligation of DNA fragments into pGEM®T-Easy (Promega) and TOPO-TA (Life Technologies) was performed according to the manufacturer's instructions. Ligations into other plasmids were performed at room temperature overnight with 1 µl of T4 DNA Ligase (New England Biolabs) and 1 x concentration T4 ligation buffer (New England Biolabs).

2.6.3 Enzymatic reaction clean-up

To purify DNA from restriction digest reactions, the MinElute reaction cleanup kit (Qiagen) was used according to the manufacturer's instructions.

2.7 DNA gel electrophoresis

DNA fragments were separated on 1 - 2% (w/v) agarose gels stained with 1 x SYBR Safe (Invitrogen) in TAE buffer [0.04 M Tris-acetate, 1 mM EDTA (pH 8.0)]. DNA samples were run alongside 100 bp or 1 kbp DNA molecular weight markers (New England Biolabs or Invitrogen) to allow estimation of the size of DNA bands. DNA was visualised using a G-BOX gel documentation system (Syngene) and Genesnap software.

2.8 DNA sequencing

DNA was sequenced by using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's instructions. Thermocycling was performed in a BioRAD C1000 Thermal Cycler with conditions as follows: 94°C for 5 minutes, followed by 30 cycles of 94°C for 10 seconds, 50°C for 5 seconds, and 60°C for 3 minutes. Products were cleaned either with the

DyeEx 2.0 spin kit (QIAGEN) or by ethanol precipitation, and sequenced at the Sequencing and Genotyping Facility, Department of Pathology, The University of Melbourne.

2.9 Bacterial growth curves

2.9.1 Standard growth curves

Bacterial growth kinetics were measured using a Pharmacia Biotech Ultrospec 3000 spectrophotometer. Overnight cultures of strains were standardised to an equal OD_{600} before being diluted 1 in 500 in LB broth and incubated at 37°C with shaking at 200 ± 20 rpm for 8 hours, with OD_{600} being measured at least once per hour until reaching a plateau.

2.9.2 Low-volume growth curves

E. coli H10407 grown as above in LB broth with potential inhibitor compounds at a concentration of 10 µM and a final concentration of 1% (v/v) DMSO and dispensed in 1 ml volumes into 1.5 ml capacity tubes. Triplicate cultures were incubated at 37^{o}C, and at each timepoint, tubes were mixed thoroughly and the OD_{600} was measured using a Pharmacia Biotech Ultrospec 3000 spectrophotometer. Readings were taken at least once per hour until reaching a plateau.

2.10 Assays of colonisation factor production

2.10.1 Heat extraction of surface proteins

Overnight cultures of H10407 and derivatives were diluted 1:500 in CFA broth and incubated at 37°C. The cultures were grown without shaking, but flasks were gently swirled every hour, to avoid bacteria settling at the bottom of the flask. After reaching an OD_{600} of approximately 0.6, 10 ml of each sample were centrifuged for 10 minutes, and the pellet resuspended in 150 µl PBS. Samples were held at 60°C for 30 minutes, with vigorous mixing every 5 minutes. Samples were then centrifuged at 2700 x *g f*or 10 minutes, and the supernatant was collected as the final protein sample.

2.10.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

16.25 µl of a heat-extracted surface protein preparation was mixed in a 25 µl reaction with 1 × NuPAGE LDS sample buffer (Life Technologies) and 1 × NuPAGE sample reducing agent (Life Technologies) and held at 70°C for 10 min before being electrophoresed on a 12% Bis-Tris polyacrylamide gel. Gels were run in MOPS gel running buffer with NuPAGE antioxidant (Life technologies) in an XCell SureLock Mini-Cell electrophoresis system, at 180 V. Proteins were run alongside a Novex® sharp pre-stained protein standard (Life Technologies). To detect protein bands, gels were stained overnight using Kang's colloidal stain, which contains 0.02% (w/v) Coomassie brilliant blue G-250, 5% (w/v) aluminium sulphate 14-hydrate, 10% (v/v) ethanol, 2% (v/v) orthophosphoric acid, and destained using 10% (v/v) ethanol and 2% (v/v) orthophosphoric acid.

2.10.3 Haemagglutination of chicken RBC

H10407 and derivatives were grown overnight at 37°C on CFA agar plates. Colonies were swabbed from the plates and suspended in low ionic-strength media (LIM, containing 200 mM sorbitol, 50 mM mannose and 25 mM 2-(N-morpholino)ethanesulfonic acid (MES)) to an OD_{600} absorbance reading of 4. Chicken whole blood in Alsever's solution was diluted 1 in 5 in LIM. On glass slides, 20 µl of suspended bacteria were mixed with 20 µl chicken blood in LIM and gently stirred. Agglutination was judged visually after an incubation of approximately 5 minutes at room temperature, with periodic gentle rocking.

2.11 DNA transformation

2.11.1 TSS transformation of *E. coli*

Chemically competent strains of *E. coli* were transformed using the "transformation and storage solution" (TSS) method as described by Chung in 1989 (213) except that 50 - 70 µl TSS enhance (100 mM KCl, 30 mM $CaCl_2$, 50 mM $MgCl_2$) was added to the reaction tube.

2.11.2 Preparation of electrocompetent cells

Overnight cultures of *E. coli* H10407 were diluted 1 in 50 in SOB and grown at 37°C, with shaking, to log phase. When electroporating with Kan^R fragments for λ-red recombinase gene deletion, 0.2% L-arabinose was added after the first hour of growth, growth media were supplemented with 100 µg/ml ampicillin, and incubated at 30°C. Cultures at log phase were washed twice in ice-cold sterile water, and then once with ice-cold filter-sterilised 10% (v/v) glycerol, with centrifugation at 4,000 x *g* at 4°C between washes. Cells were finally resuspended highly concentrated in ice-cold filter-sterilised 10% (v/v) glycerol. Electrocompetent cells were used immediately.

2.11.3 Electroporation of *E. coli* H10407

DNA to be transformed was added to between 40 µl and 100 µl of electrocompetent cells in a cold electroporation cuvette and electroporated using a Micropulse electroporator (BioRAD) at 1.80 kV for 5 ms (default setting Ec1). They were then collected from the cuvette using 1 ml SOC broth and incubated at 37°C with shaking for 1 hour. When electroporating with Kan^R fragments for λ-red recombinase gene deletion, 0.2% L-arabinose was added to the SOC, and incubation was at 30°C. Transformants were plated onto selective media and incubated overnight at 37°C.

2.12 Construction of bacterial gene deletion mutants and complemented strains

2.12.1 The λ-red recombinase method of gene deletion

Deletion mutants H10407Δ*rns* and H10407Δ*cfaABCE* were constructed using the λ-red recombinase method of gene deletion (207). In this system, the presence of pKD46 (Amp^R) allows the replacement of a target gene with linear DNA electroporated into *E. coli* cells, based on homologous flanking regions. pKD46 encodes the λ-Red recombinase system, containing gene products Gam, which inhibits host cell exonucleases to disrupt the degradation of linear DNA, and

Bet and Exo, which act to promote the recombination of homologous sequences. The system is induced by L-arabinose, and the plasmid is temperature sensitive for convenient curing. Figure 2.1. is a schematic diagram describing the process of constructing a *cfaABCE* deletion mutant of H10407. Other gene deletions were performed using the same general method.

2.12.2 Generation of linear DNA fragments for gene deletions

A gene encoding kanamycin resistance (*kan*^{R}) was used to replace genes targeted for deletion. Oligonucleotides were designed with approximately 50 bp homologous to the gene targeted for deletion at the 5' end, and approximately 20 bp at the 3' end complementary to pKD4. These oligonucleotides were used in a PCR reaction to amplify a kanamycin resistance gene from pKD4, yielding a linear fragment of DNA including *kan*^{R}, flanked by 50 bp sequences homologous to the gene targeted for deletion. The primer pairs rnskoF and rnskoR, or cfAkoF and cfaEkoR, were used to generate DNA for the deletion of *rns* and *cfaABCE*, respectively (table 2.3).

2.12.3 Gene deletions

Linear DNA for gene deletion was introduced into H10407(pKD46) by electroporation. The λ-Red Recombinase system encoded on pKD46 was induced with 0.2% L-arabinose to encourage recombination of the fragment into the H10407 genome. Deletion mutants were selected by growth with kanamycin, and confirmed by PCR and sequencing using primer pairs pKD4Fs and CfaDseqR, and pKD4Rs and CfaDseqF, to amplify and sequence the DNA surrounding the *rns* deletion, and CfaAseqF and pKD4Rs, and pKD4Fs and CfaEseqR, to amplify and sequence the DNA surrounding the *cfaABCE* deletion (figure 2.2, table 2.3). pKD46 was "cured" by growing the recipient bacteria at 37°C.

2.12.4 Complementation plasmids

To complement the *rns* gene deletion, H10407Δ*rns* was transformed with pRns. pRns was provided by Dr. Ji Yang, (The University of Melbourne) and has been used in a previous publication (208). This plasmid is a derivative of pACYC184 (New England Biolabs) which carries the gene

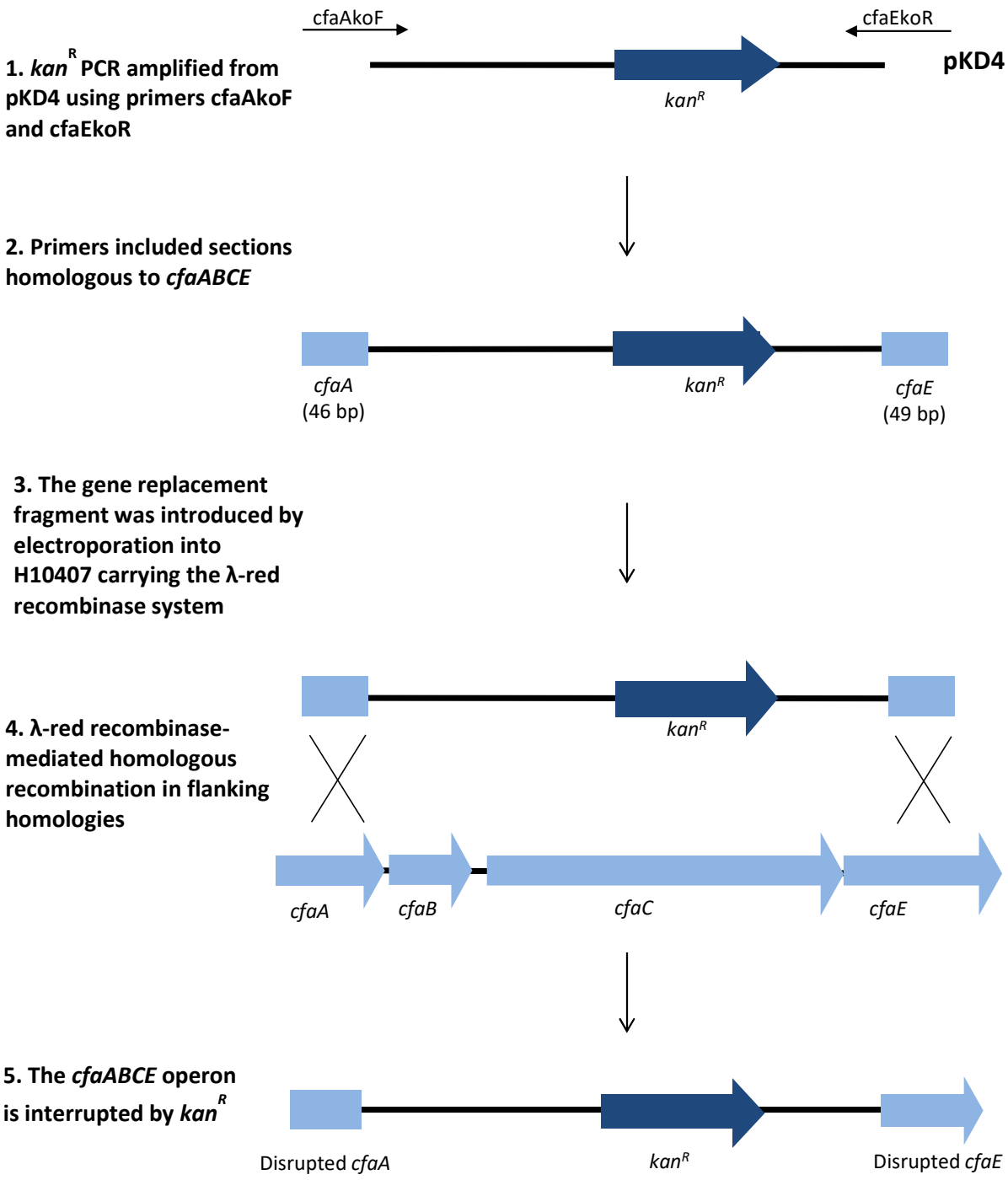


Figure 2.1. Schematic diagram of the construction of a *cfaABCE* deletion mutant of H10407 using the λ-red recombinase method of gene deletion

Primers were created with approximately 50 bp homologous to *cfaA* and *cfaE.* These primers were used in a PCR reaction with pKD4, producing a linear fragment including the kanamycin resistance gene flanked by regions homologous to *cfaA* and *cfaE*. This fragment was electroporated into *E. coli* H10407(pKD46). The λ-red recombinase system mediates homologous recombination, replacing *cfaABCE* with the kanamycin resistance fragment. Mutants were selected for by growth with kanamycin.

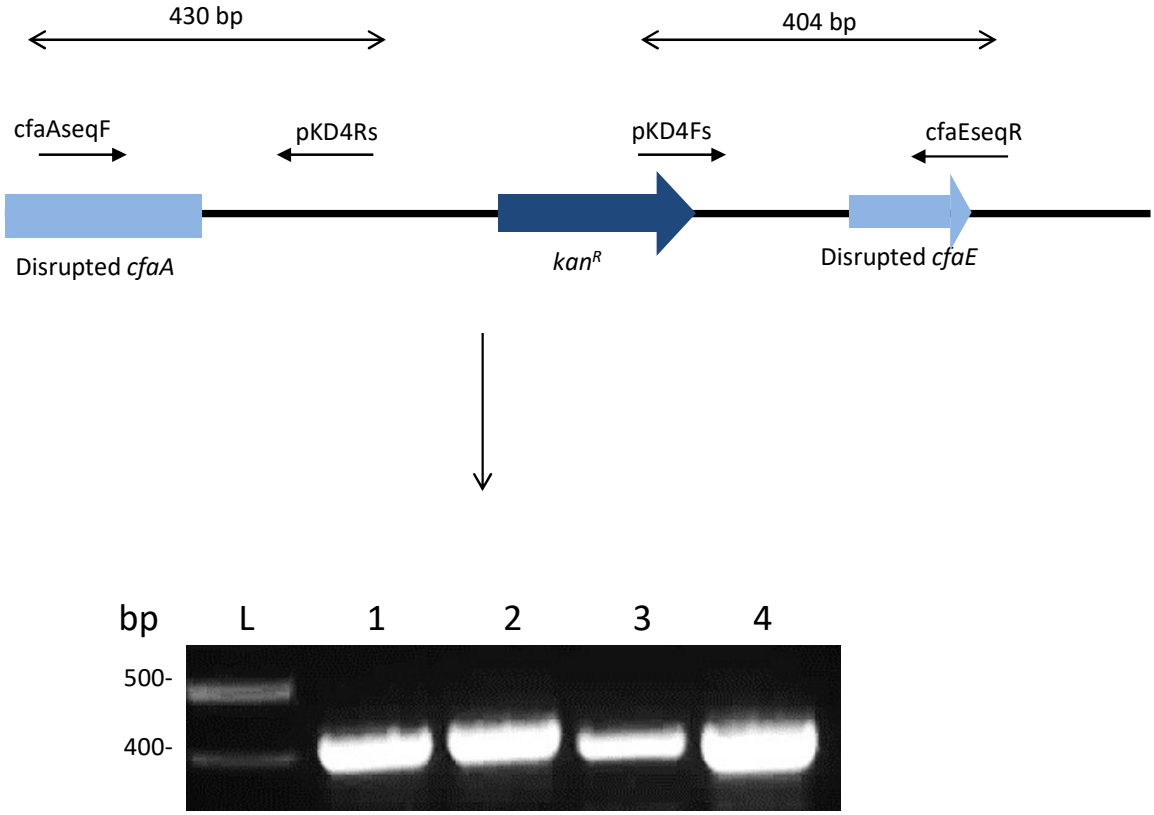


Figure. 2.2. Confirmation of two colonies of H10407Δ*cfaABCE* by using agarose gel electrophoresis of PCR products

Primers cfaAseqF and cfaEseqR are specific to regions of *cfaA* and *cfaE* which were not replaced by the kanamycin insert. Primers pKD4Rs and pKD4Fs are specific to regions within the kanamycin insert. Lanes: L, 100 bp DNA ladder; 1-2, PCR products produced using primers cfaAseqF and pKD4Rs; 3-4, PCR products produced using primers pKD4Fs and cfaEseqR. The bands were then sequenced to confirm that the kanamycin insert had replaced *cfaABCE*.

encoding *rns*.

2.13 Construction of promoter-*lacZ* transcriptional fusions

Promoter-*lacZ* reporters for P*yghJ* and P*cfa* were provided by Dr. Ji Yang (The University of Melbourne). To construct promoter-*lacZ* fusion plasmids for P*agn43,* P*etpB* and P*rtr*, these promoter and Prtr.F and Prtr.R, respectively (table 2.3). PCR products were cloned into pGEM®T-Easy (Promega) according to the manufacturer's instructions. The resultant plasmids were transformed into *E. coli* MC4100 and confirmed by PCR to contain the promoter region of interest, using the same primers as those used to amplify the promoter region. Plasmid pMU2385 carries the coding region of the *lacZ* gene downstream of a cloning site (table 2.2). Promoter regions were digested from pGEM®T-Easy using the restriction enzymes BamHI and HindIII or XbaI (New England Biolabs or Promega), and ligated into pMU2385 (that had been cut with the same enzymes) to form promoter-*lacZ* transcriptional fusion plasmids pMU2385::P*agn43,* pMU2385::P*yghJ,* and pMU2385::P*rtr*. *E. coli* MC4100 was then transformed with each promoter-*lacZ* fusion plasmid, as well as with either pACYC184 or pRns, to create promoter-*lacZ* reporter strains. Figure 2.3 is a schematic diagram describing the construction of pMU2385::P*etpB.* Other promoter-*lacZ* transcriptional fusions plasmids were constructed using the same general method.

2.14 Isolation of MBP:Rns translational fusion protein

2.14.1 Construction of an MBP::Rns translational fusion protein

To construct the MBP::Rns fusion protein, the coding region of Rns, flanked by restriction digest sequences for BamHI and HindIII, was amplified from a preparation of pRns using primers MBPRns.F and MBPRns.R (table 2.3). The cleaned PCR product was cloned into the commercial vector TOPO-TA® using the TOPO®-TA Cloning kit (Life Technologies) according to the manufacturer's instructions. TOPO-TA was transformed into Top10 chemically competent cells (Life Technologies, table 2.1). The presence of *rns* was confirmed by PCR and sequencing, and then was

1. PCR amplification of the *etpB* promoter region using primers PetpB.F and PetpB.R (table 2.3). This introduced restriction digest sequences for BamHI and HindII flanking the amplified region

2. The pGEM-T-Easy vector was digested with BamHI and HindIII, which each digest exclusively within the *lacZ* gene

3. The *etpB* promoter region was cloned into pGEM-T-Easy vector plasmid to form pGEM-T-Easy::P*etpB*

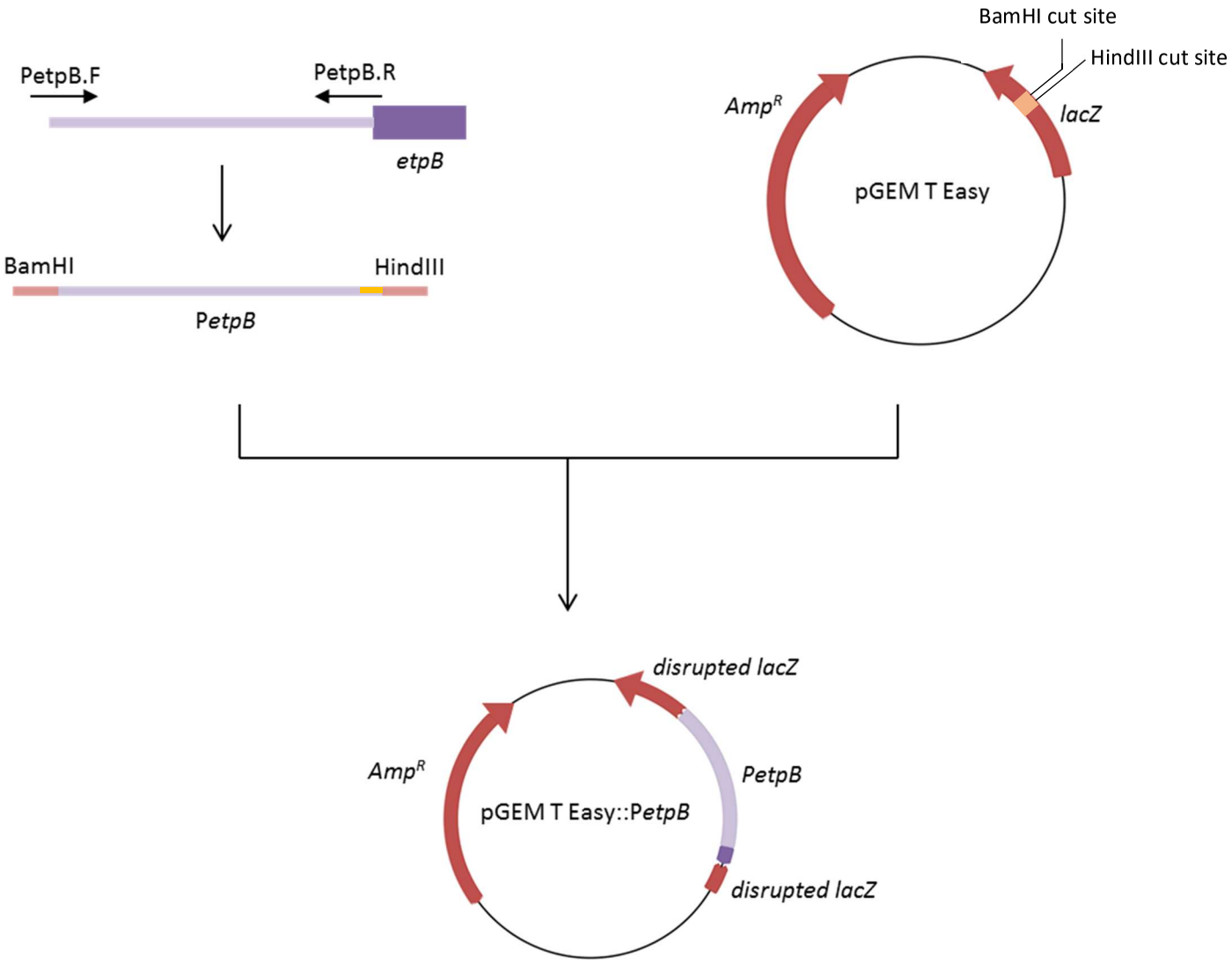


Figure. 2.3. Schematic diagram describing the construction of pMU2385::P*etpB*
Primers were created with restriction digest sequences at the 5' end. These primers were used in a PCR to produce a linear fragment of DNA consisting of the promoter region of *etpB,* flanked by restriction digest sequences BamHI and HindIII. This segment was ligated into the multiple cloning site of plasmid pGEM T Easy, producing the plasmid pGEM T Easy::P*etpB.* This plasmid was cloned into *E. coli* MC4100 for storage and to allow large amounts of this DNA segment to be produced.

1. PCR amplification of the *etpB* promoter region using primers PetpB.F and PetpB.R (table 2.3). This introduced restriction digest sequences for BamHI and HindII flanking the amplified region

2. The pGEM-T-Easy vector was digested with BamHI and HindIII, which each digest exclusively within the *lacZ* gene

3. The *etpB* promoter region was cloned into pGEM-T-Easy vector plasmid to form pGEM-T-Easy::P*etpB*

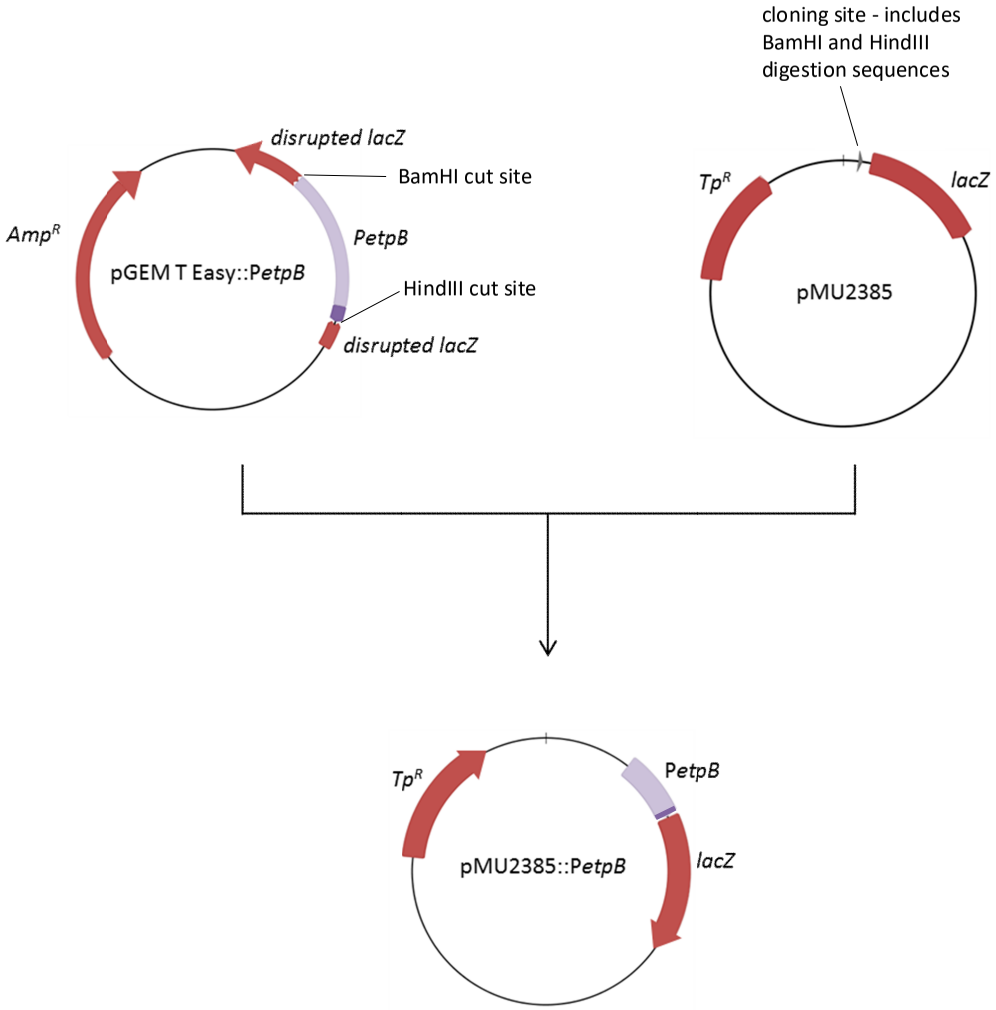


Figure. 2.3 (cont.). Schematic diagram describing the construction of pMU2385::P*etpB*
The promoter region of *etpB* was cut from pGEM T Easy::P*etpB* by digestion with BamHI and HindIII, and ligated into pMU2385 to produce the plasmid pM2385::P*etpB*. This plasmid was transformed into *E. coli* MC4100 for storage.

restriction digested with BamHI-HF and HindIII, gel purified, and cloned into pMAL-c2x in-frame with MBP, creating a fusion protein (table 2.2) (figure 2.4). The resultant plasmid, pMAL-c2x::*rns*, was transformed into *E. coli* BL21 DE3 for protein expression.

2.14.2 Purification of MBP::Rns

E. coli strain BL21 DE3 carrying pMAL-c2X::*rns* was subcultured 1 in 250 in 20 ml LB and grown at 30°C to an OD_{600} of approximately 1. The media was then supplemented with 300 µM IPTG and incubated overnight at 16°C to induce the production of MBP::Rns. These cultures were centrifuged at 4000 x *g* at 4°C. All further work was performed on ice. Cells were resuspended in Buffer A (10 mM Tris-Cl, 200 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol) and sonicated to release intracellular proteins. The suspension was centrifuged at 6700 x *g* at 4°C and the supernatant collected as the protein sample containing MBP::Rns. The sample was column purified on an amylose resin, washed multiple times with buffer A, and eluted using an elution buffer, which was Buffer A containing with 4 mg/ml D-maltose. MBP::Rns was isolated in ten fractions and the concentration was estimated by a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2.15 mRNA sequencing

2.15.1 mRNA isolation and purification

E. coli H10407Δ*rns*(pACYC184) and H10407Δ*rns*(pACYC184::*rns*) were subcultured 1 in 100 into LB with 10 µg/ml chloramphenicol, with or without 45 mM bicarbonate, and incubated at 37°C with shaking to an OD_{600} of approximately 1.0. Two volumes of RNAprotect (QIAGEN) was added to one volume of culture, and samples were incubated at room temperature for 10 minutes. They were then centrifuged at 4000 x *g* for 20 minutes. RNA was extracted using the FastRNA Pro Blue Kit (QBiogene) according to the manufacturer's instructions, except that after the addition of chloroform, 350 µl of the upper phase was added to 35 µl sodium acetate and 875 µl cold 100% ethanol, and held at −20°C overnight. Samples were then centrifuged at 4000 x *g* for 15 minutes at 4°C. Supernatant was removed and the pellet left to air dry for approximately 45 minutes before

1. The coding region of *rns* was amplified using primers MBPRns.F and MBPRns.R. This introduced restriction digest sequences for BamHI and HindII flanking the amplified region

2. Digestion

3. *rns* was ligated into the TOPO-TA vector plasmid to form TOPO-TA:: *rns*

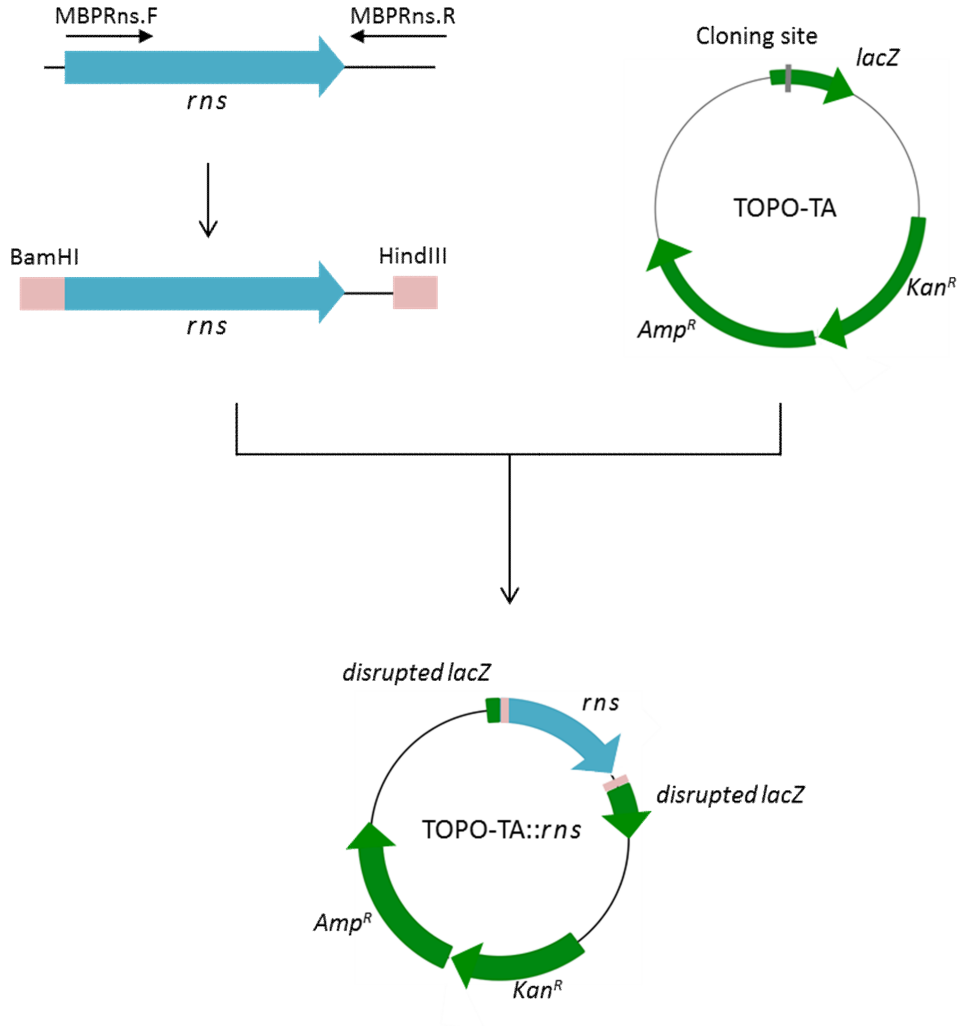


Figure. 2.4. Schematic diagram describing the construction of the MBP::*rns* fusion plasmid
Primers we created with restriction digest sequences at the 5' end. These primers were used in a PCR to produce a linear fragment of DNA consisting of the coding region of rns, flanked by restriction digest sequences for BamHI and HindIII. This fragment was cloned into the cloning site of plasmid TOPO-TA after digestion with BamHI and HindIII, to produce plasmid TOPO-TA::rns.

4. *rns* was isolated from TOPO-TA ::*rns* by digesting with BamHI and HindII

5. Plasmid pMAL-c2x was also digested with BamHI and HindIII

6. *rns* was ligated into the cloning site of pMAL-c2x, to form an ORF encoding an MBP::Rns fusion

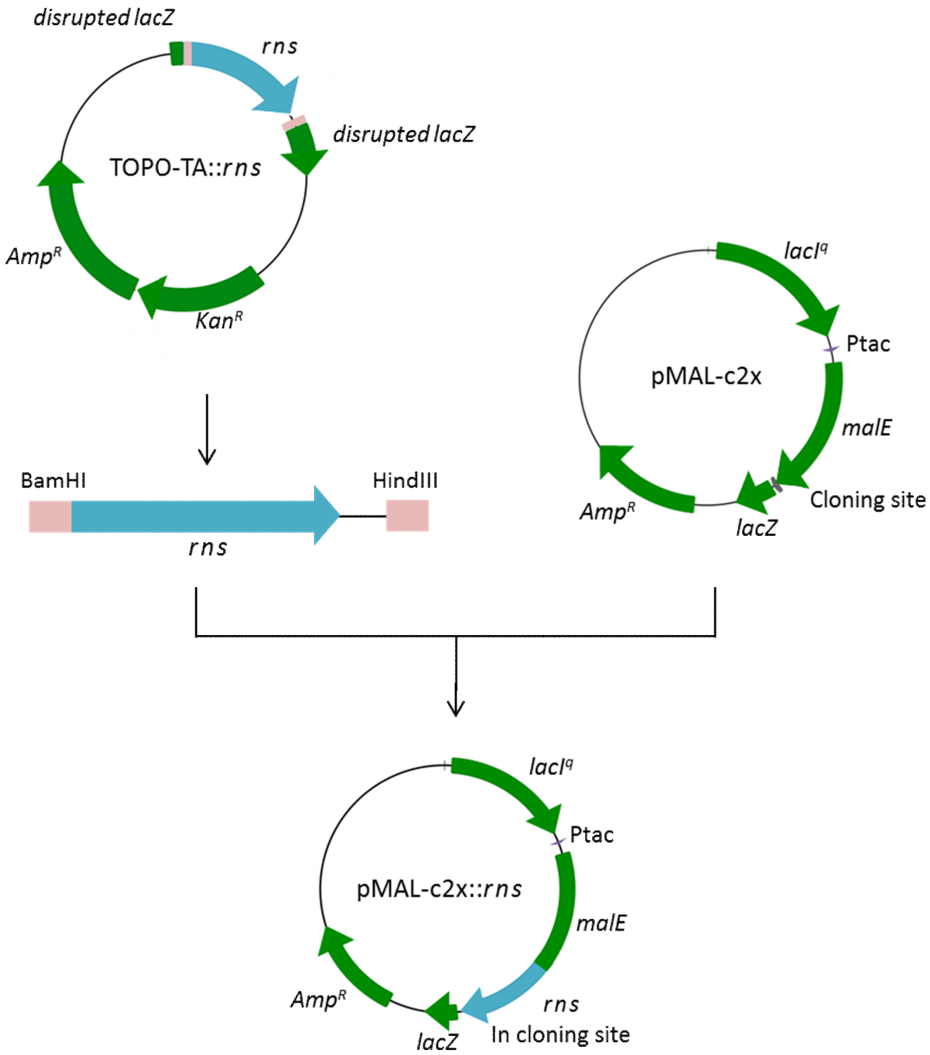


Figure 2.4 (cont.). Schematic diagram describing the construction of the MBP::*rns* fusion plasmid
TOPO-TA::rns and pMAL-c2x were then digested by restriction enzymes BamHI and Hind III, and the rns coding region was ligated into pMAL-c2x, producing the plasmid pMAL-c2x::rns.

being resuspended in 87.5 µl nuclease-free water. The sample was treated with 2.5 µl DNase I and 10 µl RDD buffer from the RNase-Free DNase Set (QIAGEN), and cleaned using the RNeasy MinElute cleanup kit (QIAGEN) according to the manufacturer's instructions, with a final elution volume of 34 µl in RNase-free water. Ribosomal RNA depletion was performed using the Ribo-Zero™ Magnetic Kit (Gram-Negative Bacteria) (Epicentre) according to the manufacturer's instructions. The samples were then re-purified using the RNeasy MinElute cleanup kit (QIAGEN) with a final elution volume of 12 µl in RNase-free water. The samples were tested for residual ribosomal RNA by the Australian Genome Research Facility (Parkville, Vic, Australia) before sequencing.

2.15.2 mRNA sequencing

RNA sequencing was performed using the Illumina Hi-seq 2000, by the Australian Genome Research Facility (Parkville, Vic, Australia).

2.16 β-galactosidase reporter assays

2.16.1 β-galactosidase reporter assays

Overnight cultures of promoter-*lacZ* fusion reporter strains were diluted 1 in 50 in 5 ml fresh LB broth in the presence of chloramphenicol and trimethoprim, with or without 45 mM bicarbonate. Cultures were grown at 37°C with shaking, in a 10 ml test tube, to an OD_{600} of approximately 0.6. 200 µl of these cultures were added to 800 µl Z buffer, containing the following: 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10mM KCl, 1 mM $MgSO_4$, β-mercaptoethanol at a concentration of 2.7 ml per litre. β-mercaptoethanol was added immediately before use. Cells were lysed with chloroform and 0.1% SDS, and each tube was vigorously mixed for 20-30 seconds and allowed to equilibrate to 30°C before beginning the assay. 200 µl of 4 mg/ml ONPG was added to each tube and the time required for the reaction to yield a yellow colour was measured. At this point, 500 µl of 1 M sodium carbonate was added to stop the reaction. The final Ab_{420} and Ab_{550} absorption readings of each tube was taken, and the level of promoter activity was measured in Miller Units using the following equation:

$$\text{Miller Units} = [1000\times \frac{(\text{Ab}_{420} - 1.75\times\text{Ab}_{550})}{t\times\ v\times\text{OD}_{600}}]$$

Where *t* = time in minutes and *v* = the volume of cells used (ml).

2.16.2 Modified β-galactosidase reporter assays

Promoter-*lacZ* fusion reporter strains were grown overnight in 1 ml LB in 1.5 ml tubes, without shaking, in the presence of chloramphenicol and trimethoprim, with or without 45 mM bicarbonate, and with test chemicals dissolved in 1% DMSO when required. Z-buffer was added to 100 or 200 µl of cultures to make a final reaction volume of 1 ml, and the assay was performed as in section 2.16.1 above.

2.16.3 Low-volume β-galactosidase reporter assays

Promoter-*lacZ* fusion reporter strains were grown overnight in 200 µl LB in 1.5 ml-capacity tubes, without shaking, in the presence of chloramphenicol and trimethoprim, with 45 mM bicarbonate, and with test chemical dissolved 1% in DMSO when required. 800 µl Z-buffer was added to cultures to make a final reaction volume of 1 ml, and the assay was performed as in section 2.16.1 above.

2.17 Electrophoretic mobility shift assay (EMSA)

2.17.1 Amplification of DNA fragments for EMSA

The DNA fragments agn43Control.EMSA, agn43.EMSA, etpB.EMSA and yghJ.EMSA were generated by PCR from *E. coli* H10407 genomic DNA using the primer pairs agn43EMSAControl.F and agn43EMSAControl.R; agn43.EMSA.F and agn43.EMSA.R; etpB.EMSA.F and etpB.EMSA.R; and yghJ.EMSA.F and yghJ.EMSA.R (table 2.3). The PCR products were gel purified and cloned into pGEM®T-Easy (Promega) according to the manufacturer's instructions, for storage and amplification. The resultant plasmids were transformed into *E. coli* MC4100, and the fragments were sequenced.

2.17.2 EMSA

Primers agn43.EMSA.F, agn43EMSAControl.F, etpB.EMSA.F and yghJ.EMSA.F were labelled at their 5' end using [γ-^{32}P]ATP and T4 polynucleotide kinase, before being used to amplify DNA fragments for EMSA in a PCR reaction with the corresponding reverse primer (table 2.3). These ^{32}P-labelled fragments, were incubated at a concentration of approximately 3 nM with varying amounts of MBP::Rns at 25°C for 30 mins in binding buffer [10 mM Tris HCl (pH 7.4), 100 mM KCl, 0.1 mM DTT, 0.01% Triton X-100, 1 mM EDTA, 100 µg/ml BSA, 5 ng/µL poly(dI-dC), 10% (v/v) glycerol]. A final concentration of 40 mM bicarbonate was also included in the binding buffer. Binding was assessed after electrophoresis on 5% native polyacrylamide gels at 4°C for 12 h, at 10 V/cm, in a buffer of 50 mM TrisCl (pH8.0), 50 mM boric acid, 1 mM MgCl2, 40 mM NaHCO3, and 6% glycerol.

2.18 Screening for chemical compounds that inhibit the activity of Rns

2.18.1 High-throughput screening of a library of chemical compounds

Screening of the Chembridge Microformats chemical library was performed in 96-well trays. P*cfa-lacZ* reporter strains were diluted 1 in 100 in LB with chloramphenicol and trimethoprim. 100 µl of culture was dispensed into each well of a 96-well tray. 5 µl of chemical compounds at 10 mM in 1% DMSO were dispensed into each well, for a final concentration of 500 µM. Trays were incubated overnight at 37^{o}C in a humidified chamber. 8 µl of 6 mg/ml lysozyme was then added to each well, and trays were incubated at room temperature for 20 to 40 minutes before the screening assay was performed using a FLUOstar Omega Plate Reader (BMG Labtech). The light output from each well was measured following the addition of 25 µl of Beta-Glo® reagent (promega) (figure 2.5). Alongside each test plate, a control plate carrying an unrelated *lacZ* reporter strain but the same chemical compounds was also included to identify compounds that were broadly active, or affected the growth of *E. coli*.

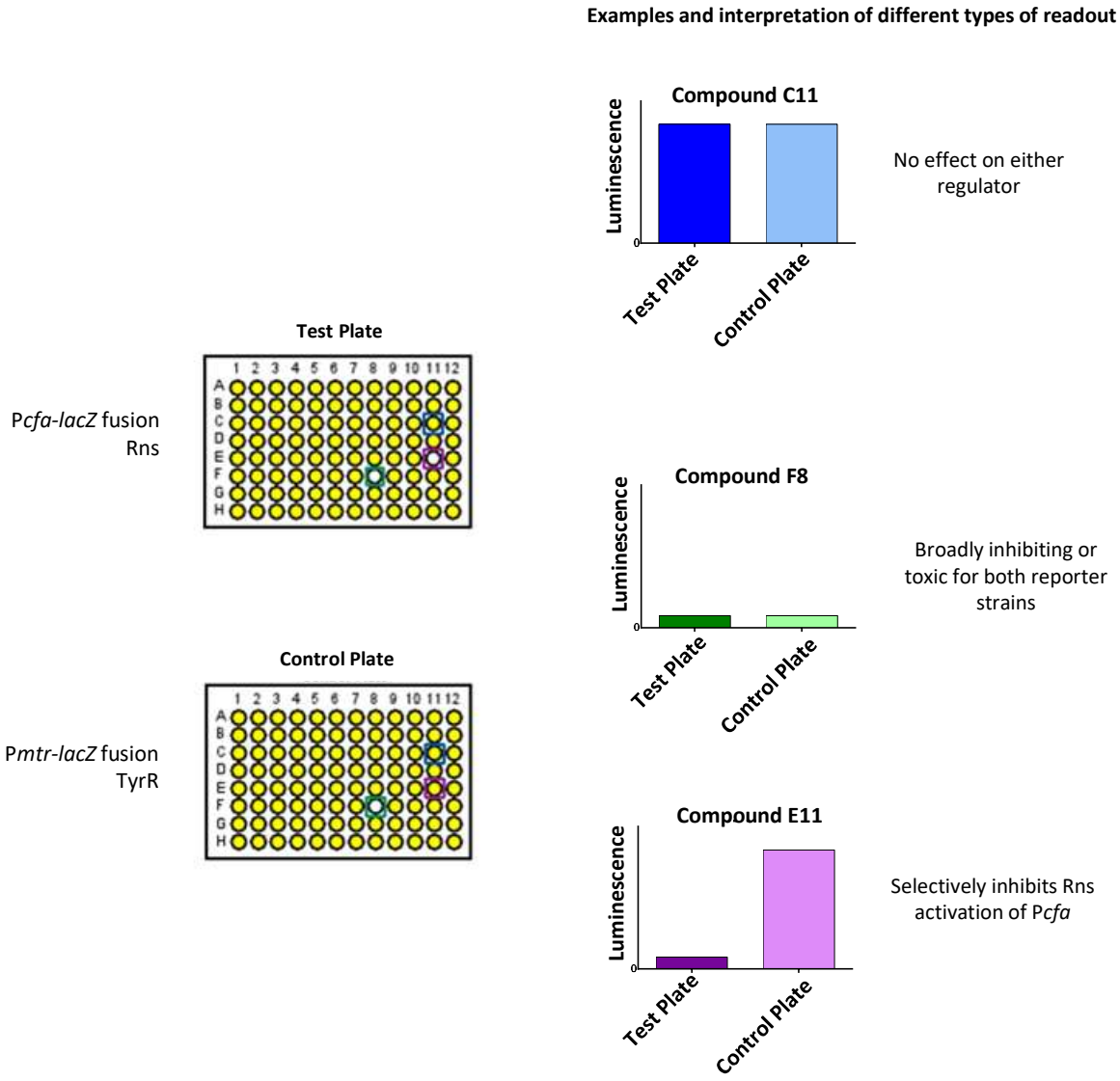


Figure. 2.5. Illustration of high-throughput screening assay for an Rns inhibitor.
Two identical 96-well trays containing a different chemical compound in each well were used to culture *E. coli* MC4100 carrying either a P*cfa-lacZ* fusion reporter plasmid, and another plasmid encoding Rns, or a P*mtr-lacZ* fusion reporter plasmid, and another plasmid encoding TyrR. On the addition of Beta-Glo® (Promega), light output corresponds to the presence of LacZ, and therefore to the transcription level from the promoter being tested. Compounds which caused a decrease in Rns-activated promoter activity, but had no effect on the control regulator, were selected for further analysis.

2.18.2 Virtual screening of chemical compounds

All virtual screening was performed by Dr. Jessica Holien at the St Vincents Institute of Medical Research. To establish a virtual model of Rns, HHpred was used to determine the closest homologue of Rns with a solved crystal structure, based on the Hidden Markov Model (214, 215). This model showed that ToxT from *V. cholerae* is the closest relative of Rns whose structure has been determined. MODELLER v 9.10 was used to generate the model of Rns based on a PIR-alignment with ToxT, which was then assessed with Procheck (216-218). Site ID (sybylx2.0) and fred_receptor (Openeye) were used to identify pockets within the Rns structure that may be targets for binding by chemical compounds. A library of virtual chemicals from multiple databases were filtered using Filter 2.1.1 (Openeye) to yield approximately 4 million compounds with drug-like qualities. Fred 3.0.1 (Openeye) was used to fit each compound into the potential binding pockets of Rns, and the top 500 hits were analysed visually using Vida 4.3.1 (Openeye) and sybylx2.0.

3 The Rns regulon in ETEC strain H10407

3.1 Introduction

AraC-like regulators commonly activate the transcription of a number of target genes to regulate a coordinated response to an environmental signal. Close homologues of Rns include AggR of enteroaggregative *E. coli* and RegA of *C. rodentium*, which both control the expression of multiple virulence genes (143, 146). We hypothesised that Rns may have a similarly broad effect on the ETEC genome.

Rns was originally discovered as an activator of the colonisation factors CS1 and CS2, and later other CFs (112, 113). Investigations of the *rns* promoter region led to the discovery that Rns is also an auto-activator - activating transcription from its own promoter (138). A consensus binding sequence was proposed, and searching the *E. coli* H10407 genome for this sequence revealed *cexE*, an Rns-activated homologue of dispersin, and also *nlpA*, an Rns-repressed membrane protein, which may be involved in the formation of outer membrane vesicles (116, 118). However, as Rns binding sites are quite diverse, it is difficult to predict Rns target genes solely by searching the genome for consensus sequences (figure 1.5) (116, 118, 129, 130). In an attempt to overcome this obstacle, Munson et al (117) used an MBP-Rns affinity column to isolate DNA fragments from ETEC strain C921b-1. This led to the identification of *yiiS*, which is involved in the *E. coli* stress response (117).

Although these experimental approaches uncovered individual Rns target genes, no comprehensive characterisation of the entire Rns regulon has been performed, and many Rns targets may yet be undiscovered. In this chapter, I describe the use of transcriptomics to investigate the role Rns plays in modulating gene expression by *E. coli* H10407.

Virulence regulators must respond to environmental signals to ensure that virulence is activated when required. One such signal is bicarbonate ions, which are abundant in the proximal small intestine, and which act as co-factors for virulence regulators of various intestinal

pathogens (127). In the case of RegA from *C. rodentium*, binding to bicarbonate induces a change in conformation to an active state (figure 3.1) (126). Accordingly, I also determined if bicarbonate ions play a similar role in the activation of Rns.

3.2 Results

3.2.1 Production of CFA/I by *E. coli* H10407 and its derivatives

3.2.1.1 Surface proteins expressed by *E. coli* H10407 and its derivatives

Rns is required for the expression of the *cfaABCE* operon, encoding CFA/I (113). To confirm this, surface protein expression by *E. coli* H10407Δ*rns* was compared with that of wild-type *E. coli* H10407, *E. coli* H10407Δ*cfaABCE*, and *E. coli* H10407Δ*rns*(pRns). SDS-PAGE was performed on heat-extracted surface proteins from these strains (figure 3.2). A band of approximately 17 kDa, which corresponds to the expected size of CfaB, the major subunit of CFA/I, was detected in wild type *E. coli* H10407 but not in *E. coli* H10407Δ*rns* or*E. coli* H10407Δ*cfaABCE*. In the lane containing proteins from *E. coli* H10407Δ*rns*(pRns), the CfaB band was far more prominent than in the proteins from wild type *E. coli* H10407, suggesting that overexpression of Rns led to overexpression of CFA/I proteins.

The 17 kDa band from proteins from wild type *E. coli* H10407 was excised and sequenced by the use of mass spectroscopy at the Walter and Eliza Hall Institute for Medical Research. The results confirmed that CfaB was the predominant protein making up this band.

3.2.1.2 Haemagglutination by *E. coli* H10407 and its derivatives

To investigate if a decrease in CFA/I production causes a measurable loss of cell adherence by *E. coli* H10407, the ability of *E. coli* H10407 and its derivatives to agglutinate chicken red blood cells (RBCs) was investigated. This phenotype has previously been shown to be mediated by CFA/I (219). Wild-type *E. coli* H10407 caused haemagglutination, but this ability was absent from both *E. coli* H10407Δ*cfaABCE* and *E. coli* H10407Δ*rns* (figure 3.3). The agglutination phenotype was restored

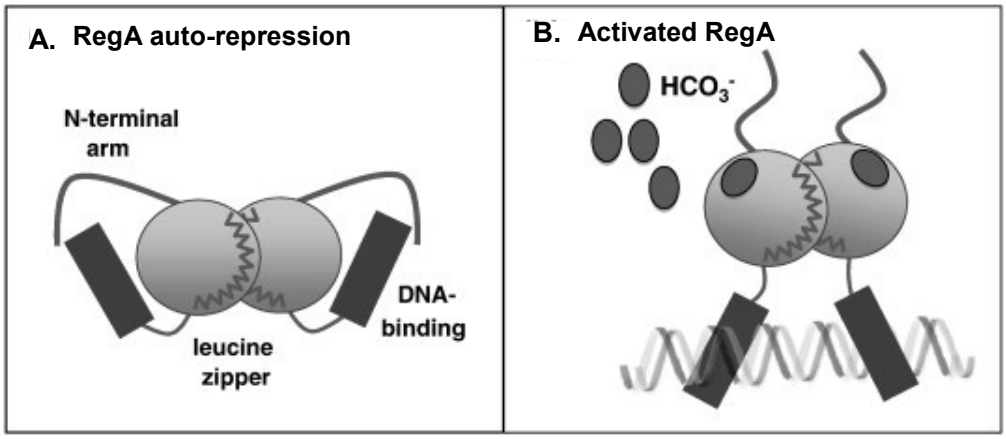


Figure from Yang, et al.,2009 (126)

Figure 3.1. Bicarbonate ions relieve auto-repression of RegA

A model of activation of RegA by bicarbonate ions. A) In the absence of bicarbonate, the N-terminal arm of RegA occludes its DNA-binding domain, thus inhibiting its ability to activate transcription from target promoters. B) The binding of bicarbonate ions triggers a change in the conformation of RegA, allowing the DNA-binding domain to bind target gene sequences and activate transcription.

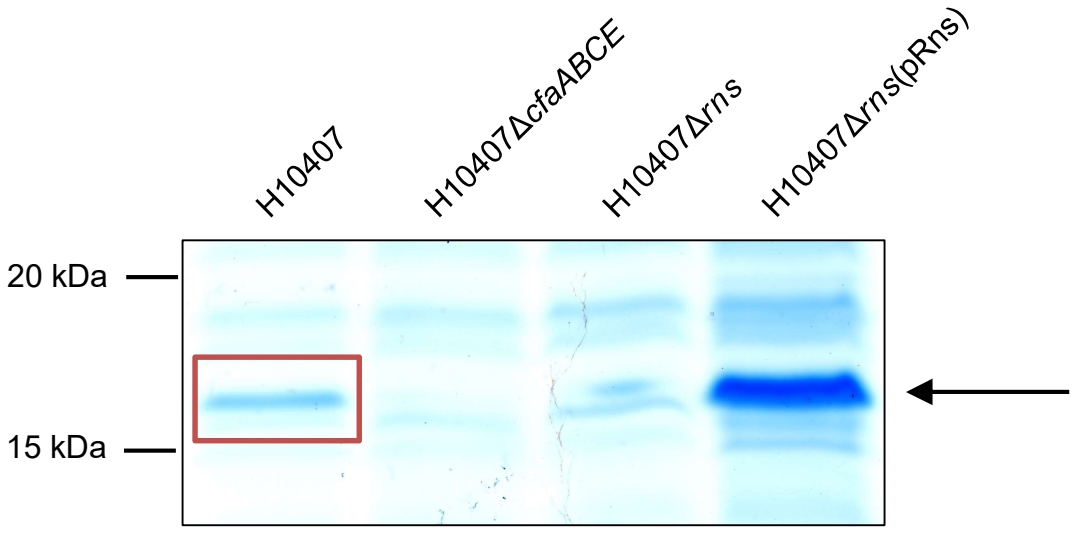
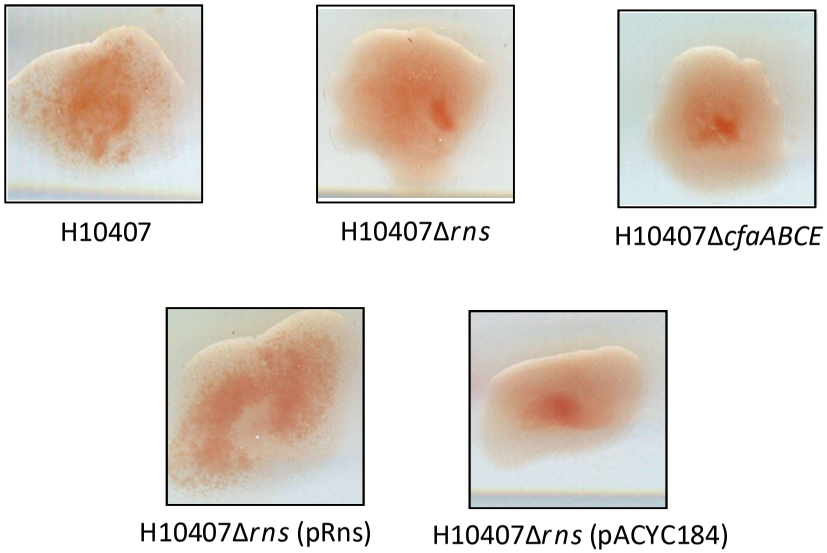


Figure 3.2. Surface protein expression by *E. coli* H10407 and its derivative strains

SDS-PAGE gel showing Kang's-colloidal-stained, heat-extracted surface proteins of *E. coli* H10407 and its derivatives after growth in CFA medium. The band corresponding to CfaB (~17 kDa) is indicated by an arrow. This band is present in surface proteins extracted from wild type H10407, but missing from protein preparations of both H10407Δ*cfaABCE* and H10407Δ*rns,* and restored and overexpressed in H10407Δ*rns*(pRns). The band excised for sequencing is indicated within a red box.



	Haemagglutination
H10407	+
H10407Δ*rns*	-
H10407Δ*cfaABCE*	-
H10407Δ*rns*(pRns)	+
H10407Δ*rns*(pACYC184)	-

Figure 3.3. Haemagglutination of chicken red blood cells (RBCs) by *E. coli* H10407 and its derivatives

Haemagglutination of RBCs by *E. coli* H10407, H10407Δ*rns,* H10407Δ*cfaABCE*, H10407Δ*rns*(pRns), and H10407Δ*rns*(pACYC184). Bacteria were grown on CFA agar, resuspended in LIM and mixed on a glass slide with an equal volume of chicken whole blood, also diluted 1 in 5 in LIM. Wild type H10407 agglutinated RBCs. This phenotype was lost from H10407Δ*cfaABCE* or H10407Δ*rns,* and was restored by trans-complementing *rns* on plasmid, pRns, but not by the empty vector plasmid, pACYC184.

in *E. coli* H10407Δ*rns*(pRns), but not in *E. coli* H10407Δ*rns* carrying the empty vector plasmid, pACYC184.

3.2.2 Effect of bicarbonate on Rns activation of the *cfa* promoter (P*cfaA)*

β-galactosidase reporter assays were performed using *lacZ*-fusions of the *cfaA* promoter, co-transformed into *E. coli* MC4100 with either pRns or pACYC184. The basal level of activity of P*cfaA* in the absence of Rns was 54 ± 1.6 (mean ± SEM) Miller units (MU). In the presence of Rns, transcription increased more than 10-fold, to 571 ± 9 MU (figure 3.4). In the presence of Rns and 45 mM bicarbonate, promoter activity increased approximately 20-fold, to 1071 ± 25. Bicarbonate alone had no significant effect on transcription in the absence of Rns, with promoter activity of 74 ± 10 MU.

3.2.3 Differential expression of the *E. coli* H10407 genome in response to Rns

To measure differential gene expression in response to Rns and screen the *E. coli* H10407 genome for previously unidentified genes targeted by Rns, transcriptome sequencing was performed on strains *E. coli* H10407Δ*rns* (pACYC184) and *E. coli* H10407Δ*rns*(pRns). *E. coli* H10407Δ*rns*(pRns) was used in preference to wild type *E. coli* H10407 as the Rns-positive strain in order to magnify the effect of Rns and thus improve the sensitivity of the screen.

Duplicate cultures *E. coli* H10407Δ*rns* (pACYC184) and *E. coli* H10407Δ*rns*(pRns) were grown in the presence of 45 mM bicarbonate, and mRNA was isolated and sequenced. Short reads were aligned to the *E. coli* H10407 genome (Genbank accession no. NC_017723.1) by the Australian Genome Research Facility (AGRF) and the levels of gene expression in the presence and absence of Rns were calculated.

3.2.3.1 Previously known Rns targets identified by mRNA sequencing

Known Rns targets, *cfaABCE* and *cexE,* were identified by mRNA sequencing as being activated by Rns (table 3.1). These results also indicated that the genes for the *aatPABC* secretion

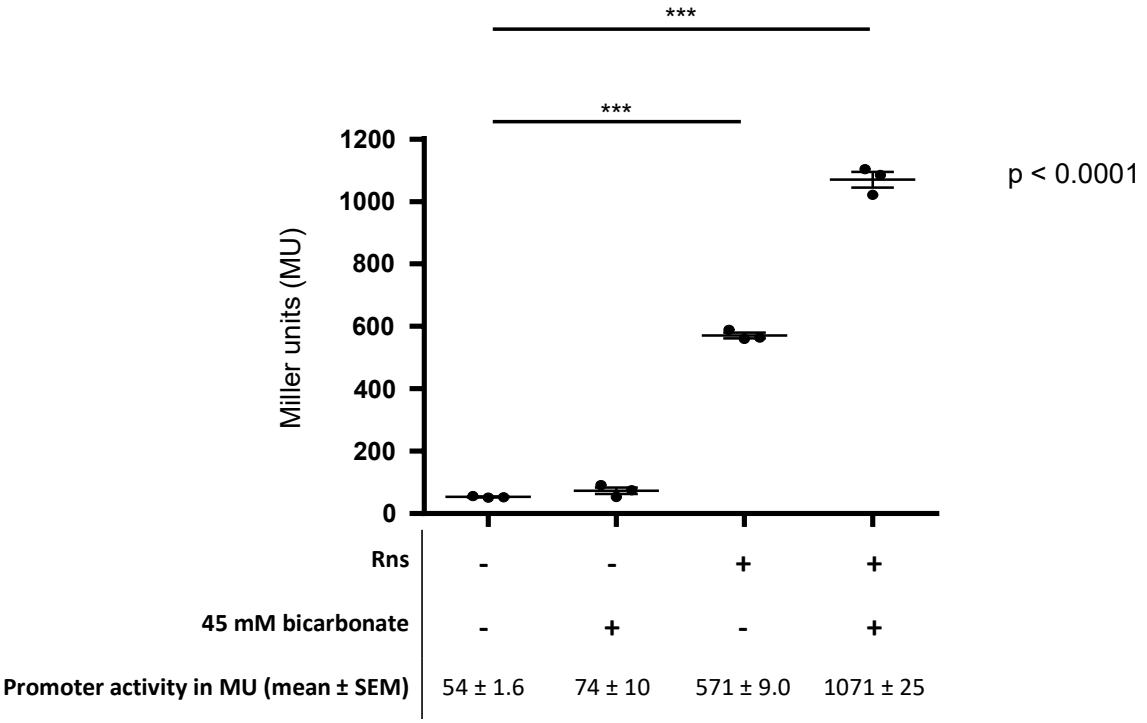


Figure 3.4. Effect of bicarbonate on Rns activation of P*cfaA*

Results of β-galactosidase reporter assays on the activity of P*cfaA* in the presence or absence of Rns and 45 mM bicarbonate. Promoter activity increased approximately 10-fold in the presence of Rns, and approximately 20-fold in the presence of Rns and 45 mM bicarbonate. No increase in promoter activity occurred due to bicarbonate in the absence of Rns. Error bars indicate standard error of the mean, n=3. p value calculated using Student's t-test.

Table 3.1. Genes upregulated in the presence of Rns: Previously known Rns targets

Gene	Function	Activation (fold change)	False discovery rate
cfaA	CFA/I colonisation factor	413	5.06 E-10
cfaB	"	393	2.24 E-08
cfaC	"	187	3.48 E-09
cfaD	"	7.1	4.26 E-06
cfaE	"	50	2.83 E-08
cexE	Dispersin	917	5.86 E-09
aatP	ABC transporter	111	4.43 E-07
aatA	"	36	2.43 E-07
aatB	"	19	4.43 E-07
aatC	"	9.5	9.08 E-05

system were transcribed in an operon from the Rns-activated *cexE* promoter. Although these genes were not previously known to be Rns targets, Rns activation of the *cexE* promoter has been described previously (116) and was not explored further. For this experiment, *rns* was provided on a plasmid and was transcribed from a constitutively active promoter. Accordingly, we did not analyse the expression of *rns* itself.

Previous studies have identified two other chromosomal genes as being regulated by Rns. These are *yiiS* (annotated as ETEC_4191), involved in the sigma factor σ^E stress response of *E. coli*, which was activated approximately 2-fold by Rns, and *nlpA* (ETEC_3954), involved in the formation of outer membrane vesicles, reported as being repressed by Rns and its homologues (117, 118, 139, 140). In our mRNA sequencing experiments, no significant difference in the transcription of these genes was seen in response to Rns, and expression of these genes was generally low. For *yiiS*, transcription was at a rate of 75 ± 7.21 reads per million in the absence of Rns, and 89.95 ± 47.16 reads per million in the presence of Rns, and *nlpA* transcription was 10.65 ± 1.48 reads per million in the absence of Rns and 9.15 ± 2.33 reads per millions in the presence of Rns (table 3.2).

3.2.3.2 Previously unknown Rns-activated genes identified through mRNA sequencing

mRNA sequencing identified 14 genes that were activated at least four-fold in the presence of Rns and had not previously been described as being Rns-activated (table 3.3). This included 5 putative virulence factors. The cut-off value of four-fold was chosen to allow us to focus on the most significant changes, and we screened for RNA-sequencing results with a false discovery rate (FDR) of 0.01 or lower. The FDR is an adjusted p-value which reduces the rate of false-positives when analysing large data sets. It is calculated by analysing the distribution of p-values to estimate what proportion of significant results are likely to be false.

The putative virulence factors identified were: two copies of an Rns transcriptional repressor, ETEC_p948_0450 and ETEC_p948_1070 (hereafter referred to as *rtr2* and *rtr*, respectively), two copies of a gene encoding Antigen 43, annotated *agn43* and *flu*, and *etpBAC*,

Table 3.2. mRNA sequencing shows no differential expression of *nlpA* and *yiiS* in the presence of Rns

Annotation in Genbank	Gene	False discovery rate	Fold-change in response to Rns	Reads per million in the absence of Rns (mean ± SD)	Reads per million in the presence of Rns (mean ± SD)
ETEC_4191	*yiiS*	0.55	1.46	75 ± 7.21	89.95 ± 47.16
ETEC_3954	*nlpA*	0.89	- 1.09	10.65 ± 1.48	9.15 ± 2.33

Table 3.3 Genes upregulated in the presence of Rns: Previously unknown Rns targets

Gene	Function	Activation (fold change)	False discovery rate
ETEC_p948_0450 (rtr2)	Rns transcriptional repressor	124	1.40 E-04
ETEC_p948_1070 (rtr)	"	110	3.49 E-08
agn43	Antigen 43	21	3.60 E-08
flu	"	7.8	1.13 E-06
ETEC_3214	Unknown	18	1.12 E-07
ETEC_p948_1060	ISSfl4 transposase	13	3.60 E-07
ETEC_p948_0010	IS66-family transposase	4.5	1.22 E-03
traM	Plasmid mobilisation	9.8	1.14 E-04
cstA	Putative carbon starvation protein	6.6	1.36 E-04
etpB	EtpA adhesin	6.0	4.09 E-05
etpA	"	4.4	6.03 E-04
etpC	"	5.1	1.53 E-04
ompC	Outer membrane protein	5.8	7.26 E-05
ETEC_0521	Adenine phosphoribosyltransferase	4.2	7.93 E-03

which encode a secreted adhesin.

During this work, a paper was published describing the activity of YghJ, a protein which assists in ETEC adhesion by degrading host intestinal mucins, thus improving the delivery of LT to host cells (80). The authors of this reported that while non-pathogenic *E. coli* also express YghJ, pathogenic strains typically produce it in greater amounts. This led me to consider whether Rns could be the mechanism behind this over-production. Accordingly, I also investigated *yghJ* as a potential Rns target using β-galactosidase reporter assays and electrophoretic mobility shift assays (EMSA).

Antigen 43 is a surface protein which is present in both commensal and pathogenic strains of *E. coli*, and is encoded by duplicate genes, *agn43* and *flu*, on the chromosome of *E. coli* H10407. Both copies were upregulated (21- and 7.8-fold, respectively) in the samples containing Rns. For the experiments detailed in sections 3.2.6, 3.2.9, and 3.2.10, I focused on *agn43,* which was more highly activated in the presence of Rns than *flu*.

In the case of the Rns transcriptional repressors, both *rtr* and *rtr2* were activated more than 100-fold in the presence of Rns. These genes are part of a duplicated section of the plasmid p948, consisting of *rtr* and *rns* (figure 3.5). A non-functional copy of the gene encoding Rns, *rns2* is located immediately downstream of the *cfaABCE* operon. It has 93% pairwise identity with *rns*, but contains multiple stop codons. The functional *rns* gene is located elsewhere on the same plasmid (figure 3.5). For the purposes of this study, I chose to focus on *rtr*, rather than *rtr2*, because it is encoded alongside the functional copy of *rns*. *rtr* and *rtr2* share 92.4% pairwise identity at the nucleotide level (183 of 198 base pairs are identical), and the proteins they encode share 87.7% identity (57 of 65 amino acids are identical).

Due to their putative or confirmed role in pathogenesis, the three newly discovered Rns targets, *agn43*, *rtr*, and *etpBAC,* were selected for further investigation using β-galactosidase reporter assays and EMSA, as discussed below.

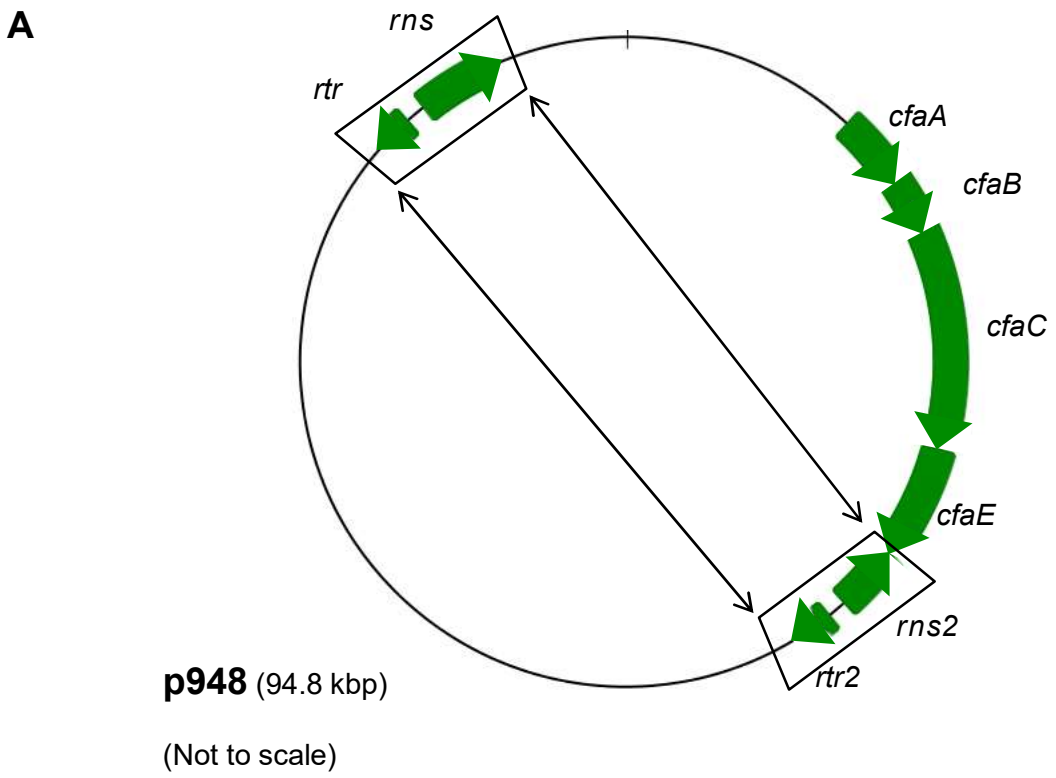


Figure 3.5. Duplication of genes *rns* and *rtr* in plasmid p948

A) Diagram of the *cfaABCE*, *rns* and *rtr* genes encoded on p948. A DNA sequence of approximately 1.5 kbp, including *rns* and *rtr,* is duplicated, with one copy immediately downstream of *cfaABCE* and one elsewhere on the plasmid. The copy of *rns (rns2),* situated immediately downstream of *cfaABCE,* contains multiple stop codons.

B

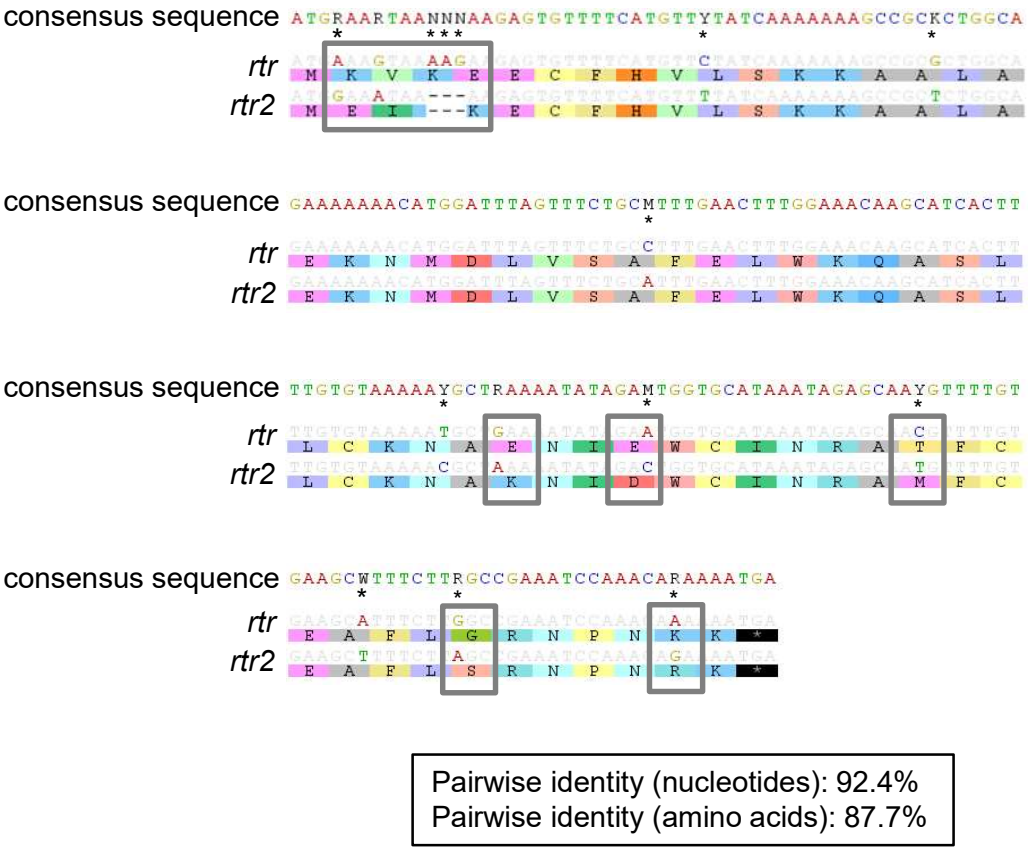


Figure 3.5 (cont.). Duplication of genes *rns* and *rtr* in plasmid p948

B) DNA and protein sequence alignment of *rtr* and *rtr2*. Base pairs that are not identical between the two sequences are marked with an asterisk, and are listed in the consensus sequence using the IUPAC nucleotide ambiguity codes. Amino acids that are not identical between the two sequences are boxed.

Other newly identified Rns target genes included a gene of unknown function annotated as ETEC_3214, multiple transposases on p948, *traM, cstA*, *ompC,* and a protein resembling an adenine phosphoribosyltransferase (table 3.3).

The upstream DNA regions of these putative targets were searched for potential Rns-binding sites using a 'tolerant' consensus sequence generated from 11 experimentally confirmed Rns binding sites. This sequence incorporated every nucleotide occurring in each of the known Rns binding sites into a single sequence using the IUPAC nucleotide ambiguity code. That sequence was "BRWWWDHDTATY". This is the same sequence proposed by Pilonieta et. al., 2007 (116). Numbering from the ATG start codon of each gene, potential Rns binding sites were found 136 base pairs (bp) upstream of ETEC_3214; 297 bp upstream of ETEC_p948_1060; 374 bp upstream of *traM;* and 227 bp and 259 bp upstream of *ompC* (figure 3.6).

3.2.3.3 Previously unknown Rns repressed genes identified through mRNA sequencing

Ten genes were downregulated at least four-fold in the presence of Rns (table 3.4). Five of these are genes involved in chemotaxis and the synthesis of flagella (*tap*, *tar*, *cheA*, *flgA*, and *fliC)*, suggesting a cessation of bacterial motility of the bacteria as they enter the host and adhere to the intestinal epithelium. Two hypothetical genes, ETEC_1274 and ETEC_1275, also showed significantly lower expression in the presence of Rns. A BLASTn search of the NCBI database found that ETEC_1274 has homologues throughout the *Enterobacteriaceae*, but none encode a protein with a known function. ETEC_1275 encodes a 114 amino acid protein with homology to an OmpA membrane domain, and includes a glycine zipper, suggesting that it exists as an oligomer in the membrane (220). Other repressed genes include those encoding the inner membrane protein, YhaO, which is conserved throughout the *Enterobacteriaceae*, L-asparaginase II, and a ferredoxin-type protein.

The upstream regions of each of these genes were searched for potential Rns binding sites, using the same sequence indicated above. Numbering from the ATG start codon of each gene,

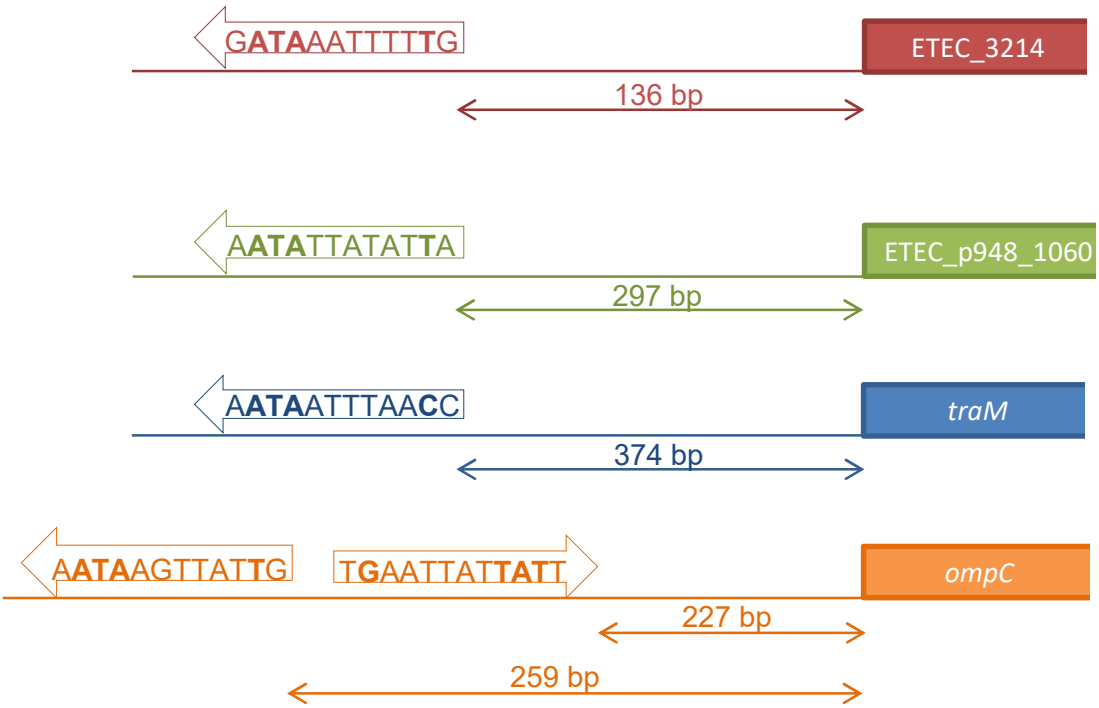


Figure 3.6. Potential Rns binding sites: genes upregulated in the presence of Rns

Diagrams depicting potential Rns binding sites in the regulatory region of genes which were upregulated in the presence of Rns. The sequences matching the Rns consensus binding sequence are listed inside an arrow depicting the direction of the binding site. The distance in base pairs (bp) between the potential binding site and the start of the coding region of each gene is shown. The A/G and TAT nucleotides which makes contact with the Rns protein is depicted in boldface type. Figures shown are not to scale.

Table 3.4. Genes downregulated in the presence of Rns: Previously unknown Rns targets

Gene	Function	Repression (fold change)	False discovery rate
tap	Methyl-accepting chemotaxis protein IV	7.1	3.41 E-4
tar	Methyl-accepting chemotaxis protein II	6.0	6.72 E-5
cheA	Chemotaxis sensor kinase protein	5.4	8.30 E-5
flgA	Flagellar basal body P-ring protein	5.8	5.26 E-5
fliC	Flagellin	5.2	6.52 E-5
yhaO	Inner membrane transport	5.5	9.26 E-4
ETEC_3147	L-asparaginase II	4.8	5.96 E-3
ETEC_1274	Unknown	4.6	2.03 E-3
ETEC_1275	Unknown (membrane associated)	4.5	3.81 E-3
ETEC_2342	Ferredoxin-type protein	4.1	5.01 E-3

potential Rns binding sites were found 142 bp upstream of ETEC_3147; 447 bp upstream of ETEC_1275; and 201, 205 (two overlapping sequences) and 344 bp upstream of ETEC_2342 (figure 3.7).

Because my focus was the activation of virulence mechanisms, these findings were not explored further. In future, promoter-*lacZ* fusions may clarify whether the repression seen here is due to the direct action of Rns, or if it is a downstream effect.

3.2.4 Investigation of previously unidentified Rns target promoters

For any gene that is more highly expressed in the presence of Rns, transcriptional activation may be due to Rns acting on the promoter region directly, or increased expression may be a downstream effect of Rns-mediated action on other genes. To determine if Rns acts directly on the promoter regions of newly discovered Rns targets, *lacZ* fusions were created using the promoter regions of *agn43*, *rtr*, *etpB*, and *yghJ* (see methods section 2.13).

3.2.5 Direct activation of the *rtr* promoter (P*rtr*) by Rns

Direct Rns activation of P*rtr* was measured by using β-galactosidase reporter assays (figure 3.8). A *lacZ* fusion was constructed using the DNA sequence from 427 bp upstream to 73 bp downstream of the ATG start codon of *rtr*. Promoter activity increased from 1.8 ± 1.2 (mean ± SEM, n ≥6) MU in the absence of both Rns and 45 mM bicarbonate to 846 ± 81 MU in the presence of Rns ($p < 0.0001$ compared to in the absence of Rns), and to 1430 ± 103 MU in the presence of both Rns and bicarbonate ($p < 0.0001$ compared to in the absence of Rns). This represents activation of approximately 400-fold in response to Rns alone and approximately 700-fold in response to Rns plus bicarbonate. Promoter activity in the presence of both Rns and bicarbonate was approximately 600-fold higher than in the presence of bicarbonate alone. There was no significant increase in promoter activity in response to bicarbonate in the absence of Rns (2.3 ± 2.0 MU, $p=0.83$).

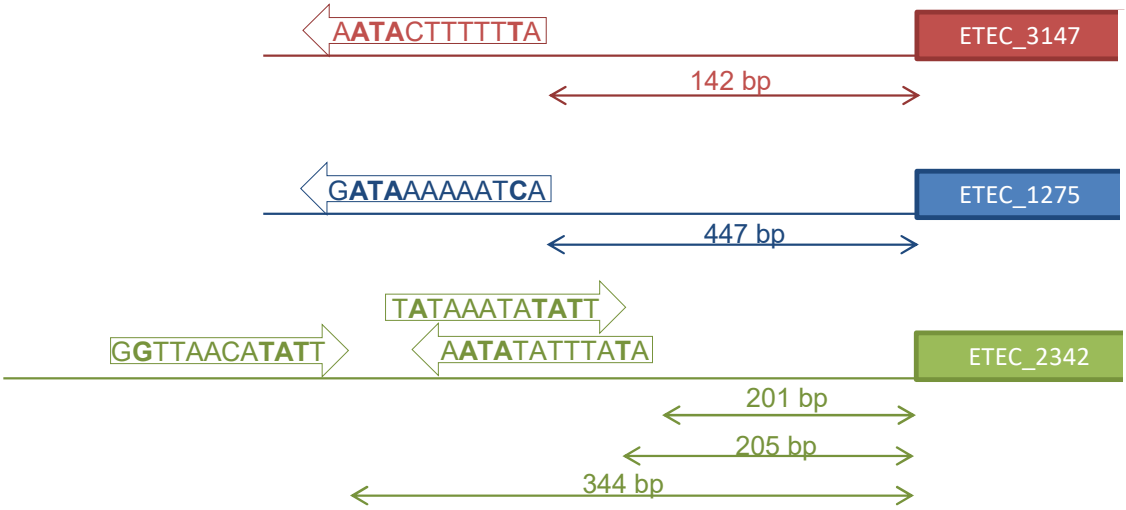
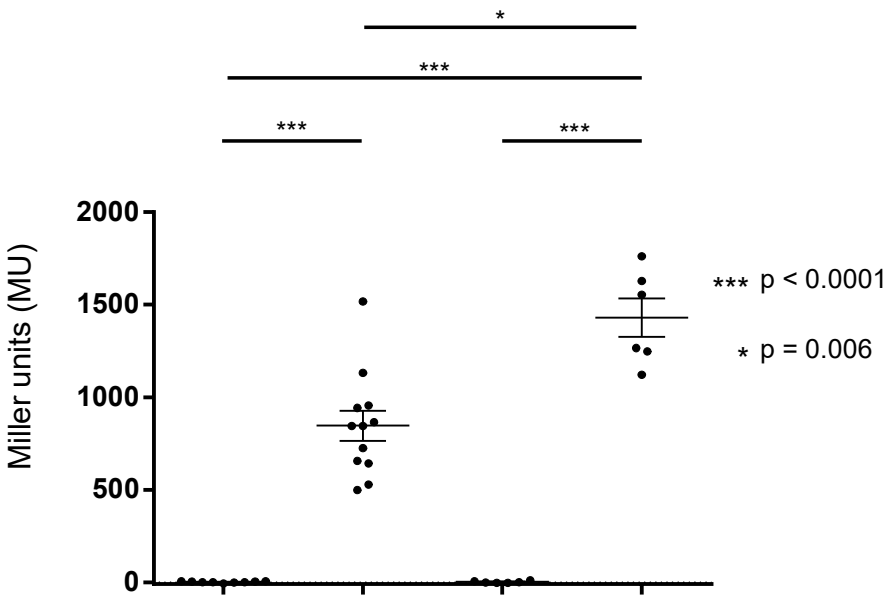


Figure 3.7. Potential Rns binding sites: genes downregulated in the presence of Rns

Diagrams depicting potential Rns binding sites in the regulatory region of genes which are repressed in the presence of Rns. The sequences matching the Rns consensus binding sequence are listed inside an arrow depicted the direction of the binding site. The distance in base pairs (bp) between the potential binding site and the start of the coding region of each gene is shown. The A/G and TAT nucleotides which makes contact with the Rns protein is depicted in boldface type. Figures shown are not to scale.



Rns	-	+	-	+
45 mM bicarbonate	-	-	+	+
Promoter activity in MU (mean ± SEM)	1.8 ± 1.2	846 ± 81	2.3 ± 2.0	1430 ± 103

Figure 3.8. Activation of P*rtr* by Rns

Results of β-galactosidase reporter assays on the activity of P*rtr* in response to Rns, in the presence or absence of 45 mM bicarbonate. P*rtr* activity increased approximately 400-fold in the presence of Rns, and approximately 700-fold in the presence of Rns plus 45 mM bicarbonate. No increase in promoter activity was mediated by bicarbonate in the absence of Rns. Error bars indicate standard error of the mean, n≥6. p values were calculated using Student's t-test.

3.2.6 The activity of the *agn43* promoter (P*agn43*) in *E. coli* strains MC4100 and JW3933-3

The activity of P*agn43* was measured by using a β-galactosidase reporter assay (figure 3.9). A *lacZ* fusion was constructed using the DNA sequence from 287 bp upstream to 703 bp downstream of the TSS of P*agn43*. No changes in expression levels from the *agn43* promoter were detected in the presence of Rns, even when 45 mM bicarbonate was also present (figure 3.9). Moreover, expression was low under all conditions: 12 ± 2.6 MU in the absence of both Rns and bicarbonate, 18 ± 1.2 in the presence of Rns, 25 ± 1.6 in the presence of bicarbonate only, and 22 ± 1.9 in the presence of both Rns and bicarbonate. P*agn43* is known to be phase-variable, and the low levels of transcription indicated that the promoter may not have been active in these experiments. To determine if Rns had an effect on the *agn43* promoter when it was in a 'switched-on' state, we transformed the P*agn43-lacZ* fusion reporter plasmid into *E. coli* strain JW3933-3 (table 2.1). This strain is a Δ*oxyR* mutant *E. coli*, in which *agn43* is phase-locked 'on'. In this background, transcription from P*agn43* was high, confirming that the reporter construct included a functional promoter. However, there was no change in promoter activity in the presence or absence of Rns (data not shown).

3.2.7 Direct activation of the *etpB* promoter (P*etpB*) by Rns

Direct Rns activation of P*etpB* was measured by using a β-galactosidase reporter assay (figure 3.10). A *lacZ* fusion was constructed using the DNA sequence from 947 bp upstream to 76 bp downstream of the ATG start codon of *etpB*. Promoter activity increased from 50 ± 5.2 MU in the absence of both Rns and 45 mM bicarbonate to 1831 ± 170 MU in the presence of Rns ($p < 0.0001$), and to 2853 ± 344 MU in the presence of both Rns and bicarbonate ($p < 0.0001$). This represents an activation of approximately 30-fold in response to Rns alone, and 60-fold in response to Rns and bicarbonate. The 30-fold increase in response to Rns was also seen when comparing promoter activity in the presence of bicarbonate alone (104 ± 18 MU) to promoter activity in the presence of both Rns and bicarbonate ($p < 0.0001$). A 2-fold increase in promoter activity was detected in response to bicarbonate in the absence of Rns ($p = 0.02$).

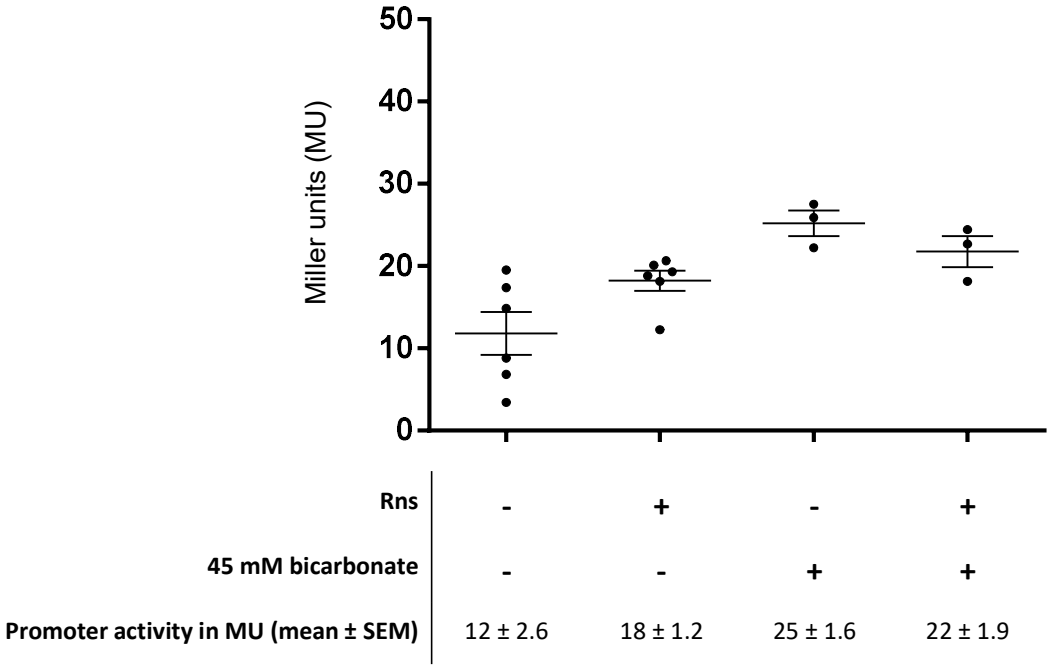
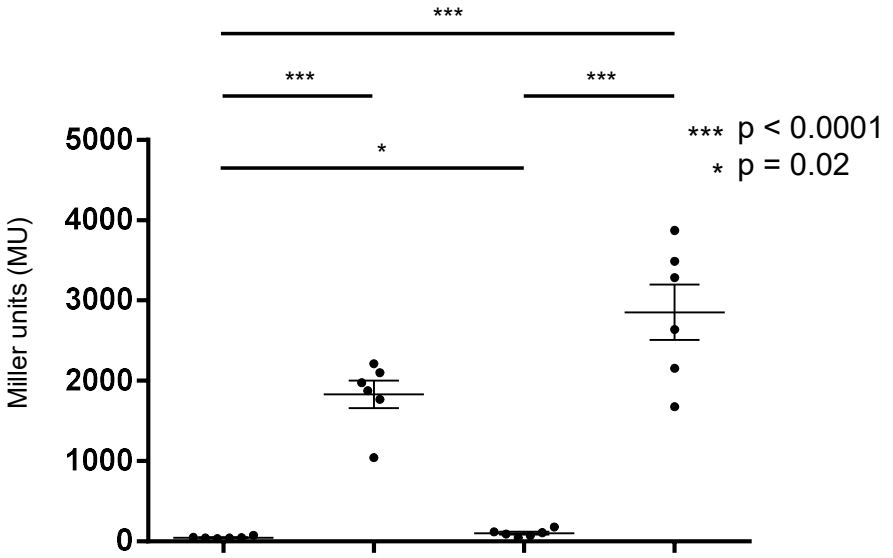


Figure 3.9. Examination of the activation of P*agn43* by Rns

Results of β-galactosidase reporter assays on the activity of P*agn43* in response to Rns, in the presence or absence of 45 mM bicarbonate. No significant change in promoter activity was detected in the presence of Rns and 45 mM bicarbonate. Error bars indicate standard error of the mean, n≥3.



Rns	-	+	-	+
45 mM bicarbonate	-	-	+	+
Promoter activity in MU (mean ± SEM)	50 ± 6.2	1831 ± 170	104 ± 18	2853 ± 344

Figure 3.10. Activation of P*etpB* by Rns

Results of β-galactosidase reporter assays on the activity of P*etpB* in response to Rns, in the presence or absence of 45 mM bicarbonate. P*etpB* activity increased approximately 30-fold in the presence of Rns, and approximately 60-fold in the presence of Rns plus 45 mM bicarbonate. Error bars indicate standard error of the mean, n≥6. p values are calculated using Student's t-test.

3.2.8 Direct activation of the *yghJ* promoter (P*yghJ*) by Rns and bicarbonate

Direct Rns activation of P*yghJ* was measured by using a β-galactosidase reporter assay (figure 3.11). A *lacZ* fusion was constructed using the DNA sequence from 321 bp upstream to 301 bp downstream of the TSS of *yghJ*. Activity from P*yghJ* in the absence of both Rns and 45 mM bicarbonate was 120 ± 12 MU. The presence of Rns increased promoter activity 2-fold, to 244 ± 10 MU ($p = 0.01$). In the presence of Rns and bicarbonate, promoter activity increased 3-fold, to 368 ± 24 MU ($p = < 0.0001$). Bicarbonate alone increased promoter activity approximately 1.5-fold to 192 ± 23 ($p < 0.0001$). This 1.5-fold increase in response to bicarbonate was also seen when comparing promoter activity in the presence of Rns and bicarbonate to the promoter activity in response to Rns-only ($p < 0.0001$). Promoter activity in the presence of both Rns and bicarbonate was 2-fold higher than in the presence of bicarbonate alone ($p = 0.0003$).

3.2.9 Amplification of promoter regions for EMSA

To determine if Rns binds to the promoter regions of *agn43*, *etpB*, and *yghJ*, EMSA were performed. Regions of approximately 300 bp were chosen to evaluate Rns binding to the promoter region of target genes. We preferentially selected upstream regions with high AT content, especially those which include potential Rns binding sites. In figure 3.12, primer binding sites are marked on the gene sequence, along with the start of the coding region of each gene, and, when known, the TSS.

3.2.10 Binding of Rns to P*agn43*

Despite the fact that there was no change in the promoter activity of P*agn43* in response to Rns in *E. coli* strains MC4100 and JW3933-3, it remained possible that Rns does play a role in *agn43* regulation in ETEC strain H10407. To investigate this possibility, EMSA was performed to evaluate the ability of MBP::Rns to bind to the promoter region of *agn43* (figure 3.13). MBP::Rns did not bind P*agn43*, (139 upstream to 199 downstream of the TSS) at concentrations up to 400 nM. This inability to bind the promoter region of *agn43* provided further evidence that *agn43* is not a direct target of

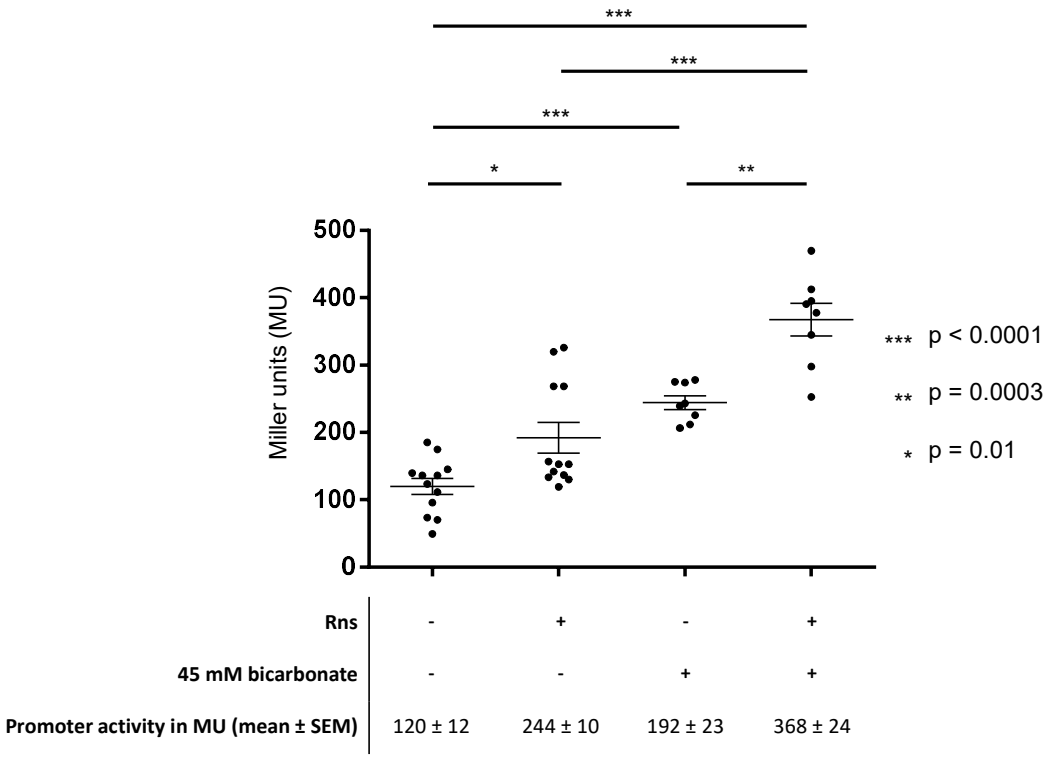


Figure 3.11. Activation of P*yghJ* by Rns and bicarbonate

Results of β-galactosidase reporter assays on the activity of P*yghJ* in response to Rns, in the presence or absence of 45 mM bicarbonate. In response to the presence of both Rns plus 45 mM bicarbonate, promoter activity increased approximately 3-fold. The presence of bicarbonate alone led to an approximately 1.5-fold increase in promoter activity. In the presence of bicarbonate, the addition of Rns increased promoter activity approximately 2-fold. Error bars indicate standard error of the mean, n≥8. p values were calculated using Student's t-test.

A

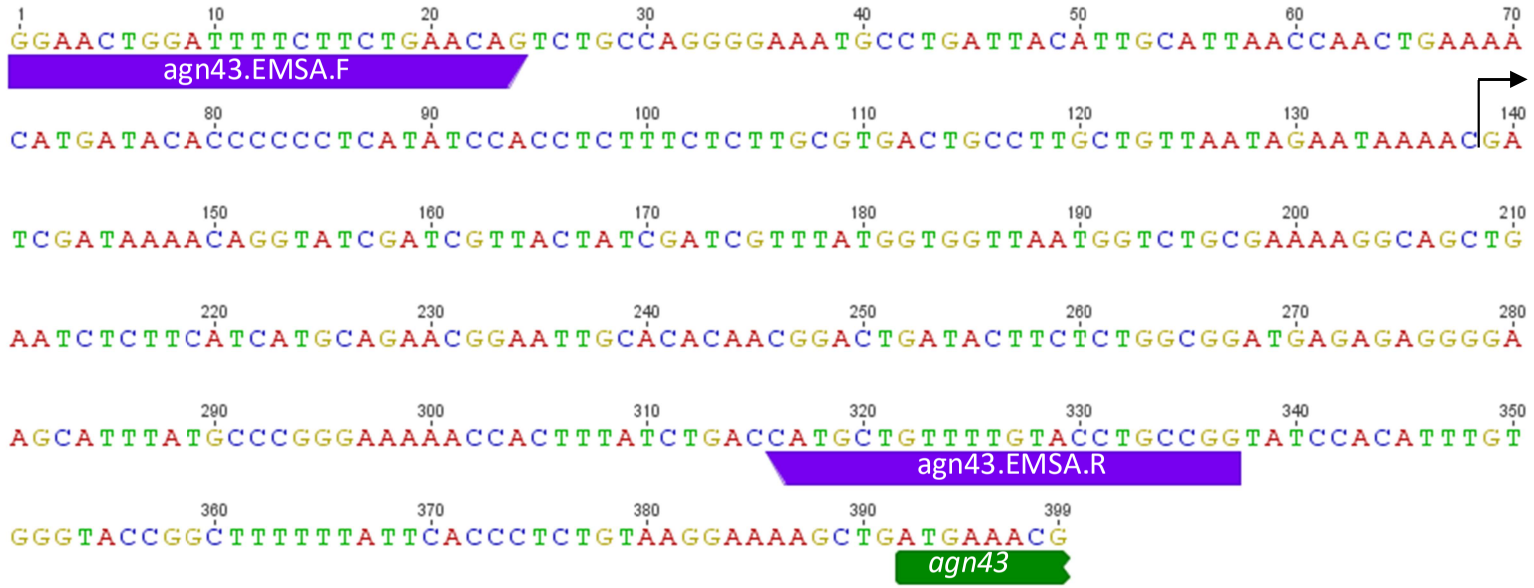


Figure 3.12. Promoter regions amplified for EMSA

Nucleotide sequences, transcription start sites, and binding sites for primers used to amplify the promoter region for EMSA of the *agn43* regulatory region. Primers are marked in purple, coding regions are marked in green, and transcription start sites (where known) are marked with an arrow. A) DNA sequence, primers and transcription start sites of P*agn43.*

B

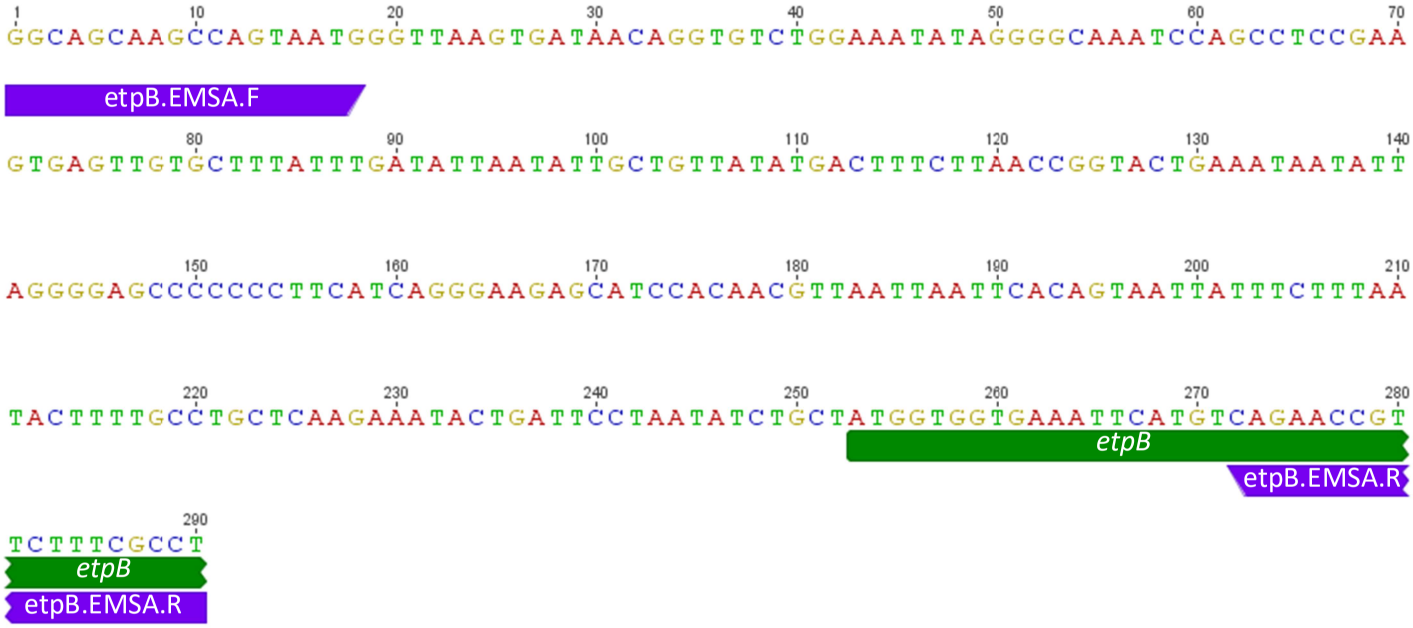


Figure 3.12. Promoter regions amplified for EMSA (cont.)

B) Nucleotide sequences, transcription start sites, and binding sites for primers used to amplify the promoter region for EMSA of the *etpB* regulatory region.

C

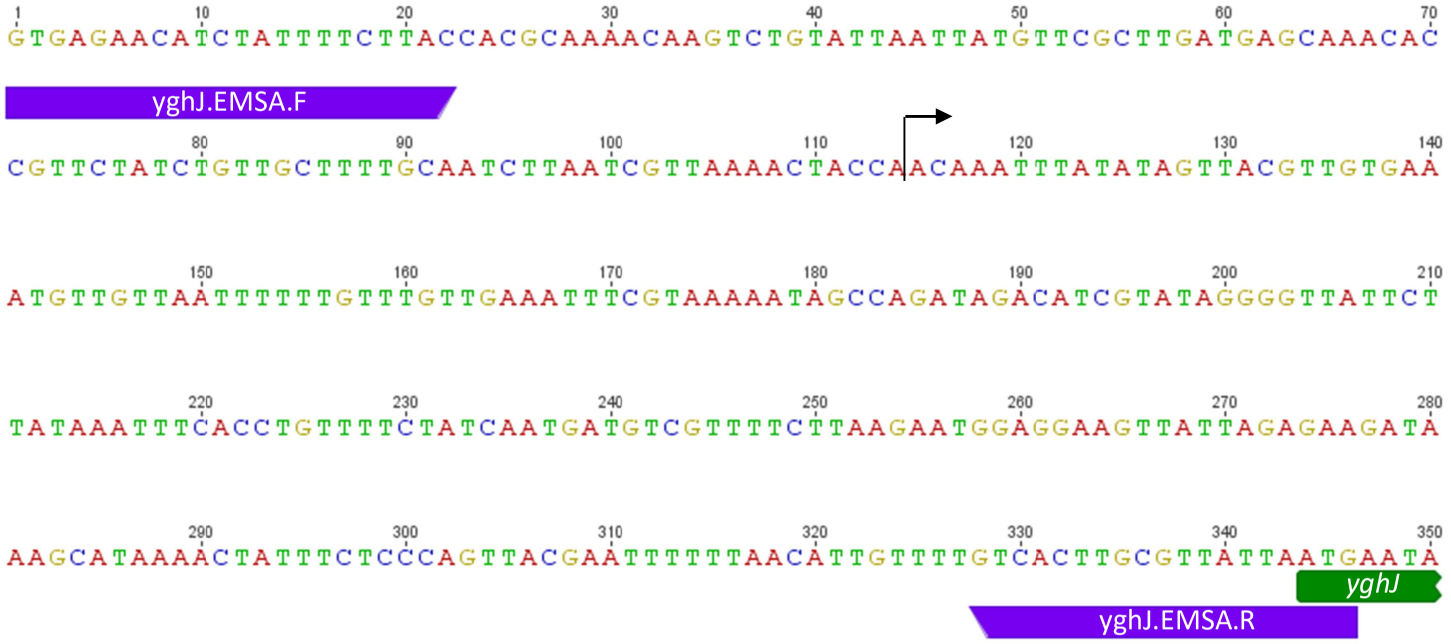


Figure 3.12. Promoter regions amplified for EMSA (cont.)

C) Nucleotide sequences, transcription start sites, and binding sites for primers used to amplify the promoter region for EMSA of the *yghJ* regulatory region.

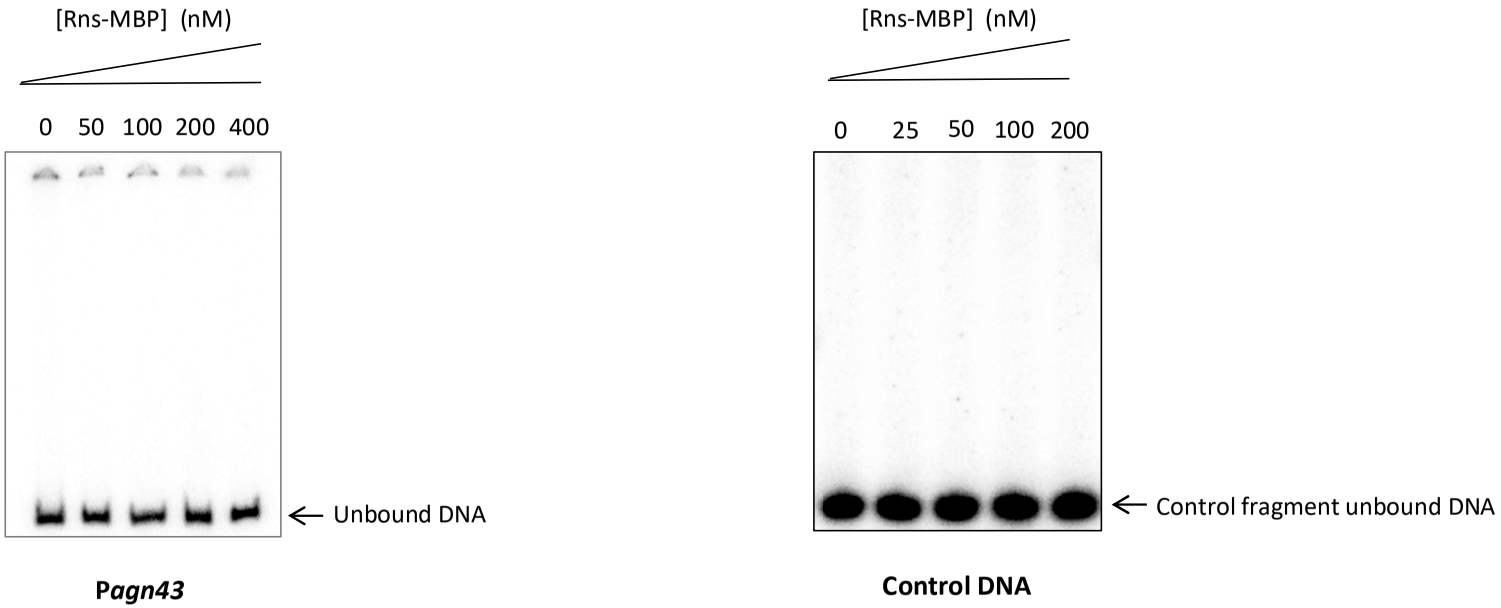


Figure 3.13. Use of EMSA to assess Rns binding to P*agn43*

EMSA showing lack of binding of Rns to the *agn43* promoter region, P*agn43*, at MBP::Rns concentrations up to 400 nM. The ^{32}P-labelled *agn43* promoter region or a control fragment of DNA from the coding region of *agn43* was incubated with various concentrations of MBP::Rns before separation on 5% native polyacrylamide gels.

Rns activation.

3.2.11 Binding of Rns to P*etpB*

EMSA was performed to evaluate the ability of Rns to bind to the DNA sequence from 252 bp upstream to 38 bp downstream of the ATG start codon of *etpB*. The assay showed that MBP::Rns specifically bound to P*etpB* at concentrations as low as 25 nM, with secondary bands appearing at concentrations of 50 nM and above (figure 3.14). A control region of DNA taken from a sequence internal to the coding region of *agn43* was not bound by MBP::Rns at concentrations up to 200 nM. A search of the *etpB* promoter region for an Rns binding sequence identified two potential binding sites, with the sequence "G**A**TATTAA**TAT**T" 152 bp upstream of ATG, and "TGAAATAA**TAT**T" 112 bp upstream of ATG (figure 3.15).

3.2.12 Binding of Rns to P*yghJ*

EMSA was also used to evaluate the ability of MBP::Rns to bind to the DNA sequence from 144 bp upstream to 232 bp downstream of the TSS of *yghJ* (figure 3.16). Faint, incomplete binding of MBP::Rns to P*yghJ* was visible at concentrations of 25 nM and above, consistent with the β-galactosidase results which indicated a low-level activation of this promoter. Searching the promoter region for consensus Rns binding sites found no exact matches to the preferred Rns binding sequence of "BRWWWDHDTATY", but many AT-rich areas including the minimum consensus sequence RN_6TAT. Within the region upstream of *yghJ*, the CG content is approximately 30% (figure 3.17).

3.3 Discussion

3.3.1 Virulence gene regulation by ETEC

Like all living things, bacteria must be able to detect and respond to their environment. Intestinal pathogens such as ETEC are exposed to many different environments: inside the host in

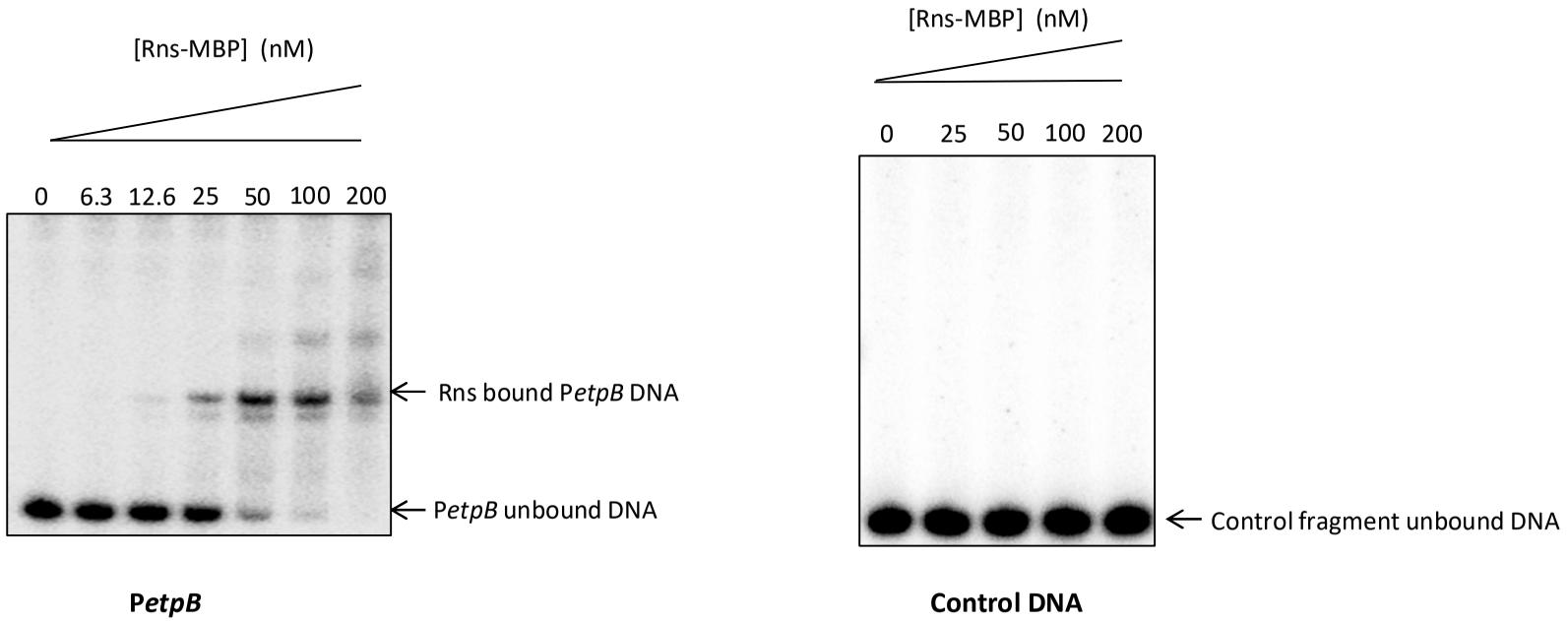


Figure 3.14. Use of EMSA to assess Rns binding to P*etpB*

EMSA showing Rns binding to the *etpB* promoter region, P*etpB,* at MBP::Rns concentrations as low as 25 nM. The ^{32}P-labelled *etpB* promoter region or a control fragment of DNA from the coding region of *agn43* was incubated with various concentrations of MBP::Rns before separation on 5% native polyacrylamide gels.

Figure 3.15. Possible Rns binding sites in the promoter region of *etpB*

Nucleotide sequences, transcription start sites, and binding sites for primers used to amplify the promoter region for EMSA of the *etpB* regulatory region. Primers are marked in purple, the coding region is marked in green, and potential Rns binding sites are marked in red.

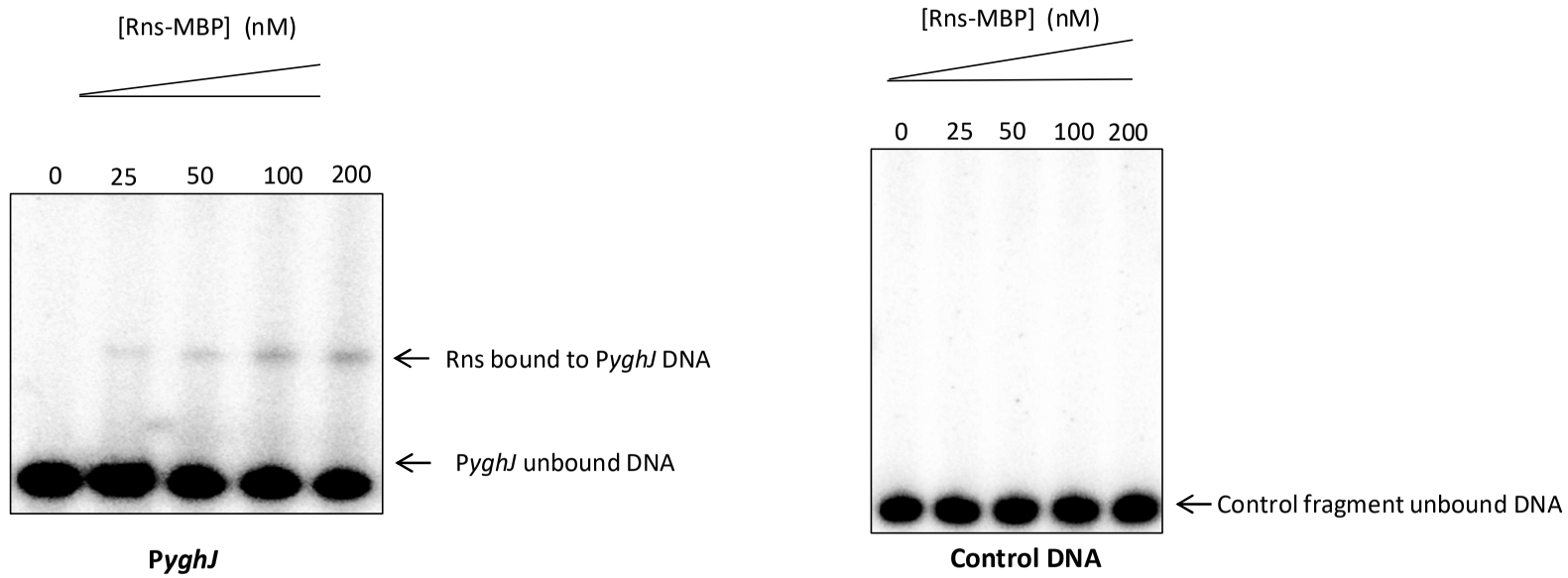


Figure 3.16. Use of EMSA to assess Rns binding to P*yghJ*

EMSA showing weak Rns binding to the *yghJ* promoter region, P*yghJ*, at MBP::Rns concentrations from 25 nM to 200 nM. The ^{32}P-labelled *yghJ* promoter region or a control fragment of DNA from the coding region of *agn43* was incubated with various concentrations of MBP::Rns before separation on 5% native polyacrylamide gels.

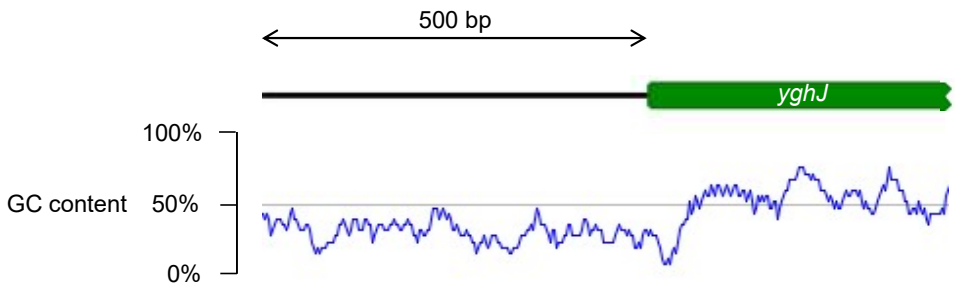


Figure 3.17. The GC content of the *yghJ* promoter region

Graph depicting the GC content of the *yghJ* promoter region. The GC content was calculated using Geneious® V7.1.7, with a sliding window size of 25 bp, and remains below 50% for 500 bp upstream of the ATG start site of *yghJ.*

the stomach and intestines, as well as outside the host, in waste systems, food, water etc. Virulence mechanisms of ETEC are required only when the bacterium is in the proximal small intestine, the site where the enterotoxins exert the strongest effect (86). ETEC must be able to detect when it enters the small intestine, and regulate its virulence genes to ensure they are transcribed when their products are most needed. Rns is an AraC-like regulator that activates virulence genes of ETEC, predominantly by de-repression of H-NS.

H-NS is a global gene silencer that forms filaments along AT-rich regions of DNA and prevents the binding of RNA-polymerase to inhibit gene transcription (87). H-NS can be displaced by the preferential binding of virulence activators, such as Rns, to allow transcription of target genes. In this way, Rns activates virulence gene transcription.

Previous studies have shown that Rns activates the transcription of many CFs, as well as the gene for a dispersin-like protein, *cexE* (116). Using mRNA-sequencing, I identified 36 genes as being significantly transcriptionally altered in the presence of Rns. Four of the upregulated genes, those that related to virulence, were chosen for further investigation into the nature of this activation. For the others, the upstream regulatory region was screened *in silico* for potential Rns binding sites. Overall, the true response of *E. coli* H10407 to Rns included the upregulation of multiple virulence genes, and the downregulation of several genes involved in motility and chemotaxis.

3.3.2 Investigation of the Rns regulon

3.3.2.1 mRNA-sequencing

Previous investigations into the Rns regulon have been limited to studying a handful of genes at once, using in silico screening, *lacZ* fusion constructs, or DNA binding studies to identify novel targets (112, 116-118). In this chapter, I describe the use of mRNA sequencing to examine the entire transcriptome of *E. coli* H10407 and compare it to that of an *rns* deletion mutant, *E. coli* H10407Δ*rns*. Prior to the widespread use of mRNA-sequencing, the standard technique for

quantification of transcription was microarray analysis. Compared to microarray, mRNA-sequencing can quantify over a broader range and provides more complete coverage of the genome. The dynamic range of RNA sequence data covers up to 5 orders of magnitude, compared to microarray which can only discern approximately a hundred-fold change between the lower and upper limits of quantification (221).

Data analysis is the major obstacle to overcome with mRNA-sequencing technology. A number of software programs have been developed to deal with the large amount of data generated from transcriptome studies, and the techniques originally developed to analyse microarray data have had to be modified to interpret mRNA-sequencing results. For this study, data analysis software Voom and Limma, created by the Victorian Bioinformatics Consortium, were used.

3.3.2.2 Choice of experimental strains

In my mRNA-sequencing experiments, I compared transcripts from *E. coli* H10407Δ*rns*(pACYC184) to those from *E. coli* H10407Δ*rns*(pRns), rather than the wild type *E. coli* H10407. *E. coli* H10407Δ*rns*(pRns) expresses Rns from a constitutively active promoter on a plasmid with a copy number of approximately 12. This ensured a production of sufficient Rns to mildly amplify its effects and ensure that we identified as many Rns targets as possible. To stimulate Rns activity, strains were grown at 37°C, with 45 mM bicarbonate added to growth media. However, there are likely to be many other environmental signals which affect the Rns virulence response, and incubation in culture media in a laboratory environment does not closely mimic a human small intestine. A mild over-expression of Rns on pACYC184 may compensate for some of that loss of signal.

Since we altered the expression of *rns* for the sake of the experiment, our mRNA-sequencing results could not be used to analyse the auto-activation of *rns*. The deletion and trans-complementation of *rns* did not include any modification of the coding or promoter region of *rtr*, and so the expression of this gene should be unaffected.

3.3.3 Gene activation by Rns

3.3.3.1 CFA/I-mediated haemagglutination by an Rns mutant

CFA/I was the second CF to be identified as being transcriptionally activated by Rns (113). Production of CFs is the major reason we consider Rns to be a good drug target, as CFs are essential virulence factors of ETEC. As such, we needed to demonstrate that inhibiting the activity of Rns is sufficient to eradicate the binding phenotype of a CF. CFA/I is able to agglutinate chicken RBCs (219). The *E. coli* H10407 deletion mutant, H10407Δ*rns,* completely lost the ability to haemagglutinate, as did *E. coli* H10407Δ*cfaABCE* (figure 3.3). This, combined with surface protein analysis showing a strong reduction in CfaB expression in *E. coli* H10407Δ*rns* (figure 3.2), shows that the sharp reduction in the transcription of *cfaABCE* in the absence of Rns does indeed translate to the loss of a CFA/I-mediated phenotype, and validates Rns as a drug target.

3.3.3.2 The ANR family of negative regulators

A recent study into the regulon of Rns homologue AggR by Santiago *et al* (141) investigated a small coding region divergently transcribed from *aggR* that they dubbed '*aar*' for AggR-activated regulator (141). This gene is transcriptionally activated by AggR and the protein it encodes inhibits AggR transcription. Santiago *et al* hypothesised that Aar acts as a time-delayed negative feedback system which curtails the AggR response once the production of Aar reaches a tipping point, which in their gene expression experiments was approximately 4 hours after the beginning of the AggR response. They found many homologues of Aar throughout *Enterobacteriaceae*, particularly in various pathotypes of *E. coli*. They called this family of proteins 'ANR' for AraC-like negative regulators.

The self-activating nature of Rns-like regulators allows for a rapid and intense response to environmental stimuli, but this positive feedback loop must, at some point, be halted. Santiago *et al* proposed a model wherein the expression of an ANR is closely linked to the expression of a cognate activator in a way which ensures that both are transcribed simultaneously to allow the virulence

response to be stopped after a set time, and give the bacteria a chance to detach from the intestinal epithelium.

Two ANR-family proteins are present in the H10407 genome. These are *rtr* and *rtr2*, which are present in plasmid p948 along with *rns* (figure 3.5). In my mRNA-sequencing experiments, both were upregulated more than 100 fold. The genes *rns* and *rtr* are only 197 bp apart, and are divergently transcribed, sharing a promoter region. Two Rns binding sites have been previously confirmed in this region through DNase I footprinting and uracil interference assays during investigations of the *rns* promoter (130), and I did not repeat these experiments.

The amino acid sequences of Rtr and Rtr2 share 87.7% identity (figure 3.5). The paper in which ANR family of regulators were identified showed that both of these proteins can repress the AggR-mediated activation of a target gene (141). This is in contrast to *rns* and *rns2*, as *rns2* is inactive due to mutations which have introduced multiple stop codons into the coding region. It is interesting that this means that *E. coli* H10407 carries two functional copies of *rtr*, but only one of *rns*. Other fully sequenced ETEC strains, E23477A and E1392/75, each carry only one copy of *rtr*.

3.3.3.3 EtpBAC

EtpA is a secreted non-fimbrial adhesin which acts during the early stages of intestinal colonisation by ETEC (222). EtpC controls glycosylation of EtpA, and EtpB is an outer membrane protein through which EtpA is secreted. After secretion, EtpA detaches from the bacterial cell surface and binds to a host enterocyte. It then 'anchors' passing ETEC cells to the host by binding to the exposed ends of flagella (70).

In both the mRNA-sequencing and β-galactosidase reporter assays, Rns strongly activated the transcription of P*etpB*. EMSA showed clear, strong binding of Rns to P*etpB*, with a second shifted band appearing at higher concentrations of Rns, suggesting a second binding site within the EMSA region. Two strong possibilities exist for Rns binding sites in this region (figure 3.15). These results show that EtpBAC is an important part of the virulence regulon of Rns. A recent review into novel

antigens for ETEC vaccine development suggests that EtpA is a widespread antigen, present in over 70% of ETEC strains from multiple sites (223). Antibodies against EtpA or flagella have been effective at reducing colonisation by ETEC, indicating that EtpA is important in pathogenesis (71, 224). This finding helps to solidify the role of Rns as an activator of the colonisation processes of ETEC.

3.3.3.4 YghJ

YghJ is a metalloprotease secreted by both pathogenic and commensal strains of *E. coli* (80, 225, 226). Host mucins form a protective layer over the intestinal epithelium, limiting access by bacteria, and YghJ degrades these mucins, assisting in the colonisation of commensals and pathogens alike. A recent study into the effects of YghJ suggested that pathogenic strains of *E. coli* secrete greater levels of YghJ (80). Of the strains tested, the ETEC strains H10407, B7A, and E24377/A, all of which carry Rns, displayed far higher YghJ expression than the commensal *E. coli* strains HS, Nissle, and MG1655, which do not encode an Rns homologue. Additionally, the *yghJ* promoter is known to be under the control of H-NS repression (227). For these reasons, I wondered whether Rns may be responsible for this upregulation. In the mRNA sequencing data, the expression of *yghJ* was very low regardless of the presence or absence of Rns. β-galactosidase reporter assays showed that Rns does indeed activate P*yghJ* approximately 2-fold (figure 3.11). These assays also revealed a 1.5-fold activation contributable to bicarbonate ions. These effects combine to a 3-fold activation in the presence of both bicarbonate and Rns.

The promoter region of *yghJ* does not contain the Rns binding consensus sequence "BRWWWDHDTATY". However, this region does have a low GC-percentage (approximately 30%), and EMSA confirms that weak Rns binding does occur (figure 3.16). Moreover, there are 8 sites that meet the minimum Rns consensus sequence "RN_6TAT" (figure 3.18). In future, DNase I footprinting can more precisely identify the site at which Rns contacts the DNA.

3.3.3.5 Antigen 43: indirect activation by Rns

Antigen 43 causes cell-cell aggregation and is involved in the early stages of biofilm

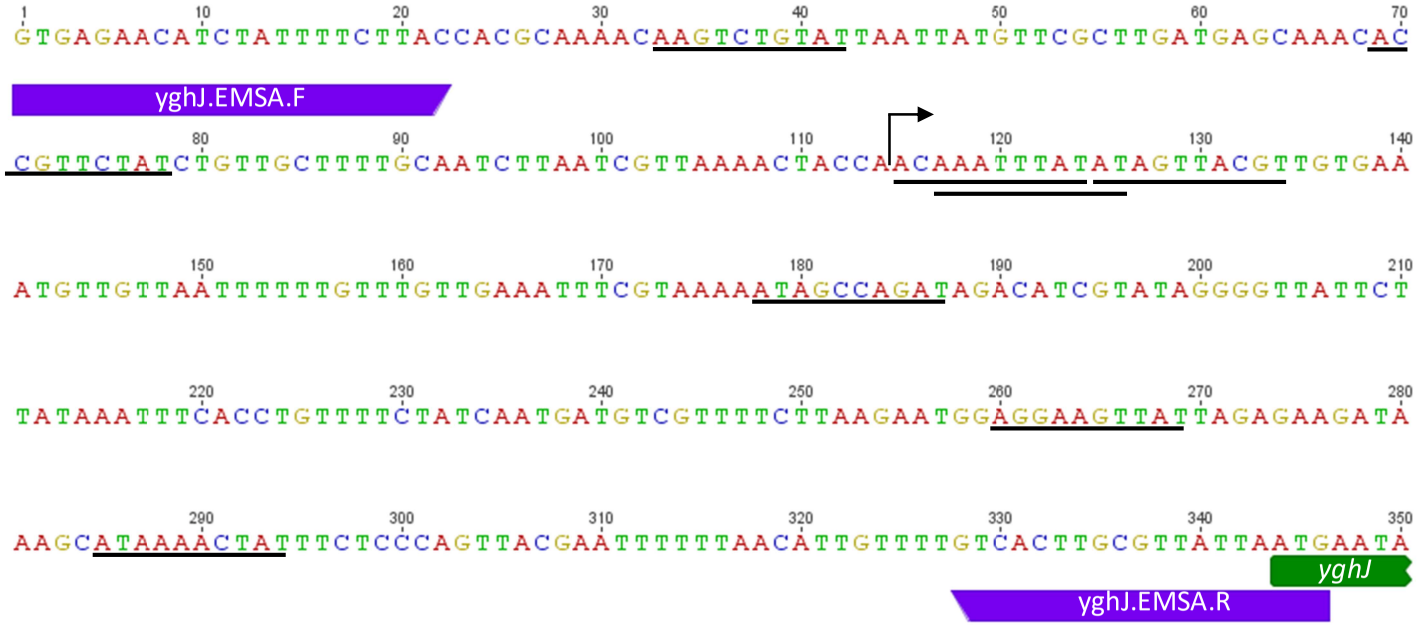


Figure 3.18. Potential Rns binding sites in the promoter region of *yghJ*

Nucleotide sequences, transcription start sites, and binding sites for primers used to amplify the promoter region for EMSA of the *yghJ* regulatory region. Primers are marked in purple, the coding region is marked in green, and the transcription start site is marked with an arrow. Sites matching the minimum Rns consensus sequence RN_6TAT are underlined.

formation by uropathogenic *E. coli* (228). It has been linked to virulence in uropathogenic *E. coli*, where it promotes long-term persistence in the bladder (229). Despite being expressed in intestinal *E. coli*, antigen 43 does not appear to bind directly to intestinal epithelial cells (230).

The gene encoding antigen 43 is phase-variable, switching 'on' or 'off' based on the methylation state of three GATC sequences in the promoter region. When these sites are unmethylated, OxyR is able to bind to them, inhibiting RNA-polymerase from transcribing the genes while also preventing Dam methylation (231). If the GATC sites are methylated by Dam, OxyR is unable to bind and transcription occurs.

I performed β-galactosidase reporter assays on P*agn43-lacZ* fusions in *E. coli* MC4100, but promoter activity was very low, and no Rns effect was observed. To determine if Rns has an effect on *agn43* transcription on top of phase variation, I then performed the same assays in *E. coli* JW3933-3, an *oxyR* deletion mutant strain in which antigen 43 is phase-locked 'on'. In this strain, the promoter was active, but there was still no change in expression in response to Rns.

Two copies of genes for antigen 43 are present on the chromosome of H10407. These are known as *agn43* and *flu*. My mRNA-sequencing results indicated that both of these genes were upregulated in the presence of Rns. The coding region of these genes have 93.9% sequence identity. For this study, I chose the copy annotated as *agn43* as the main focus. The close similarity of these genes means that there is likely to be some inaccuracy in the mRNA-sequencing data analysis, due to the unavoidable difficulties in assigning transcript reads to repeated regions. In the future it may be useful to investigate the effect of Rns on P*flu*, but the high similarity of the regulatory regions (approximately 90% sequence identity) makes it unlikely that there will be any large difference in the results. In neither copy of the gene are there any Rns consensus binding sites upstream, and the promoter regions are not AT-rich, further suggesting that Rns does not bind here. For these reasons, it is likely that Rns has only an indirect effect on *agn43* transcription.

3.3.3.6 ETEC_3214

The high level of upregulation of (approximately 18-fold) and the presence of a potential Rns binding site, "C**A**AAAATT**TAT**C" (figure 3.6), suggest that ETEC_3214 may be a case of direct activation by Rns.

Results from BLAST on the NCBI databases showed that the gene annotated as ETEC_3214 is present in the genomes of multiple strains of *E. coli,* including: H10407; SEC470, isolated from a piglet with diarrhoea in China; ATCC 25922, a serotype O6 reference strain; Nissle 1917, an *E. coli* probiotic strain sold as "Mutaflor"; CFT073, a uropathogenic reference strain; and ABU83972, Di2 and D i14, three isolates from a longitudinal study of UPEC in humans and a dog. No homologues in the databases have any proposed function.

Because it is present in non-pathogenic *E. coli*, ETEC_3214 may not be directly related to virulence. However, it may improve the bacterium's ability to live inside a mammalian host (human, pig, or dog), and/or it may only contribute to virulence when a particular set of other genes are also present. Further studies are required to characterise the function of this protein.

3.3.3.7 Transposases

ETEC_p948_0010, an IS66-family transposase, and ETEC_p948_1060, an ISSfl4 transposase, are located with a large number of other transposases that flank *rns* and *rtr* in p948. This region of the plasmid is a duplication of the end of the *cfaABCE* operon, where a non-functional copy of *rns* is encoded with an ISSfl4 transposase nearby (figure 3.19). There is a strong Rns consensus binding site 297 bp upstream of ETEC_p948_1060 (figure 3.6), which was upregulated 13-fold. This may be direct activation by Rns. ETEC_p948_0010 was also activated, but to a lesser degree (approximately 4.5-fold), and does not have any obvious Rns binding sites within the promoter region, suggesting that the activation observed is an indirect response.

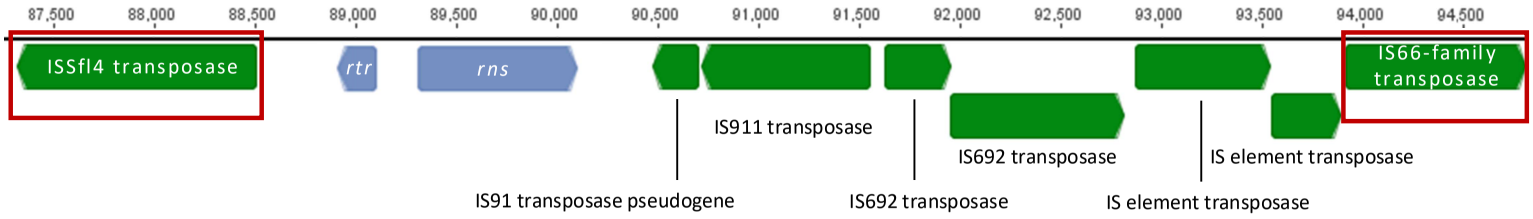


Figure 3.19. Transposases surrounding *rtr* and *rns*

Diagram of the transposases in the genomic region surrounding *rtr* and *rns* in ETEC strain H10407. The transposases which were upregulated in the presence of Rns are boxed in red. Numbering according to the plasmid sequence available from the National Center for Biotechnology Information, accession number NC_017724.1.

3.3.3.8 TraM

TraM is a regulator of F-plasmid mobility that is able to sense bacterial mating and initiate transfer of the IncF plasmid (232). TraM is encoded on both p666 and p948, (these genes share 79.2% sequence identity at the base pair level, and 79.5% identity of amino acids), but only the p666 copy was upregulated by Rns sufficiently to meet our cut-off values for inclusion (table 3.3). The copy of *traM* on p948 was upregulated 2.8-fold in the presence of Rns, with an FDR of 6.21E-3. A potential Rns binding site is located 374 bp upstream of the coding region of *traM* on p666, with an identical sequence 382 bp upstream of the gene on p948. This sequence "G**G**TTAAAT**TAT**T" is very AT-rich, but has guanine instead of adenine at the second position. This substitution is present in 2 of the 11 experimentally determined Rns binding sites. The fact that both copies were activated, albeit to small degree, suggests that this activation may be direct. Further evidence for this is the fact that *traM* is repressed by H-NS (233). The increased expression of *traM* and genes encoding transposases suggest that a secondary role of Rns may be to increase gene mobility. This makes some intuitive sense, as virulence genes are often mobile elements which arrive in the genome through horizontal gene transfer. The intestine is a rich environment with many different genes, strains, and species of bacteria living in close proximity, sharing the same ecological niche, and thus is a prime location for horizontal gene transfer (234).

3.3.3.9 CstA

The carbon starvation protein, CstA, is a peptide transporter regulated by the cAMP-CRP complex (235), and was upregulated 6.6-fold in H10407 expressing Rns. The *cstA* gene is not related to virulence, and the lack of Rns consensus binding sites in the promoter region indicates that this is unlikely to be a direct target of Rns. However, there is a region of low GC-percentage (33%) from approximately 190 to 300 bp upstream of the coding region, and there are multiple sites within this region that fit the minimum Rns consensus sequence "RN_6TAT". As such, further investigation is required to determine if this gene is targeted by Rns directly.

3.3.3.10 OmpC

The mRNA-sequencing results show a 5.8-fold activation of *ompC* by Rns. OmpC is an outer membrane protein involved in the non-specific transport of small hydrophilic compounds (236). OmpC and OmpF are two major outer membrane porins of *E. coli*, and the ratio of the two varies depending on the environment. OmpF is larger and less stringent than OmpC (237). OmpC is the dominant porin when *E. coli* is in the intestinal tract due to the high osmotic pressure and the abundance of potentially harmful factors, so it would make sense, teleologically speaking, for *ompC* to be under Rns control (238). There are two potential Rns binding sites in the *ompC* promoter region, at 227 bp and 259 bp upstream of the coding region (figure 3.6), so it is possible that Rns has a direct effect on the transcription of *ompC*. It has been suggested that H-NS represses OmpC production indirectly through the activity of stationary-phase sigma factor, RpoS (239). However, Rns had no effect on the expression of RpoS (FDR=0.33) in my mRNA-sequencing experiments, apparently ruling this out as a mechanism by which Rns upregulates *ompC*.

3.3.3.11 Adenine phosphoribosyltransferase

Adenine phosphoribosyltransferase, an enzyme that converts adenine to adenosine-monophosphate (AMP), which can then be converted to ATP, was upregulated 4.2-fold in the presence of Rns. This gene is highly unlikely to be a direct Rns target, as the upstream region is not AT-rich, contains no Rns consensus binding sequence, and no RN_6TAT sequences.

3.3.3.12 YiiS

YiiS is a part of the stress response in *E. coli* (139). In my mRNA-sequencing experiments, there was no significant change in *yiiS* expression in the presence of Rns, despite past publications having identified *yiiS* as an Rns target (117). In a study by Munson *et al*, DNA from ETEC strain C921b-1 (a strain which encodes Rns and CS1) was passed through an MBP-Rns affinity column to isolate DNA fragments that bound to Rns. These isolated fragments were then tested for their ability to bind to MBP-Rns by using EMSA. The fragment with the strongest MBP-Rns binding was

sequenced and found to include a 227 bp sequence from the promoter region of *yiiS*. β-galactosidase reporter assays performed in this study showed 1.6 to 2.3-fold more promoter activity in the presence of Rns. In contrast, my mRNA-sequence data showed no significant increase in *yiiS* expression in response to Rns (FDR = 0.55). The gene *yiiS* in encoded on the chromosome of C921b-1, the sequence of which is not available. (Only the virulence plasmid pCoo has been sequenced from this strain). Nevertheless, the sequence of *yiiS* and the surrounding region is identical in the published sequences of *E. coli* H10407 and *E. coli* MC4100, and this is the sequence I used for my analysis. The region of DNA Munson *et al* (117) used for their experiments has a GC-percentage of 34%, making it a relatively AT-rich region. Within this region, no strong Rns binding sites, matching the sequence BRWWWDHDTATY, are present. Using instead the minimum Rns consensus sequence, "RN_6TAT", two sequences can be found: "A**A**ATAAAA**TAT**G" and "A**A**GAGATA**TAT**T" (figure 3.20). Given the results obtained by Munson *et al*, (117) it is likely that Rns has some ability to activate this promoter under certain conditions, but that these results were not consistent with our mRNA-sequencing experiments. Alternatively, the gene sequence within the regulatory region of *yiiS* in C921b-1 may contain mutations which allow Rns to act in this region. Another possibility is that a separate mechanism of regulation is obscuring Rns regulation of *yiiS* in *E. coli* H10407, but the genetic region used for EMSA and *lacZ* reporter studies by Munson *et al* did not include the sequence required for this higher level of regulation to occur.

3.3.4 Gene repression by Rns

3.3.4.1 NlpA

Only one prior study has reported gene repression by Rns (118). In this study, Rns was found to bind to the *nlpA* gene and prevents the formation of the RNA-polymerase complex at the TSS (118). Thus Rns can mediate gene repression by binding to a positive regulatory region, rather than by overcoming H-NS mediated regulation, the typical method of gene activation. NlpA is involved in the production of outer membrane vesicles, which may be one method of LT delivery (140, 240).

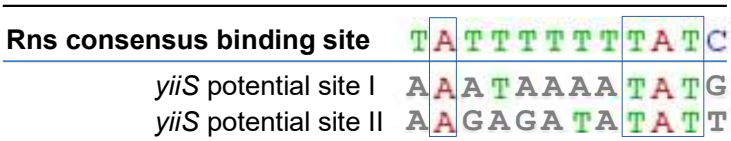


Figure 3.20. Potential Rns binding sites in the promoter region of *yiiS*

Consensus sequence of Rns binding sites and potential Rns binding sites in the promoter region of *yiiS*. The RN_6TAT sequence that makes contact with the Rns protein is boxed. Nucleotides which match the consensus sequence are shown in colour; while mismatches are shown in grey.

Interestingly, this suggests that Rns may act to reduce the delivery of LT to host enterocytes. This could be related to the temporal spacing of pathogenesis, as Rns acts to increase colonisation before toxin delivery begins. In my mRNA-sequencing results however, Rns had no effect on transcription of *nlpA* (FDR = 0.89). Rns may have some ability to repress transcription from this promoter under conditions which were not consistent with our mRNA-sequencing experiments.

3.3.4.2 Motility and chemotaxis

Many of the genes downregulated in the presence of Rns are part of three genetic regions within *E. coli* that control the production and motion of flagella. Five of these genes; *tar*, *tap*, *cheA*, *flgA* and *fliC*, are downregulated greater than 4-fold in the presence of Rns, as shown and discussed in section 3.2.3.3, table 3.4. An additional 20 genes from these regions were downregulated to a lesser extent (table 3.5). The layout of these regions is depicted in figure 3.21.

Flagellar operons are classified as class 1, class 2, or class 3 (241). The only class 1 promoter in *E. coli* controls the transcription of *flhDC*. FlhDC then activates class 2 promoters, which leads to the production of FliA, a sigma-factor that activates class 3 promoters. This step-by-step process ensures the efficient construction of flagella. Proteins involved in constructing the basal body and hook of the flagella are produced first, and only once that machinery is in place do the bacteria produce proteins that make up of the shaft of the flagella and those which control its operation (241).

H-NS is involved in the regulation of *flhDC*, the master regulator that controls the expression of the many flagellar and other operons (242, 243). This interaction is complex, with H-NS either repressing or activating transcription depending on whether it is bound upstream or downstream of the ATG translation start site (243). H-NS also represses the transcription of *hdfR*, a LysR family regulator which itself represses *flhDC* (244). Finally, H-NS is thought to play a structural role, interacting with the flagella rotor FliG subunit (245). Through these mechanisms, the net effect of H-NS is to increase motility of the bacterium (243).

Table 3.5. Genes relating to motility downregulated in the presence of Rns

Gene	Function	Repression (fold change)	False discovery rate
flgM	Negative regulator of flagellin synthesis	3.3	1.27 E-03
flgA	Flagellar basal body P-ring protein	5.8	5.26E-05
flgB	Flagellar basal-body rod protein	2.8	1.39 E-03
flgC	Flagellar basal-body rod protein	2.4	2.36 E-03
flgD	Basal-body rod modification protein	2.2	5.56 E-03
flgE	Flagellar hook protein	2.4	4.22 E-03
flgK	Flagellar hook-associated protein 2	2.3	6.04 E-03
flgL	Flagellar hook-associated protein 3	2.4	4.94 E-03
flhA	Flagellar biosynthesis protein	2.5	3.24 E-03
tar	Methyl-accepting chemotaxis protein II	6.0	6.72E-05
tap	Methyl-accepting chemotaxis protein IV	7.1	3.41E-04
cheA	Chemotaxis sensor kinase protein	5.4	8.30E-05
cheW	Chemotaxis protein	3.2	5.35 E-03
motA	Chemotaxis protein	2.9	6.09 E-03
fliA	RNA-pol sigma factor for flagellar operon	3.5	1.04 E-04
fliC	Flagellin	5.2	6.52E-05
fliD	Filament cap	3.9	5.15 E-05
fliT	Flagellar protein	3.1	5.21 E-03
fliE	Flagellar hook-basal body complex protein	3.0	4.94 E-03
fliF	Flagellar M-ring protein	2.9	1.39 E-03
fliG	Flagellar motor switch protein	2.4	3.81 E-03
fliH	Flagellar assembly protein	2.8	3.81 E-03
fliI	Flagellum-specific ATP synthase	2.2	9.88 E-03
fliK	Flagellar hook-length control protein	2.6	2.86 E-03
fliL	Flagellar protein	2.4	7.20 E-03

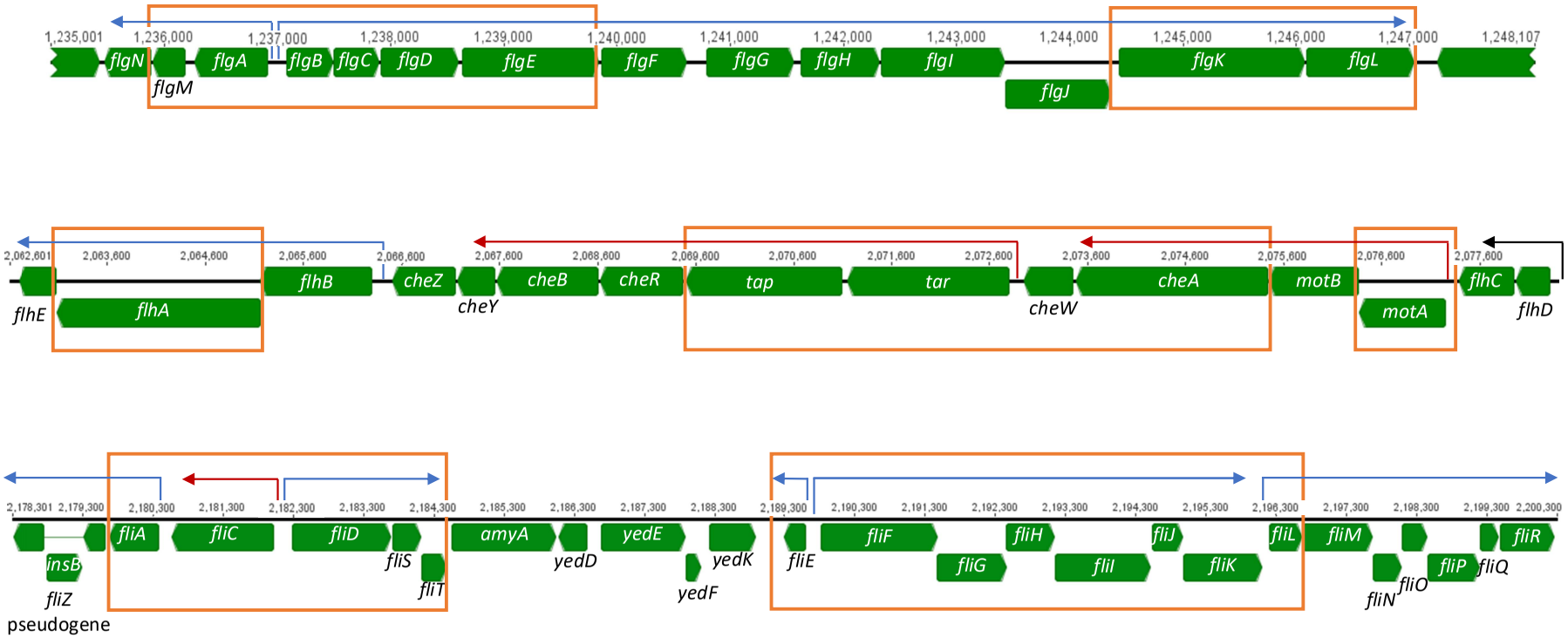


Figure 3.21 Downregulation flagellar genes by Rns

The figure shows the three genomic regions encoding the flagella of *E. coli*. Class one operons are indicated with a black arrow, class two operons are indicated with a blue arrow, and class three operons are indicated with a red arrow. Genes which were downregulated in the presence of Rns are boxed in orange.

Given this, it is tempting to propose a model where Rns removes H-NS repression of *hdfR*, thus inhibiting the transcription of *flhDC*. However, neither *flhDC* nor *hdfR* were differentially expressed in the presence of Rns (FDR=0.33 and 0.94, respectively). None of the flagellar operons are AT-rich, so it is unlikely that Rns can bind to these regions directly. Instead, a more complex mechanism or intermediary regulator is likely to be involved.

I have discussed above how Rns upregulates the expression of EtpA, and that this adhesin contributes to ETEC colonisation of the small intestine by binding to flagella on the bacterial cell surface. It appears paradoxical, then, that Rns would cause a downregulation of flagella. However, the downregulation of flagella is a lot weaker than the upregulation of EtpA, and so it is likely that the surface expression of flagella remains sufficient for EtpA adherence function to occur. The downregulation seen here of genes relating to motility may be the beginning of a response which becomes stronger as colonisation progresses, so that EtpA is able to bind to flagella before they degrade.

3.3.4.3 YhaO

YhaO is an inner membrane protein that is common in *E. coli* and its close relatives (246). The upstream region of *yhaO* contains no Rns consensus binding sequences, no "RN_6TAT" sites, and the GC-percentage higher than 40%, indicating that this is unlikely to be a direct effect of Rns on gene transcription.

3.3.4.4 ETEC_3147: L-asparaginase II

L-asparaginase II is an enzyme which converts L-asparagine into aspartic acid. A potential Rns binding site is located 142 bp upstream of the coding region of ETEC_3147 (figure 3.7). Further investigation is required to determine if Rns binds to this site to inhibit transcription directly.

3.3.4.5 ETEC_1274 and ETEC_1275

ETEC_1274 has homologues throughout the *Enterobacteriaceae*, but none encode a protein with a known function. ETEC_1275 encodes a 114 amino acid protein with homology to OmpA (220).

These genes are repressed approximately 4.5-fold in the presence of Rns. A potential Rns binding site is located upstream of the coding region of ETEC_1275, and ETEC_1274 is encoded immediately downstream, most likely transcribed in an operon from the ETEC_1275 promoter. A putative binding site, with the sequence "T**G**ATTTTT**TAT**", contains many thymine residues, which is generally a marker of a strong Rns binding site, but it has a guanine in place of adenine at the "**R**N_6TAT" binding sequence. This substitution is present in only two of the 11 experimentally confirmed Rns binding sites (figure 1.4). The binding site itself is located 447 bp upstream of the coding region of ETEC_1275. In the only confirmed instance of Rns gene repression, Rns acts by binding to the ATG translational start site. It is unclear how Rns could act as a transcriptional repressor while binding further upstream. Further investigation is required to confirm the nature of Rns repression of this site.

3.3.4.6 ETEC_2342: Ferredoxin-type protein

Ferredoxin is an electron transport protein involved in cell metabolism pathways. There are two strong Rns binding sites approximately 200 bp upstream of the coding region of ETEC_2342, and a weaker potential site 344 bp upstream (figure 3.7). Further investigation is required to determine if this is a case of direct repression by Rns.

3.3.5 Bicarbonate

Bicarbonate ions are released in high concentrations by the duodenum to counteract the acid secretion of the stomach. RegA of *C. rodentium* appears to use bicarbonate ions as a localisation signal to detect its position within the host. The mechanism of this interaction is that in the absence of bicarbonate, the N-terminal arm of RegA occludes its DNA-binding domain, making it unable to bind to DNA and activate transcription from its target promoters. The binding of bicarbonate ions as a co-factor triggers a change in conformation of RegA, which results in the N-terminal arm releasing the DNA binding domain (figure 3.1). In the case of RegA, this auto-inhibition is quite strong, and RegA has very little effect in the absence of bicarbonate. Rns is also activated by bicarbonate ions,

but to a lesser extent. In most cases, the addition of 45 mM bicarbonate approximately doubled the effect of Rns.

In the case of P*yghJ*, bicarbonate increased promoter activity independently of Rns. This was only a 1.5-fold increase, but was statistically significant and was seen consistently with or without Rns. High levels of bicarbonate are a good indicator of being in the small intestine, and it is not surprising that multiple regulatory systems may have evolved to use it. mRNA-sequencing results from our laboratory suggest that bicarbonate has a large effect on the *E. coli* H10407 transcriptome, including activating transcription of the heat-stable enterotoxin genes, *hstA* (3.6-fold, FDR=2.69E-4) and *hstB* (6.1-fold, FDR=1.15E-3). These ETEC enterotoxins are major virulence factors, which do not appear to be regulated by Rns. These findings suggest that there is another virulence regulation system in ETEC that may also respond to bicarbonate.

3.3.6 Future lines of investigation

3.3.6.1 DNase I footprinting and uracil interference assays

EMSA studies are able to demonstrate binding of the Rns protein to a segment of DNA approximately 300 bp long. In the promoter region of *etpB* are two potential Rns binding sites within the sequence of DNA used for EMSA (figure 3.15). In the promoter region of *yghJ*, where the demonstrated binding by Rns was weak, no strong Rns consensus binding sequence was found. DNase I footprinting would allow us to home in on the specific site of Rns binding to the DNA of each of these promoters (247). Uracil interference assays can then determine which individual thymine residues make contact with Rns (248). These assays will be particularly useful in identifying binding sites in P*yghJ*, where we were unable to predict Rns binding sites by analysis of the sequence alone.

3.3.6.2 β-galactosidase reporter assays in H10407

During this study, β-galactosidase reporter assays were performed using promoter-*lacZ* fusion plasmids, transformed into *E. coli* MC4100. It is possible that certain factors that are specific to *E. coli* H10407 may affect the activity of Rns, explaining some of the differences seen between the

results of mRNA-sequencing and β-galactosidase reporter assays. Transforming the promoter-*lacZ* fusions into H10407 and repeating the β-galactosidase reporter assays in the natural host of Rns may yield slightly different results.

3.3.6.3 Phenotype assays in H010407Δrns

Haemagglutination assays demonstrated that Rns is vital for CFA/I production and function. Having identified Rns target genes, it remains to be seen how much the phenotypes conferred by these genes will be affected when Rns is inhibited or removed from the genome. Assays to detect the activity of YghJ, EtpBAC, and other newly discovered targets should be performed in H10407 and H10407Δ*rns*. A loss of phenotype in H10407Δ*rns* will demonstrate that the removal of Rns activation of a particular target gene reduces its activity. If complementing H10407Δ*rns* with the target gene, transcribed from a constitutive promoter, restores the phenotype, this will show that the loss of phenotype is due primarily to the loss of Rns activation of the individual target gene, and not due to the decreased expression of other Rns targets.

YghJ activity can be measured in multiple ways. The contribution it makes to adhesion can be demonstrated by examining ileal sections of mice infected with H10407, as performed by Luo *et al* during their investigation of the role of YghJ (80). They discovered that, without YghJ, H10407 were less effective at forming close interactions with mouse intestinal epithelium. However, CFA/I is also controlled by Rns and may also effect H10407 interactions with enterocytes. Thus, the lack of CFA/I expression by H10407Δ*rns* may render H10407 unable to bind mouse intestinal epithelium, regardless of the presence or absence of YghJ, (although there is no evidence that CFA/I binds well to the mouse intestine). A more specific way to measure YghJ activity is to assay for the ability of H10407 strains to degrade MUC2 (80). This can be measured by incubating H10407 strains in media containing MUC2 and comparing the build-up of MUC2 on the ETEC cell surface, or by using treated supernatant from H10407 strains to degrade MUC2 from LS174T cells (80).

In analysing the loss of the EtpBAC associated phenotype in H10407Δ*rns*, I expect difficulties will arise in attempting to distinguish the effects of reduced EtpBAC expression from those of reduced CFA/I expression, as both of these are important adhesins. Instead of measuring a phenotype, we could directly observe EtpA production by using immunogold staining and electron microscopic examination to look for EtpA adhering to the tips of flagella of *E. coli* H10407, H10407Δ*etpBAC* and H10407Δ*rns*, a technique employed by Roy *et. al.* (70). In addition, *cfaABCE* deletion mutants can be used as the base strain for these experiments, so that CFA/I is not present in any of the test strains.

3.3.6.4 Novel Rns target genes

For the purposes of this study, I have focused on the newly identified Rns targets that have at least a putative connection to virulence. I identified many genes which were upregulated in the presence of Rns, but did not investigate them further. I also did not study downregulated genes. These genes are all possible subjects of further study, using the assays described in this chapter to measure direct Rns binding to and activation of promoter regions, as well as the potential loss of phenotype in H10407Δ*rns*.

3.3.6.5 The effect of bicarbonate ions

This work has hinted at a larger role bicarbonate ions may play in the modulation of the ETEC transcriptome. My data suggests that Rns is not the only virulence gene regulator in ETEC which responds to bicarbonate. H10407 encodes two copies of the heat-stable toxin, the 'human'-type STh, encoded by *hstA*, and the 'porcine'-type STp, encoded by *hstB*. Both genes were upregulated in the presence of bicarbonate, despite Rns having no effect on the expression of either gene. The Rns target YghJ was also significantly affected by bicarbonate ions, independently of Rns. A potential future line of investigation could be to identify other virulence regulators encoded by ETEC, which also respond to bicarbonate and may control the expression of other virulence factors, such as the enterotoxins.

3.3.7 Conclusion

Having developed a wealth of data on the Rns regulon in H10407, I believe that Rns is a legitimate drug target for the prevention and/or treatment of ETEC-induced diarrhoea. Not only CFs, but also EtpB and YghJ, are directly transcriptionally activated by Rns. Moreover, Antigen 43, an autotransporter with a potential role in virulence, is also indirectly upregulated in the presence of Rns.

In the next chapter, I identify small chemical compound inhibitors of Rns, which may be developed into a therapeutic agent against ETEC-induced diarrhoea.

4 A small molecule inhibitor of Rns

4.1 Introduction

ETEC-induced diarrhoea is a major health concern, both for travellers and young children in developing countries. Although antibiotics can be used to treat ETEC infections, antibiotic resistance in ETEC is widespread, with recent studies reporting that at least 39% of isolates exhibit multi-drug resistance (169-174). Overuse of antibiotics contributes to the growing problem of resistance, and antibiotic treatment has side-effects, including diarrhoea, and significantly disrupts commensal microbial communities (249).

It is well established that ETEC requires a functional CF to cause disease, and, in some cases, CF-based immunisation is protective against infections with ETEC strains which carry the same CF (188, 250, 251). However, the wide antigenic variety of CFs expressed by human ETEC makes it difficult to produce a broadly protective vaccine based on these antigens. Instead of targeting individual CFs for inhibition, I have focused on the transcriptional activator Rns, which controls the production of around half of all CFs, as well as several accessory virulence determinants (112, 113, 116). Because ETEC cannot colonise the small intestine and deliver enterotoxins to induce diarrhoea without the expression of a CF, a chemical compound which is able to inhibit the activity of Rns could render many strains of ETEC non-virulent. Inhibitors of virulence activators have been shown to prevent or treat intestinal infection in mice infected with *C. rodentium* and in a mouse model of *V. cholerae* (194, 196). The global virulence activators of these pathogens, RegA and ToxT respectively, are closely related to Rns, and we hypothesised that a similar strategy to those used to discover inhibitors of these regulators may be effective in the development of pharmaceuticals that target ETEC.

In this chapter, I describe the use of a high-throughput *in vitro* screening (HTS) method which I used in combination with computer modelling to identify a chemical compound that inhibits

the activity of Rns. I also investigated other potential inhibitor compounds, which could be developed into novel antimicrobials for the prevention and treatment of ETEC-induced diarrhoea.

4.2 Results

4.2.1 The effect of the RegA inhibitor, Regacin, on Rns-mediated activation of the *cfaA* promoter, P*cfaA*

Previous work in our lab revealed a compound that inhibits the activity of RegA, an Rns homologue that regulates virulence in the mouse pathogen *C. rodentium (194)*. Because of the similarity between Rns and RegA, we wanted to see if this inhibitor, designated Regacin, is also active against Rns. To this end, β-galactosidase reporter assays were performed using a P*cfaA-lacZ* fusion, co-transformed into *E. coli* MC4100(pRns) (see methods section 2.13). DMSO at a final concentration of 1% in growth media was used as a solvent for Regacin. P*cfaA* activity in the presence of 1% DMSO alone was 1,366 ± 73 MU (mean ± SEM, n=2), and was not significantly reduced in the presence of 20 µM or 100 µM Regacin (1,221 ± 283 MU and 1,370 ± 88 MU, respectively) (figure 4.1).

4.2.2 Screening for potential Rns inhibitors

To identify novel Rns inhibitors, I employed two complementary methods to screen chemical compounds: a high-throughput *in vitro* screen, and a virtual screen using computer modelling to predict compounds that may bind and inhibit the Rns protein. The latter was performed by Dr. Jessica Holien at the St Vincent's Institute of Medical Research. Figure 4.2 is a flowchart summarising the outcome of these screens, which are described in the following sections.

4.2.2.1 High-throughput in vitro screen for an Rns inhibitor

To search for an Rns inhibitor, we screened the Chembridge Microformat library (ChemBridge Corp.), a commercial collection of small chemical compounds with drug-like properties, for their ability to inhibit Rns-mediated activation of P*cfaA*. Screening was performed in 96-well

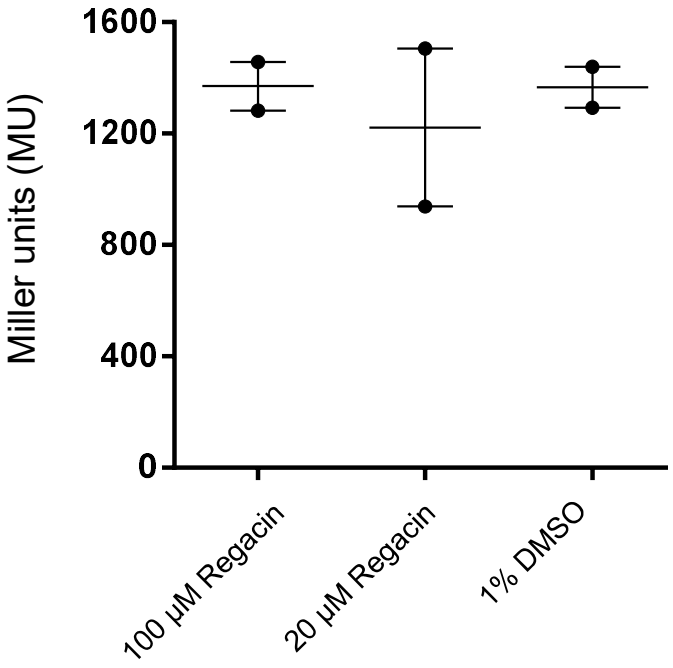


Figure 4.1. Effect of Regacin on Rns-mediated activation of *PcfaA*

Results of β-galactosidase reporter assays on the activation of P*cfa* by Rns in the presence of 20 µM and 100 µM Regacin in 1% DMSO, or a DMSO control, showed that Regacin had no effect on Rns-mediated activation of P*cfaA*. Error bars indicate standard error of the mean, n=2.

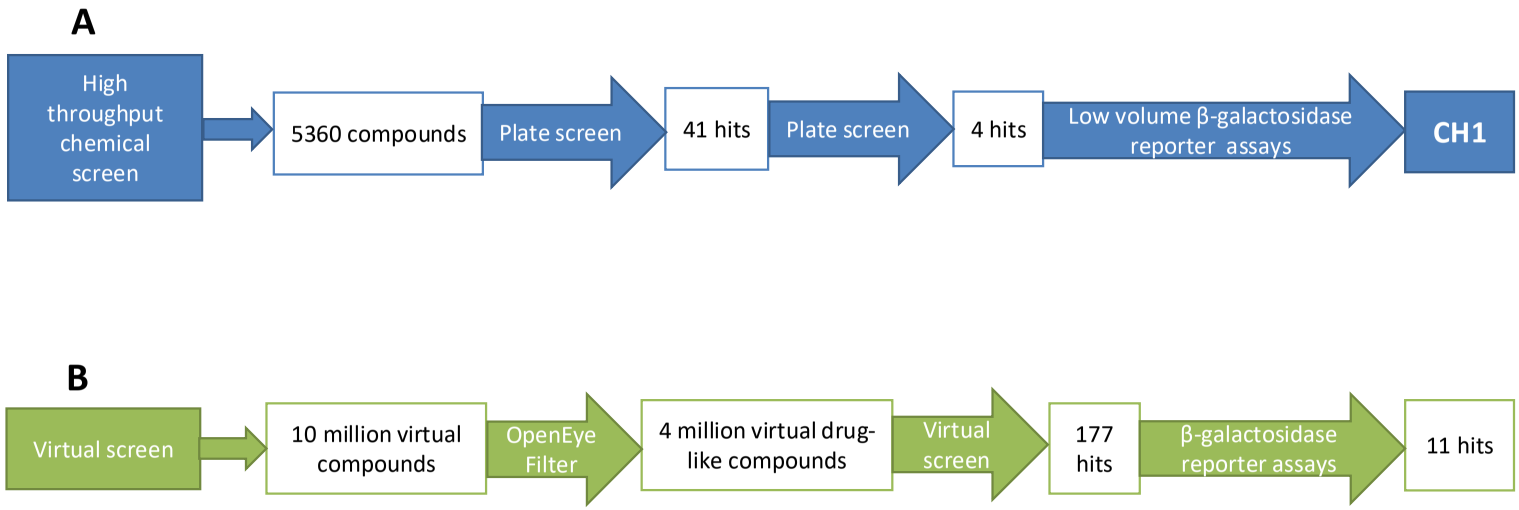


Figure 4.2. Summary of the screening methods used to identify inhibitors of Rns

The flowchart depicts the use of two separate screening methods to identify potential Rns inhibitor compounds. A) 5360 compounds were screened in a high-throughput chemical screen, 41 of which demonstrated some ability to inhibit Rns activity without affecting the control. Four of these hits were reproducible, and were tested again in low volume β-galactosidase reporter assays. One of these was selected as the initial hit compound, CH1. B) A virtual library of approximately 10 million compounds was filtered to approximately 4 million compounds with drug-like properties using Filter 2.1.1 (OpenEye Scientific). From this subset, 189 compounds were selected to be tested in vitro after virtual screening using Sybylx 2.0 (Certara), Fred 3.0.1 (OpenEye Scientific), and Vida 4.2.1 (OpenEye Scientific). 11 of these compounds showed some ability to inhibit Rns, but none of these was as effective as CH1, and they were not investigated further.

microtitre trays, using the P*cfaA-lacZ* fusion co-transformed into *E. coli* MC4100(pRns), and the method described in section 2.18.1. A control plate using MC4100 carrying a reporter system for an unrelated regulator/target promoter-*lacZ* pair, MrkH and P*mrkA-lacZ*, was included, to ensure that the inhibitory effects of the selected compounds were specific for Rns (252). MrkH activates the transcription of type 3 fimbriae in *Klebsiella pneumoniae*, and is not related to Rns (253).

5,360 compounds from the Chembridge library were screened using this method. Of these, 41 compounds, at a concentration of 100 µM inhibited the Rns-mediated activation of *PcfaA* to below 67% of a DMSO control, but not the control regulator. Re-testing these 41 compounds reduced the list to only four repeatable hits, which we designated D5, F2, A6, and E7.

The ability of these compounds to inhibit Rns-mediated activation of P*cfaA* was then tested in low volume β-galactosidase assays, in the presence of bicarbonate (figure 4.3). P*cfaA* activity in the reporter strain was measured after growth in the presence of 100 µM of each compound in a final concentration of 1% DMSO, or a DMSO control. P*cfaA* activity in the absence of Rns was measured using the P*cfaA-lacZ* fusion co-transformed into *E. coli* MC4100(pACYC184).

Only compound D5 was found to reliably inhibit the action of Rns. P*cfaA* activity in the absence of Rns was 247 ± 1 MU (mean ± SEM, n=2). In the presence of Rns, P*cfaA* activity was 1,915 ± 90 MU (n=5). Compounds F2, A6 and E7 had no significant effect on this level of expression; with promoter activity of 1,543 ± 119 MU (n=4), 1,763 ± 203 MU (n=4), and 1,875 ± 4 MU (n=2), respectively. Compound D5, hereafter referred to as CH1, decreased promoter activity approximately 3-fold, to 601 ± 191 MU (n=4, p = 0.0003).

4.2.2.2 The construction of a computer model of Rns

Computer modelling allows extremely large libraries of chemical compounds to be screened for activity *in silico* before they are tested experimentally. To screen compounds for their ability to bind Rns, it was necessary to obtain a 3-dimensional model of the protein. Ideally, a crystal structure is required as the basis for the computer modelling of a protein structure. Since the crystal structure

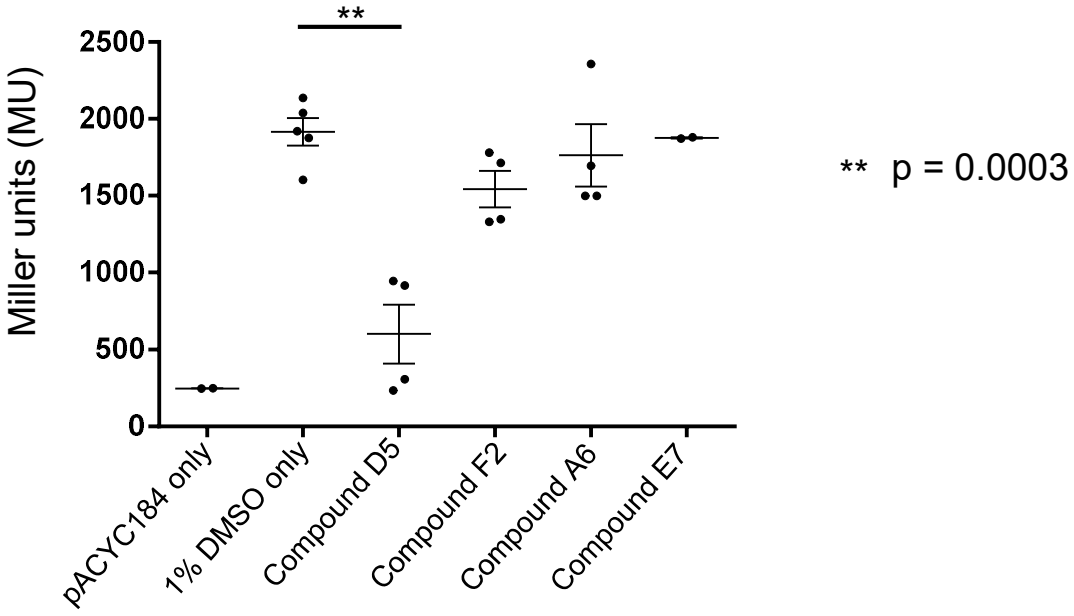


Figure 4.3. In vitro evaluation of compounds identified in a screening assay as having inhibitory effects on Rns

Four hit compounds from the high-throughput chemical screening, designated D5, F2, A6, and E7, were tested for their effect on Rns-mediated activation of P*cfaA* in a β-galactosidase reporter assay, compared to a DMSO control and to the activity of P*cfaA-lacZ* in the absence of Rns. Error bars show standard error of the mean, n ≥ 2. p value calculated using Student's t-test.

for Rns has not yet been solved, we based the model on that of a close homologue. Work performed by Dr. Holien of the St. Vincent's Institute of Medical Research determined that ToxT, the lead virulence regulator in *V. cholerae*, was the closest Rns homologue with a solved crystal structure, and a model of Rns was created based on this homology. This model of Rns, illustrated in figure 4.4, was used for all *in silico* experiments.

4.2.2.3 Identification of potential Rns inhibitors using a virtual screen

Dr. Holien of the St. Vincent's Institute of Medical Research then filtered virtual libraries containing approximately ten million chemical compounds down to approximately 4 million compounds with drug-like properties, using Filter 2.1.1 (OpenEye Scientific). Molecular modelling programs Sybylx 2.0 (Certara) and fred_receptor (OpenEye Scientific) were used to identify potential binding pockets within the Rns 3D model. Two major pockets were selected, one central to the protein (depicted in magenta in figure 4.4), and one between the dimerisation and DNA-binding domains (depicted in green in figure 4.4). Fred 3.0.1 (OpenEye Scientific) was used to dock each compound into each binding site and rank the energy of each interaction. The top 500 hits were visually inspected in Vida 4.2.1 (OpenEye Scientific) and from these 177 compounds chosen for *in vitro* testing by Dr. Holien, based on the results and the commercial availability of the compounds.

These 177 compounds were screened for their ability to inhibit Rns-mediated activation of P*cfaA.* Compounds were used at a concentration of 100 µM in a modified β-galactosidase reporter assay (described in section 2.16.2), using the P*cfaA-lacZ* fusion co-transformed into *E. coli* MC4100(pRns) (table 4.1). This initial screen identified 11 compounds that inhibited Rns-mediated activation of P*cfaA* by around half; five of these below 40% of a DMSO control. The top five compounds were retested using the same assay, but none of these results could be consistently repeated. I therefore focused on CH1, identified by HTS, for the following experiments.

4.2.3 The effect of CH1 on the growth of *E. coli* H10407

The growth of *E. coli* H10407 was measured in the presence of 1% DMSO, containing

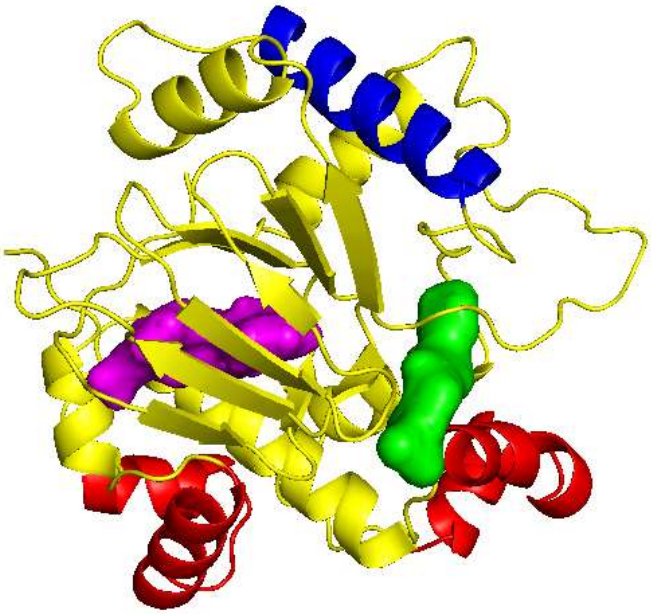


Figure 4.4. Pockets in the Rns structure targeted for chemical compound docking

A model of Rns with two potential binding pockets highlighted. Rns is in yellow with the DNA binding helices in red, and the putative dimerization domain in blue. In magenta is a central pocket, and the in green is the pocket between the dimerization and DNA binding domains of Rns. These pockets were used as targets in the virtual screen for potential Rns inhibitors.

Table 4.1. Inhibition of Rns activation of P*cfaA* by chemical compounds

Compounds which reduced the Rns-mediated activation of P*cfaA* to less than 40% of a DMSO control are boxed in red.

Compound #	Inhibition (Fold change)
1	0.94
2	1.09
3	0.67
4	0.39
5	0.80
6	0.53
7	0.96
8	0.86
9	0.97
10	0.66
11	0.53
12	0.63
13	0.81
14	1.33
15	1.26
16	1.15
17	1.39
18	1.15
19	1.17
20	1.16
21	1.33
22	0.86
23	0.73
24	1.39
25	1.37
26	1.45
27	1.27
28	1.54
29	1.30
30	1.40
31	1.09
32	1.04
33	0.13
34	1.15
35	1.00
36	0.94
37	0.76
38	0.64
39	1.16
40	0.73
41	0.78
42	1.31
43	0.82
44	1.33
45	0.13

Compound #	Inhibition (Fold change)
46	1.05
47	0.80
48	1.02
49	0.10
50	0.64
51	0.74
52	0.81
53	0.79
54	0.74
55	0.82
56	0.84
57	0.71
58	0.52
59	0.65
60	0.57
61	1.42
62	1.44
63	1.27
64	1.42
65	1.19
66	1.26
67	1.13
68	1.07
69	1.11
70	1.55
71	0.63
72	1.15
73	2.41
74	1.23
75	1.21
76	1.12
77	1.35
78	1.56
79	1.80
80	1.35
81	1.29
82	1.22
83	1.64
84	1.37
85	0.88
86	0.97
87	0.87
88	0.83
89	0.85
90	0.89

Table 4.1. (continued)

Compound #	Inhibition (Fold change)
91	0.88
92	0.52
93	1.01
94	0.97
95	1.00
96	0.92
97	0.53
98	0.96
99	0.77
100	0.70
101	0.69
102	0.59
103	0.93
104	0.90
105	1.16
106	0.82
107	0.93
108	1.22
109	0.59
110	0.79
111	0.45
112	1.00
113	0.66
114	0.95
115	0.67
116	0.72
117	0.41
118	0.67
119	1.03
120	0.59
121	0.13
122	0.45
123	1.18
124	0.76
125	1.00
126	1.05
127	0.78
128	1.23
129	1.13
130	0.65
131	1.05
132	0.96
133	0.96
134	0.99
135	1.11

Compound #	Inhibition (Fold change)
136	1.11
137	0.87
138	0.60
139	0.99
140	0.41
141	0.84
142	0.81
143	0.41
144	0.42
145	2.16
146	2.86
147	2.93
148	2.73
149	3.53
150	2.10
151	3.69
152	3.38
153	3.96
154	4.09
155	2.10
156	1.40
157	1.96
158	0.94
159	0.65
160	1.19
161	0.97
162	1.06
163	0.62
164	0.69
165	0.91
166	0.77
167	0.70
168	1.89
169	0.85
170	0.85
171	0.76
172	0.65
173	0.52
174	0.87
175	0.61
176	0.71
177	0.89

200 µM, 30 µM, or no CH1, as detailed in sections 2.9.1. OD_{600} readings were taken every half hour for 8 hours, and the growth kinetics were plotted. CH1 had no effect on the growth of H10407, at either concentration (figure 4.5).

4.2.4 Determination of the IC_{50} of CH1 for Rns-mediated activation of P*cfaA*

An IC_{50} is the concentration at which an inhibitor exerts 50% of its maximum effect. The IC_{50} for CH1 on Rns-mediated activation of *PcfaA* was measured using a modified β-galactosidase reporter assay (see methods sections 2.16.2) at a range of inhibitor concentrations (figure 4.6). Because we wanted to investigate the effect of CH1 on Rns-mediated gene activation more broadly, rather than focus on a single target promoter, we also investigated the effect of CH1 on the Rns-mediated activation of P*etpB*. Plasmids carrying P*etpB-lacZ* or P*cfaA-lacZ* were transformed into *E. coli* MC4100(pRns). Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points, and calculate the IC_{50}.

The result showed that CH1 inhibited Rns-mediated activation of P*cfaA* and P*etpB* with an IC_{50} of 1.02 µM and 2.96 µM, respectively. During these experiments, I observed that although CH1 was soluble in 100% DMSO at high concentrations (over 40 mM), it precipitated out of solution after overnight incubation in *E. coli* cultures at a concentration of 50 µM in 1% DMSO.

4.2.5 A model of CH1 binding to Rns

In work performed by Dr. Holien, the Sybylx2.1 (Certara) 'Surflex-Dock Geom mode' feature was used to dock CH1 into the potential binding pockets (protomols) of Rns. The SiteID search identified 9 potential pockets, including the two used to initially screen the chemical library for putative inhibitors. Three of these were sufficiently large to fit CH1 and were accessible from outside the protein. These included the two pockets used in the virtual screen, and another pocket near the dimerisation domain.

Two major qualities were considered when determining the likelihood of each binding prediction. These were the highest scoring solution, and the cluster percentage. The solution score is

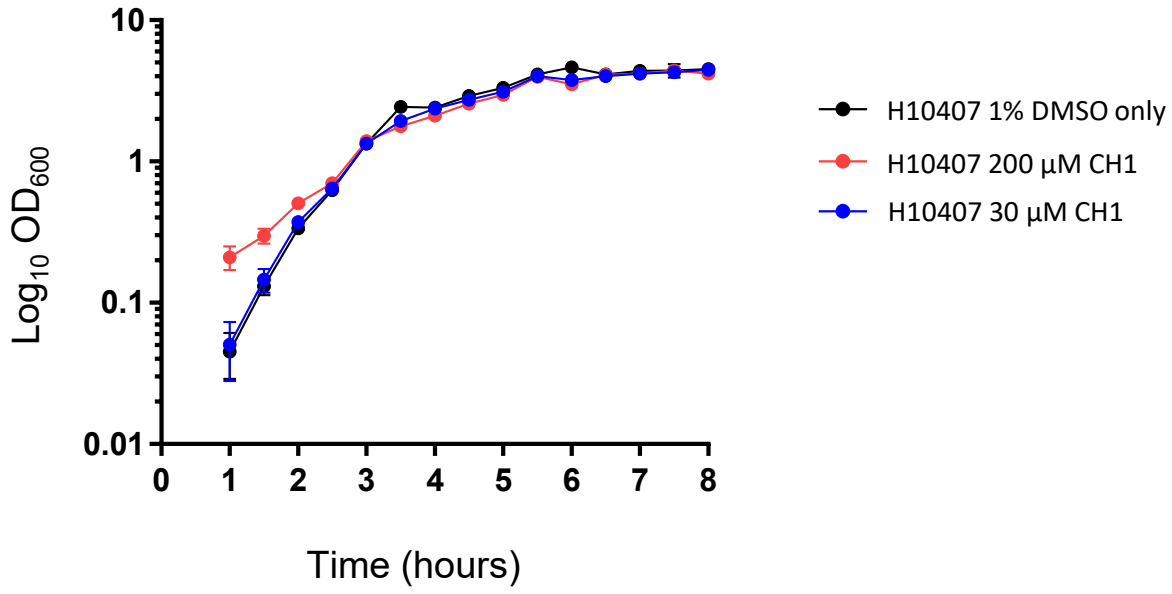


Figure 4.5. Effect of CH1 on the growth of H10407

The growth kinetics of H10407 was measured in the presence of 1% DMSO (—●—), or 200 µM (—●—) or 30 µM (—●—) CH1 in 1% DMSO. The results are shown as OD_{600} readings at half hour intervals over 8 hours. CH1 had no effect on the growth rate of *E. coli* H10407 at either concentration tested. Error bars indicate standard error of the mean, n = 2.

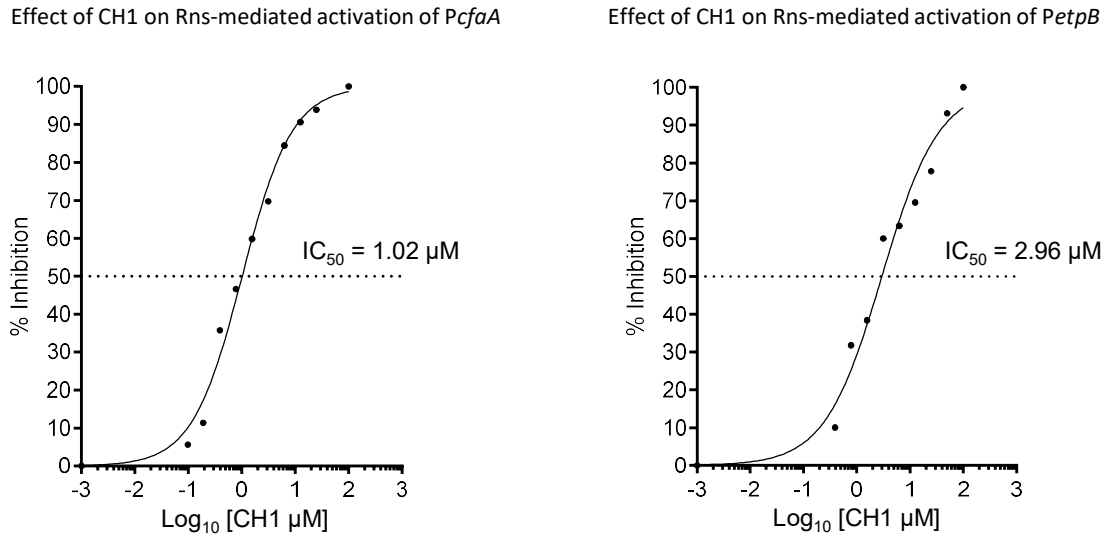


Figure 4.6. Determination of the IC_{50} of CH1 on Rns-mediated activation of P*cfaA* and P*etpB*

Results of β-galactosidase reporter assays showing P*cfaA* and *PetpB* activation by Rns in the presence of various concentrations of CH1. Inhibition is shown as a percentage, where 100% is the maximum inhibitory activity, and 0% is equal to the β-galactosidase activity of the promoter in the presence of Rns and the absence of any inhibitor. The IC_{50} is the concentration at which CH1 exhibits half of its maximum inhibitory effect, and was 1.02 µM when measuring Rns-mediated activation of P*cfaA,* and 2.96 µM when measuring Rns-mediated activation of P*etpB*. Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points and calculate the $IC_{50.}$

determined from the chemical energy of the binding solution, so that a lower binding energy gives a higher scoring solution, corresponding to a more energetically favourable state. Clustering is a measure of how many binding solutions are in the same pose in the protomol. A high level of clustering suggests that the binding solution is more likely to be real, with multiple solutions for the compound to bind in one protomol. If there is no clustering, the binding solution may be due to random chance, rather than being a real, probable interaction.

The docking model of CH1 binding to Rns in protomols two and three was provided by Dr. Holien, and is depicted in figure 4.7. Protomol one had the highest energy score, but no clustering of solutions, indicating that this binding is unlikely to be real (table 4.2). Protomol two has the second highest score, and there was significant clustering, making it the more likely binding solution. Protomol two is the pocket between the dimerisation and DNA binding domains, depicted in green in figure 4.7. When docked into protomol three, depicted in black in figure 4.7, a functional group of CH1 was outside the binding pocket and exposed to the solvent. This is energetically undesirable, making this solution unlikely despite the presence of two clusters at 35%.

4.2.6 Virtual screen for CH1 analogues

After identifying CH1, we returned to the computer modelling software to search for analogues of CH1 that could also act as inhibitors. The aim was twofold: to collect data that could be used to improve the binding model of the compound docked into Rns, and to identify other compounds that act in the same way as CH1, but may be more effective or have more desirable biochemical properties. To address the first goal, when screening for CH1 analogues we did not limit the search to chemicals likely to be suitable as a drug. Instead, Dr. Holien investigated the entire library of approximately 10 million compounds.

A 2D Unity similarity search was performed by Dr. Holien, using a Tanimoto similarity value of 65%, was performed using Sybyl2.1. From this process, 33 chemicals, designated SVI-2898 to SVI-2931, were selected for *in vitro* testing (table 4.3). These compounds were purchased and modified

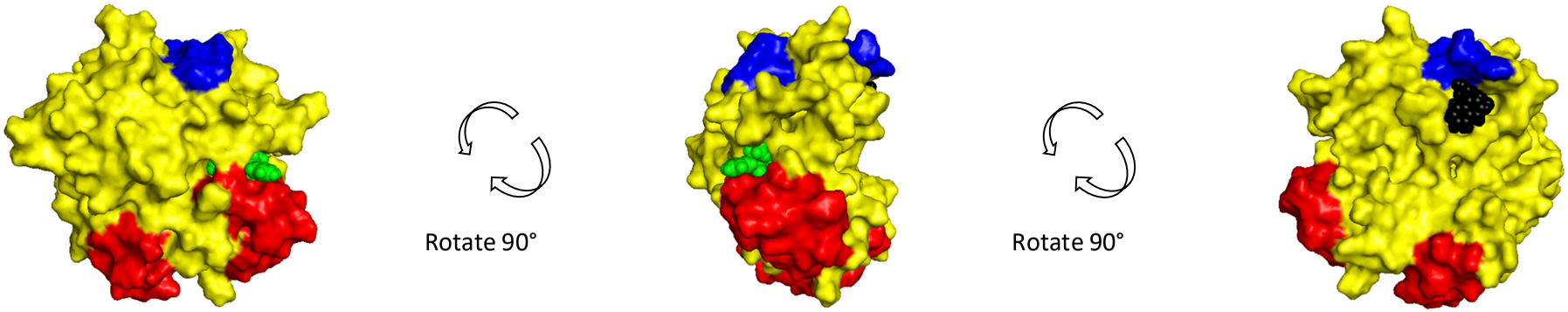


Figure 4.7. Model of possible binding sites for CH1

The Rns protein is represented as a yellow surface filling model, with the DNA-binding helices highlighted in red, and the putative dimerisation domain in blue. Potential inhibitor binding sites are depicted in green (protomol two) and black (protomol three).

Table 4.2 Possible binding conformations of CH1 docking into Rns

Protomol #	Highest Scoring Solution	Cluster %	Location of protomol	Colour in figure 4.7
1	7.4403	None	Near dimerisation domain	N/A
2	5.4875	70	Between dimerisation and DNA-binding domains	Green
3	5.1938	2 x 35%	Near dimerisation domain	Black

Table 4.3. Fold-change of Rns-mediated activation of P*cfaA* in response to chemical compounds

Compounds which the Rns-mediated activation pf *PcfaA* to less than 50% of a DMSO control are boxed in red.

Compound #	Fold change
SVI-2898	0.76
SVI-2899	0.98
SVI-2900	0.75
SVI-2901	0.76
SVI-2902	0.50
SVI-2903	0.74
SVI-2904	0.70
SVI-2905	1.07
SVI-2906	0.71
SVI-2907	0.64
SVI-2908	0.62
SVI-2909	1.08
SVI-2910	0.94
SVI-2912	0.64
SVI-2913	0.87
SVI-2914	0.35
SVI-2915	1.09
SVI-2916	0.57
SVI-2917	0.47
SVI-2918	0.47
SVI-2919	0.45
SVI-2920	0.80
SVI-2921	1.01
SVI-2922	0.22
SVI-2923	0.64
SVI-2924	0.96
SVI-2925	0.23
SVI-2926	1.08
SVI-2927	1.26
SVI-2928	0.50
SVI-2929	0.72
SVI-2930	0.79
SVI-2931	0.86

β-galactosidase reporter assays (see methods section 2.16.2) were performed to measure Rns-mediated activation of *PcfaA-lacZ* in the presence of 100 µM of each compound. Six compounds were able to inhibit promoter activity to less than half of a DMSO control (table 4.3). These were then tested for their ability to inhibit the Rns-mediated activation of P*cfaA* at concentrations ranging from 500 µM to 244 nM in doubling dilutions (figure 4.8). Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points for each compound. Two compounds, SVI-2922 and SSVI-2925, were found to have efficacy comparable to that of CH1. These two compounds are referred to as CH2 and CH3, respectively, hereafter.

Figure 4.9 is a flowchart summarising the process of identifying compounds CH2 and CH3.

4.2.7 Determination of the IC_{50} of CH2 and CH3

The IC_{50} of CH2 and CH3 was measured using modified β-galactosidase reporter assays at a range of inhibitor concentrations from 5 µM to 200 mM (figure 4.10). Both the P*etpB-lacZ* and P*cfaA-lacZ* fusions, co-transformed into *E. coli* MC4100(pRns), were used for determination of the IC_{50}. Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points, and calculate the IC_{50}. These analyses revealed that CH2 inhibited Rns-mediated activation of P*cfaA* and P*etpB* with an IC_{50} of 0.49 µM and 2.26 µM, respectively (figure 4.10), and that CH3 inhibited Rns-mediated activation of P*cfaA* and P*etpB* with an IC_{50} of 6.55 µM and 7.14 µM, respectively (figure 4.11). These data are summarised in table 4.4.

4.2.8 Investigation of the effect of CH2 and CH3 on the growth of *E. coli*

Low volume growth curves (see methods section 2.9.2) were performed on *E. coli* H10407 to determine if the Rns inhibitors affect growth in stationary media. H10407 was grown in LB in 1 ml volumes without shaking, in the presence of 10 µM CH1, CH2, or CH3 in 1% DMSO, or a DMSO control. No effect on growth was observed for any inhibitor (figure 4.12).

4.2.9 Effect of the Rns inhibitors on the activity of TyrR

To determine the specificity of CH1, CH2 and CH3, β-galactosidase assays were performed

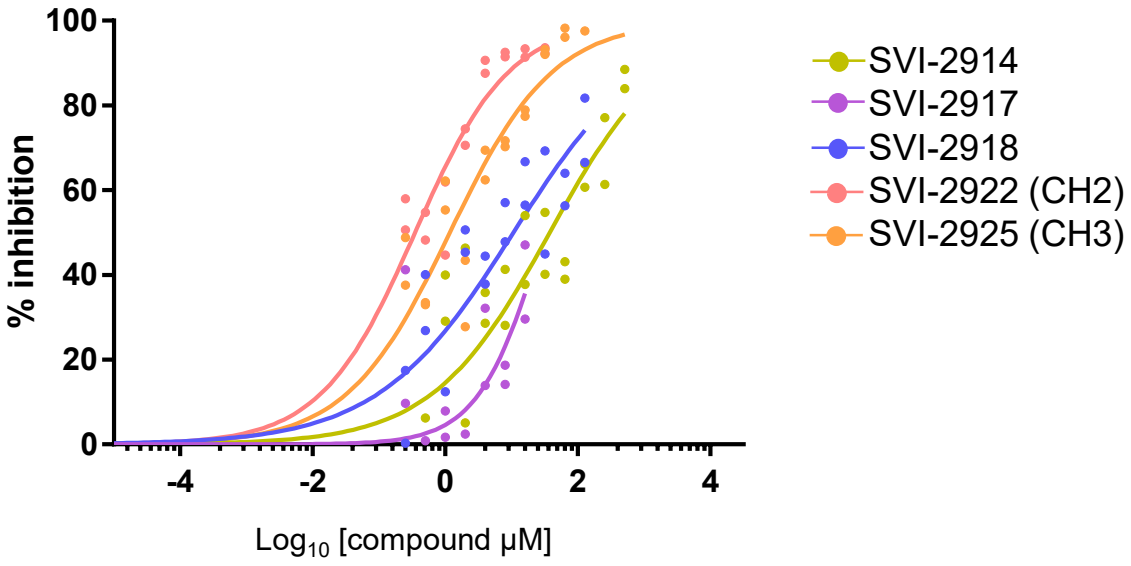


Figure 4.8. Dose-response curves of the effect of five CH1 analogues on Rns-mediated activation of P*cfaA*

Dose-response curves showing the ability of five CH1 analogues to inhibit Rns-mediated activation of P*cfa* measured by β-galactosidase activity. Inhibition is listed as a percentage, where 100% is equal to the β-galactosidase activity in the absence of Rns, and 0% is equal to the β-galactosidase activity in the presence of Rns. Each data point is represented. Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points, and calculate the $IC_{50.}$

Figure 4.9. Summary of the screen for CH1 analogues

Flowchart depicting the process used to identify CH1 analogues, CH2 and CH3. Rns inhibitor CH1 was identified through a high-throughput chemical screen. A virtual library of 10 million compounds was then screened for CH1 analogues using Sybylx 2.0 (Certara). 33 analogues were tested in β-galactosidase reporter assays, and the six most effective inhibitors in these assays were tested for their activity against Rns at various concentrations. Two of these were selected as the hit compounds CH2 and CH3.

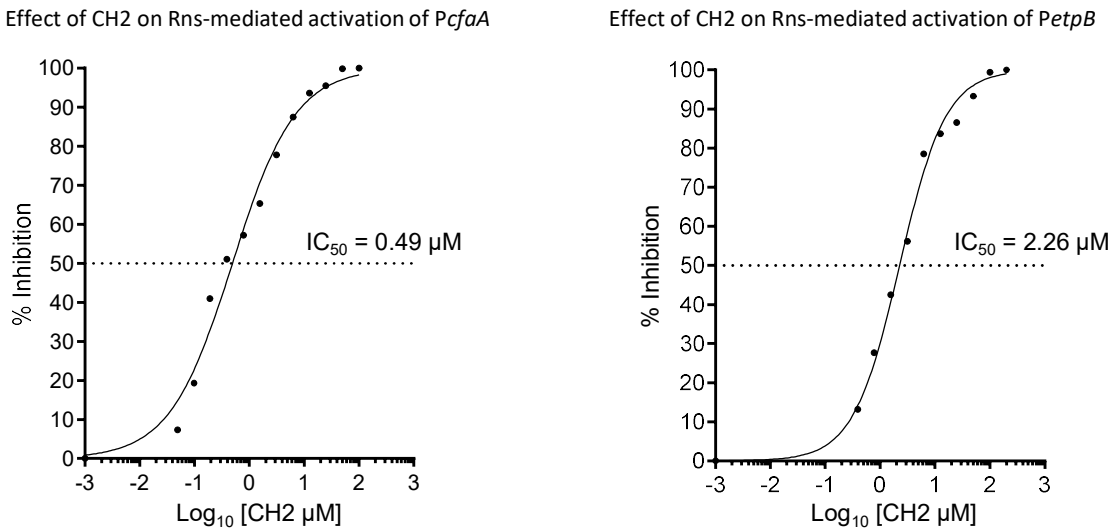


Figure 4.10. Determination of the IC_{50} of CH2

Results of β-galactosidase reporter assays showing P*cfaA* and *PetpB* activation by Rns in the presence of various concentrations of CH2. Inhibition is shown as a percentage, where 100% is the maximum inhibitory activity, and 0% is equal to the β-galactosidase activity of the promoter in the presence of Rns and the absence of any inhibiting compound. The IC_{50} is the concentration at which CH2 exhibits half of its maximum inhibitory effect, and is 0.49 µM when measuring Rns-mediated activation of P*cfaA,* and 2.26 µM when measuring Rns-mediated activation of P*etpB*. Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points, and calculate the $IC_{50.}$

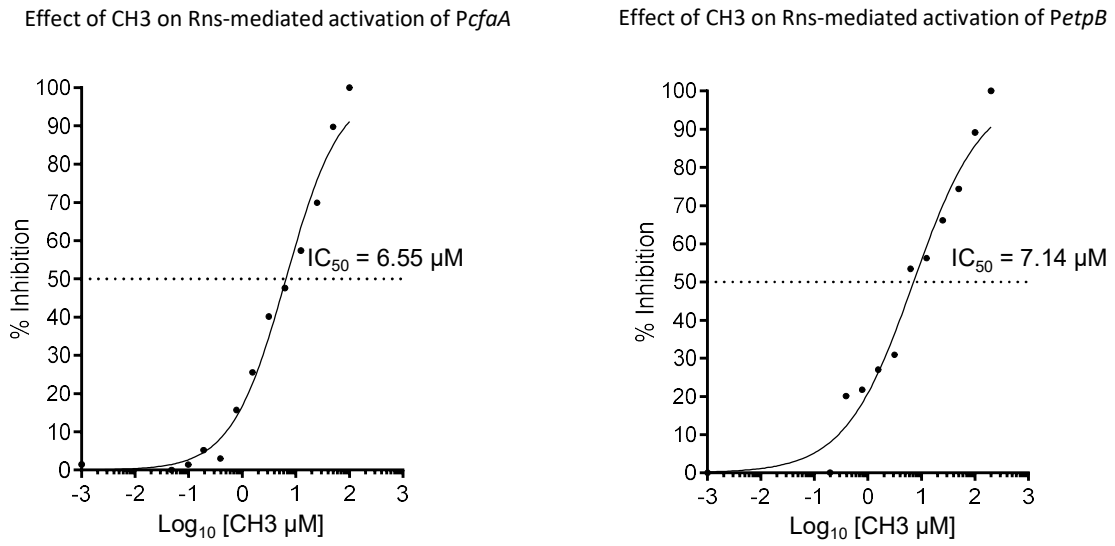


Figure 4.11. Determination of the IC_{50} of CH3

Results of β-galactosidase reporter assays showing P*cfaA* and *PetpB* activation by Rns in the presence of various concentrations of CH3. Inhibition is shown as a percentage, where 100% is the maximum inhibitory activity, and 0% is equal to the β-galactosidase activity of the promoter in the presence of Rns and the absence of any inhibiting compound. The IC_{50} is the concentration at which CH3 exhibits half of its maximum inhibitory effect, and is 6.55 µM when measuring Rns-mediated activation of P*cfaA,* and 7.14 µM when measuring Rns-mediated activation of P*etpB*. Prism 6.01 (GraphPad) was used to fit a sigmoidal variable slope curve to the plotted data points, and calculate the $IC_{50.}$

Table 4.4. The IC_{50} of CH1, CH2 and CH3 for Rns-mediated activation of P*cfaA* and P*etpB*

	IC_{50} P*cfaA*	IC_{50} P*etpB*
CH1	1.018 µM	2.960 µM
CH2	0.4907 µM	2.263 µM
CH3	6.55 µM	7.140 µM

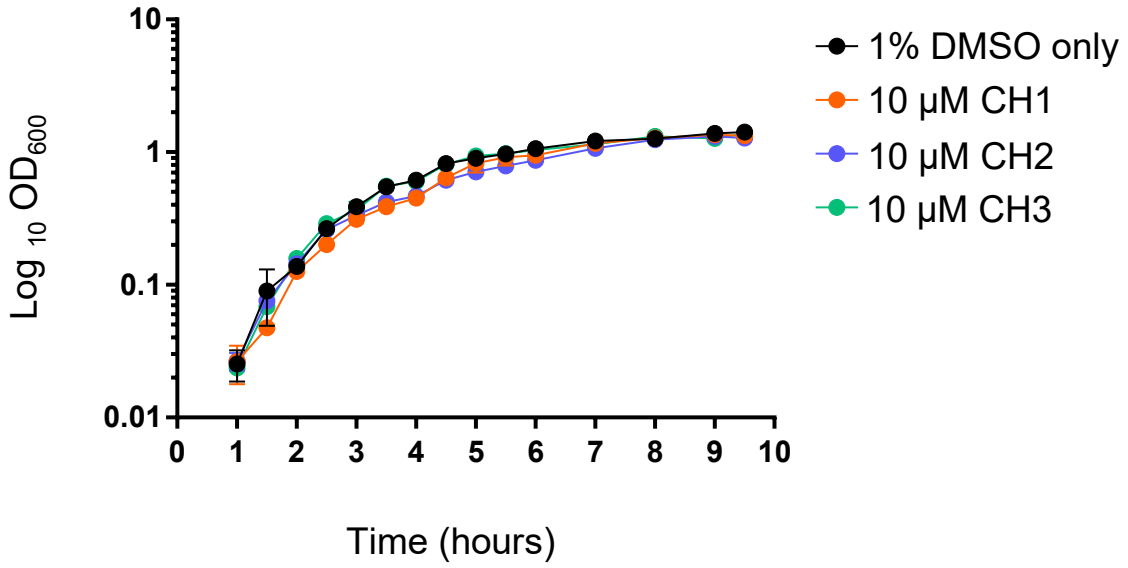


Figure 4.12. Effect of CH1, CH2 and CH3 on the growth of H10407

The growth kinetics of H10407 was measured in the presence of 1% DMSO (—●—), or 10 µM CH1 (—●—), CH2 (—●—) or CH3 (—●—) in 1% DMSO. The results are shown as OD_{600} readings at half hour intervals over 9.5 hours. No compound had an effect on the growth of *E. coli* H10407. Error bars indicate standard error of the mean, n = 3.

using a P*mtr-lacZ* fusion, co-transformed into MC4100(pTyrR) (194). TyrR activates the transcription of a tryptophan-specific transport system *E. coli*, and is not related to Rns (254). Bacteria were grown with compounds CH1 or CH2, or CH3 (figure 4.13). As CH1 and CH2 were not soluble at concentrations higher than 50 µM, they were tested at this concentration. CH3 remained soluble at a higher concentration, and was tested at 400 µM. No inhibition of TyrR was seen in response to any inhibitor at the concentrations tested.

4.2.10 Effect of CH1, CH2 and CH3 on the expression of CFA/I by *E. coli* H10407

All of the above investigations of CH1, CH2 and CH3 were performed in *E. coli* MC4100, using *lacZ*-fusion plasmids. To determine if the candidate inhibitors had an effect on the expression of CFA/I by *E. coli* H10407, SDS-PAGE was performed on heat-extracted surface proteins from *E. coli* H10407, H10407Δ*cfaABCE*, H10407Δ*rns,* H10407Δ*rns*(pRns), grown in 1% DMSO, and *E. coli* H10407 grown in the presence of 10 µM of compounds CH1, CH2 and CH3 in 1% DMSO (figure 4.14). A band of approximately 17 kDa, which corresponds to the expected size of CfaB, the major subunit of CFA/I fimbriae, was detected in H10407, but absent from both the *cfaABCE* and *rns* deletion mutants; as expected, it was more prominent in H10407Δ*rns*(pRns). The 17 kDa band from the lane of the gel containing proteins from H10407Δ*rns*(pRns) was excised and sequenced by mass spectroscopy at the Walter and Eliza Hall Institute for Medical research. Protein sequencing confirmed that CfaB was the predominant protein in this band.

The CfaB band was fainter in protein preparations from H10407 grown in the presence of 10 µM CH2 and CH3. CH1 present at a concentration of 10 µM did not have an obvious effect on CfaB production in H10407 (figure 4.14).

4.3 Discussion

4.3.1 Rns as a target for the development of a virulence gene inhibitor

Current treatment for ETEC focuses primarily on rehydration, with antibiotics given in extreme cases (31). Resistance to antibiotics is increasing within the ETEC pathotype, and novel

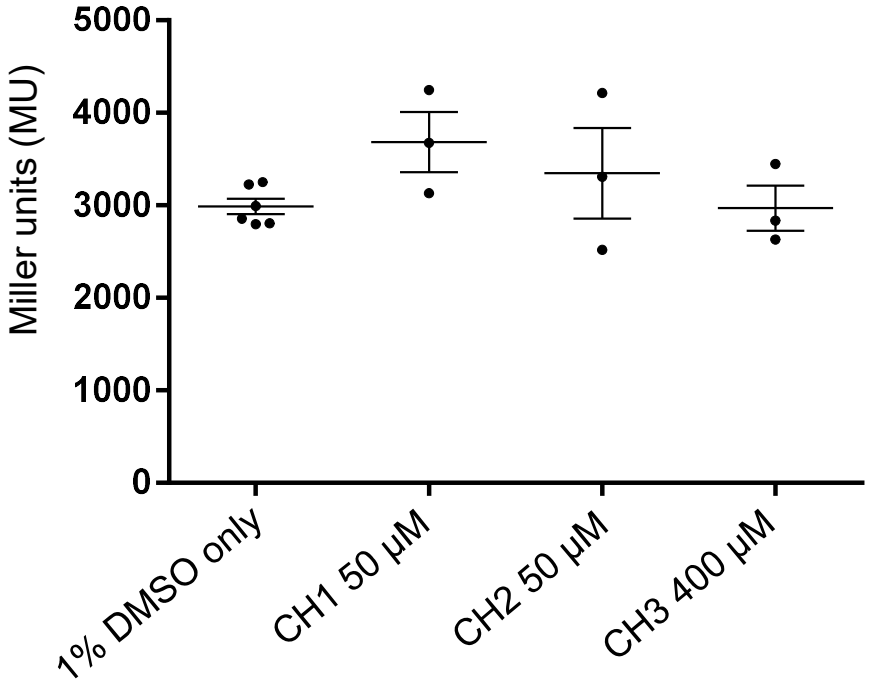


Figure 4.13. Effect of CH1, CH2 and CH3 on TyrR-mediated activation of P*mtr*

Results of β-galactosidase reporter assays on the activation of P*mtr* by TyrR in the presence of 50 µM CH1 and CH2 in 1% DMSO, 400 µM of compound CH3 in 1% DMSO, or a DMSO control showed that no compound had an effect on TyrR-mediated activation of P*mtr*. Error bars indicate mean and standard error of the mean, n ≥ 3.

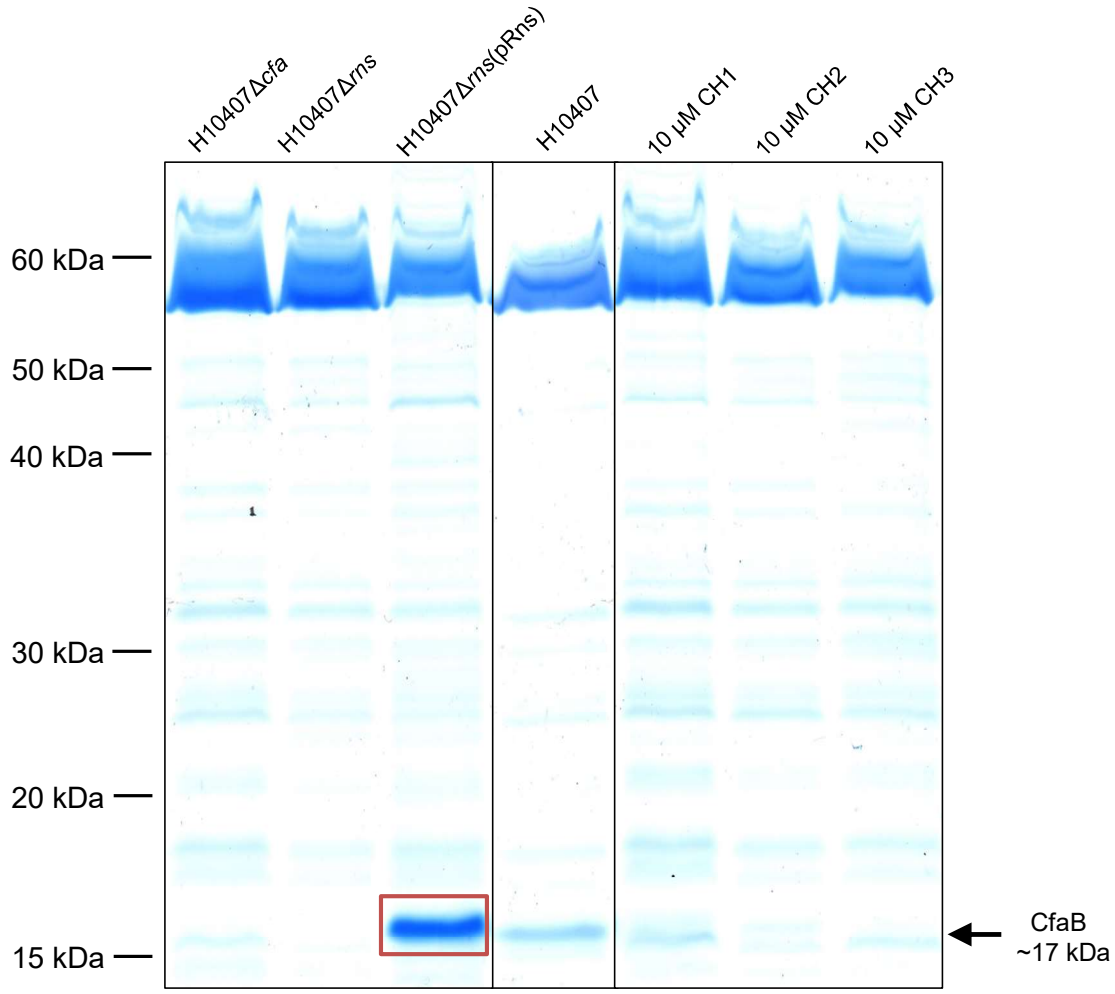


Figure 4.14. Surface protein expression by *E. coli* H10407 grown in the presence of Rns inhibitors

SDS-PAGE showing Kang's-colloidal-stained, heat-extracted, surface proteins of *E. coli* H10407 and its derivative strains grown in the presence of 1% DMSO, and H10407 grown in the presence of 10 µM of compounds CH1, CH2 or CH3 in 1% DMSO. The band corresponding to CfaB (~17 kDa) is indicated with an arrow. This band is present in surface proteins extracted from wild type H10407, but missing from protein preparations of both H10407Δ*cfaABCE* and H10407Δ*rns,* and restored in H10407Δ*rns*(pRns). The band is faint in the samples grown with 10 µM CH2 and CH3.

therapeutics are required in the absence of a vaccine (177). In chapter three, I investigated the Rns regulon of *E. coli* H10407, and found that as well as the previously described Rns targets, the CFs and CexE, Rns was also responsible for the transcriptional activation of putative virulence genes *etpBAC*, *yghJ*, *rtr*, and, indirectly, *agn43*. This makes Rns an attractive drug target for the treatment and prevention of ETEC infection. If activation of the Rns regulon is inhibited, ETEC will be unable to colonise the small intestine and cause disease.

Inhibitors of virulence gene regulators, such as Regacin and Virstatin, have been a popular topic of research in recent years, as they have several advantages over traditional antibiotics (194, 196). Because they target virulence mechanisms specifically, the selection pressure for the development of resistance is far lower than it would be for a drug that kills the bacteria or interfering with their growth. The specificity of inhibition also means that commensal organisms living in the intestine should not be harmed by treatment with the virulence inhibitor. This has multiple benefits. Damage to the intestinal microbiota is responsible for many of the side-effects of antibiotics, and so virulence inhibitors should cause fewer side-effects (255-257). Second, any resistance which may develop against these drugs will be beneficial only to the targeted pathogen. Commensal organisms will have no incentive to harbour resistance genes, reducing the chance that they will act as a reservoir for the horizontal transfer of these genes to pathogens.

In this chapter, I described the use of two complementary methods (a high-throughput screen and a virtual screen) to investigate libraries of chemical compounds and find those which inhibit target gene activation by Rns. The Rns inhibitor CH1, and its analogues CH2 and CH3, were identified, and their properties investigated.

4.3.2 High-throughput in vitro screen for an Rns inhibitor

The HTS I used is able to screen thousands to tens of thousands of compounds for their ability to inhibit Rns activity *in vitro*. The aim of this screen was to identify a compound that inhibits Rns specifically. To that end, each compound during the HTS was also tested for its ability to inhibit

gene activation by an unrelated regulator, MrkH. The latter is a transcriptional regulator which controls the expression of type 3 fimbriae by *Klebsiella pneumoniae* (252). If a compound reduced the activity of both Rns and MrkH, then it was either broadly inhibiting or toxic to the cell, and was not considered further. All compounds were tested at a concentration of 10 µM, a concentration which has been successfully used in our lab for this type of screen in the past (194).

For the HTS, I initially included any compound that reduced the Rns-mediated activation of P*cfaA* to below 67% of a DMSO control. Hits were rare, and so this generous threshold was chosen so as not to exclude potential hit compounds from being detected. Repeating the assay for the 41 potential hit compounds identified initially showed that only four compounds were reliably able to reduce Rns-mediated activation of PcfaA to below 67% of a DMSO control.

Further testing was performed with the compounds at 100 µM to increase our ability to detect inhibition, and using low volume β-galactosidase assays as a practical measure to conserve our stock of the compounds. In these tests, one compound had a highly significant effect on the Rns-mediated activation of P*cfaA*, inducing an approximately 3-fold reduction in promoter activity. This compound was CH1. A second compound, F2, reduced promoter activity to approximately 80% of a DMSO control, with a p value = 0.038 (figure 4.3).

4.3.3 Virtual screen for an Rns inhibitor

A virtual screen can be of a much larger scale than an HTS, as it predicts the ability of millions of compounds to bind to selected target sites within a protein. Before the screen can begin, the protein must be accurately modelled, and its functional regions must be mapped. Once this process is complete, we can select potential binding pockets that are likely to inhibit the activity of the protein if bound by a small chemical compound. The actual screen itself requires extensive processing power, taking weeks to be completed. The targeted nature of a virtual screen means that only compounds predicted to bind to the selected regions of the protein are assayed. This is in contrast to HTS, which has the possibility to identify compounds that work in unexpected ways.

The virtual screen was conducted on a model of Rns. The model was based on the crystal structure of ToxT, which shows 18.1% sequence identity to Rns (214). Two pockets in the Rns protein were identified as possible compound binding sites and were targeted for the virtual screen (figure 4.4). One of these was central to the protein (depicted in magenta in figure 4.4), and the other was located between the dimerisation and DNA-binding domains (depicted in green). Both of these functions are vital for the function of AraC-like regulators, and both have been successfully targeted by virulence inhibitors in other pathogens (124, 258, 259). The dimerisation domain is the site where Virstatin binds ToxT, inhibiting its function, while Regacin inhibits RegA through binding of the DNA-binding domain (194, 197).

Four million drug-like compounds were screened against each site and the top 500 hits kept for analysis. These compounds were then analysed visually and computationally, resulting in a shortlist of 177 structurally diverse compounds that I subjected to biological analysis. In modified β-galactosidase reporter assays, 11 of these 177 compounds decreased Rns-mediated activation of P*cfaA* to less than 50% of a DMSO control. Unfortunately, none of these 11 compounds was reliably able to reduce Rns-mediated activation of P*cfaA* over multiple repeated assays. This may have been due to the degradation of the compounds over time, or may reflect a weak ability to inhibit Rns-mediated activation of P*cfaA*. By this point in time, I had identified CH1 as a strong candidate for Rns inhibition. I decided to focus on CH1, and not explore the potential hit compounds from the initial virtual screen any further at this stage.

4.3.4 Docking of CH1 into binding pockets of Rns

After CH1 was identified through the HTS, Dr. Holien performed computer modelling to predict the interactions between it and Rns. Knowing how a compound docks into a binding pocket on Rns allows us to deliberately modify that compound in ways which are likely to improve binding affinity. Knowing which sections of the compound are essential for binding also allows us to target

other regions of the compound for modifications to improve the solubility, reduce toxicity, or alter other biochemical properties of the compound, without interfering with its binding to Rns.

For the virtual screen, two binding pockets on the Rns structure were targeted. One of these sites is central to the protein, depicted in magenta in figure 4.4, and is likely to impact the conformation of the dimerisation domain. The other is between the dimerisation and DNA-binding domains, and is depicted in green in figure 4.4. Compounds binding to these pockets are expected to inhibit Rns by interfering with dimerisation and/or binding to DNA, both of which are essential for target promoter activation by Rns. When CH1 was identified and the likely position of binding to Rns was modelled, two potential binding sites were discovered (figure 4.7). One of these binding conformations, depicted in green in figure 4.7, binds to one of the pockets in Rns that was targeted in the virtual screen (also depicted in green in figure 4.4). This site is the likely putative binding site for CH1. Figure 4.15 is a side-by-side comparison of the binding pockets targeted in the virtual screen and the potential binding conformations of CH1. If the green site is indeed the predominant site of binding by CH1, then the compound is most likely acting by inhibiting the ability of Rns to bind to target DNA. The other potential CH1 binding site, depicted in black in figure 4.7, is closer to the putative dimerisation domain of Rns and was not one of the sites targeted in the virtual screen.

4.3.5 CH1 analogues

One way to attempt to improve the binding affinity of CH1 to Rns is to make small modifications to either the compound or the protein, and see how those changes affect the inhibitory activity of the compound. Mutagenesis of Rns followed by β-galactosidase reporter assays to determine any changes in the ability of CH1 to inhibit Rns activity can narrow down which residues are contacting the compound, and help to determine which of the putative CH1 binding sites is correct, as well as which residues are essential for contact with the compound.

The same idea can be used to determine which regions of CH1 contact the Rns protein. To further investigate the binding properties of CH1, a 2-dimensional search for CH1 analogues was

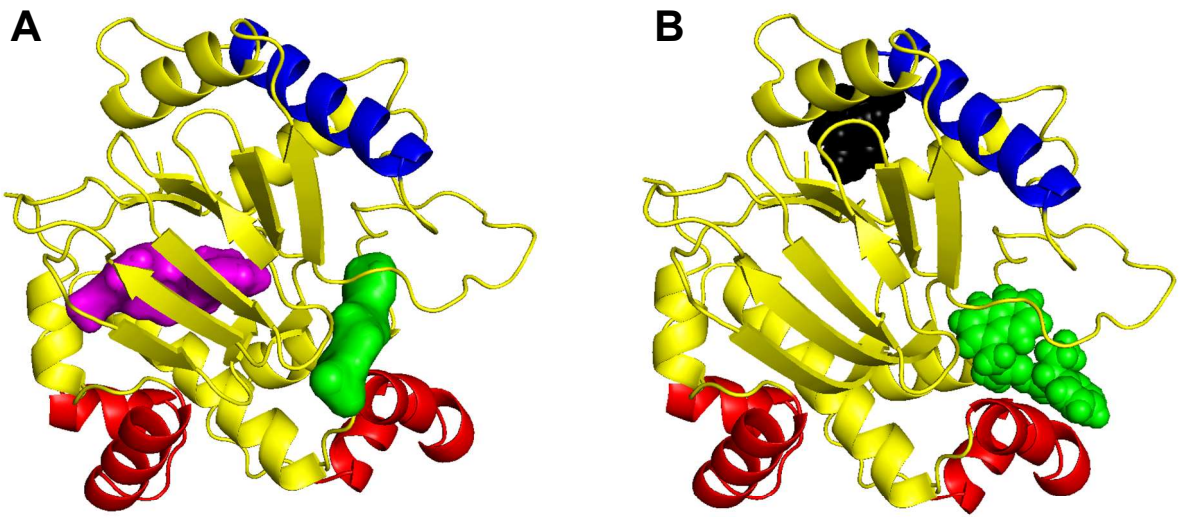


Figure 4.15. Pockets in the Rns structure targeted for compound docking and possible binding sites of CH1

The Rns protein is represented in yellow, with the DNA-binding helices in red and the putative dimerization domain in blue. A) Depiction of the binding pockets, in green and magenta, used as targets in the VS. B) Depiction of the possible binding sites of CH1, in green and black.

performed. Dr. Holien screened the virtual library of compounds for those which share the basic structure of CH1, but with small modifications, such as the presence or arrangement of functional groups. These analogues were identified by calculating the Tanimoto similarity coefficient of each compound in the library to CH1. This is a figure that compares the number of substructural fragments in common between two compounds to those that are unique to each, to yield a similarity score between 0% and 100% using the equation:

$$\text{Tanimoto similarity coefficient} = \frac{c}{a+b-c}$$

Where 'a' is the number of substructural fragments present in one compound alone, 'b' is the number of substructural fragments present in the other compound alone, and 'c' is the number of substructural fragments that are common to both compounds.

No CH1 analogues were found with a Tanimoto similarity coefficient greater than 80%, and only 13 were found with a Tanimoto similarity coefficient greater than 75%. Reducing the threshold for analogue discovery to 65%, there were more than 50 hit compounds. For convenience, 33 compounds that were available from suppliers were obtained for *in vitro* testing. In modified β-galactosidase reporter assays, six compounds reduced promoter activity to below 50% of a DMSO control.

CH1 analogues were also investigated for the purpose of improving the accuracy of the binding model, so that we could continue investigations with a group of similar compounds, rather than a single hit. These six compounds were tested at various concentrations for their ability to inhibit Rns-mediated activation of P*cfaA* (figure 4.8). One compound, SVI-2919, returned consistently erratic results, and was not included in further investigations. In figure 4.8, 100% inhibition refers to the promoter activity of a DMSO control. This allows for visualisation of the maximum efficacy of each compound as well as the concentrations at which they exert their effect. This means that these curves cannot be used to calculate the IC_{50}, because the IC_{50} is defined as 50%

of the maximum effect of that compound, not of complete inhibition. Compounds SVI-2914, SVI-2917, and SVI-2918 became insoluble at higher concentrations, before their dose-response curves reached a plateau. CH2 and CH3 were selected for further characterisation, as they exerted the greatest inhibitory effect at the lowest concentration, and they were the most soluble.

4.3.6 Dose-response curves of the effect of CH1, CH2, and CH3 on Rns-mediated activation of P*cfaA* and P*etpB*

The *in vitro* efficacy of an inhibitor is measured by the IC_{50}, the concentration at which the compound exerts 50% of its maximum effect. The dose-response curves used to measure the IC_{50}s of CH1, CH2 and CH3 end just before reaching a clear plateau (figures 4.6, 4.10, and 4.11). This is because they became insoluble at concentrations above those plotted. This may have compromised the accuracy when calculating the IC_{50}, artificially decreasing the value and leading to an overestimation of the effect of the compounds, particularly for CH3 (figure 4.11).

Interestingly, the IC_{50} for each compound differed depending on which of the two Rns target promoters was used for the assay. The most striking example was CH2, which has an IC_{50} of 0.49 µM against Rns-mediated activation of P*cfaA*, and of 2.26 µM against Rns-mediated activation of P*etpB*. This is mostly likely due to slight differences in the way Rns binds to each of these promoters, so that the change in Rns that is induced by inhibitor binding has a stronger inhibiting effect on the binding to some promoters than to others. For all three inhibitors, the IC_{50}s were higher when measuring the effect on Rns-mediated activation on P*etpB* than on P*cfaA*. During further development of an Rns inhibitor, it will be interesting to measure the effects of the compounds on Rns-mediated activation of a variety of target promoters, and to see how these changes affect the phenotypes mediated by these Rns target genes.

4.3.7 The effect of CH1, CH2, and CH3 on CFA/I production of *E. coli* H10407

Figure 4.14 depicts the expression of CfaB by *E. coli* H10407 grown in the presence of CH1, CH2 and CH3. Despite IC_{50} calculations estimating that CH1 has a stronger effect on transcription of the *cfaABCE* operon than CH3, examination of the protein profiles suggest that CH1 actually had the least effect on CfaB production, and CH2 was the most effective. It is not entirely unexpected that the IC_{50} calculated using promoter-*lacZ* fusions does not correspond exactly to the final protein production. The majority of the work in this study was performed in *E. coli* MC4100, using promoter-*lacZ* fusion constructs, and there may be a different activity profile when the compounds are applied in the Rns native ETEC background. The inhibitor compounds must reach a high enough intracellular concentration to have an effect on Rns, and *E. coli* H10407 may have different levels of uptake and efflux of the compounds than *E. coli* MC4100. The abundance of Rns and of target promoter DNA will also differ between *E. coli* H10407 and the *E. coli* MC4100 reporter strains. Finally, other regulatory mechanisms specific to H10407 that interact with the Rns response could also account for some of the differences in these results. In the future, this work must be repeated in *E. coli* H10407.

4.3.8 Future lines of investigation

4.3.8.1 Improving the computer model of Rns and inhibitors

The more accurate our computer model of Rns, the better we can understand the interactions between Rns and its inhibitors. We can then use this information to improve the inhibitors. The results of β-galactosidase reporter assays measuring the ability of each of the CH1 analogues to inhibit Rns activity can be used to improve the accuracy of the model. We can also gain information about the way in which various inhibitor compounds dock into binding pockets of Rns by using mutations to the protein. Amino acid changes in the regions to which an inhibitor compound is expected to bind can help us to confirm experimentally the docking site identified *in silico*. In addition, solving the crystal structure of Rns itself will help us improve upon the computer model, which is currently based on the crystal structure of ToxT.

4.3.8.2 Determining the effect of inhibitor compounds on H10407 phenotypes

Improving the efficacy of inhibitor compounds with further rounds of model refinement and testing of analogues will be performed using the same *in vitro* assays as described in this chapter. Additionally, assays should be performed to determine the loss of Rns-mediated phenotypes of H10407 in response to the inhibitor compounds. The production of CFA/I by H10407 in the presence of inhibitor compounds was measured qualitatively in figure 4.14, and CH2 was found to have a strong ability to reduce CfaB production. As well as CFA/I production, other assays should be employed to measure the phenotypes controlled by Rns and see how they are affected by inhibitor compounds. These assays could include adherence to CaCo-2 cells (mediated by CFA/I), secretion of EtpA, degradation of MUC2 (mediated by YghJ), and biofilm formation (mediated by Agn43). Since the IC_{50}s of the compounds varied when measuring Rns-mediated activation of different promoters, it is possible that different compounds will have subtly different effects on the H10407 virulence phenotype.

4.3.8.3 The mechanism of action of CH1, CH2, and CH3

Inhibitors of Rns are likely to function by interfering with its dimerisation or its ability to bind target DNA. To determine if the compounds inhibit DNA binding by Rns, EMSA can be performed to measure Rns binding to P*cfaA* and/or other Rns target promoters in the presence and absence of the inhibitors. If the compound inhibits the ability of Rns to bind to its target DNA, the gel-shift of the promoter DNA in response to Rns will be impeded at high concentrations of compound. This method has been used previously in our lab to identify the mechanism of action of Regacin upon RegA (194).

To determine if the compounds are inhibiting Rns by blocking dimerisation, ultracentrifugation can be performed to measure the sedimentation coefficient of Rns in the presence and absence of the inhibitors. Rns is generally thought to exist as a dimer in solution, although this stance has been questioned (123, 125). If Rns exists as monomers in the presence of inhibitor compounds, this would indicate that the compound is inhibiting the dimerisation of Rns.

4.3.8.4 In vivo testing of Rns inhibitors

There are at least three potential mouse models of Rns-mediated infection that could be used to test inhibitor compounds *in vivo*. *E. coli* does not thrive in the mouse intestine, primarily due to commensal bacteria that produce short-chain fatty acids which inhibit *E. coli* growth (260). Mice that have been pre-treated with streptomycin to reduce the number and influence of these commensal bacteria are susceptible to colonisation by ETEC, but this colonisation doesn't cause disease. Moreover, ETEC CFs are species-specific, meaning that ETEC colonisation of the mouse intestine isn't necessarily indicative of infection in humans (261).

Another potential Rns-dependent animal model uses the natural mouse pathogen *C. rodentium*. The virulence activator of *C. rodentium*, RegA, is closely related to Rns, and preliminary work in our lab has established than Rns can restore the mouse colonisation phenotype lost in a *regA* deletion mutant of *C. rodentium* (figure 4.16). This model would allow us to see the effect of an Rns inhibitor in a model where Rns controls virulence through the transcriptional activation of multiple target genes.

Both of these models are useful for determining the ability of the compounds to access intestinal bacteria and affect Rns-mediated colonisation and/or virulence *in vivo*. Unfortunately, neither can be used to investigate ETEC virulence processes specifically.

A third mouse model of ETEC infection is performed in neonatal mice. The commensal bacteria that inhibit *E. coli* growth have not yet become established in the intestines of one-day old mice, and as such they are susceptible to ETEC-induced disease. This model has been used to test vaccines targeting ETEC by the vaccination of pregnant mice and subsequent challenge of their offspring with ETEC strain H10407, and could also be used to test an Rns inhibitor (262).

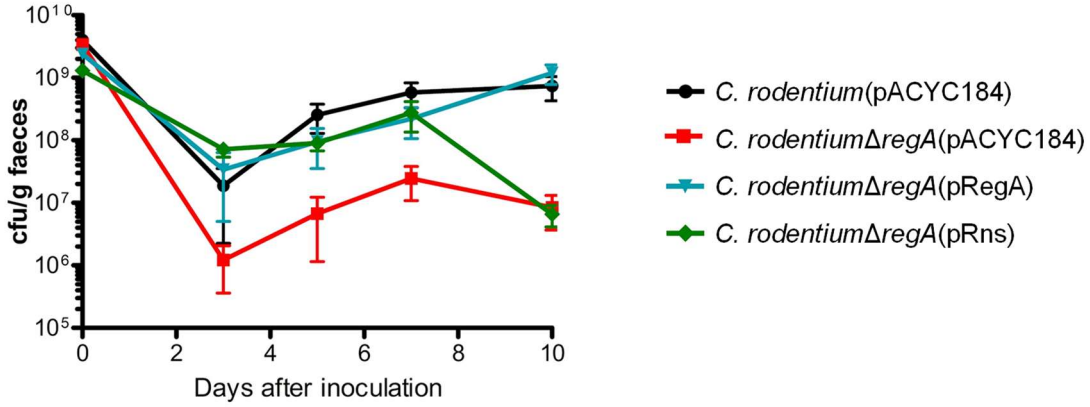


Figure 4.16. Colonisation of the mouse intestine by *C. rodentium*

The ability of *C. rodentium* to colonise the mouse intestine was measured by calculating the cfu/g of bacteria excreted in the faeces of mice orally infected with 10^6 cfu *C. rodentium*(pACYC184) (●), *C. rodentium*Δ*regA*(pACYC184) (■), *C. rodentium*Δ*regA*(pRegA) (▼), or *C. rodentium*Δ*regA*(pRns)(◆). The *regA* deletion mutant of *C. rodentium* was attenuated in its ability to colonise the mouse intestine. This ability was restored by the trans-complementation of *regA* on a plasmid. Trans-complementation of *rns* on a plasmid completely restored colonisation by *C. rodentium*Δ*regA* for the first seven days of infection, but not at day ten.

4.3.9 The potential of an Rns inhibitor for therapeutic use

ETEC is a major cause of diarrhoea in children under the age of five who live in developing countries, and for travellers to these regions. Both of these groups would benefit from the development of an effective inhibitor of Rns.

4.3.9.1 An Rns inhibitor as prophylaxis

It is likely that virulence inhibitors would work best when taken prophylactically, especially since Rns is primarily involved in the early colonisation stage of ETEC infection. Using traditional antibiotics as prophylaxis is not encouraged due to the strong selective pressure it applies for the development of resistance, not only of ETEC but of all gut microbiota. Because virulence inhibitors do not exercise the same widespread selective pressure and act on a highly specific target, an Rns inhibitor could be used more widely.

Travellers at risk of ETEC-induced diarrhoea could take an Rns inhibitor for the entire duration of their trip, protecting them from colonisation of the small intestine by ETEC. Travellers could safely stop taking the inhibitor once they are no longer in a region to which ETEC is endemic. Alternatively, an Rns inhibitor could be used as post-exposure prophylaxis. In this scenario, it would be taken by travellers only when they believe they may have been exposed to ETEC, such as after drinking water or eating raw food in areas with poor sanitation.

For children living in areas where ETEC is endemic, it is not feasible to be under constant treatment with virulence inhibitors. Instead, an Rns inhibitor would be best utilised to limit the spread of an outbreak. The inhibitor could be given to any children who are in close contact with an individual sick with diarrhoea, and taken for the duration of the exposure. It would also be applicable in situations where large numbers of people are forced into close proximity, such as in refugee camps, cruise ships, military excursions, etc.

4.3.9.2 An Rns inhibitor as treatment

An Rns inhibitor may also be effective as treatment for patients who already have symptoms of an ETEC infection. Experiments in mice have shown that Regacin, the virulence inhibitor of *C. rodentium*, can reduce infection in mice if taken 12 hours after exposure (194). The same is true for Virstatin inhibition of ToxT in a mouse model of *V. cholerae* infection (196). An Rns inhibitor which can reverse an established ETEC infection could be given to any patient presenting with watery diarrhoea in a region where ETEC is endemic.

4.3.10 Conclusion

The work presented in this thesis adds to a growing body of investigation into virulence inhibitors. As the problem of antimicrobial resistance continues to rise, it is vitally important that more research is focused on alternative mechanisms of bacterial control. My work provides an early exploration into the potential for the prevention and/or treatment of ETEC infections through the targeting of a transcriptional regulator of virulence.

The road to drug development is a long one. We must first improve our understanding of the results presented here, by refining the computer model of Rns, and by identifying exactly how the inhibitor compounds interact with the protein. Once this work is complete, there are several layers of complexity which must be considered before a viable drug is developed. The interaction of the Rns regulatory system with others specific to ETEC, as well as strain variation of these systems, must be explored. My final experiment, which looked at the protein expression profile of *E. coli* H10407 in response to the three compounds CH1, CH2, and CH3, already suggests that the effects of the inhibitors on the ETEC phenotype may not precisely mimic the results seen in a reporter construct system.

After a family of inhibitors has been well characterised in *E. coli* H10407, we must then begin to consider how they will act within a mammalian host. The safety or toxicity of the compounds must be assessed, and the compounds selected or modified to reduce any negative effects they may

have on the human body. Once in vivo experiments begin, the complication of bioavailability becomes an issue. A drug must be able to reach the same sites where an infectious agent has colonised the host. In this regard, it is an advantage to be working on a pathogen that colonises the intestine, as a compound can be taken orally and will not have to make its way from the digestive tract to a distant site of infection. Still, we do have to consider how the human digestive system might break down the compound, and how well the compound is able to penetrate the mucus layer and actually contact the bacteria in the small intestine.

All this means that CH1, CH2 and CH3 are unlikely to be the actual chemicals sold to travellers or given out by aid workers to combat ETEC-induced diarrhoea. Instead, they are intended as a jumping-off point, the first generation in a line of compounds that may one day become a licensed medicine. Despite the vast amount of work that still needs to be done, this thesis describes a very important first step. These proof of concept experiments have now been successful in three pathogenically similar organisms, *C. rodentium*, *V. cholerae*, and ETEC, early successes which are vital to encourage funding organisations and/or pharmaceutical companies to consider this approach to pathogen control a valid tactic, worthy of significant investment.

5 References

1. **Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, Cousens S, Mathers C, Black RE.** 2015. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. The Lancet **385:**430-440.
2. **Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, Mathers C, Black RE, Child Health Epidemiology Reference Group of WHO, Unicef.** 2012. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The Lancet **379:**2151-2161.
3. **Naghavi M, Wang H, Lozano R, Davis A, Liang X, Vollset S, Ozgoren A.** 2015. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet **385:**117-171.
4. **Mata L.** 1992. Diarrheal disease as a cause of malnutrition. Am J Trop Med Hyg **47:**16-27.
5. **Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM.** 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. The Lancet **382:**209-222.
6. **Allocati N, Masulli M, Alexeyev MF, Di Ilio C.** 2013. *Escherichia coli* in Europe: an overview. Int J Environ Res Public Health **10:**6235-6254.
7. **Steffen R, Hill DR, DuPont HL.** 2015. Traveler's diarrhea: a clinical review. J Am Med Assoc **313:**71-80.
8. **Huang DB, Okhuysen PC, Jiang ZD, DuPont HL.** 2004. Enteroaggregative *Escherichia coli:* an emerging enteric pathogen. Am J Gastroenterol **99:**383-389.
9. **Page AV, Liles WC.** 2013. Enterohemorrhagic *Escherichia coli* Infections and the Hemolytic-Uremic Syndrome. Med Clin North Am **97:**681-695, xi.
10. **Gonzales L, Joffre E, Rivera R, Sjoling A, Svennerholm AM, Iniguez V.** 2013. Prevalence, seasonality and severity of disease caused by pathogenic *Escherichia coli* in children with diarrhoea in Bolivia. J Med Microbiol **62:**1697-1706.
11. **Black RE, Merson MH, Rowe B, Taylor PR, Abdul Alim AR, Gross RJ, Sack DA.** 1981. Enterotoxigenic *Escherichia coli* diarrhoea: acquired immunity and transmission in an endemic area. Bull World Health Organ **59:**263-268.
12. **Stoll BJ, Svennerholm AM, Gothefors L, Barua D, Huda S, Holmgren J.** 1986. Local and systemic antibody responses to naturally acquired enterotoxigenic *Escherichia coli* diarrhea in an endemic area. The Journal of Infectious Diseases **153:**527-534.
13. **Pitzurra R, Steffen R, Tschopp A, Mutsch M.** 2010. Diarrhoea in a large prospective cohort of European travellers to resource-limited destinations. BMC Infect Dis **10:**231.
14. **Steffen R.** 2005. Epidemiology of traveler's diarrhea. Clin Infect Dis **41 Suppl 8:**S536-540.
15. **Belderok SM, van den Hoek A, Kint JA, Schim van der Loeff MF, Sonder GJ.** 2011. Incidence, risk factors and treatment of diarrhoea among Dutch travellers: reasons not to routinely prescribe antibiotics. BMC Infect Dis **11:**295.
16. **Kittitrakul C, Lawpoolsri S, Kusolsuk T, Olanwijitwong J, Tangkanakul W, Piyaphanee W.** 2015. Traveler's Diarrhea in Foreign Travelers in Southeast Asia: A Cross-Sectional Survey Study in Bangkok, Thailand. Am J Trop Med Hyg **93:**485-490.
17. **Spiller R, Garsed K.** 2009. Postinfectious irritable bowel syndrome. Gastroenterology **136:**1979-1988.

18. **Okhuysen PC, Jiang ZD, Carlin L, Forbes C, DuPont HL.** 2004. Post-diarrhea chronic intestinal symptoms and irritable bowel syndrome in North American travelers to Mexico. Am J Gastroenterol **99:**1774-1778.
19. **Kolenda R, Burdukiewicz M, Schierack P.** 2015. A systematic review and meta-analysis of the epidemiology of pathogenic *Escherichia coli* of calves and the role of calves as reservoirs for human pathogenic *E. coli*. Frontiers in cellular and infection microbiology **5:**23.
20. **Chan G, Farzan A, DeLay J, McEwen B, Prescott JF, Friendship RM.** 2013. A retrospective study on the etiological diagnoses of diarrhea in neonatal piglets in Ontario, Canada, between 2001 and 2010. Can J Vet Res **77:**254-260.
21. **Mohlatlole R, Madoroba E, Muchadeyi F, Chimonyo M, Kanengoni A, Dzomba E.** 2013. Virulence profiles of enterotoxigenic, shiga toxin and enteroaggregative *Escherichia coli* in South African pigs. Trop Anim Health Prod **45:**1399-1405.
22. **de la Fé Rodríguez PY, Martin LOM, Muñoz E, Imberechts H, Butaye P, Goddeeris B, Cox E.** 2013. Several enteropathogens are circulating in suckling and newly weaned piglets suffering from diarrhea in the province of Villa Clara, Cuba. Trop Anim Health Prod **45:**435-440.
23. **Okello E, Moonens K, Erume J, De Greve H.** 2015. Enterotoxigenic *Escherichia coli* strains are highly prevalent in Ugandan piggeries but disease outbreaks are masked by antibiotic prophylaxis. Trop Anim Health Prod **47:**117-122.
24. **Ruan X, Zhang W.** 2013. Oral immunization of a live attenuated *Escherichia coli* strain expressing a holotoxin-structured adhesin-toxoid fusion (1FaeG-FedF-LTA(2):5LTB) protected young pigs against enterotoxigenic *E. coli* (ETEC) infection. Vaccine **31:**1458-1463.
25. **Hur J, Lee JH.** 2012. Comparative evaluation of a vaccine candidate expressing enterotoxigenic *Escherichia coli* (ETEC) adhesins for colibacillosis with a commercial vaccine using a pig model. Vaccine **30:**3829-3833.
26. **Melkebeek V, Goddeeris BM, Cox E.** 2013. ETEC vaccination in pigs. Vet Immunol Immunopathol **152:**37-42.
27. **Wolf MK.** 1997. Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*. Clin Microbiol Rev **10:**569-584.
28. **von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffre E, Corander J, Pickard D, Wiklund G, Svennerholm AM, Sjoling A, Dougan G.** 2014. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. Nat Genet **46:**1321-1326.
29. **Escobar-Paramo P, Clermont O, Blanc-Potard AB, Bui H, Le Bouguenec C, Denamur E.** 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol Biol Evol **21:**1085-1094.
30. **Steinsland H, Lacher DW, Sommerfelt H, Whittam TS.** 2010. Ancestral lineages of human enterotoxigenic *Escherichia coli*. J Clin Microbiol **48:**2916-2924.
31. **Qadri F, Svennerholm AM, Faruque AS, Sack RB.** 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev **18:**465-483.
32. **Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, Petty NK, Mahon V, Brinkley C, Hobman JL, Savarino SJ, Turner SM, Pallen MJ, Penn CW, Parkhill J, Turner AK, Johnson TJ, Thomson NR, Smith SG, Henderson IR.** 2010. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. J Bacteriol **192:**5822-5831.
33. **Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J.** 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol **190:**6881-6893.

34. **Gaastra W, Svennerholm AM.** 1996. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). Trends Microbiol **4:**444-452.
35. **Isidean SD, Riddle MS, Savarino SJ, Porter CK.** 2011. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. Vaccine **29:**6167-6178.
36. **Bodero MD, Harden EA, Munson GP.** 2008. Transcriptional regulation of subclass 5b fimbriae. BMC Microbiol **8:**180.
37. **Anantha RP, McVeigh AL, Lee LH, Agnew MK, Cassels FJ, Scott DA, Whittam TS, Savarino SJ.** 2004. Evolutionary and functional relationships of colonization factor antigen I and other class 5 adhesive fimbriae of enterotoxigenic *Escherichia coli*. Infect Immun **72:**7190-7201.
38. **Gaastra W, Sommerfelt H, van Dijk L, Kusters JG, Svennerholm AM, Grewal HM.** 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.
39. **Li YF, Poole S, Nishio K, Jang K, Rasulova F, McVeigh A, Savarino SJ, Xia D, Bullitt E.** 2009. Structure of CFA/I fimbriae from enterotoxigenic *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **106:**10793-10798.
40. **Andersson M, Bjornham O, Svantesson M, Badahdah A, Uhlin BE, Bullitt E.** 2012. A structural basis for sustained bacterial adhesion: biomechanical properties of CFA/I pili. J Mol Biol **415:**918-928.
41. **Mol O, Oudega B.** 1996. Molecular and structural aspects of fimbriae biosynthesis and assembly in *Escherichia coli*. FEMS Microbiol Rev **19:**25-52.
42. **Chattopadhyay S, Tchesnokova V, McVeigh A, Kisiela DI, Dori K, Navarro A, Sokurenko EV, Savarino SJ.** 2012. Adaptive evolution of class 5 fimbrial genes in enterotoxigenic *Escherichia coli* and its functional consequences. J Biol Chem **287:**6150-6158.
43. **Levine MM, Ristaino P, Marley G, Smyth C, Knutton S, Boedeker E, Black R, Young C, Clements ML, Cheney C, et al.** 1984. Coli surface antigens 1 and 3 of colonization factor antigen II-positive enterotoxigenic *Escherichia coli*: morphology, purification, and immune responses in humans. Infect Immun **44:**409-420.
44. **Ghosal A, Bhowmick R, Banerjee R, Ganguly S, Yamasaki S, Ramamurthy T, Hamabata T, Chatterjee NS.** 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:**2125-2135.
45. **Roy SP, Rahman MM, Yu XD, Tuittila M, Knight SD, Zavialov AV.** 2012. Crystal structure of enterotoxigenic *Escherichia coli* colonization factor CS6 reveals a novel type of functional assembly. Mol Microbiol **86:**1100-1115.
46. **Taniguchi T, Fujino Y, Yamamoto K, Miwatani T, Honda T.** 1995. Sequencing of the gene encoding the major pilin of pilus colonization factor antigen III (CFA/III) of human enterotoxigenic *Escherichia coli* and evidence that CFA/III is related to type IV pili. Infect Immun **63:**724-728.
47. **Nishimura LS, Giron JA, Nunes SL, Guth BE.** 2002. Prevalence of enterotoxigenic *Escherichia coli* strains harboring the longus pilus gene in Brazil. J Clin Microbiol **40:**2606-2608.
48. **Porter CK, Riddle MS, Tribble DR, Louis Bougeois A, McKenzie R, Isidean SD, Sebeny P, Savarino SJ.** 2011. A systematic review of experimental infections with enterotoxigenic *Escherichia coli* (ETEC). Vaccine **29:**5869-5885.
49. **Aimoto S, Takao T, Shimonishi Y, Hara S, Takeda T, Takeda Y, Miwatani T.** 1982. Amino-acid sequence of a heat-stable enterotoxin produced by human enterotoxigenic *Escherichia coli*. Eur J Biochem **129:**257-263.
50. **Chan SK, Giannella RA.** 1981. Amino acid sequence of heat-stable enterotoxin produced by *Escherichia coli* pathogenic for man. J Biol Chem **256:**7744-7746.

51. **Shimonishi Y, Hidaka Y, Koizumi M, Hane M, Aimoto S, Takeda T, Miwatani T, Takeda Y.** 1987. Mode of disulfide bond formation of a heat-stable enterotoxin (STh) produced by a human strain of enterotoxigenic *Escherichia coli*. FEBS Lett **215:**165-170.
52. **Schulz S, Green CK, Yuen PS, Garbers DL.** 1990. Guanylyl cyclase is a heat-stable enterotoxin receptor. Cell **63:**941-948.
53. **Chao AC, de Sauvage FJ, Dong YJ, Wagner JA, Goeddel DV, Gardner P.** 1994. Activation of intestinal CFTR Cl- channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. EMBO J **13:**1065-1072.
54. **Currie MG, Fok KF, Kato J, Moore RJ, Hamra FK, Duffin KL, Smith CE.** 1992. Guanylin: an endogenous activator of intestinal guanylate cyclase. Proceedings of the National Academy of Sciences of the United States of America **89:**947-951.
55. **Marx UC, Klodt J, Meyer M, Gerlach H, Rosch P, Forssmann WG, Adermann K.** 1998. One peptide, two topologies: structure and interconversion dynamics of human uroguanylin isomers. J Pept Res **52:**229-240.
56. **Gill DM, Richardson SH.** 1980. Adenosine diphosphate-ribosylation of adenylate cyclase catalyzed by heat-labile enterotoxin of *Escherichia coli*: comparison with cholera toxin. J Infect Dis **141:**64-70.
57. **Sixma TK, Kalk KH, van Zanten BA, Dauter Z, Kingma J, Witholt B, Hol WG.** 1993. Refined structure of *Escherichia coli* heat-labile enterotoxin, a close relative of cholera toxin. J Mol Biol **230:**890-918.
58. **Mudrak B, Kuehn MJ.** 2010. Heat-labile enterotoxin: beyond GM1 binding. Toxins (Basel) **2:**1445-1470.
59. **Taxt A, Aasland R, Sommerfelt H, Nataro J, Puntervoll P.** 2010. Heat-stable enterotoxin of enterotoxigenic *Escherichia coli* as a vaccine target. Infect Immun **78:**1824-1831.
60. **Zenser TV, Metzger JF.** 1974. Comparison of the action of *Escherichia coli* enterotoxin on the thymocyte adenylate cyclase-cyclic adenosine monophosphate system to that of cholera toxin and prostaglandin E1. Infect Immun **10:**503-509.
61. **Elsinghorst EA, Kopecko DJ.** 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli*. Infect Immun **60:**2409-2417.
62. **Fleckenstein JM, Kopecko DJ, Warren RL, Elsinghorst EA.** 1996. Molecular characterization of the tia invasion locus from enterotoxigenic *Escherichia coli*. Infect Immun **64:**2256-2265.
63. **Mammarappallil JG, Elsinghorst EA.** 2000. Epithelial cell adherence mediated by the enterotoxigenic *Escherichia coli* Tia protein. Infect Immun **68:**6595-6601.
64. **Elsinghorst EA, Weitz JA.** 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:**3463-3471.
65. **Sherlock O, Vejborg RM, Klemm P.** 2005. The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. Infect Immun **73:**1954-1963.
66. **Lindenthal C, Elsinghorst EA.** 2001. Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestine epithelial cells. Infect Immun **69:**52-57.
67. **Lindenthal C, Elsinghorst EA.** 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. Infect Immun **67:**4084-4091.
68. **Espert SM, Elsinghorst EA, Munson GP.** 2011. The tib adherence locus of enterotoxigenic *Escherichia coli* is regulated by cyclic AMP receptor protein. J Bacteriol **193:**1369-1376.
69. **Meli AC, Kondratova M, Molle V, Coquet L, Kajava AV, 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:**143-154.
70. **Roy K, Hilliard GM, Hamilton DJ, Luo J, Ostmann MM, Fleckenstein JM.** 2009. Enterotoxigenic *Escherichia coli* EtpA mediates adhesion between flagella and host cells. Nature **457:**594-598.

71. **Roy K, Hamilton D, Ostmann MM, Fleckenstein JM.** 2009. Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic *Escherichia coli* in a murine model. Vaccine **27:**4601-4608.
72. **Sheikh A, Luo Q, Roy K, Shabaan S, Kumar P, Qadri F, Fleckenstein JM.** 2014. Contribution of the highly conserved EaeH surface protein to enterotoxigenic *Escherichia coli* pathogenesis. Infect Immun **82:**3657-3666.
73. **Fleckenstein JM, Munson GM, Rasko DA.** 2013. Enterotoxigenic *Escherichia coli*: Orchestrated host engagement. Gut Microbes **4:**392-396.
74. **Gutierrez D, Pardo M, Montero D, Onate A, Farfan MJ, Ruiz-Perez F, Del Canto F, Vidal R.** 2015. TleA, a Tsh-like autotransporter identified in a human enterotoxigenic *Escherichia coli* strain. Infect Immun **83:**1893-1903.
75. **Savarino SJ, Fasano A, Watson J, Martin BM, Levine MM, Guandalini S, Guerry P.** 1993. Enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 represents another subfamily of *E. coli* heat-stable toxin. Proceedings of the National Academy of Sciences of the United States of America **90:**3093-3097.
76. **Savarino SJ, McVeigh A, Watson J, Cravioto A, Molina J, Echeverria P, Bhan MK, Levine MM, Fasano A.** 1996. Enteroaggregative *Escherichia coli* heat-stable enterotoxin is not restricted to enteroaggregative *E. coli*. J Infect Dis **173:**1019-1022.
77. **McGuckin MA, Linden SK, Sutton P, Florin TH.** 2011. Mucin dynamics and enteric pathogens. Nat Rev Microbiol **9:**265-278.
78. **Patel SK, Dotson J, Allen KP, Fleckenstein JM.** 2004. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. Infect Immun **72:**1786-1794.
79. **Kumar P, Luo Q, Vickers TJ, Sheikh A, Lewis WG, Fleckenstein JM.** 2014. EatA, an immunogenic protective antigen of enterotoxigenic *Escherichia coli*, degrades intestinal mucin. Infect Immun **82:**500-508.
80. **Luo Q, Kumar P, Vickers TJ, Sheikh A, Lewis WG, Rasko DA, Sistrunk J, Fleckenstein JM.** 2014. Enterotoxigenic *Escherichia coli* secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. Infect Immun **82:**509-521.
81. **Dharmani P, Srivastava V, Kissoon-Singh V, Chadee K.** 2009. Role of intestinal mucins in innate host defense mechanisms against pathogens. J Innate Immun **1:**123-135.
82. **Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, Boutin A, Rose DJ, Blattner FR, Nataro JP.** 2003. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. J Biol Chem **278:**45680-45689.
83. **Sheikh J, Czeczulin JR, Harrington S, Hicks S, Henderson IR, Le Bouguenec C, Gounon P, Phillips A, Nataro JP.** 2002. A novel dispersin protein in enteroaggregative *Escherichia coli*. J Clin Invest **110:**1329-1337.
84. **Fleckenstein JM, Lindler LE, Elsinghorst EA, Dale JB.** 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:**2766-2774.
85. **Michie KA, Boysen A, Low HH, Moller-Jensen J, Lowe J.** 2014. LeoA, B and C from enterotoxigenic Escherichia coli (ETEC) are bacterial dynamins. PLoS One **9:**e107211.
86. **Krause WJ, Cullingford GL, Freeman RH, Eber SL, Richardson KC, Fok KF, Currie MG, Forte LR.** 1994. Distribution of heat-stable enterotoxin/guanylin receptors in the intestinal tract of man and other mammals. J Anat **184 (Pt 2):**407-417.
87. **Lim CJ, Lee SY, Kenney LJ, Yan J.** 2012. Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. Sci Rep **2:**509.
88. **Ochman H, Lawrence JG, Groisman EA.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.
89. **Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JC.** 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog **2:**e81.

90. **Stoebel DM, Free A, Dorman CJ.** 2008. Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. Microbiology **154:**2533-2545.
91. **Porter ME, Dorman CJ.** 1994. A role for H-NS in the thermo-osmotic regulation of virulence gene expression in *Shigella flexneri*. J Bacteriol **176:**4187-4191.
92. **Yang J, Tauschek M, Strugnell R, Robins-Browne RM.** 2005. The H-NS protein represses transcription of the *eltAB* operon, which encodes heat-labile enterotoxin in enterotoxigenic *Escherichia coli*, by binding to regions downstream of the promoter. Microbiology **151:**1199-1208.
93. **Haycocks JR, Sharma P, Stringer AM, Wade JT, Grainger DC.** 2015. The molecular basis for control of ETEC enterotoxin expression in response to environment and host. PLoS Pathog **11:**e1004605.
94. **Trachman JD, Maas WK.** 1998. Temperature regulation of heat-labile enterotoxin (LT) synthesis in *Escherichia coli* is mediated by an interaction of H-NS protein with the LT A-subunit DNA. J Bacteriol **180:**3715-3718.
95. **Kansal R, Rasko DA, Sahl JW, Munson GP, Roy K, Luo Q, Sheikh A, Kuhne KJ, Fleckenstein JM.** 2013. Transcriptional modulation of enterotoxigenic *Escherichia coli* virulence genes in response to epithelial cell interactions. Infect Immun **81:**259-270.
96. **Notley-McRobb L, Death A, Ferenci T.** 1997. The relationship between external glucose concentration and cAMP levels inside *Escherichia coli*: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. Microbiology **143 (Pt 6):**1909-1918.
97. **Green J, Stapleton MR, Smith LJ, Artymiuk PJ, Kahramanoglou C, Hunt DM, Buxton RS.** 2014. Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. Curr Opin Microbiol **18:**1-7.
98. **Bodero MD, Munson GP.** 2009. Cyclic AMP receptor protein-dependent repression of heat-labile enterotoxin. Infect Immun **77:**791-798.
99. **Gonzales L, Ali ZB, Nygren E, Wang Z, Karlsson S, Zhu B, Quiding-Jarbrink M, Sjoling A.** 2013. Alkaline pH Is a signal for optimal production and secretion of the heat labile toxin, LT in enterotoxigenic *Escherichia coli* (ETEC). PLoS One **8:**e74069.
100. **Kim SH, Lee BY, Lau GW, Cho YH.** 2009. IscR modulates catalase A (KatA) activity, peroxide resistance and full virulence of *Pseudomonas aeruginosa* PA14. J Microbiol Biotechnol **19:**1520-1526.
101. **Miller HK, Kwuan L, Schwiesow L, Bernick DL, Mettert E, Ramirez HA, Ragle JM, Chan PP, Kiley PJ, Lowe TM, Auerbuch V.** 2014. IscR is essential for *yersinia pseudotuberculosis* type III secretion and virulence. PLoS Pathog **10:**e1004194.
102. **Lim JG, Choi SH.** 2014. IscR is a global regulator essential for pathogenesis of *Vibrio vulnificus* and induced by host cells. Infect Immun **82:**569-578.
103. **Haines S, Arnaud-Barbe N, Poncet D, Reverchon S, Wawrzyniak J, Nasser W, Renauld-Mongenie G.** 2015. IscR Regulates Synthesis of Colonization Factor Antigen I Fimbriae in Response to Iron Starvation in Enterotoxigenic *Escherichia coli*. J Bacteriol **197:**2896-2907.
104. **Skaar EP.** 2010. The battle for iron between bacterial pathogens and their vertebrate hosts. PLoS Pathog **6:**e1000949.
105. **Olier M, Rousseaux S, Piveteau P, Lemaitre JP, Rousset A, Guzzo J.** 2004. Screening of glutamate decarboxylase activity and bile salt resistance of human asymptomatic carriage, clinical, food, and environmental isolates of *Listeria monocytogenes*. Int J Food Microbiol **93:**87-99.
106. **Pumbwe L, Skilbeck CA, Nakano V, Avila-Campos MJ, Piazza RM, Wexler HM.** 2007. Bile salts enhance bacterial co-aggregation, bacterial-intestinal epithelial cell adhesion, biofilm formation and antimicrobial resistance of *Bacteroides fragilis*. Microb Pathog **43:**78-87.
107. **Schuhmacher DA, Klose KE.** 1999. Environmental signals modulate ToxT-dependent virulence factor expression in *Vibrio cholerae*. J Bacteriol **181:**1508-1514.

108. **Sahl JW, Rasko DA.** 2012. Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A. Infect Immun **80:**1232-1242.
109. **Sully EK, Malachowa N, Elmore BO, Alexander SM, Femling JK, Gray BM, DeLeo FR, Otto M, Cheung AL, Edwards BS, Sklar LA, Horswill AR, Hall PR, Gresham HD.** 2014. Selective chemical inhibition of agr quorum sensing in *Staphylococcus aureus* promotes host defense with minimal impact on resistance. PLoS Pathog **10:**e1004174.
110. **O'Loughlin CT, Miller LC, Siryaporn A, Drescher K, Semmelhack MF, Bassler BL.** 2013. A quorum-sensing inhibitor blocks *Pseudomonas aeruginosa* virulence and biofilm formation. Proc Natl Acad Sci U S A **110:**17981-17986.
111. **Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL, Mekalanos JJ.** 2002. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. Proc Natl Acad Sci U S A **99:**3129-3134.
112. **Caron J, Coffield LM, Scott JR.** 1989. A plasmid-encoded regulatory gene, rns, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **86:**963-967.
113. **Caron J, Scott JR.** 1990. A *rns*-like regulatory gene for colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. Infect Immun **58:**874-878.
114. **Favre D, Ludi S, Stoffel M, Frey J, Horn MP, Dietrich G, Spreng S, Viret JF.** 2006. Expression of enterotoxigenic *Escherichia coli* colonization factors in *Vibrio cholerae*. Vaccine **24:**4354-4368.
115. **Munson GP, Holcomb LG, Scott JR.** 2001. Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. Infect Immun **69:**186-193.
116. **Pilonieta MC, Bodero MD, Munson GP.** 2007. CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic *Escherichia coli*. J Bacteriol **189:**5060-5067.
117. **Munson GP, Holcomb LG, Alexander HL, Scott JR.** 2002. In vitro identification of Rns-regulated genes. J Bacteriol **184:**1196-1199.
118. **Bodero MD, Pilonieta MC, Munson GP.** 2007. Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic *Escherichia coli*. J Bacteriol **189:**1627-1632.
119. **Savelkoul PH, Willshaw GA, McConnell MM, Smith HR, Hamers AM, van der Zeijst BA, Gaastra W.** 1990. Expression of CFA/I fimbriae is positively regulated. Microb Pathog **8:**91-99.
120. **Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL.** 1997. Arac/XylS family of transcriptional regulators. Microbiol Mol Biol Rev **61:**393-410.
121. **Martin RG, Rosner JL.** 2001. The AraC transcriptional activators. Curr Opin Microbiol **4:**132-137.
122. **Rhee S, Martin RG, Rosner JL, Davies DR.** 1998. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. Proceedings of the National Academy of Sciences of the United States of America **95:**10413-10418.
123. **Basturea GN, Bodero MD, Moreno ME, Munson GP.** 2008. Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding. J Bacteriol **190:**2279-2285.
124. **Mahon V, Smyth CJ, Smith SGJ.** 2010. Mutagenesis of the Rns regulator of enterotoxigenic *Escherichia coli* reveals roles for a linker sequence and two helix–turn–helix motifs. Microbiology **156:**2796-2806.
125. **Mahon V, Fagan RP, Smith SG.** 2012. Snap denaturation reveals dimerization by AraC-like protein Rns. Biochimie **94:**2058-2061.
126. **Yang J, Dogovski C, Hocking D, Tauschek M, Perugini M, Robins-Browne RM.** 2009. Bicarbonate-mediated stimulation of RegA, the global virulence regulator from *Citrobacter rodentium*. J Mol Biol **394:**591-599.
127. **Yang J, Tauschek M, Robins-Browne RM.** 2011. Control of bacterial virulence by AraC-like regulators that respond to chemical signals. Trends Microbiol **19:**128-135.

128. **Murphree D, Froehlich B, Scott JR.** 1997. Transcriptional control of genes encoding CS1 pili: negative regulation by a silencer and positive regulation by Rns. J Bacteriol **179:**5736-5743.
129. **Munson GP, Scott JR.** 1999. Binding site recognition by Rns, a virulence regulator in the AraC family. J Bacteriol **181:**2110-2117.
130. **Munson GP, Scott JR.** 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:**1391-1402.
131. **Jordi BJ, Dagberg B, de Haan LA, Hamers AM, van der Zeijst BA, Gaastra W, Uhlin BE.** 1992. The positive regulator CfaD overcomes the repression mediated by histone-like protein H-NS (H1) in the CFA/I fimbrial operon of *Escherichia coli*. The EMBO Journal **11:**2627-2632.
132. **Yang J, Hart E, Tauschek M, Price GD, Hartland EL, Strugnell RA, Robins-Browne RM.** 2008. Bicarbonate-mediated transcriptional activation of divergent operons by the virulence regulatory protein, RegA, from *Citrobacter rodentium*. Mol Microbiol **68:**314-327.
133. **de Haan LA, Willshaw GA, van der Zeijst BA, Gaastra W.** 1991. The nucleotide sequence of a regulatory gene present on a plasmid in an enterotoxigenic *Escherichia coli* strain of serotype O167:H5. FEMS Microbiol Lett **67:**341-346.
134. **Honarvar S, Choi BK, Schifferli DM.** 2003. Phase variation of the 987P-like CS18 fimbriae of human enterotoxigenic *Escherichia coli* is regulated by site-specific recombinases. Mol Microbiol **48:**157-171.
135. **Valvatne H, Sommerfelt H, Gaastra W, Bhan MK, Grewal HM.** 1996. Identification and characterization of CS20, a new putative colonization factor of enterotoxigenic *Escherichia coli*. Infect Immun **64:**2635-2642.
136. **Gomez-Duarte OG, Chattopadhyay S, Weissman SJ, Giron JA, Kaper JB, Sokurenko EV.** 2007. Genetic diversity of the gene cluster encoding longus, a type IV pilus of enterotoxigenic *Escherichia coli*. J Bacteriol **189:**9145-9149.
137. **Del Canto F, Botkin DJ, Valenzuela P, Popov V, Ruiz-Perez F, Nataro JP, Levine MM, Stine OC, Pop M, Torres AG, Vidal R.** 2012. Identification of Coli Surface Antigen 23, a novel adhesin of enterotoxigenic *Escherichia coli*. Infect Immun **80:**2791-2801.
138. **Froehlich B, Husmann L, Caron J, Scott JR.** 1994. Regulation of rns, a positive regulatory factor for pili of enterotoxigenic *Escherichia coli*. J Bacteriol **176:**5385-5392.
139. **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.
140. **Schwechheimer C, Kulp A, Kuehn MJ.** 2014. Modulation of bacterial outer membrane vesicle production by envelope structure and content. BMC Microbiol **14:**324.
141. **Santiago AE, Ruiz-Perez F, Jo NY, Vijayakumar V, Gong MQ, Nataro JP.** 2014. A large family of antivirulence regulators modulates the effects of transcriptional activators in Gram-negative pathogenic bacteria. PLoS Pathog **10:**e1004153.
142. **Nataro JP, Yikang D, Yingkang D, Walker K.** 1994. AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. J Bacteriol **176:**4691-4699.
143. **Hart E, Yang J, Tauschek M, Kelly M, Wakefield MJ, Frankel G, Hartland EL, Robins-Browne RM.** 2008. RegA, an AraC-like protein, is a global transcriptional regulator that controls virulence gene expression in *Citrobacter rodentium*. Infect Immun **76:**5247-5256.
144. **Higgins DE, Nazareno E, DiRita VJ.** 1992. The virulence gene activator ToxT from *Vibrio cholerae* is a member of the AraC family of transcriptional activators. J Bacteriol **174:**6974-6980.
145. **Dorman CJ.** 1992. The VirF protein from *Shigella flexneri* is a member of the AraC transcription factor superfamily and is highly homologous to Rns, a positive regulator of virulence genes in enterotoxigenic *Escherichia coli*. Mol Microbiol **6:**1575.

146. **Morin N, Santiago AE, Ernst RK, Guillot SJ, Nataro JP.** 2013. Characterization of the AggR regulon in enteroaggregative *Escherichia coli*. Infect Immun **81:**122-132.
147. **DiRita VJ, Parsot C, Jander G, Mekalanos JJ.** 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. Proceedings of the National Academy of Sciences of the United States of America **88:**5403-5407.
148. **Binder HJ, Brown I, Ramakrishna BS, Young GP.** 2014. Oral rehydration therapy in the second decade of the twenty-first century. Current gastroenterology reports **16:**376.
149. **Pierce NF, Banwell JG, Rupak DM, Mitra RC, Caranasos GJ, Keimowitz RI, Mondal A, Manji PM.** 1968. Effect of intragastric glucose-electrolyte infusion upon water and electrolyte balance in Asiatic cholera. Gastroenterology **55:**333-343.
150. **Curran PF.** 1960. Na, Cl, and water transport by rat ileum in vitro. J Gen Physiol **43:**1137-1148.
151. **Levy JA, Bachur RG, Monuteaux MC, Waltzman M.** 2013. Intravenous dextrose for children with gastroenteritis and dehydration: a double-blind randomized controlled trial. Ann Emerg Med **61:**281-288.
152. **Thillainayagam AV, Hunt JB, Farthing MJ.** 1998. Enhancing clinical efficacy of oral rehydration therapy: is low osmolality the key? Gastroenterology **114:**197-210.
153. **Sazawal S, Black RE, Bhan MK, Bhandari N, Sinha A, Jalla S.** 1995. Zinc supplementation in young children with acute diarrhea in India. N Engl J Med **333:**839-844.
154. **Bhutta ZA, Bird SM, Black RE, Brown KH, Gardner JM, Hidayat A, Khatun F, Martorell R, Ninh NX, Penny ME, Rosado JL, Roy SK, Ruel M, Sazawal S, Shankar A.** 2000. Therapeutic effects of oral zinc in acute and persistent diarrhea in children in developing countries: pooled analysis of randomized controlled trials. Am J Clin Nutr **72:**1516-1522.
155. **Scrimgeour AG, Lukaski HC.** 2008. Zinc and diarrheal disease: current status and future perspectives. Curr Opin Clin Nutr Metab Care **11:**711-717.
156. **Walker CL, Black RE.** 2010. Zinc for the treatment of diarrhoea: effect on diarrhoea morbidity, mortality and incidence of future episodes. Int J Epidemiol **39 Suppl 1:**i63-69.
157. **Berni Canani R, Buccigrossi V, Passariello A.** 2011. Mechanisms of action of zinc in acute diarrhea. Current Opinion in Gastroenterology **27:**8-12.
158. **Canani RB, Cirillo P, Buccigrossi V, Ruotolo S, Passariello A, De Luca P, Porcaro F, De Marco G, Guarino A.** 2005. Zinc inhibits cholera toxin-induced, but not *Escherichia coli* heat-stable enterotoxin-induced, ion secretion in human enterocytes. The Journal of Infectious Diseases **191:**1072-1077.
159. **Hoque KM, Rajendran VM, Binder HJ.** 2005. Zinc inhibits cAMP-stimulated Cl secretion via basolateral K-channel blockade in rat ileum. Am J Physiol Gastrointest Liver Physiol **288:**G956-963.
160. **Kantele A.** 2015. A call to restrict prescribing antibiotics for travellers' diarrhea--Travel medicine practitioners can play an active role in preventing the spread of antimicrobial resistance. Travel Med Infect Dis **13:**213-214.
161. **Ruiz J, Pons MJ.** 2013. Prevention of travellers' diarrhoea: where and who? Lancet Infect Dis **13:**911-912.
162. **Salam I, Katelaris P, Leigh-Smith S, Farthing MJ.** 1994. Randomised trial of single-dose ciprofloxacin for travellers' diarrhoea. The Lancet **344:**1537-1539.
163. **Taylor DN, Bourgeois AL, Ericsson CD, Steffen R, Jiang ZD, Halpern J, Haake R, Dupont HL.** 2006. A randomized, double-blind, multicenter study of rifaximin compared with placebo and with ciprofloxacin in the treatment of travelers' diarrhea. Am J Trop Med Hyg **74:**1060-1066.
164. **Zanger P, Nurjadi D, Gabor J, Gaile M, Kremsner PG.** 2013. Effectiveness of rifaximin in prevention of diarrhoea in individuals travelling to south and southeast Asia: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Infect Dis **13:**946-954.

165. **Pakyz AL.** 2005. Rifaximin: a new treatment for travelers' diarrhea. Ann Pharmacother **39:**284-289.
166. **DuPont HL.** 2016. Review article: the antimicrobial effects of rifaximin on the gut microbiota. Aliment Pharmacol Ther **43 Suppl 1:**3-10.
167. **Rodas C, Mamani R, Blanco J, Blanco JE, Wiklund G, Svennerholm AM, Sjoling A, Iniguez V.** 2011. Enterotoxins, colonization factors, serotypes and antimicrobial resistance of enterotoxigenic *Escherichia coli* (ETEC) strains isolated from hospitalized children with diarrhea in Bolivia. Braz J Infect Dis **15:**132-137.
168. **Meng CY, Smith BL, Bodhidatta L, Richard SA, Vansith K, Thy B, Srijan A, Serichantalergs O, Mason CJ.** 2011. Etiology of diarrhea in young children and patterns of antibiotic resistance in Cambodia. Pediatr Infect Dis J **30:**331-335.
169. **Talukdar PK, Rahman M, Rahman M, Nabi A, Islam Z, Hoque MM, Endtz HP, Islam MA.** 2013. Antimicrobial Resistance, Virulence Factors and Genetic Diversity of *Escherichia coli* Isolates from Household Water Supply in Dhaka, Bangladesh. PLoS One **8:**e61090.
170. **Pazhani GP, Chakraborty S, Fujihara K, Yamasaki S, Ghosh A, Nair GB, Ramamurthy T.** 2011. QRDR mutations, efflux system & antimicrobial resistance genes in enterotoxigenic *Escherichia coli* isolated from an outbreak of diarrhoea in Ahmedabad, India. The Indian Journal of Medical Research **134:**214-223.
171. **Haghi F, Zeighami H, Hajiahmadi F, Khoshvaght H, Bayat M.** 2014. Frequency and antimicrobial resistance of diarrhoeagenic *Escherichia coli* from young children in Iran. J Med Microbiol **63:**427-432.
172. **Nontongana N, Sibanda T, Ngwenya E, Okoh AI.** 2014. Prevalence and Antibiogram Profiling of *Escherichia coli* Pathotypes Isolated from the Kat River and the Fort Beaufort Abstraction Water. Int J Environ Res Public Health **11:**8213-8227.
173. **Guerra JA, Romero-Herazo YC, Arzuza O, Gómez-Duarte OG.** 2014. Phenotypic and Genotypic Characterization of Enterotoxigenic *Escherichia coli* Clinical Isolates from Northern Colombia, South America. BioMed Research International **2014:**236260.
174. **Medina AM, Rivera FP, Pons MJ, Riveros M, Gomes C, Bernal M, Meza R, Maves RC, Huicho L, Chea-Woo E, Lanata CF, Gil AI, Ochoa TJ, Ruiz J.** 2015. Comparative analysis of antimicrobial resistance in enterotoxigenic *Escherichia coli* isolates from two paediatric cohort studies in Lima, Peru. Trans R Soc Trop Med Hyg **109:**493-502.
175. **Kartsev NN, Fursova NK, Pachkunov DM, Bannov VA, Eruslanov BV, Svetoch EA, Dyatlov IA.** 2015. Molecular characterization of enterotoxin-producing *Escherichia coli c*ollected in 2011-2012, Russia. PLoS One **10:**e0123357.
176. **Mendez Arancibia E, Pitart C, Ruiz J, Marco F, Gascon J, Vila J.** 2009. Evolution of antimicrobial resistance in enteroaggregative *Escherichia coli* and enterotoxigenic *Escherichia coli* causing traveller's diarrhoea. J Antimicrob Chemother **64:**343-347.
177. **WHO.** 2014. Antimicrobial resistance: global report on surveillance. World Health Organ Tech Rep Ser.
178. **da Hora VP, Conceicao FR, Dellagostin OA, Doolan DL.** 2011. Non-toxic derivatives of LT as potent adjuvants. Vaccine **29:**1538-1544.
179. **Jertborn M, Ahren C, Holmgren J, Svennerholm AM.** 1998. Safety and immunogenicity of an oral inactivated enterotoxigenic *Escherichia coli* vaccine. Vaccine **16:**255-260.
180. **Clemens JD, Sack DA, Harris JR, Chakraborty J, Neogy PK, Stanton B, Huda N, Khan MU, Kay BA, Khan MR, et al.** 1988. Cross-protection by B subunit-whole cell cholera vaccine against diarrhea associated with heat-labile toxin-producing enterotoxigenic *Escherichia coli*: results of a large-scale field trial. J Infect Dis **158:**372-377.
181. **Qadri F, Wenneras C, Ahmed F, Asaduzzaman M, Saha D, Albert MJ, Sack RB, Svennerholm A.** 2000. Safety and immunogenicity of an oral, inactivated enterotoxigenic *Escherichia coli* plus cholera toxin B subunit vaccine in Bangladeshi adults and children. Vaccine **18:**2704-2712.

182. **Sack DA, Shimko J, Torres O, Bourgeois AL, Francia DS, Gustafsson B, Karnell A, Nyquist I, Svennerholm AM.** 2007. Randomised, double-blind, safety and efficacy of a killed oral vaccine for enterotoxigenic *E. coli* diarrhoea of travellers to Guatemala and Mexico. Vaccine **25:**4392-4400.
183. **Lundgren A, Leach S, Tobias J, Carlin N, Gustafsson B, Jertborn M, Bourgeois L, Walker R, Holmgren J, Svennerholm AM.** 2013. Clinical trial to evaluate safety and immunogenicity of an oral inactivated enterotoxigenic *Escherichia coli* prototype vaccine containing CFA/I overexpressing bacteria and recombinantly produced LTB/CTB hybrid protein. Vaccine **31:**1163-1170.
184. **Lundgren A, Bourgeois L, Carlin N, Clements J, Gustafsson B, Hartford M, Holmgren J, Petzold M, Walker R, Svennerholm AM.** 2014. Safety and immunogenicity of an improved oral inactivated multivalent enterotoxigenic *Escherichia coli* (ETEC) vaccine administered alone and together with dmLT adjuvant in a double-blind, randomized, placebo-controlled Phase I study. Vaccine **32:**7077-7084.
185. **Harro C, Sack D, Bourgeois AL, Walker R, DeNearing B, Feller A, Chakraborty S, Buchwaldt C, Darsley MJ.** 2011. A combination vaccine consisting of three live attenuated enterotoxigenic *Escherichia coli* strains expressing a range of colonization factors and heat-labile toxin subunit B is well tolerated and immunogenic in a placebo-controlled double-blind phase I trial in healthy adults. Clin Vaccine Immunol **18:**2118-2127.
186. **Turner AK, Stephens JC, Beavis JC, Greenwood J, Gewert C, Randall R, Freeman D, Darsley MJ.** 2011. Generation and characterization of a live attenuated enterotoxigenic *Escherichia coli* combination vaccine expressing six colonization factors and heat-labile toxin subunit B. Clin Vaccine Immunol **18:**2128-2135.
187. **Darsley MJ, Chakraborty S, DeNearing B, Sack DA, Feller A, Buchwaldt C, Bourgeois AL, Walker R, Harro CD.** 2012. The oral, live attenuated enterotoxigenic *Escherichia coli* vaccine ACE527 reduces the incidence and severity of diarrhea in a human challenge model of diarrheal disease. Clin Vaccine Immunol **19:**1921-1931.
188. **Ruan X, Knudsen DE, Wollenberg KM, Sack DA, Zhang W.** 2014. Multiepitope Fusion Antigen Induces Broadly Protective Antibodies That Prevent Adherence of *Escherichia coli* Strains Expressing Colonization Factor Antigen I (CFA/I), CFA/II, and CFA/IV. Clin Vaccine Immunol **21:**243-249.
189. **Ruan X, Robertson DC, Nataro JP, Clements JD, Zhang W, Group STTVC.** 2014. Characterization of heat-stable (STa) toxoids of enterotoxigenic *Escherichia coli* fused to double mutant heat-labile toxin peptide in inducing neutralizing Anti-STa antibodies. Infect Immun **82:**1823-1832.
190. **Ruan X, Sack DA, Zhang W.** 2015. Genetic fusions of a CFA/I/II/IV MEFA (multiepitope fusion antigen) and a toxoid fusion of heat-stable toxin (STa) and heat-labile toxin (LT) of enterotoxigenic *Escherichia coli* (ETEC) retain broad anti-CFA and antitoxin antigenicity. PLoS One **10:**e0121623.
191. **Ofek I, Zafriri D, Goldhar J, Eisenstein BI.** 1990. Inability of toxin inhibitors to neutralize enhanced toxicity caused by bacteria adherent to tissue culture cells. Infect Immun **58:**3737-3742.
192. **Stewardson AJ, Huttner B, Harbarth S.** 2011. At least it won't hurt: the personal risks of antibiotic exposure. Curr Opin Pharmacol **11:**446-452.
193. **Allen RC, Popat R, Diggle SP, Brown SP.** 2014. Targeting virulence: can we make evolution-proof drugs? Nat Rev Microbiol **12:**300-308.
194. **Yang J, Hocking DM, Cheng C, Dogovski C, Perugini MA, Holien JK, Parker MW, Hartland EL, Tauschek M, Robins-Browne RM.** 2013. Disarming bacterial virulence through chemical inhibition of the DNA binding domain of an AraC-like transcriptional activator protein. J Biol Chem **288:**31115-31126.

195. **Deng W, Li Y, Vallance BA, Finlay BB.** 2001. Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. Infect Immun **69:**6323-6335.
196. **Hung DT, Shakhnovich EA, Pierson E, Mekalanos JJ.** 2005. Small-Molecule Inhibitor of *Vibrio cholerae* Virulence and Intestinal Colonization. Science **310:**670-674.
197. **Shakhnovich EA, Hung DT, Pierson E, Lee K, Mekalanos JJ.** 2007. Virstatin inhibits dimerization of the transcriptional activator ToxT. Proceedings of the National Academy of Sciences of the United States of America **104:**2372-2377.
198. **Shakhnovich EA, Sturtevant D, Mekalanos JJ.** 2007. Molecular mechanisms of virstatin resistance by non-O1/non-O139 strains of *Vibrio cholerae*. Mol Microbiol **66:**1331-1341.
199. **Clemmer KM, Bonomo RA, Rather PN.** 2011. Genetic analysis of surface motility in *Acinetobacter baumannii*. Microbiology **157:**2534-2544.
200. **Nait Chabane Y, Mlouka MB, Alexandre S, Nicol M, Marti S, Pestel-Caron M, Vila J, Jouenne T, De E.** 2014. Virstatin inhibits biofilm formation and motility of *Acinetobacter baumannii*. BMC Microbiol **14:**62.
201. **Withey JH, Nag D, Plecha SC, Sinha R, Koley H.** 2015. Conjugated linoleic acid reduces cholera toxin production in vitro and in vivo by inhibiting *Vibrio cholerae* ToxT activity. Antimicrob Agents Chemother doi:10.1128/AAC.01029-15.
202. **Plecha SC, Withey JH.** 2015. Mechanism for inhibition of *Vibrio cholerae* ToxT activity by the unsaturated fatty acid components of bile. J Bacteriol **197:**1716-1725.
203. **Skerman FJ, Formal SB, Falkow S.** 1972. Plasmid-associated enterotoxin production in a strain of *Escherichia coli* isolated from humans. Infect Immun **5:**622-624.
204. **Casadaban MJ.** 1976. Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J Mol Biol **104:**541-555.
205. **Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol **185:**60-89.
206. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2:**2006 0008.
207. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America **97:**6640-6645.
208. **Tan A, Yang J, Tauschek M, Praszkier J, Robins-Browne RM.** 2011. Autogenous transcriptional regulation of the regA gene, encoding an AraC-Like, essential virulence regulator in *Citrobacter rodentium*. J Bacteriol **193:**1777-1782.
209. **Praszkier J, Wilson IW, Pittard AJ.** 1992. Mutations affecting translational coupling between the *rep* genes of an IncB miniplasmid. J Bacteriol **174:**2376-2383.
210. **Evans DG, Evans DJ, Jr., Tjoa W.** 1977. Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor. Infect Immun **18:**330-337.
211. **Law CW, Chen Y, Shi W, Smyth GK.** 2014. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **15:**R29.
212. **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK.** 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res **43:**e47.
213. **Chung CT, Niemela SL, Miller RH.** 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proceedings of the National Academy of Sciences of the United States of America **86:**2172-2175.
214. **Lowden MJ, Skorupski K, Pellegrini M, Chiorazzo MG, Taylor RK, Kull FJ.** 2010. Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes.

Proceedings of the National Academy of Sciences of the United States of America **107:**2860-2865.
215. **Soding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33:**W244-248.
216. **Šali A, Potterton L, Yuan F, van Vlijmen H, Karplus M.** 1995. Evaluation of comparative protein modeling by MODELLER. Proteins: Structure, Function, and Bioinformatics **23:**318-326.
217. **Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Kourtesis P, Ledley RS, Suzek BE, Vinayaka CR, Zhang J, Barker WC.** 2003. The Protein Information Resource. Nucleic Acids Res **31:**345-347.
218. **Laskowski RA, MacArthur MW, Moss DS, Thornton JM.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography **26:**283-291.
219. **Sedlock DM, Bartus HF, Zajac I, Actor P.** 1981. Analysis of parameters affecting the hemagglutination activity of *Escherichia coli* possessing colonization factor antigens: improved medium for observing erythrocyte agglutination. J Clin Microbiol **13:**301-308.
220. **Kim S, Jeon T-J, Oberai A, Yang D, Schmidt JJ, Bowie JU.** 2005. Transmembrane glycine zippers: Physiological and pathological roles in membrane proteins. Proceedings of the National Academy of Sciences of the United States of America **102:**14278-14283.
221. **Wang Z, Gerstein M, Snyder M.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10:**57-63.
222. **Fleckenstein JM, Roy K, Fischer JF, Burkitt M.** 2006. Identification of a two-partner secretion locus of enterotoxigenic *Escherichia coli*. Infect Immun **74:**2245-2258.
223. **Fleckenstein J, Sheikh A, Qadri F.** 2014. Novel antigens for enterotoxigenic *Escherichia coli* vaccines. Expert Review of Vaccines **13:**631-639.
224. **Roy K, Hamilton D, Allen KP, Randolph MP, Fleckenstein JM.** 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:**2106-2112.
225. **Baldi DL, Higginson EE, Hocking DM, Praszkier J, Cavaliere R, James CE, Bennett-Wood V, Azzopardi KI, Turnbull L, Lithgow T, Robins-Browne RM, Whitchurch CB, Tauschek M.** 2012. The type II secretion system and its ubiquitous lipoprotein substrate, SslE, are required for biofilm formation and virulence of enteropathogenic *Escherichia coli*. Infect Immun **80:**2042-2052.
226. **Decanio MS, Landick R, Haft RJ.** 2013. The non-pathogenic *Escherichia coli* strain W secretes SslE via the virulence-associated type II secretion system beta. BMC Microbiol **13:**130.
227. **Yang J, Baldi DL, Tauschek M, Strugnell RA, Robins-Browne RM.** 2007. Transcriptional regulation of the *yghJ-pppA-yghG-gspCDEFGHIJKLM* cluster, encoding the type II secretion pathway in enterotoxigenic *Escherichia coli*. J Bacteriol **189:**142-150.
228. **Heras B, Totsika M, Peters KM, Paxman JJ, Gee CL, Jarrott RJ, Perugini MA, Whitten AE, Schembri MA.** 2014. The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping. Proceedings of the National Academy of Sciences of the United States of America **111:**457-462.
229. **Ulett GC, Valle J, Beloin C, Sherlock O, Ghigo JM, Schembri MA.** 2007. Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a role in long-term persistence in the urinary tract. Infect Immun **75:**3233-3244.
230. **de Luna M, Scott-Tucker A, Desvaux M, Ferguson P, Morin NP, Dudley EG, Turner S, Nataro JP, Owen P, Henderson IR.** 2008. The *Escherichia coli* biofilm-promoting protein Antigen 43 does not contribute to intestinal colonization. FEMS Microbiol Lett **284:**237-246.
231. **Haagmans W, van der Woude M.** 2000. Phase variation of Ag43 in *Escherichia coli:* Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. Mol Microbiol **35:**877-887.

232. **Penfold SS, Simon J, Frost LS.** 1996. Regulation of the expression of the traM gene of the F sex factor of *Escherichia coli*. Mol Microbiol **20:**549-558.
233. **Will WR, Lu J, Frost LS.** 2004. The role of H-NS in silencing F transfer gene expression during entry into stationary phase. Mol Microbiol **54:**769-782.
234. **Huddleston JR.** 2014. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. Infection and drug resistance **7:**167-176.
235. **Schultz JE, Matin A.** 1991. Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*. J Mol Biol **218:**129-140.
236. **Nakae T.** 1976. Identification of the outer membrane protein of *E. coli* that produces transmembrane channels in reconstituted vesicle membranes. Biochem Biophys Res Commun **71:**877-884.
237. **Nikaido H, Rosenberg EY.** 1983. Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins. J Bacteriol **153:**241-252.
238. **Nikaido H.** 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev **67:**593-656.
239. **Pratt LA, Hsing W, Gibson KE, Silhavy TJ.** 1996. From acids to *osmZ*: multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli*. Mol Microbiol **20:**911-917.
240. **Horstman AL, Kuehn MJ.** 2000. Enterotoxigenic *Escherichia coli* secretes active heat-labile enterotoxin via outer membrane vesicles. J Biol Chem **275:**12489-12496.
241. **Fitzgerald DM, Bonocora RP, Wade JT.** 2014. Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. PLoS Genet **10:**e1004649.
242. **Soutourina OA, Bertin PN.** 2003. Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol Rev **27:**505-523.
243. **Soutourina O, Kolb A, Krin E, Laurent-Winter C, Rimsky S, Danchin A, Bertin P.** 1999. Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. J Bacteriol **181:**7500-7508.
244. **Ko M, Park C.** 2000. H-NS-Dependent regulation of flagellar synthesis is mediated by a LysR family protein. J Bacteriol **182:**4670-4672.
245. **Paul K, Carlquist WC, Blair DF.** 2011. Adjusting the spokes of the flagellar motor with the DNA-binding protein H-NS. J Bacteriol **193:**5914-5922.
246. **Daley DO, Rapp M, Granseth E, Melen K, Drew D, von Heijne G.** 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. Science **308:**1321-1323.
247. **Galas DJ, Schmitz A.** 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res **5:**3157-3170.
248. **Pu WT, Struhl K.** 1992. Uracil interference, a rapid and general method for defining protein-DNA interactions involving the 5-methyl group of thymines: the GCN4-DNA complex. Nucleic Acids Res **20:**771-775.
249. **Modi SR, Collins JJ, Relman DA.** 2014. Antibiotics and the gut microbiota. J Clin Invest **124:**4212-4218.
250. **Evans DG, Satterwhite TK, Evans DJ, DuPont HL.** 1978. Differences in serological responses and excretion patterns of volunteers challenged with enterotoxigenic *Escherichia coli* with and without the colonization factor antigen. Infect Immun **19:**883-888.
251. **Yang X, Thornburg T, Holderness K, Suo Z, Cao L, Lim T, Avci R, Pascual DW.** 2011. Serum Antibodies Protect against Intraperitoneal Challenge with Enterotoxigenic *Escherichia coli*. J Biomed Biotechnol **2011:**632396.
252. **Wilksch JJ, Yang J, Clements A, Gabbe JL, Short KR, Cao H, Cavaliere R, James CE, Whitchurch CB, Schembri MA, Chuah ML, Liang ZX, Wijburg OL, Jenney AW, Lithgow T, Strugnell RA.** 2011. MrkH, a novel c-di-GMP-dependent transcriptional activator, controls

Klebsiella pneumoniae biofilm formation by regulating type 3 fimbriae expression. PLoS Pathog **7:**e1002204.
253. **Johnson JG, Murphy CN, Sippy J, Johnson TJ, Clegg S.** 2011. Type 3 fimbriae and biofilm formation are regulated by the transcriptional regulators MrkHI in *Klebsiella pneumoniae*. J Bacteriol **193:**3453-3460.
254. **Heatwole VM, Somerville RL.** 1991. The tryptophan-specific permease gene, *mtr*, is differentially regulated by the tryptophan and tyrosine repressors in *Escherichia coli* K-12. J Bacteriol **173:**3601-3604.
255. **Bartlett JG.** 2002. Clinical practice. Antibiotic-associated diarrhea. N Engl J Med **346:**334-339.
256. **McFarland LV.** 1998. Epidemiology, risk factors and treatments for antibiotic-associated diarrhea. Dig Dis **16:**292-307.
257. **Hogenauer C, Hammer HF, Krejs GJ, Reisinger EC.** 1998. Mechanisms and management of antibiotic-associated diarrhea. Clin Infect Dis **27:**702-710.
258. **Prouty MG, Osorio CR, Klose KE.** 2005. Characterization of functional domains of the *Vibrio cholerae* virulence regulator ToxT. Mol Microbiol **58:**1143-1156.
259. **Childers BM, Weber GG, Prouty MG, Castaneda MM, Peng F, Klose KE.** 2007. Identification of residues critical for the function of the *Vibrio cholerae* virulence regulator ToxT by scanning alanine mutagenesis. J Mol Biol **367:**1413-1430.
260. **Mushin R, Dubos R.** 1965. Colonization of the mouse intestine with *Escherichia coli*. J Exp Med **122:**745-757.
261. **Cassels FJ, Wolf MK.** 1995. Colonization factors of diarrheagenic *E. coli* and their intestinal receptors. J Ind Microbiol **15:**214-226.
262. **Luiz WB, Rodrigues JF, Crabb JH, Savarino SJ, Ferreira LC.** 2015. Maternal vaccination with a fimbrial tip adhesin and passive protection of neonatal mice against lethal human enterotoxigenic *Escherichia coli* challenge. Infect Immun doi:10.1128/IAI.00858-15.