Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Characterisation of the Rns transcriptional regulator of enterotoxigenic *Escherichia coli* and regulation of CS1 fimbrial expression

A dissertation presented for the degree of Doctor of Philosophy, in the Faculty of Engineering, Mathematics and Science, Trinity College Dublin

By

Vivienne Mahon

2008

School of Genetics and Microbiology,
Moyne Institute of Preventive Medicine,
Trinity College Dublin

Declaration

I, Vivienne Mahon, declare that the work presented herein represents my own work, except where duly acknowledged in text, and has not been previously presented for a higher degree at this or any other University. I agree that the Library of Trinity College Dublin may lend or copy this thesis upon request.

Vivienne Mahon

Vivienne Mahon

Summary

Enterotoxigenic *Escherichia coli* (ETEC) is a major cause of diarrhoea amongst children in developing countries and travellers to such countries. Virulence in ETEC is dependent on both toxin production and adherence of the bacteria to the host intestine. This adherence is an important event in the establishment of infection and is mediated by adhesins such as fimbriae. ETEC produces several distinct types of fimbriae including CS1, CS2 and CFA/I. The Rns transcriptional regulator activates expression of CS1 and CS2 fimbriae.

The techniques of scanning linker mutagenesis and random mutagenesis were employed to gain an insight into the structure-function relationship of the Rns protein. The Rns mutants generated by these techniques were assayed for their ability to activate expression from the *coo* (CS1) promoter and, in a select number of cases, their ability to activate the *rns* promoter and to bind to DNA. The N-terminal half of Rns was found to be largely tolerant of mutation, although a small number of pentapeptide insertions in this region significantly reduced the activity of the protein. The C-terminal half of Rns was found to be critical for protein function and to be involved in DNA binding. The Rns protein is predicted to contain a disordered region comprising amino acids 100-104. Further mutagenesis of this region revealed that its predicted disorder appears to be crucial for optimal Rns activity. It is proposed that this area may act as an interdomain flexible linker sequence.

Analysis of purified Rns by gel filtration chromatography and experiments performed with a LexA-Rns fusion indicate that the Rns protein dimerises. Cross-linking studies were initially impeded by the apparent aggregation of Rns, but also provided evidence that this protein forms dimers.

Mutations of the DNA sequence within Rns binding sites at the *coo* promoter were demonstrated to reduce Rns binding and Rns-dependent activation of *coo* transcription, thus confirming the importance of specific residues within these sites. These mutations, as well as loss of DNA sequences upstream of the *coo* promoter, were found to result in elevated levels of basal *coo* transcription, suggesting that this promoter may be subject

to complex regulation involving additional, as yet unidentified factors. The nucleoid-associated protein H-NS was revealed to bind directly to the *coo* promoter region and repress *coo* transcription. Rns was found to activate *coo* transcription even in the absence of H-NS, thus indicating that Rns does not act solely as an anti-repressor. Under the conditions tested, Rns did not displace H-NS from *coo* promoter DNA, while H-NS was found to be capable of displacing Rns.

Study of the thermoregulation of CS1 expression revealed that H-NS is not required for thermal control, although in its absence a degree of fimbrial production occurs at 20°C. Transcription from the *coo* promoter was demonstrated to be influenced by temperature. In contrast, *rns* transcription and Rns protein levels were found to be largely insensitive to temperature. Finally, H-NS-mediated repression of *rns* transcription and binding of H-NS to *rns* promoter DNA was observed to occur.

In conclusion, the results presented in this thesis define regions of Rns that are critical to its function and provide evidence that Rns is a dimer. This study has also provided further insight into the regulation of CS1 fimbrial expression.

Acknowledgements

Firstly I would like to thank Dr. Stephen Smith for being an excellent supervisor and for all of his help, advice and patience over the past few years. It was his encouragement that gave me the confidence to apply for this project and to persevere during the tougher times. I am really grateful to him for all that he has done for me throughout my time in his lab.

Special thanks to all of the former members of the Smith lab; I have so many fond memories of our time together as a little family. Matthew, thank you for all of the laughs and chats, you were a great lab mate and are a great friend. I really missed you during the last year. Robert, thanks for your valuable help and for your patience with my endless e-mails. Gwen, I loved working with you, even though you weren't a Finnish boy! To our medical lab mates after the move: thanks Mike for all of the entertaining and adventure-filled stories and thank you to Mary for all of the coffees and for always providing a helpful, caring ear. Thanks also to the girls around the corner, Sarah and Niamh, who were always up for a distracting gossip!

Thank you to everyone in the Moyne, past and present. The brilliant people and fantastic nights out kept me coming back every Friday evening, and will continue to do so long after I have finished. Thanks especially to Helen, Joan, Rebecca and Mark. We started together and we got through it together. You are wonderful friends and my PhD years would have been a lot less enjoyable without you.

Thanks also to the victorious Lab Rats tag rugby team for being just about the only people who managed to get me out of the house while I was writing up!

Finally, I would like to thank my family, especially Caroline for going away for a year so that I could take over her room while I was writing. You can have it back now, I never want to go in there again! Most importantly, thank you so much to my parents, your continual love and support helped me more than I can say, especially during the write up. Your constant reassurance kept me going, I couldn't have done it without you.

Table of Contents

List of Abbreviations

Abbreviation	Full name
DMSO	N,N-dimethylsulphoxide
dNTP	deoxynucleoside triphosphate
EDTA	Ethylenediaminetetraacetic acid
HEPES	*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
HRP	horseradish peroxidase
IPTG	isopropylthiogalactoside
MBP	maltose-binding protein
MCS	multiple-cloning site
MES	2-(*N*-morpholino)ethanesulfonic acid
OD	Optical density
ONPG	*o*-nitrophenyl β-D-galactopyranoside
PBS	phosphate-buffered saline
PIPES	Piperazine-1,4-bis(2-ethansulfonic acid)
PVDF	Polyvinylidene difluoride
SDS-PAGE	Sodium dodecylsulphate-polyacrylamide gel electrophoresis
TBE	Tris-borate-EDTA buffer
X-Gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside

List of Figures

List of Tables

Chapter 1 General Introduction

1.1 *Escherichia coli*, commensal and pathogen

Escherichia coli is the predominant facultative anaerobe of the normal intestinal microflora of humans, other mammals and birds. Typically the gastrointestinal tract of human infants is colonised by *E. coli* within hours of birth and from then on both bacteria and host usually coexist peacefully and can, on occasion, derive mutual benefit (202). These commensal strains of *E. coli* rarely cause disease, although if the host is immunosuppressed or the gastrointestinal barriers are breached then infection may occur. However, the acquisition of virulence determinants by horizontal gene transfer led to the evolution of several highly adapted *E. coli* clones that are capable of causing a range of diseases even in healthy hosts. Each pathogenic category, or pathotype, of *E. coli* represents a family of clones that shares a specific combination of virulence traits, which act in concert to produce a distinctive mechanism of pathogenesis (202). Three general clinical conditions result from infection with *E. coli* pathotypes: (i) urinary tract infections and (ii) bacteraemia/sepsis/meningitis, which are caused by extraintestinal pathogenic strains, and (iii) enteric/diarrhoeal diseases, which are caused by intestinal pathogenic strains. There are at least six well characterised pathotypes of intestinal *E. coli* pathogenic strains: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and enterotoxigenic *E. coli* (ETEC). The extraintestinal pathotypes comprise uropathogenic *E. coli* (UPEC) and neonatal-meningitic *E. coli* (NMEC).

E. coli pathogenesis is a multi-step process comprising (i) adhesion to, and colonisation of, a mucosal surface, (ii) avoidance of host defences, (iii) multiplication, and (iv) damage to the host (274). Adhesins and toxins are two major categories of virulence factors required for *E. coli* pathogenesis. Adhesins commonly take the form of fimbriae; polymeric protein structures typically 1-2 μm in length, found at the bacterial cell surface. Adhesins enable the pathogens to bind closely to sites not usually inhabited by *E. coli*, such as the small intestine, and to resist removal by peristalsis. A wide range of secreted toxins and effector proteins then mediate host cell damage by disrupting critical eukaryotic processes including ion secretion, protein synthesis and cytoskeletal function. It was generally thought that, with the exception of EIEC, the pathogenesis of intestinal *E. coli* strains did not involve an intracellular stage. However, there have been reports describing the invasion of cultured human epithelial

cells by strains of EPEC (258), ETEC (111), EHEC (287) and EAEC (23). In most cases the role of cell invasion in the pathogenesis of these strains is undefined, but it is postulated that cell penetration could represent an evasion and/or persistence strategy.

The pathogenic diversity of *E. coli* strains is reflected in the genomic variability observed amongst different strains. The genomes of natural *E. coli* isolates can vary in size by up to 1 Mb (26). The genome of a commensal *E. coli* K-12 isolate, MG1655, was found to be 4.64 Mb in size while the genomes of UPEC and EHEC isolates were reported to be 4.94 and 5.53 Mb in size respectively (32, 43, 301). The EHEC genome contains more than 1.4 Mb of DNA that is not present in *E. coli* K-12 (83). As mentioned above, this genetic diversity is due in part to the acquisition of foreign DNA encoding virulence traits by pathogenic *E. coli* strains. Virulence factors are often encoded on mobile genetic elements including plasmids, bacteriophages and pathogenicity islands. Genetic information has also been lost during the evolution of pathogenic strains; 0.53 Mb of *E. coli* K-12-specific DNA sequences are absent from the EHEC genome (83). Horizontal gene transfer, along with the loss of DNA regions, resulted in the mosaic genome structure of *E. coli* pathotypes in which novel genetic sequences are present within the common backbone of the *E. coli* genome (422). Significant variation is not only observed between the genomes of pathogenic and commensal strains of *E. coli*, but also between the genomes of different pathotypes. A genomic comparison of strains of EHEC, UPEC and *E. coli* K-12 demonstrated that the three strains shared only 39.2% of their combined set of proteins (422). Therefore *E. coli* strains display great genome plasticity and genetic diversity.

In keeping with their association with mobile genetic elements, *E. coli* virulence genes are often regulated by pathogen-specific regulators that are also encoded on plasmids or pathogenicity islands. These regulators are frequently members of the AraC family of proteins (see Section 1.5). However, *E. coli* pathotypes also exploit global regulatory proteins found in commensal strains, such as IHF (integration host factor), Fis (factor for inversion stimulation) and H-NS (histone-like nucleoid structuring protein), to control the expression of their horizontally acquired virulence genes. This reflects the remarkable adaptability and versatility of *E. coli*, which has led to its success as a pathogen.

1.1.1 Extraintestinal *E. coli* pathotypes

1.1.1.1 Uropathogenic *E. coli* (UPEC)

UPEC are responsible for the majority of urinary tract infections (UTIs), causing 70-90% of the estimated 150 million UTIs diagnosed annually (388). UTIs, which include cystitis (bladder infection) and pyelonephritis (kidney infection), vary in disease severity and location and can also be sporadic, recurrent or chronic. Several virulence factors have been described for UPEC including iron uptake systems (such as aerobactin), adhesins and cytotoxins (440). Among different subgroups of UPEC these virulence factors are present in varying percentages, therefore no single phenotypic profile is responsible for UTIs (197, 198). The attachment of UPEC to uroepithelial cells is mediated by adhesins including type 1, P (Pap; pilus associated with pyelonephritis), S, M and Dr fimbriae (196). Type 1 fimbriae, which are present in most *E. coli* strains, are the most commonly expressed virulence factor of UPEC (162). These fimbriae are involved in the adherence of UPEC to bladder epithelial cells (224). P fimbriae play a role in the attachment of UPEC to kidney epithelium and are possessed by 80% of pyelonephritis strains (215). In addition to the diversity of virulence genes found in UPEC, there is also variation in the expression of these genes. Type 1 fimbrial expression is regulated by an invertible element (see Section 1.4.2) which is generally switched to the "ON" position among bladder UPEC isolates and strains from cystitis patients (161). In contrast, pyelonephritis strains of UPEC tend to have their invertible element in the "OFF" position (161). Coordinated expression of fimbriae is due to cross regulation between the P and type 1 fimbrial operons. Initially it was found that P fimbrial expression turns off type 1 fimbrial expression via the action of the PapB regulatory protein (437). Subsequently type 1 fimbrial expression was also found to coordinately influence the expression of P fimbriae in an inverse manner, thus implying that there may be communication between virulence genes to aid sequential colonisation of various urinary tract tissues (376). The toxins produced by UPEC include the well known haemolysin and cytotoxic necrotizing factor 1 (CNF1), which are directly cytotoxic to host tissues (125, 204), and the more recently identified secreted autotransporter toxin (Sat), a vacuolating cytotoxin that damages kidney epithelium during upper UTIs (163). In addition to causing host damage, CNF1 is thought to facilitate UPEC survival during the acute inflammatory response (76).

1.1.1.2 Neonatal-meningitic *E. coli* (NMEC)

NMEC is the most common cause of gram-negative meningitis and sepsis in neonates, which are associated with high morbidity and mortality. Expression of the K1 capsule is widespread among NMEC strains. Approximately 40% of *E. coli* sepsis isolates and 80% of *E. coli* meningitic isolates were found to have the K1 capsular serotype (214, 331). The pathogenesis of *E. coli* K1 meningitis is a multi-step process in which the bacteria colonise mucosal surfaces (generally of the upper respiratory or gastrointestinal tract) and then invade the bloodstream where they survive and multiply, resulting in bacteraemia (208). NMEC then penetrates the blood brain barrier and invades the central nervous system leading to inflammation and neuronal injury (208). The K1 capsule contributes to the virulence of NMEC as it enhances serum resistance and has antiphagocytic properties (5). Several other virulence factors of NMEC have been identified. To traverse the blood brain barrier NMEC must first bind to brain microvascular endothelial cells (BMECs). Afimbrial and fimbrial adhesins such as OmpA (313), type 1 fimbriae (398) and S fimbriae (292, 392) are implicated in *E. coli* K1 binding to BMECs. However, the role of S fimbriae in NMEC pathogenesis is not entirely clear as, in contrast to earlier findings, it was subsequently reported that these fimbriae are not directly involved in the adherence of *E. coli* K1 to BMECs (421). Proteins demonstrated to be required for successful invasion of BMECs by *E. coli* K1 include OmpA (314), AslA (181), TraJ (15), CNF1 (207) and Ibe proteins (185, 186). Recently, the Hek protein was identified as being involved in the colonisation and invasion of gastrointestinal cells by *E. coli* K1 (123, 124). Therefore, great progress has been made over the last decade in identifying virulence factors of NMEC.

1.1.2 Intestinal *E. coli* pathotypes

1.1.2.1 Enteropathogenic *E. coli* (EPEC)

EPEC are a leading cause of infant diarrhoea in the developing world. At a cellular level, EPEC infections are associated with a distinctive attaching and effacing (A/E) histopathology characterised by intimate attachment of bacteria to intestinal epithelial cells and effacement of host microvilli (262). This is due to polymerised actin, which accumulates to produce pedestal-like structures directly underneath the attached

bacteria. The factors responsible for the formation of A/E lesions, including a type III secretion system and the proteins Tir and intimin, are encoded on the locus of enterocyte effacement (LEE) pathogenicity island (252). The expression of the genes of this pathogenicity island is controlled by the Ler protein (LEE encoded regulator) (254). Another characteristic of EPEC is the distinctive pattern of localised adherence (LA) to cultured epithelial cells exhibited by the bacteria (352). This pattern of attachment is associated with the expression of a type 4 fimbriae termed bundle-forming pili (BFP), which are involved in interbacterial adherence (144). EPEC also produce other fimbriae such as FB171 fimbriae, which may also be involved in localised adherence (145). The genes encoding BFP are present on the EPEC adherence factor (EAF) plasmid, which also contains the *per* locus (377). The PerA and PerC products of this locus are required for the activation of the *bfp* operon (see Section 1.5.3.3) and the LEE1 operon (encodes Ler) respectively (311).

EPEC pathogenesis is a complex process comprising three steps. The first step involves the initial adherence of the bacteria to intestinal epithelial cells. It is not clear which factor mediates this attachment, although several adhesins including BFP (144) and LifA (14) have been implicated. During the second step, the type III secretion system delivers a set of effector proteins into the host cell including Tir (which then inserts into the host cell membrane and becomes phosphorylated) and the Esp proteins (195). In the final step, EPEC becomes intimately attached to the host cell due to an interaction between the outer membrane protein Tir and intimin (206), which is followed by pedestal formation. The effector proteins translocated into host cells disrupt several aspects of host cell physiology including intestinal permeability, inflammation and ion secretion, thus resulting in the occurrence of diarrhoea (202).

1.1.2.2 Enterohaemorrhagic *E. coli* (EHEC)

EHEC was first recognised as a distinct *E. coli* pathotype in the early 1980s (328). EHEC infections can result in mild to bloody diarrhoea and haemorrhagic colitis. In some cases the potentially fatal complication haemolytic uremic syndrome (HUS) develops. *E. coli* O157:H7 is the dominant EHEC serotype associated with HUS in North America, northern Europe and Japan but other serotypes including O26 and O111 are prominent causative agents elsewhere in the world (109, 407). EHEC shares some similarities with EPEC such as possession of the LEE pathogenicity island and thus the

ability to form A/E lesions (274). The defining feature of EHEC however, is the production of Shiga toxins (Stx), indeed EHEC strains are thought to have evolved from an EPEC strain that acquired a Stx-encoding bacteriophage (324). Shiga toxins, of which there are two major families (Stx1 and Stx2), cause the host damage that can lead to HUS by inhibiting protein synthesis in renal endothelial cells, which results in cell death (234). The EAF plasmid is not present in EHEC therefore BFP are not produced by this pathotype. However, some EHEC serotypes contain a plasmid, pO157, which encodes a range of virulence factors including a haemolysin similar to that found in UPEC (356), and a potential adhesin named ToxB (397). Several other putative adhesins have been identified in EHEC, but the LEE-encoded protein intimin is the only factor whose involvement in adherence to epithelial cells has been clearly demonstrated (89). Sequencing of two EHEC genomes revealed the presence of at least 16 potential fimbrial operons (173, 236, 301), although it has not been established if they are functional *in vivo*. While EHEC has been reported to produce fimbriae such as long polar fimbriae (406), the *E. coli* common pilus (326) and a type 1 fimbriae homologue F9 (235), the biological significance of fimbriae to host colonisation by EHEC remains unclear. However, the recent finding that EHEC produces type 4 fimbriae termed haemorrhagic *coli* pilus that may act as intestinal colonisation factors has contributed to the developing understanding of the mechanisms involved in EHEC adherence to epithelial cells (438).

1.1.2.3 Enteroaggregative *E. coli* (EAEC)

EAEC is an important cause of diarrhoea, which is often persistent, in both children and adults worldwide. EAEC strains are very heterogeneous (108), but are defined by their characteristic pattern of aggregative adherence to cultured epithelial cells, in which bacteria attach to cell surfaces in a "stacked-brick" configuration (275). Three stages comprise EAEC pathogenesis: adherence to the intestinal mucosa; formation of a mucoid biofilm; and release of toxins that cause intestinal cell damage and, ultimately, diarrhoea (272). Adherence of EAEC is associated with the presence of fimbrial structures termed aggregative adherence fimbriae (AAF), which are distantly related to the Dr family of adhesins. Based on the sequences of their major fimbrial subunits, three main variants of AAF (AAF/I, II and III) have been described (28, 66, 273). The genes encoding the fimbrial subunits are present on large virulence plasmids called

pAA. The expression of AAF genes, and several other EAEC virulence determinants, is regulated by the AggR protein (see Section 1.5.3.4), which is also encoded on pAA (98, 278). Each AAF variant is only present in a minority of EAEC strains, suggesting that other adhesins play a role in aggregative adherence. Indeed there have been reports describing the involvement of type 1 fimbriae (263) and several outer membrane proteins in EAEC adherence (77, 385). Recently a potential fourth AAF, an adhesin termed Hda, was identified and implicated in aggregative adherence (36). A number of toxins that are associated with mucosal damage have been described for EAEC. EAST1 is a heat stable toxin similar to the ETEC ST toxin (see Section 1.1.2.6) (348), although its role in disease requires clarification, as it is also found in commensal isolates (255). The other well characterised EAEC toxins include *Shigella* enterotoxin 1 (ShET1) and plasmid-encoded toxin (Pet), a cytopathic autotransporter protein (21, 117).

1.1.2.4 Enteroinvasive *E. coli* (EIEC)

EIEC infection most commonly results in watery diarrhoea, but can occasionally lead to dysentery (274). EIEC are closely related to *Shigella* species and it has even been suggested that EIEC and *Shigella* form a single pathotype of *E. coli* (223). These microorganisms share a complex mechanism of pathogenesis that starts with their penetration of the intestinal epithelium by invading M cells (347), followed by evasion of phagocytosis by inducing apoptosis in macrophages (449), and invasion of epithelial cells via their basolateral aspect (264). The endocytosed bacteria escape from the endocytic vacuole, multiply intracellularly and are then propelled by actin tails into adjacent cells, thus resulting in dissemination of bacteria in the epithelium (293). The factors required for this system of invasion and intercellular spread are encoded by a large virulence plasmid (VP) which is a mosaic of genes from multiple origins (44). A 30 kb region of the VP, known as the entry region, codes for the secreted Ipa proteins necessary for host cell entry and the type III secretion system responsible for the translocation of these effector proteins into the cell (256). Other VP-encoded virulence determinants include SepA, a secreted serine protease (24), and Sen, an enterotoxin that may be associated with the diarrhoeal disease caused by EIEC (277). A cascade involving the products of the *virF* and *virB* genes, which are also located on the VP, mediates the regulation of the invasion genes. VirF activates expression of VirB, which then activates the expression of the invasion genes in the entry region (95). The VirF

protein of *Shigella flexneri* is discussed in Section 1.5.3.5. There is limited information concerning fimbrial adhesins of EIEC, however it was recently reported that 20% of EIEC strains tested contained the genes for long polar fimbriae (405). These fimbriae were previously demonstrated to facilitate attachment of *Salmonella* to M cells (20).

1.1.2.5 Diffusely adherent *E. coli* (DAEC)

DAEC are a heterogeneous group of isolates that are defined by their characteristic pattern of diffuse adherence to cultured epithelial cells (276). The adherence of the first class of DAEC strains is mediated by afimbrial and fimbrial adhesive structures termed Afa/Dr adhesins (365). These Afa/Dr DAEC strains have been demonstrated to be associated with UTIs (410) and have also been implicated in a number of studies as a cause of diarrhoea, particularly amongst older children (146, 351). F1845 fimbriae, which belong to the Dr family of adhesins, are present in approximately 75% of DAEC strains (31, 202). The cellular receptor of the F1845 fimbrial adhesin was identified as the membrane-associated protein DAF (decay accelerating factor) (27). DAEC strains induce the production of finger-like protrusions from the surface of infected cells that envelop and protect the bound bacteria (60). It is proposed that this phenomenon is due to the F1845-mediated binding and clustering of the DAF protein (296). A feature of some *E. coli* strains included in the second class of DAEC strains is the expression of an adhesin involved in diffuse adherence, AIDA-1 (25). AIDA-1 is an autotransporter protein that mediates several virulence related activities in addition to adhesion to epithelial cells, including aggregation and biofilm formation (368). The pathogenic mechanism of DAEC has not yet been precisely determined, but is thought to involve the disruption of brush border-associated enzymes (295).

1.1.2.6 Enterotoxigenic *E. coli* (ETEC)

ETEC is the pathotype responsible for the majority of *E. coli*-mediated cases of human diarrhoea worldwide. It is particularly prevalent amongst children in developing countries and travellers to such countries. It is estimated that annually, there are 650 million incidences of ETEC infection, resulting in 800,000 deaths in children under the age of 5 (409). The essential determinants of ETEC virulence are traditionally considered to be adherence and colonisation of the host small intestinal epithelium via

plasmid encoded colonisation factors, and subsequent release of plasmid encoded enterotoxins that induce a net secretory state leading to profuse watery diarrhoea (Fig. 1.1) (274). This view may need to be re-evaluated in light of the identification of several chromosomally encoded potential virulence determinants in ETEC (409). These recently described putative toxins include EatA, a serine protease autotransporter (294); ClyA, a pore forming cytotoxin (239); and EAST1, a heat stable toxin originally isolated from EAEC (441). However, the role of these factors in ETEC virulence has not yet been fully determined.

ETEC strains are defined by the production of one or two fully characterised enterotoxins, heat-labile toxin (LT) and/or heat-stable toxin (ST). The LT of ETEC is closely related both structurally and functionally to the cholera toxin of *Vibrio cholerae* (383). Like the oligomeric cholera toxin, LT is composed of five identical B subunits associated with one A subunit. There are two major serogroups of LT, LT-I and LT-II, but only LT-I is associated with disease in humans. Following release of LT-I, the B subunits of the toxin bind to the cell surface gangliosides GM1 and GD1b (137), facilitating the entry of the A subunit into the host cell (383). The enzymatically active A subunit transfers an ADP-ribosyl group from NAD to the α-subunit of the stimulatory G protein (G_{S_α}), which leads to irreversible stimulation of adenylate cyclase activity (Fig. 1.1 (A)). This in turn leads to overproduction of intracellular cyclic AMP (cAMP). Thus cAMP-dependent protein kinase A is activated, resulting in phosphorylation and consequent activation of chloride channels in the epithelial cell membrane, the major one being the cystic fibrosis transmembrane conductance regulator (CFTR). The resultant increased chloride ion secretion and reduced NaCl absorption leads to loss of water to the intestinal lumen and thus diarrhoea (363). Enhanced intestinal secretion may also occur via LT-mediated stimulation of the enteric nervous system and prostaglandin synthesis (274).

STs are small polypeptides comprising two unrelated classes, STa and STb, of which only STa is a proven significant human virulence factor (363). Mature STa, a cysteine rich peptide of 18 to 19 amino acids with three intramolecular disulphide bonds, is an analogue of the mammalian hormone guanylin. STa binds and activates guanylate cyclase C (361), a membrane-spanning enzyme whose endogenous agonist is guanylin (65). This results in elevated levels of intracellular cyclic GMP, which ultimately leads to activation of the CFTR chloride channel (188). Again, the net effects of enhanced

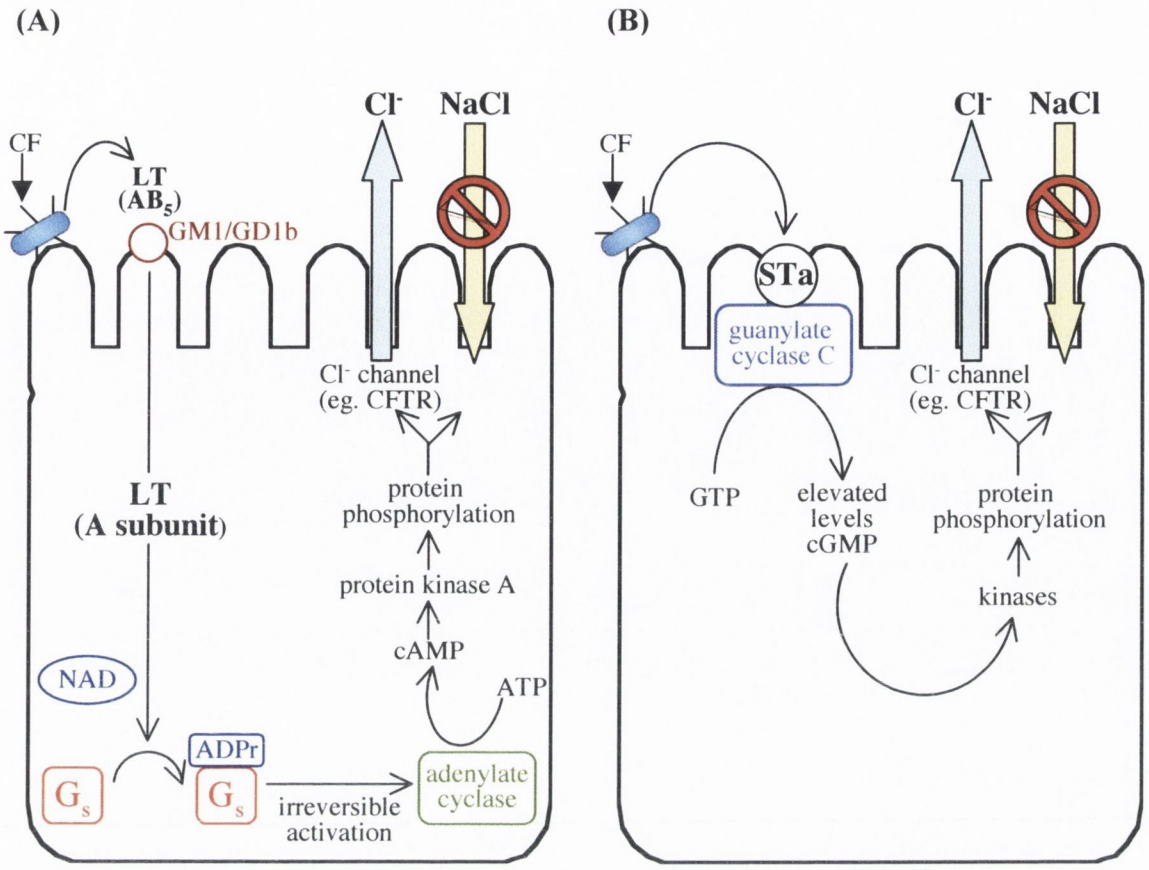

Fig. 1.1. Pathogenesis of enterotoxigenic *E. coli*. An important early step in ETEC pathogenesis is the adherence of bacteria to host intestinal epithelial cells via surface expressed colonisation factors (CFs). This is followed by the secretion of one or two enterotoxins, whose actions result in diarrhoea. **(A)** Mechanism of action of heat-labile toxin LT-I. The B subunits of the oligomeric toxin (AB_5) bind to gangliosides on the host cell surface (GM1 and GD1b). The A subunit of the toxin enters the host cell and transfers an ADP-ribosyl group (ADPr) to the α-subunit of G_s protein. This leads to an increase in cAMP, which results in phosphorylation of the cystic fibrosis membrane conductance regulator (CFTR) chloride channel. The consequent increased chloride ion (Cl^-) secretion from secretory crypt cells and decreased NaCl absorption from absorptive cells leads to diarrhoea. **(B)** Mechanism of action of heat-stable toxin STa. STa binds to and activates the membrane-spanning receptor guanylate cyclase C, resulting in an increase in cGMP levels, which again ultimately leads to diarrhoea due to increased chloride secretion and decreased NaCl absorption. Adapted from Nataro and Kaper (274).

chloride secretion and decreased NaCl absorption cause an increase in intestinal fluid secretion that manifests as diarrhoea (Fig. 1.1 (B)).

An important early step in ETEC pathogenesis, which occurs prior to the elaboration of enterotoxins, is the attachment of bacteria to the surface of the host small intestinal mucosa. This vital step in disease initiation is mediated by bacterial proteinaceous surface structures known as colonisation factors (CFs). Over 20 antigenically diverse CFs from human strains of ETEC have been described (Table 1.1) (138, 302). In addition to their antigenicity, CFs can be subdivided based on their structural morphology. Although most are fimbrial or fibrillar, some CFs have two fibrils arranged in a double helix and are classified as helical, while others are non-fimbrial (138). CFs were initially called colonisation factor antigens (CFAs), coli surface antigens (CSs) and putative colonisation factors (PCFs). Subsequently a uniform nomenclature system was adopted to designate CFs, with the exception of CFA/I, as numbered CSs according to the chronological order in which they were identified (138). CFA/II was initially thought to be a single entity (121), but was later found to consist of fine, flexible CS3 fibrillae alone or in combination with the rigid CS1 or CS2 fimbriae (374). Similarly, CFA/IV is composed of CS6 alone or in combination with either CS4 or CS5 (399). Some of the better-characterised CFs can be divided into families based on genetic relationships, for example the CFA/I-like group comprises CFA/I, CS1, CS2, CS4, CS14 and CS17 (138). The frequency with which CFs occur on ETEC strains varies geographically, however epidemiological studies revealed that approximately 75% of ETEC strains worldwide express CFA/I, CFA/II or CFA/IV, therefore these are the most prevalent CFs (433). As studies have found that antibodies directed against CFs provide protective immunity against ETEC, these prevalent CFs are key candidates for components of a much sought after ETEC vaccine (319). However it is possible that novel CFs will be identified in the future; indeed no known CFs could be identified on 20% of ETEC strains (433). While the host cell surface receptors for CFs have not been studied extensively, some are known to be oligosaccharide residues on glycoconjugates (156). The oligosaccharides present in eukaryotic cell membrane glycoconjugates vary among species, thus the interaction between CFs and their receptors confers species specificity on ETEC pathogens (138). Therefore the CFs of human ETEC isolates are distinct from those of animal isolates, such as K88, K99 and 987P fimbriae. The genes required for CF production may be encoded in operons located on plasmids that frequently also contain the enterotoxin genes. CF operons are flanked by insertion

Table 1.1 Colonisation factors of human ETEC strains

CS designation	Former designation	Structural morphology
CFA/I	CFA/I	Fimbrial
CS1	CFA/II	Fimbrial
CS2	CFA/II	Fimbrial
CS3	CFA/II	Fibrillal
CS4	CFA/IV	Fimbrial
CS5	CFA/IV	Helical
CS6	CFA/IV	Undefined
CS7	CS7	Helical
CS8	CFA/III	Fimbrial
CS10	2230	Non-fimbrial
CS11	PCF0148	Fibrillal
CS12	PCF0159	Fimbrial
CS13	PCF09	Fibrillal
CS14	PCF0166	Fimbrial
CS15	8786	Non-fimbrial
CS17	CS17	Fimbrial
CS18	PCF020	Fimbrial
CS19	CS19	Fimbrial
CS20	CS20	Fimbrial
CS21	Longus	Fimbrial
CS22	CS22	Fibrillal

Adapted from Gaastra and Svennerholm (138) and Pichel *et al.* (302).

sequences and have a significantly lower GC content, and a different codon usage, than is usual for *E. coli* (138). This implies that they were recently introduced to *E. coli* from a foreign genetic background. The biogenesis of CFs (see Section 1.2) is an energetically expensive process; therefore the expression of many CFs is dependent on a trans-acting plasmid-located gene encoding a regulator such as Rns, the activator of CS1 and CS2 expression (see Section 1.5.3.1).

The plasmid-encoded CFs are not the only adhesins of ETEC. Two chromosomal loci, *tia* and *tib*, were found to be associated with invasion and non-CF dependent adherence of ETEC to intestinal epithelial cells (111). The *tia* locus directs the

production of Tia, an outer membrane protein demonstrated to interact with a cell surface proteoglycan receptor and thus mediate bacterial attachment to epithelial cells (129). The *tib* locus is responsible for the synthesis of TibA, a large outer membrane protein that is a member of the autotransporter family (229), and is involved in binding to a specific epithelial cell receptor (228). TibA has also been implicated in biofilm formation and autoaggregation, both of which are associated with bacterial virulence (369). In addition to acting as adhesins, Tia and TibA enable ETEC to invade epithelial cells (112, 130), although the *in vivo* relevance of this invasion to ETEC pathogenesis is yet to be determined. The role of afimbrial adhesins in virulence is also unclear. However it has been postulated that while CFs may be responsible for the initial contact with the intestinal epithelium, the adherence mediated by Tia and TibA may be required for an intimate interaction between ETEC and the host cell, thus resulting in the most effective delivery of enterotoxins (409).

1.2 Fimbriae: mechanisms of biogenesis

As described in the previous section, pathogenic *E. coli* produce numerous types of fimbriae that are involved in the adherence of bacteria to host tissues and thus play a vital role in the establishment of infection. Fimbriae can be wiry and flexible; rod-shaped and rigid; or, as is the case for P fimbriae, composite structures. Several different subunit types may combine to form the fimbrial structure resulting in a heteropolymeric fibre. Alternatively fimbriae may be mainly homopolymeric. Generally, the receptor-binding adhesin is located either at the tip or along the body of the fimbrial shaft (375). Fimbriae can be categorised based on the pathways used for their assembly, of which there are four main types. An overview of each of these biogenesis mechanisms, along with representative examples, is presented below.

1.2.1 Biogenesis of type 4 fimbriae

Type 4 fimbriae, such as BFP of EPEC and toxin-coregulated pilus (TCP) of *V. cholerae*, are assembled by a multi-component machine made up of at least 12 proteins that are structurally and functionally related to the type II secretion system. The ETEC colonisation factors longus and CFA/III are also assembled via this complex pathway

(151, 396). Biogenesis of BFP is directed by a cluster of 14 genes located on the EAF plasmid of EPEC (393). The first gene of this operon, *bfpA*, encodes bundlin, the major subunit of BFP (88). Bundlin has the distinctive features of a type 4 fimbrial subunit, which include a short hydrophilic signal peptide and a hydrophobic N-terminal domain that causes the protein to be directed to the inner membrane. The remaining proteins encoded by the *bfp* operon form a hetero-oligomeric complex that spans the periplasmic space and comprises both inner and outer membrane subassemblies (62, 73). One of the inner membrane associated components of this complex, BfpP, is a pre-pilin peptidase that removes the signal peptide from bundlin to generate the mature fimbrial subunit (446). In the periplasmic space, mature subunits are assembled into a helical fimbria that is then extruded across the outer membrane to the cell surface through a multimeric channel largely formed by the BfpB component of the biogenesis complex (357). The energy for fimbrial extrusion is provided by the ATPase BfpD, which is a member of the inner membrane subassembly (63). In recent years great progress has been made in the characterisation of the BFP system, however the biogenesis of type 4 fimbriae is yet to be fully understood.

1.2.2 Extracellular nucleation-precipitation pathway

Enteric bacteria such as *E. coli* produce thin aggregative fibres termed curli, which are involved in adhesion and invasion of host cells and biofilm formation (18). The biogenesis of curli, via the extracellular nucleation pathway, is unique among bacterial fibres described to date. Two divergently transcribed operons, *csgBA* and *csgDEFG*, encode the six proteins necessary for curli assembly (168). CsgD is a transcriptional regulator of the *csgBA* operon, while the remaining proteins are fibre components or assembly factors that are translocated across the inner membrane by the general secretory pathway. Secretion of CsgA and CsgB into the extracellular milieu is dependent on the outer membrane lipoprotein CsgG (332). At the bacterial surface the nucleator protein CsgB associates with the outer membrane and triggers the polymerisation and conversion of the soluble subunit protein CsgA into insoluble cell surface fibres (169). The periplasmic proteins CsgE and CsgF interact with CsgG at the outer membrane (332). While their role in biogenesis is currently unclear, they are required for efficient curli assembly (116). Curli are members of a recently described class of fibres known as functional amyloids (132). Amyloids are associated with

neurodegenerative diseases such as Alzheimer's disease; therefore the study of curli biogenesis may provide insights into the mechanisms involved in these disorders.

1.2.3 Chaperone-usher pathway

The best-studied fimbrial biogenesis mechanism is the chaperone-usher pathway, which is involved in the assembly of more than 30 adhesive organelles in gram-negative bacteria including CS3 of ETEC and type 1 fimbriae (380). The assembly of P fimbriae of UPEC represents the paradigm for this system. The chromosomal *pap* operon contains 11 genes required for the regulation, expression and assembly of P fimbriae. These fimbriae are composite fibres comprising a thin flexible tip fibrillum joined to a rigid helical rod (218). The tip fibrillum is composed of PapE subunits and contains a single copy of the PapG adhesin at its distal end. The PapF adaptor subunit connects the adhesin to the tip fibrillum, which is itself linked via the PapK adaptor to the helical rod formed by a PapA homopolymer. PapH terminates the fimbrial structure and anchors it to the bacterial surface (45). The coordinated assembly of these subunits requires a periplasmic chaperone and an outer membrane usher. The general secretory pathway mediates the translocation of the fimbrial subunits across the inner membrane into the periplasm where they form stable complexes with the chaperone, PapD. Interaction with the chaperone facilitates proper folding of the subunit in addition to protecting the subunit from proteolytic degradation and premature assembly (219). Chaperone-subunit complexes are targeted to PapC, the outer membrane usher. PapC differentially recognises these complexes based on the final location of the subunit in the fimbrial structure (84). This results in the ordered incorporation of subunits into the growing fibre, which is translocated to the cell surface through a pore formed by the usher. The PapC usher forms dimeric channels in the outer membrane (227), however, it was recently revealed that while each of the PapC components of the dimeric complex are involved in chaperone-subunit recruitment, fibre translocation only occurs through one of the pores (325).

1.2.4 Alternate chaperone-usher pathway

Several genetically related CFs of ETEC are considered to form a family known as the CFA/I-like group (see Section 1.1.2.6). With the exception of the CS14 operon, which

contains a duplication of the fimbrial subunit gene, the operons that encode these CFs consist of only four genes that are required for fimbrial structure and assembly (8). Therefore the biogenesis of CFA/I-like fimbriae is simpler than that of other fimbriae. One family member, CS1, serves as the prototype for the study of this biogenesis system. The CS1 operon, *cooBACD* (also known as *csoBACE*) exemplifies the organisation of the four-gene bioassembly operons of the CFA/I-like group (Fig. 1.2 (A)). The second gene of the operon encodes the major subunit, CooA, which composes the body of the CS1 fimbrial structure (297). While initial evidence suggested that adherence of CS1 fimbriae is mediated by CooA (244), subsequent findings imply that the minor subunit CooD is the adhesin as a single copy of this protein is thought to be present at the fimbrial tip and mutations in CooD, but not CooA, reduce CS1-dependent adhesion (343, 344). In addition to being a fimbrial component, CooD plays an essential role in fimbrial assembly, as it is the rate limiting initiator of CooA polymerisation (345). The first gene of the CS1 operon encodes CooB, a protein thought to have a chaperone-like role as it associates with the subunit proteins in the periplasm to prevent their degradation and premature polymerisation (418). Unusually however, CooB also interacts with and stabilises CooC, an outer membrane protein proposed to be involved in the translocation of fimbrial subunits to the cell surface (343). Despite their apparent functional similarities, the CS1 assembly proteins have no significant sequence homology to the proteins involved in the classical chaperone-usher pathway (346). Therefore the CS1 biogenesis system is termed the alternate chaperone-usher pathway (380). Numerous studies have led to the following current model for CS1 fimbrial assembly (389). The four Coo proteins are transported to the periplasm, via the general secretory pathway, where CooA and CooD are bound by CooB. CooC inserts into the outer membrane and is maintained in an active conformation there by the binding of CooB. Fimbrial assembly begins when a CooB-CooD complex binds to CooC (Fig. 1.2 (B)). CooB, which is not part of the final fimbrial structure, is released and most likely recycled to the periplasm. Next a CooB-CooA complex binds to CooC at the outer membrane, leading to the addition of CooA to CooD. Sequential addition of multiple CooA subunits at the base of the growing fimbria causes the structure to extend across the outer membrane to the bacterial cell surface. Thus CS1 fimbriae are largely homopolymers of CooA with a single CooD subunit at their tip (343).

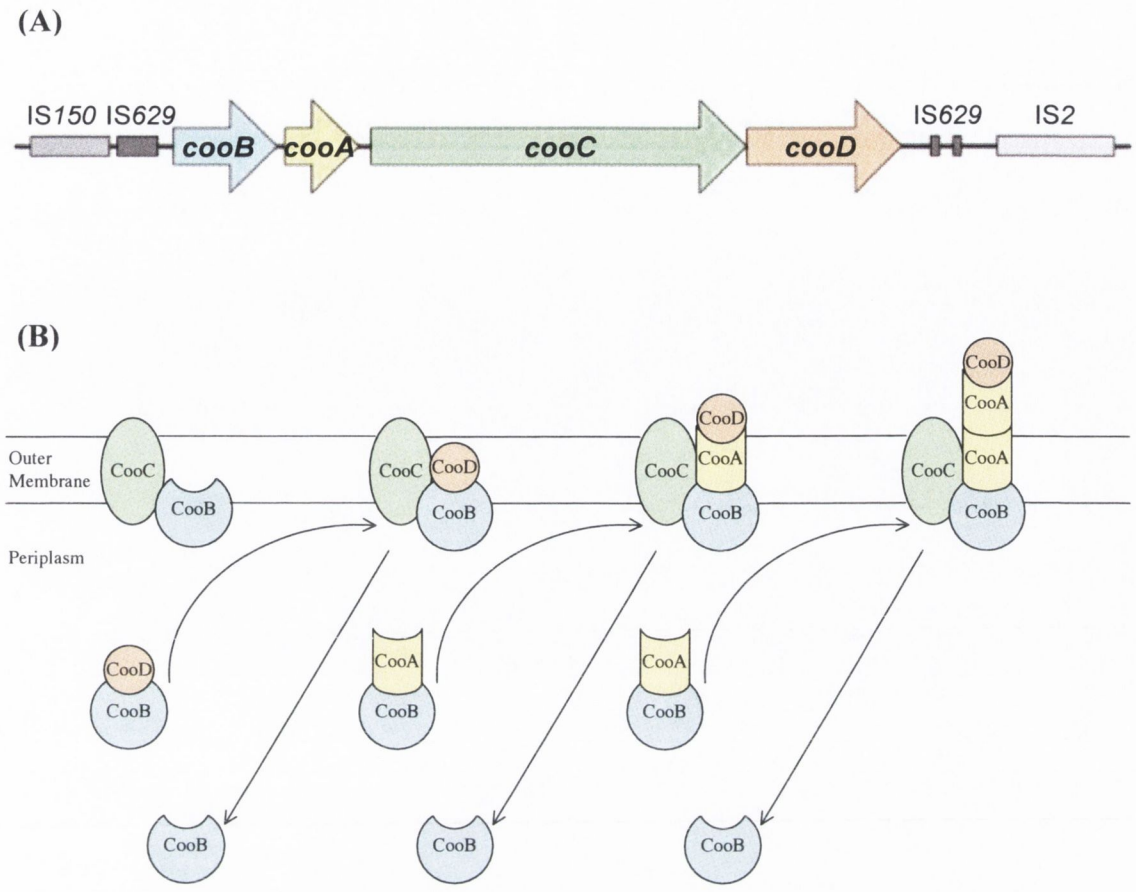

Fig. 1.2. The CS1 operon and model of assembly of CS1 fimbriae. **(A)** Map of the CS1 operon. The *coo* genes are indicated by coloured arrows. DNA regions homologous to insertion sequence elements are represented as grey boxes. **(B)** Model of the biogenesis of CS1 fimbriae via the alternate chaperone-usher pathway. The process is initiated when a CooB-CooD complex binds to the outer membrane protein CooC. CooB is displaced, either from CooC or from CooD, and is thought to be recycled to the periplasm. The repeated interaction of the CooB-CooA complexes with CooC at the outer membrane, and the resultant incorporation of the CooA major subunits, leads to the extension of the new fimbrial structure across the outer membrane. Adapted from Sakellaris and Scott (346).

1.3 Regulation of transcription initiation

To survive and succeed either as a commensal or pathogen, it is essential that bacteria effectively regulate gene expression to ensure that optimal levels of the correct proteins are produced under appropriate conditions. Bacterial gene expression is predominantly regulated at the level of transcription initiation. A multi-subunit RNA polymerase is responsible for bacterial transcription. The RNA polymerase holoenzyme consists of a core enzyme with the subunit composition $\alpha_2\beta\beta'\omega$, in complex with a σ subunit, which may be the major σ factor σ^{70}, or one of the less abundant alternative σ factors such as σ^{38}, the stress σ factor (37). Promoter recognition is a key aspect of transcription initiation and is largely mediated by the σ subunit. Domains 4 and 2 of the multi-domain RNA polymerase σ^{70} subunit contact the -35 and -10 hexamer elements of bacterial promoters respectively. Bacterial promoters may also contain an AT rich UP element upstream of the -35 hexamer, which is contacted by the C-terminal domain of the α subunit of RNA polymerase, and/or they may contain an upstream extension of the -10 hexamer that is contacted by domain 3 of σ^{70} (41). Many bacterial promoters are regulated by transcription factors; proteins that generally bind within or in the vicinity of the promoter and activate or repress its activity, often by affecting the ability of RNA polymerase to initiate transcription.

1.3.1 Transcription factors

It is estimated that there are 300-350 transcription factors in *E. coli* (300). This large number of diverse regulatory proteins enables *E. coli* to respond to a range of conditions, as they couple gene expression to environmental changes. Transcription factors can be activators (stimulate transcription by improving the affinity of a promoter for RNA polymerase, see Section 1.3.2), repressors (prevent transcription initiation, often by blocking access of RNA polymerase to the promoter), or dual regulators (activate or repress transcription depending on the target promoter) (41). Some transcription factors such as the nucleoid-associated proteins H-NS (discussed further in Section 5.1) and IHF, the cAMP receptor protein (CRP) and the leucine-responsive regulatory protein (Lrp) are termed global regulators as they affect the expression of a large number of unrelated genes (249). CRP, which is activated by binding of cAMP and is thus responsive to the cell's energy status, activates transcription from at least

100 promoters (225), while Lrp influences the transcription of more than 10% of all *E. coli* genes, including those involved in fimbrial production and amino acid metabolism (40). However, the majority of transcription factors control a limited, specific set of genes (13). In addition to controlling other genes, transcription factors frequently regulate their own expression. Negative autoregulation occurs most often, although many transcription factors are positively autoregulated (249).

Transcription factors are grouped into evolutionarily related families on the basis of sequence and structural similarities (300). Some families have been extensively studied and are thus well characterised. In addition to the AraC family (139), which is discussed further in Section 1.5, these include the LysR, CRP, OmpR and LacI families. Transcription factors are often modular with the different families sharing similar domain architectures. Approximately 75% of transcription factors are two-domain proteins consisting of a DNA-binding domain along with a second domain, typically involved in oligomerisation and/or regulation (13). This second domain is usually less well conserved. Therefore families of transcription factors are characterised by a signature sequence that usually includes the DNA binding domain (300). The most common DNA-binding motif in prokaryotic transcription factors is the helix-turn-helix (HTH), although the DNA binding of some regulators is mediated by alternative motifs such as zinc fingers and antiparallel β-sheets (300). The activity of transcription factors can be affected by modification of their regulatory domains either by phosphorylation, as in the case of the OmpR family members that function as response regulators in two component systems, or by the binding of small molecule effectors, usually metabolites (41). In addition to sharing structural properties, members of transcription factor families can also be functionally related, as they tend to regulate genes associated with similar biological functions (300).

1.3.2 Activation of transcription initiation

Simple transcription activation occurs when a single transcription factor activates a promoter that generally contains one or more defective DNA sequence elements. There are two main mechanisms by which transcription factors can function to facilitate the binding of RNA polymerase to such sub-optimal promoters. In the first mechanism, the DNA-bound activator makes direct contact with RNA polymerase and thus recruits it to the target promoter. Activators that function by this mechanism can be grouped into

two principal classes. Class I activators bind upstream of the promoter -35 hexamer and directly interact with the C-terminal domain of the RNA polymerase α subunit (αCTD). A flexible linker joins αCTD to the α N-terminal domain (αNTD), which interacts with the other RNA polymerase subunits, therefore Class I activators can bind at a range of upstream sites (41). In contrast, Class II activators are restricted to binding to sites that overlap the -35 element of the promoter as they mostly interact with domain 4 of the RNA polymerase σ subunit, although they may also make contacts with αCTD and/or αNTD (41). Some transcription factors, such as CRP, can function either as Class I or Class II activators (225). Activators that function by the second mechanism do not directly contact RNA polymerase, but instead facilitate its interaction with the target promoter by altering the conformation of the promoter DNA (41). For example the MerR protein binds between the -35 and -10 hexamers of the *merT* promoter and thus re-orientates these promoter elements to enable RNA polymerase binding (286).

Transcription activation can be more complex that the simple situations described above. Approximately 50% of genes are regulated by multiple transcription factors, usually a global regulator in addition to specific local regulators, although some promoters are affected by several global regulators (249). Promoters that undergo complex control are frequently regulated by both activators and repressors (41). In these cases an activator may be required to facilitate transcription by counteracting the effect of a repressor, and thus it actually functions as an anti-repressor. At some promoters a single interaction between RNA polymerase is insufficient for transcription initiation, thus two activators are required, both of which may be Class I activators or one can function as a Class I activator and the other as a Class II activator. Alternatively, the dependence of some promoters on more than one activator results from the co-operative binding of the activators (17). As discussed in Section 1.4, a further form of complex activation occurs at some fimbrial gene promoters.

1.4 Regulation of fimbrial expression

Fimbrial expression is affected by numerous environmental signals including temperature (see Chapter 6), osmolarity, carbon source and pH. The regulation of transcription of fimbrial genes is mediated by global regulators such as H-NS, Lrp, CRP

and IHF, and by specific local regulators that may be encoded by genes that are included in, or are physically separated from, the fimbrial operon (260). For example, both the global regulator H-NS and the specific regulator Rns (see Section 1.5.3.1) are involved in the regulation of CS1 fimbrial expression. The expression of some fimbriae is also subject to phase variation, a type of regulation in which gene expression switches between a phase ON state and a phase OFF state in a reversible and heritable manner, resulting in a phenotypically heterogeneous population of cells (413). In *E. coli,* fimbrial phase variation occurs by a DNA methylation-dependent epigenetic mechanism or by a DNA inversion mechanism, exemplified by P and type 1 fimbriae respectively. Overviews of each of these complex systems are presented in the following sections.

1.4.1 Phase variation of P fimbriae

The complex P fimbrial phase variation mechanism is based on the methylation status of two sites in the *pap* regulatory region and involves several factors including Lrp, DNA adenine methylase (Dam), CRP and the specific regulators PapI and PapB (411). Six Lrp binding sites are present in the *pap* regulatory region (Fig. 1.3). Sites 1-3 are proximal to/overlap the main promoter for the *pap* operon, while sites 4-6 are distal to this promoter (282). Sites 2 and 5 contain Dam target sequences that are designated $GATC^{prox}$ and $GATC^{dist}$ respectively. In phase OFF cells, Lrp binds to sites 1-3. Consequently *pap* transcription is repressed, methylation of $GATC^{prox}$ is blocked and, by mutual exclusion, the affinity of Lrp for sites 4-6 is reduced (177, 424). As these promoter distal sites are unoccupied, $GATC^{dist}$ becomes methylated, which further reduces the affinity of Lrp for sites 4-6 (179, 283). Switching from phase OFF to ON requires PapI, a co-regulatory protein that increases the affinity of Lrp for site 5 and, to a reduced extent, site 2 (179). Switching is also proposed to require DNA replication, as this would result in a hemimethylated $GATC^{dist}$ site and displacement of Lrp from sites 1-3 (179). This facilitates the translocation of Lrp to sites 4-6, as PapI/Lrp has a greater affinity for DNA hemimethylated at $GATC^{dist}$ than for fully methylated DNA, and methylation of the now unoccupied $GATC^{prox}$ specifically inhibits PapI/Lrp binding to sites 1-3 (179). Following a further round of replication, transition to phase ON is completed by full methylation of $GATC^{prox}$ and binding of PapI/Lrp to a nonmethylated $GATC^{dist}$ site. This binding of Lrp at sites 4-6, in addition to binding of cAMP-CRP upstream of the promoter, activates *pap* transcription leading to expression of fimbriae

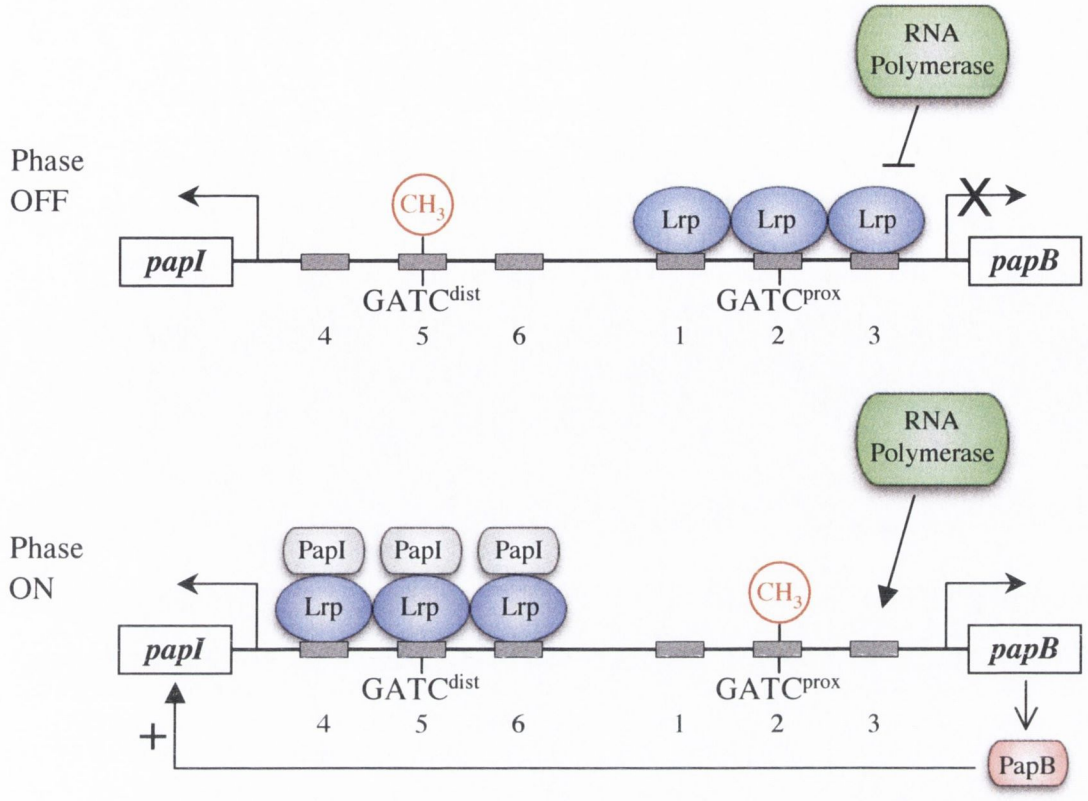

Fig. 1.3. Phase variation of the *pap* operon. Simplified views of the *pap* regulatory region in phase OFF and ON states are depicted. Lrp binding sites are numbered and shown as grey rectangles. In the phase OFF state, *pap* transcription is repressed by the co-operative binding of Lrp to sites 1-3. Sites 4-6 are unoccupied and the GATC sequence within site 5 is methylated by Dam. In the phase ON state, PapI/Lrp binds co-operatively to sites 4-6 and the GATC sequence within site 2 is open for Dam methylation as sites 1-3 are now unoccupied. Thus *pap* transcription can occur, leading to the production of PapB, which activates expression of PapI. See text for further details.

and PapB (423). Expression of PapI is activated by PapB, thus a positive feedback loop is initiated (131, 177). Incorporated into, and superimposed on the Pap switch mechanism is regulation by environmental stimuli (412). Response to stimuli such as low temperature, rich medium and envelope stress is mediated by several factors including H-NS, RimJ and the CpxAR two component system (178, 425-427).

Like P fimbriae, the phase variable expression of F1845 and S fimbriae (414), and the outer membrane protein Antigen 43 (165), also involves differential DNA methylation.

1.4.2 Phase variation of type 1 fimbriae

Type 1 fimbrial phase variation is due to the inversion of a DNA element, termed the *fim* switch, which contains the promoter for the *fim* operon (Fig. 1.4). The *fim* switch is 314 bp long and is flanked by 9 bp inverted repeats. The orientation of this switch determines whether the *fim* structural genes such as *fimA*, which encodes the major fimbrial subunit, are transcribed or not (1). Inversion is catalysed by the site-specific recombinases FimB and FimE (209). These proteins share 48% amino acid identity but have distinct activities. FimB inverts the switch in both directions with approximately equal efficiencies, while FimE promotes inversion at higher frequencies than FimB and primarily switches in the ON-to-OFF direction (251). The genes encoding the recombinases are differentially expressed under various growth conditions and the orientation of the *fim* switch is determined by whichever activity predominates. For example, *fimB* and *fimE* are differentially modulated by H-NS in a temperature dependent manner, resulting in preferential switching to the phase ON state at 37°C (289). H-NS is also proposed to affect type 1 fimbrial expression by directly interacting with the *fim* switch (285). IHF activates expression from the *fimA* promoter and (94), along with another global regulator Lrp, plays an accessory role in inversion by binding within the *fim* switch and facilitating the formation of a nucleoprotein complex that is competent for recombination (33, 34). The intracellular levels of the amino acids leucine, alanine, isoleucine and valine have been demonstrated to affect Lrp-mediated stimulation of switching and the interaction of Lrp with the *fim* switch (140, 334). Furthermore, the stress σ factor, σ^{38}, influences type 1 fimbrial expression as it has a negative regulatory role in the expression of *fimA* and *fimB* (96). Environmental signals can thus affect phase variation of type 1 fimbriae by influencing the inversion event itself, or the expression levels of the recombinases FimB and FimE (413).

1.5 The AraC family of transcriptional regulators

The defining feature of the AraC family of transcriptional regulators is a 100 amino acid region of homology that contains two predicted HTH DNA binding motifs (139). The family name derives from the AraC regulator of the L-arabinose operon, which was the first member to be identified and characterised. *In silico* analysis has led to the identification of several hundred family members in a wide range of bacterial species (104, 192). However, insolubility is a characteristic of these proteins, therefore only a fraction have been experimentally characterised. Although members of the AraC family regulate the expression of genes involved in diverse cellular processes, they can be broadly categorised as regulators of virulence genes, stress response-related genes or genes involved in metabolism, most commonly the catabolism of carbon sources (248). While there are a few exceptions (see Section 1.5.2), AraC family proteins are generally 250-300 amino acids long and comprise two structural domains (139). The conserved C-terminal domain contains the signature sequence of the family and is thus involved in DNA binding. The poorly conserved N-terminal domain is often involved in effector binding/oligomerisation, although in many cases it is of unknown function. The activity of AraC-like proteins with effector binding domains is modulated by the direct interaction of a ligand, typically a low molecular weight compound such as a sugar (304). In a small number of cases however, protein ligands have been demonstrated to influence transcriptional activation by AraC family members (74, 164). Co-crystallisation of the MarA and Rob proteins with their target promoters facilitated the study of DNA binding by AraC family members. Only the N-terminal HTH of Rob was found to be involved in specific DNA recognition (221), while MarA was demonstrated to use both HTH motifs for specific DNA contacts (327). Therefore, AraC-like proteins may exhibit two modes of DNA binding.

Most AraC family members are transcriptional activators, although a number can also repress transcription, in some cases by mediating DNA looping. The classic example is the AraC protein itself (see Section 1.5.1.1) (232). The mechanisms by which AraC family members activate transcription are often not fully understood. These proteins usually have multiple binding sites at their target promoters (139) and some, XylS and MelR for example, contact the αCTD and/or σ^{70} subunit of RNA polymerase and thus act as Class I or Class II activators (155, 340). In addition, many AraC family members antagonise repression mediated by the global regulator H-NS (see Section 5.1).

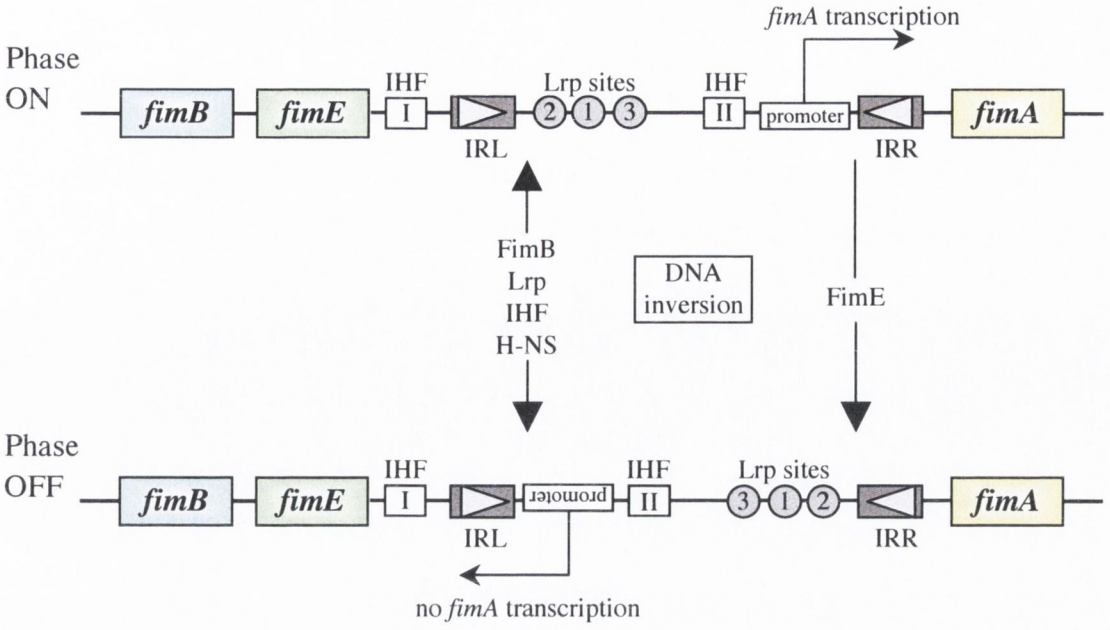

Fig. 1.4. Phase variation of type 1 fimbrial expression. The orientation of the invertible *fim* switch determines whether transcription of *fimA* (encodes the major subunit of type 1 fimbriae) occurs or not. As indicated, the *fim* switch contains the promoter of the *fim* operon and is flanked by inverted repeats (IRL and IRR). Inversion of the *fim* switch is catalysed by FimB and FimE and is influenced by Lrp, IHF and H-NS. The two binding sites for IHF (IHF I and IHF II) are shown as boxes while the three Lrp binding sites are shown as numbered circles.

1.5.1 AraC family members involved in carbon metabolism

1.5.1.1 AraC

In addition to regulating its own expression, the homodimeric AraC protein regulates operons involved in the uptake (*araE* and *araFGH*) and catabolism (*araBAD*) of L-arabinose (355). Each AraC monomer consists of an N-terminal dimerisation domain that contains an arabinose-binding pocket, and a C-terminal DNA binding domain (Fig. 1.5 (A)) (46). The mechanism by which this protein regulates the *araBAD* promoter has been most extensively studied (354). AraC represses transcription of the *araBAD* operon in the absence of arabinose and activates its expression when arabinose is present (115). A vast array of experimental data led to the proposal of the light-switch mechanism, involving the N-terminal arm of approximately 18 amino acids that extends from the dimerisation domain of the protein, as a model to explain the response of AraC to arabinose (Fig. 1.5 (B)) (350). The model proposes that in the absence of arabinose, the N-terminal arms bind to the DNA binding domains (323, 435), and hold them in a configuration that favours the binding of one AraC monomer to the I_1 DNA site upstream of the *araBAD* promoter and the other monomer to the well separated O_2 site located 210 bp away (362). This results in the formation of a DNA loop that represses transcription as it prevents RNA polymerase from accessing the *araBAD* promoter (166). When arabinose is present, the arms reposition over the bound arabinose and bind to the dimerisation domains (142), thus releasing the DNA binding domains. It is now energetically more favourable for the DNA binding domains to occupy the adjacent I_1 and I_2 DNA sites at the *araBAD* promoter from where, in combination with CRP, AraC activates transcription by directly interacting with RNA polymerase (447, 448). More recent findings have led to a slight refinement of the light switch model as experimental evidence suggests that the N-terminal arm of AraC is stably structured on the dimerisation domain in both the presence and absence of arabinose (333). It is now proposed that when arabinose is absent, the arm acts as a "gasket" between the DNA and dimerisation domains of the protein but in the presence of arabinose the arm restructures and ceases to act as a "gasket" (333). Thus the DNA binding domains are freed, enabling AraC to reposition on the DNA and activate the *araBAD* promoter.

1.5.1.2 RhaR and RhaS

The AraC family members RhaR and RhaS are activators of the L-rhamnose catabolic regulon of *E. coli*. The RhaR protein activates transcription of the *rhaSR* operon, which encodes each of these regulators (404), while the RhaS protein activates transcription of the L-rhamnose catabolic (*rhaBAD*) and transport (*rhaT*) operons (105, 415). RhaR and RhaS share 30% amino acid identity and both proteins consist of an N-terminal dimerisation/effector binding domain and a C-terminal DNA binding/transcription activation domain (428, 429). At their target promoters, RhaR and RhaS bind as dimers to sites overlapping the -35 hexamer and contact the σ^{70} subunit of RNA polymerase to activate transcription (428). RhaS may also contact αCTD (183). Maximal activation of all three *rha* operons requires CRP in addition to RhaR or RhaS (105, 182, 415). Both RhaR and RhaS only activate transcription in the presence of L-rhamnose. It was recently reported that L-rhamnose binding induces a structural change in the proteins that is proposed to result in increased DNA binding for RhaS and improved contact with RNA polymerase for RhaR (211).

1.5.1.3 XylS

The *Pseudomonas putida* XylS protein activates transcription of the *meta*-cleavage pathway operon, which encodes the enzymes involved in the catabolism of benzoates and alkylbenzoates (194). XylS becomes transcriptionally active in the presence of effectors such as 3-methlybenzoate (3MB), a pathway substrate (320). When XylS is overproduced, however, it can activate transcription even in the absence of benzoate effectors (257). It has been proposed that the N-terminal domain of XylS acts as an intramolecular repressor by interacting with the C-terminal domain and preventing it from binding to DNA (201). Recently it was reported that binding of the 3MB effector releases this intramolecular repression (86), in addition to its previously described role in stimulating XylS dimerisation (339). XylS activates transcription by recruiting RNA polymerase to the target promoter (86), possibly by directly contacting αCTD (340).

1.5.2 AraC family members involved in stress response

MarA, SoxS and Rob are members of a subgroup of the AraC family of transcriptional regulators, which was recently found to also include TetD (158, 248).

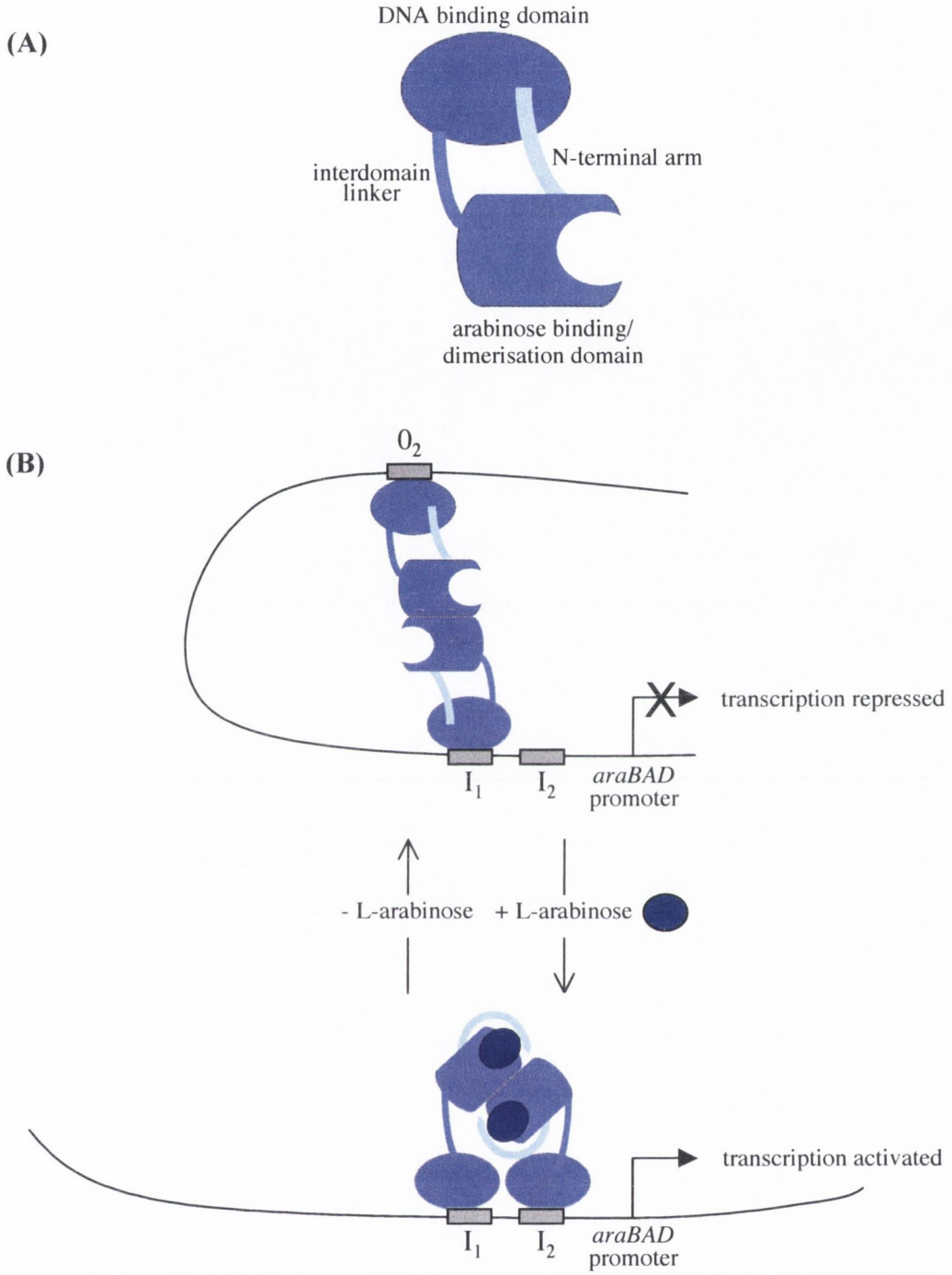

Fig. 1.5. AraC and the light switch mechanism of regulation of the *araBAD* promoter. **(A)** Schematic diagram of the domains of an AraC monomer. **(B)** Simplified view of the regulation of *araBAD* transcription. In the absence of L-arabinose the dimeric AraC protein binds to the 0_2 and I_1 sites and a repression loop is formed. In the presence of L-arabinose the AraC dimer binds to the I_1 and I_2 sites and transcription from the *araBAD* promoter is then activated. Adapted from Schleif (354).

These proteins regulate a set of genes, known as the *marA/soxS/rob* regulon, that are involved in mediating tolerance of organic solvents and resistance to multiple antibiotics, superoxides and heavy metals (16, 305). They are unusual members of the AraC family, as MarA and SoxS consist only of the conserved DNA binding domain, and although Rob does have a second domain of unknown function, its DNA binding domain is located in the N-terminal portion of the protein (139). MarA, SoxS and Rob are highly homologous and regulate their target genes by binding to degenerate asymmetric sites (marboxes) upstream of each promoter (246). The orientation of the marbox relative to the promoter determines if the proteins act as Class I or Class II activators, or as repressors of transcription (246, 358). The extent to which the transcription of the regulon genes is affected depends on the concentration of MarA, SoxS and Rob, and on the relative affinity of the proteins for each promoter (245, 247).

1.5.3 AraC family members involved in virulence

1.5.3.1 Rns

In 1989 a plasmid-located gene was identified as being required for the expression of CS1 and CS2 fimbriae in ETEC and was thus named *rns*, for regulation of CS1 and CS2 (49). This gene was found to encode a regulatory protein that was included in the AraC family of transcription factors due to its homology to the other family members known at the time. The Rns protein is a 265 amino acid polypeptide with a molecular mass of ~30 kDa and a highly basic pI of 10.1, consistent with a role in DNA binding (49). Rns is encoded on a plasmid that, in some ETEC strains, also encodes CS3 fimbriae in addition to LT and/or ST (265), This plasmid is distinct from the pCoo plasmid that encodes the *coo* genes for CS1 fimbriae (135). Codons rarely used in *E. coli* occur frequently in the *rns* gene, which also has a comparatively low GC content of 28% (49). The average GC content of *E. coli* genes is 50%, thus it appears that *rns* was recently acquired from an organism with a low GC content, for example a species of *Campylobacter*, *Mycoplasma*, or *Clostridium*. However, hybridisation to a *rns* probe was not detected in such species, therefore the origin of *rns* remains unknown (50).

When the major promoter of the *coo* operon was identified it was confirmed to be activated by Rns, but was also found to be negatively regulated by H-NS (270). It was

demonstrated that Rns could counteract H-NS mediated repression of *coo* transcription (see Chapter 5), which was reported to be largely dependent on the presence of a negative regulatory region downstream of the *coo* promoter (270). Subsequently, Rns was found to activate the expression of the *coo* genes directly by binding to two sites upstream of the *coo* promoter. The precise locations of the Rns binding sites were identified by DNase I footprinting (268). Site I extends from position -129 to -93 (numbering relative to the transcription start site) and site II extends from position -63 to -23 (Fig. 1.6). As site II overlaps the -35 hexamer it has been suggested that Rns may activate transcription from the *coo* promoter by forming direct contacts with RNA polymerase, but this has not yet been verified to occur. Rns has a similar affinity for sites I and II and can occupy each site alone or bind to both simultaneously (268). Maximal activation of *coo* promoter transcription is dependent on both sites I and II as single nucleotide alterations in either site decreased Rns-dependent expression from the promoter (268). Identification of the nucleotides that Rns interacts with at sites I and II revealed that Rns binding sites are asymmetric (268). Rns is proposed to interact with DNA in a manner typical of AraC family members such as MarA, i.e. bind along one face of the DNA helix, using both of its predicted HTH motifs to contact different sets of nucleotides in two adjacent regions of the major groove (268).

Rns activates its own expression (134). Positive regulation of *rns* transcription by Rns was reported to involve two regions of DNA, one upstream and one downstream of the *rns* promoter (134). Repression of *rns* transcription by an unknown factor, speculated to be H-NS, was also found to occur and it was postulated that Rns acted, possibly indirectly, to overcome this repression (134). However, a later study employed DNase I footprinting to demonstrate that Rns binds directly to the *rns* promoter region (269). Rns has a highly unconventional arrangement of binding sites at its own promoter (Fig. 1.7). Rns binds to a site far upstream of the *rns* promoter, site 1 centred at -227 (relative to the transcription start site), and to two sites downstream of the promoter, site 2 centred at +43 and site 3 centred at +82 (269). Despite their unusual locations, mutations within sites 1 and 3 eliminated Rns-dependent transcription from the *rns* promoter, thus both of these sites are required for positive autoregulation by Rns (269). The role of site 2 is unclear, as it was not found to be critical for activation of the *rns* promoter. Rns binding to site 2 is suggested to enable co-operative binding of Rns to site 3, which is essential for activation (269). Occupancy of these binding sites by Rns was demonstrated to facilitate the binding and formation of a productive open

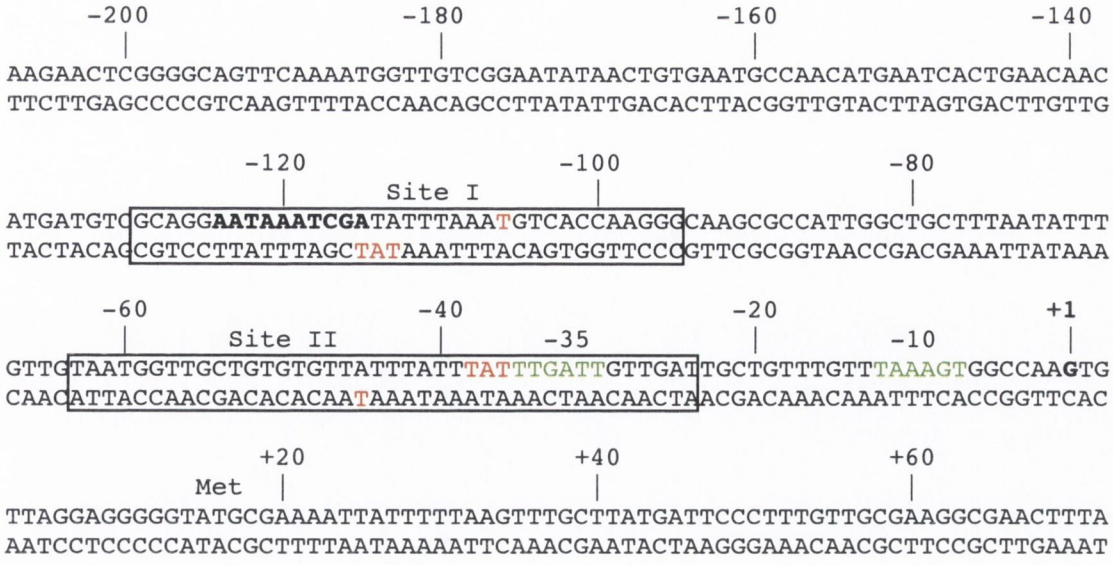

Fig. 1.6. DNA sequence of the *coo* promoter region. Numbering is relative to the transcription start site, which is indicated as +1. The -35 and -10 hexamers of the *coo* promoter are coloured green. The Rns binding sites identified by DNase I footprinting in reference 268 are boxed. Within the Rns binding sites the conserved adenosine nucleotide, and the thymine nucleotides found to be critical for Rns binding in reference 268, are coloured red (see Chapter 5). The potential high affinity H-NS binding site present within Rns binding site I is shown in bold (see Chapter 5).

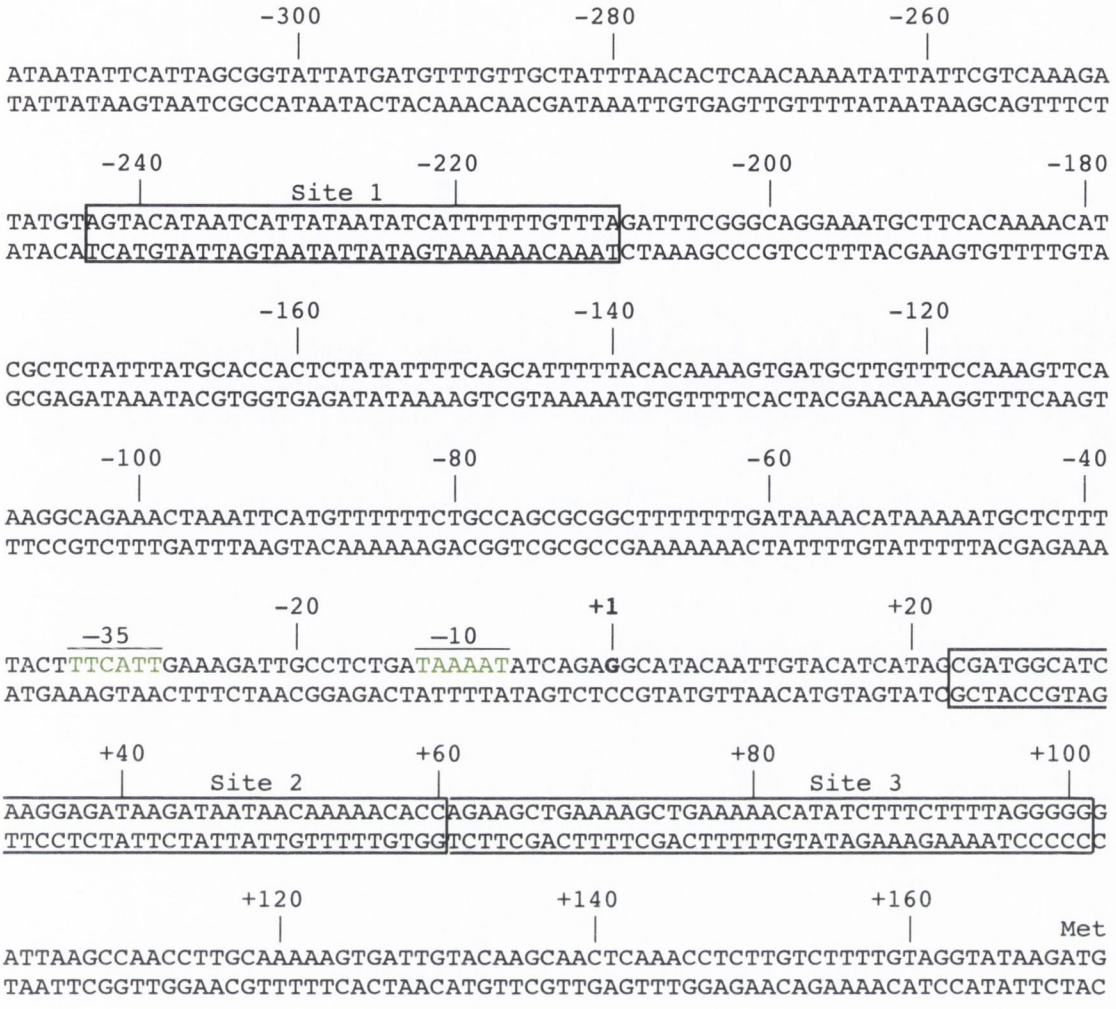

Fig. 1.7. DNA sequence of the *rns* promoter region. Numbering is relative to the transcription start site, which is indicated as +1. The -35 and -10 hexamers of the *rns* promoter are coloured green. The Rns binding sites identified by DNase I footprinting in reference 269 are boxed.

complex at the *rns* promoter by RNA polymerase (269). There is speculation that Rns activates *rns* transcription by directly interacting with RNA polymerase and that this interaction may involve looping of DNA to bring distally bound Rns into proximity with RNA polymerase, however this has not been confirmed to occur (269).

Activation of transcription by binding to sites downstream of the target promoter is extremely unusual for prokaryotic transcription factors as downstream binding sites are associated with repressors of transcription. Based on the unprecedented arrangement of binding sites at its own promoter, it is suggested that Rns may represent a novel class of prokaryotic activators (269). There are a few other bacterial regulators, including DnaA (148), phosphorylated PhoP (231), MetR (61) and LadA (417), that positively affect transcription from downstream binding sites. The mechanisms used by regulators to activate transcription from such sites are largely unknown. MetR binds upstream and downstream of the *metF* promoter to activate transcription by antagonising the effect of the MetJ repressor (61). LadA is also proposed to function as an anti-repressor (417), however it is unlikely that Rns acts in this way, as it is reported that there does not appear to be any repressor binding sites downstream of the *rns* promoter (269).

Amongst the subset of AraC family members that regulate virulence factors there is a group of proteins with strong homology to Rns (Fig. 1.8). Several of these proteins (CfaD, CsvR, AggR and PerA) also activate fimbrial expression, while VirF regulates the expression of genes required for the invasion of *S. flexneri*. Some members of this subfamily of Rns-like regulators are so closely related that they can, to varying extents, substitute for one another in the regulation of their target promoters (267, 312). The finding that AggR, CfaD and VirF could activate *rns* transcription and required Rns binding site 3 to do so, led to the proposal that Rns is the prototype for a family of regulators that are not limited to acting at promoter proximal upstream binding sites (267). Most of these Rns-related proteins, in addition to ToxT and UreR that are more distantly related to Rns, are discussed further in the following sections.

Searches for additional Rns-regulated genes revealed that Rns binds upstream and activates the expression of *yiiS*, a gene of unknown function (266), and *cexE*, which encodes an uncharacterised extracytoplasmic protein (303). Most interestingly, it was recently reported that Rns represses the expression of NlpA, an inner membrane lipoprotein thought to contribute to the production of outer membrane vesicles that deliver LT into eukaryotic cells (35). Thus Rns may be indirectly involved in restricting the release of this toxin. Rns represses *nlpA* transcription by binding immediately

downstream of the transcription start site and inhibiting the formation of an open complex at the *nlpA* promoter by RNA polymerase (35). Therefore Rns is a bifunctional transcription factor with a diverse and growing regulon.

The two predicted HTH motifs of Rns are present in its C-terminal half, leading to the assumption that the C-terminal domain of Rns is involved in DNA binding, although this has not been verified previously. Less is known about the role of the N-terminal domain of Rns. Like most other virulence regulators of the AraC family, Rns does not appear to respond to an exogenous ligand (19). Thus, while the possibility cannot be ruled out conclusively, it is unlikely that the N-terminal portion of Rns is an effector binding domain. Additionally, it was recently reported that the N-terminus of Rns does not act as a dimerisation domain (see Chapter 4 for further discussion of Rns dimerisation) (19). However, the N-terminal domain was found to be essential for transcriptional regulation and DNA binding by Rns (19). The isolation of two N-terminal point mutations that eliminated the autoregulation activity of Rns led to the proposal that the N-terminal domain of the protein may interact with the C-terminal domain in order to enable DNA binding (19). In the absence of evidence for such an interaction, the role of the N-terminal domain of Rns remains unknown. Indeed, there is much to be discovered about the structure and mechanism of action of the Rns protein.

1.5.3.2 CfaD

Expression of CFA/I fimbriae in ETEC requires two widely separated regions of DNA on the wild type plasmid NTP113 (372). The operon containing the structural and assembly genes necessary for fimbrial production is present in region 1 (200) while region 2 encodes the regulatory protein, CfaD, which activates the promoter of region 1 (349). CfaD, also named CfaR (51), shares 95% amino acid identity with Rns. These two proteins are completely functionally interchangeable and can bind to the same DNA sites (267). CfaD was found to have two binding sites upstream of the CFA/I promoter in a similar arrangement to that of the Rns sites at the *coo* promoter (303), and three predicted binding sites in the vicinity of its own promoter in an arrangement almost identical to that of Rns at the *rns* promoter (267). It is not known if CfaD is a positive autoregulator like Rns, however CfaD activates and represses, respectively, expression of the *cexE* and *nlpA* genes in a manner similar to Rns (35, 303). When activating CFA/I fimbrial expression, CfaD acts partly to counteract the repressive effect of H-NS,

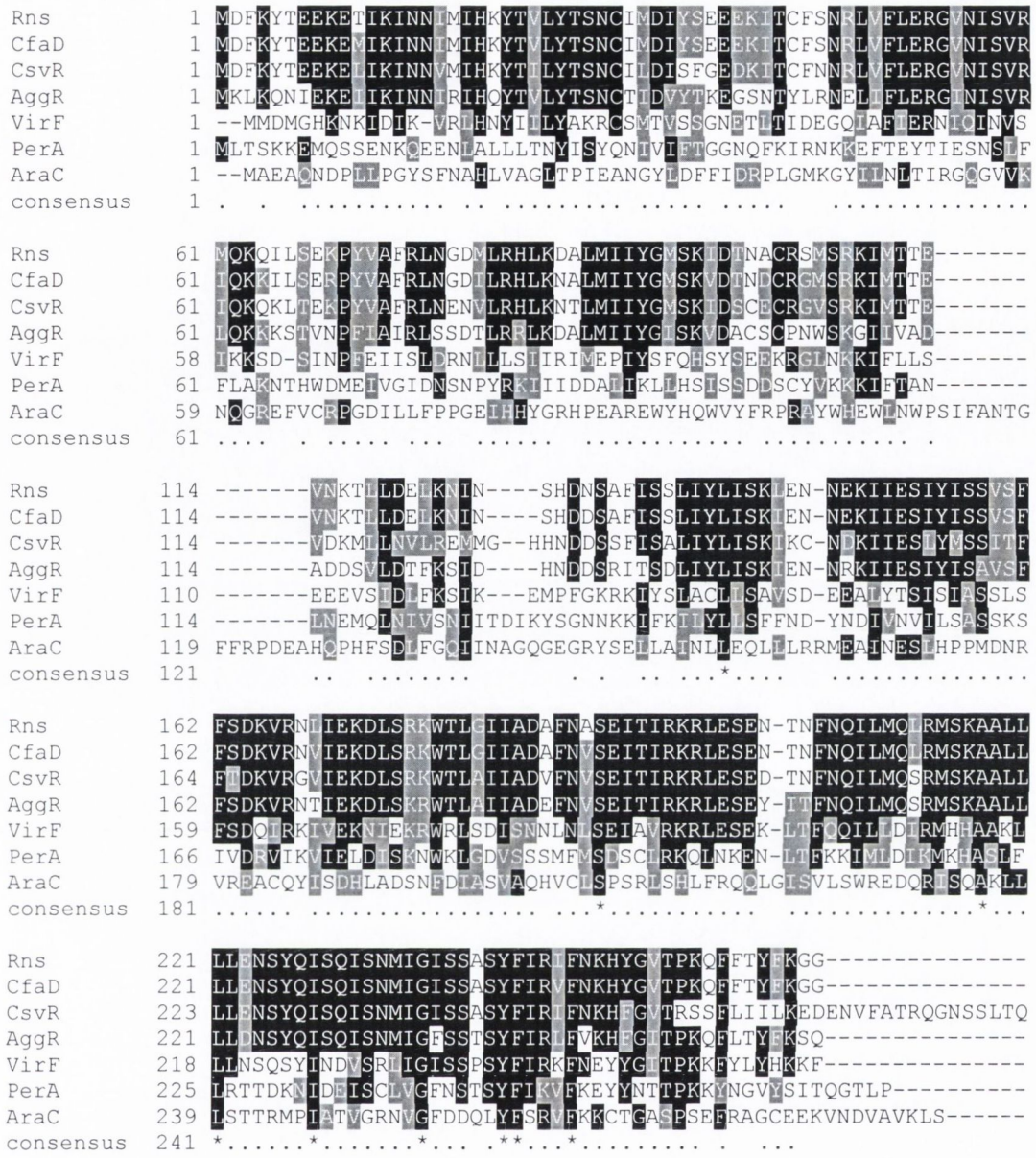

Fig. 1.8. Sequence alignment of Rns-related regulators and AraC. The alignment was carried out using the Clustal programme. Residues identical or similar to those of Rns, which are shared among four or more of the aligned proteins, are shaded in black or grey respectively. In the consensus line an asterisk marks identity and a dot marks partial identity or homology between the proteins.

which silences the CFA/I promoter (199). CfaD was the first AraC-like transcriptional activator reported to overcome H-NS-mediated repression, but it does not act only as an anti-repressor as it can also activate the CFA/I promoter in the absence of H-NS (see Chapter 5) (199). The *cfaD* gene is not transcribed at 20°C and this may be the main cause of the temperature dependent expression of CFA/I fimbriae (see Chapter 6) (199).

1.5.3.3 PerA

The *bfpTVW* operon of EPEC, which had previously been designated the *perABC* locus, was identified as being required for the activation of the *bfp* operon that encodes bundle-forming pili (see Section 1.1.2.1) (150, 401). Initially it was thought that this activation was mediated by PerA in association with the accessory proteins PerB and PerC, however it has been demonstrated that PerA alone directly activates *bfp* transcription (311). PerA also positively regulates its own expression by activating transcription from the *perABC* promoter (250). This autoactivation of the *per* operon leads to up-regulated expression of PerC, which in turn activates the Ler-encoding LEE1 operon (see Section 1.1.2.1) thus PerA is indirectly involved in the regulation of A/E lesion formation by EPEC (311). Indeed PerA is required for full EPEC virulence, as a study performed with human volunteers found that a *perA* mutant caused significantly less diarrhoea than a wild type EPEC strain (29). PerA is closely related to Rns. The two proteins share 51% amino acid similarity and like Rns at the *coo* promoter, PerA-mediated transcriptional activation involves PerA overcoming H-NS-dependent repression at the *per* and *bfp* promoters (311). However, PerA and Rns cannot cross-substitute for each other to regulate their target promoters (267).

1.5.3.4 AggR

The EAEC protein AggR was initially thought to act only as a transcriptional activator of genes involved in fimbrial biogenesis. AggR is required for the expression of AAF/I (278), AAF/II (107) and the recently identified adhesin Hda (see Section 1.1.2.3) (36). Subsequently AggR was found to positively regulate expression of *aap*, which encodes the secreted protein dispersin that promotes EAEC dispersal across the intestinal mucosa (367), and of the *aat* cluster of genes that are required for the surface translocation of dispersin (281). These factors are all encoded on the pAA virulence

plasmid of EAEC, however AggR also regulates chromosomal genes. Dudley *et al.* demonstrated that 25 contiguous genes (*aaiA-Y*), encoded within a pathogenicity island inserted at the *pheU* locus, are under the control of AggR (98). A group of these chromosomally encoded AggR-activated genes are thought to comprise a type VI secretion system that directs secretion of AaiC, however the role of most of the *aai* genes is currently unclear (98). Recently AggR was also found to activate transcription of the pAA-encoded *shf* gene, which has been implicated in biofilm formation by EAEC (136). This has lent further support to the proposal that AggR, which autoactivates its own expression, is a master regulator of virulence functions in EAEC (170).

1.5.3.5 VirF

The VirF protein is the primary regulator of the virulence gene cascade of *S. flexneri* (95). VirF directly activates transcription of *icsA*, which encodes a protein involved in intra- and intercellular spread of the bacteria, and of the gene encoding the secondary regulator, VirB, which in turn activates expression of the invasion genes (3) (see Section 1.1.2.4). The *icsA* and *virB* promoters are repressed by H-NS, and VirF acts, at least in part, to overcome this repression (68, 402). VirF is involved in the thermoregulation of virulence gene expression in *S. flexneri*, as both *virF* transcription and VirF-mediated activation of *virB* are influenced by temperature (see Chapter 6) (126, 403). In addition to temperature-dependent H-NS-mediated repression, transcription of the *virF* gene is subject to negative autoregulation (308) and activation by the nucleoid-associated proteins IHF and Fis (128, 310). Expression of *virF* is also affected by post-transcriptional RNA modification (101, 191). This complex control of VirF production is necessary due to the central role the protein plays in regulating *S. flexneri* virulence.

1.5.3.6 ToxT

ToxT directly activates transcription of the *tcp* and *ctx* operons, which encode the major virulence factors of *V. cholerae*; toxin-coregulated pilus and cholera toxin respectively (81). ToxT also directly activates several other genes (*acfA*, *acfD*, *aldA*, *tcpI* and *tagA*) whose role in virulence is unclear (81). Expression of the cytoplasmic ToxT protein is activated by the membrane-localised proteins ToxR and TcpP (217),

and repressed by H-NS (284). The *toxT* gene is located within the *tcp* operon, therefore once it is expressed ToxT then activates its own expression (443). ToxT consists of a C-terminal DNA binding domain and an N-terminal dimerisation and environmental sensing domain (317). ToxT is not known to bind an effector, however environmental signals including temperature and bile were demonstrated to repress ToxT activity (360). Both the N- and C-terminal domains of ToxT are involved in transcriptional activation (57, 317). ToxT binds to specific sites, known as toxboxes (432), upstream of its target promoters and is thought to activate transcription by directly contacting RNA polymerase subunits (187, 444). At the *ctx* promoter, and to a lesser extent at the *tcp* promoter, ToxT also acts to overcome H-NS-mediated repression (284, 444).

1.5.3.7 UreR

UreR is the transcriptional activator of the genes required for urease production in several species of *Enterobacteriaceae* (139). Urease, an important virulence factor of uropathogens, hydrolyses urea to ammonia and carbon dioxide, which leads to an increase in local pH that facilitates bacterial proliferation and urinary stone formation (59). The urease gene cluster consists of the structural and accessory genes *ureDABCEFG* in addition to the upstream and divergently transcribed *ureR* gene (67). Transcription of these genes is dependent on both UreR and the presence of urea. UreR binds urea (141), and then binds with high affinity to two sites in the *ureR-ureD* intergenic region to activate transcription from its own promoter and the *ureD* promoter (400). UreR, a dimeric protein (306), is the only AraC-like virulence regulator known to respond to an effector. Expression of the urease gene cluster is repressed by H-NS and it has been demonstrated that UreR and H-NS compete for binding to, and displace each other from, the *ureR-ureD* intergenic region (58, 307).

1.6 Scope of this thesis

The aims of the work presented in this thesis were to define the regions of the Rns transcriptional regulator of ETEC that are critical for its function and to investigate whether the protein oligomerises. The regulation of transcription from the *coo* promoter by Rns, H-NS and temperature was also examined.

Chapter 2 Materials and Methods

2.1 General Methods

2.1.1 Bacterial strains and culture conditions

2.1.1.1 Bacterial strains

All *E. coli* strains used in this study are listed in Table 2.1. Permanent stocks were maintained in L broth supplemented with 8.7% (v/v) DMSO and stored at –80°C.

2.1.1.2 Bacterial growth media

All media were prepared using Millipore 18 $M\Omega cm^{-1}$ grade water and chemicals obtained from Difco, Oxoid and Sigma. Media were sterilised by autoclaving at 120°C for 20 min prior to use. Additives not suitable for autoclaving were sterilised by filtration through 0.2 μm Millex filters (Millipore). The quantities listed below are sufficient for 1 litre of medium.

Lennox L broth and agar:

L broth and agar were used throughout this study for the routine culturing of all bacterial strains except where otherwise stated.

L broth: 10 g enzymatic digest of casein, 5 g yeast extract (low sodium), 5 g NaCl

L agar: 9.14 g enzymatic digest of casein, 4.57 g yeast extract (low sodium), 4.57 g NaCl, 13.72 g bacteriological agar

Luria broth:

Luria broth was used for the growth of cultures prior to the purification of His-tagged H-NS and to grow *E. coli* SU101 strains for use in assays of β-galactosidase activity.

Luria broth: 10 g tryptone (pancreatic digest of casein), 5 g yeast extract, 10 g NaCl

Colonisation factor antigen (CFA) medium:

Bacteria were cultured in CFA medium to study the expression of fimbriae (122).

CFA broth: 10 g Casamino acids (Difco), 1.5 g yeast extract (Difco), 50 mg $MgSO_4$, 5 mg $MnCl_2$, pH 7.4.

CFA agar: 10 g Casamino acids (Difco), 1.5 g yeast extract (Difco), 50 mg $MgSO_4$, 5 mg $MnCl_2$, 20 g agar, pH 7.4.

SOC medium:

SOC medium was used following electroporation of *E. coli* strains to increase the recovery of viable bacteria after transformation.

SOC medium: 20 g Oxoid tryptone, 5 g yeast extract, 0.5 g NaCl

After autoclaving 0.95 g $MgCl_2$, 1.2 g $MgSO_4$ and 1.8 g glucose was added.

Overnight Express™ Instant TB Medium:

Overnight Express (OnEx™) Instant TB medium is based on components that are metabolised differentially to promote growth to high density and induce protein expression from *lac* dependant promoters (394). It is composed of Terrific Broth (TB) and three other solutions: (i) an induction solution that is a blend of carbon sources (glucose, glycerol and lactose) optimised for tightly regulated uninduced growth to high cell density followed by induction with lactose and continued growth; (ii) a buffering solution that maintains pH throughout metabolic acid production and provides additional nitrogen to support increased protein synthesis; and (iii) a solution that provides a critical magnesium concentration to achieve maximum cell density. It was used for the growth of cultures prior to protein purification.

MacConkey Salicin Medium:

Bacteria were grown on MacConkey agar plates containing 1% salicin (a β-glucoside) to confirm that they were *hns* mutants. Generally *E. coli* cannot ferment β-glucosides because the *bgl* operon is cryptic. Mutations in the *hns* gene result in a β-glucoside-fermenting phenotype (78) that causes the bacteria to appear red on MacConkey salicin agar as opposed to white (non-fermentative).

MacConkey salicin agar 40 g MacConkey agar base (Difco), after autoclaving 10 g of salicin was added

TABLE 2.1. *Escherichia coli* strains used in this study

Strain	Relevant details	Reference or source
DH5α	F′ *endA1 hsdR17* ($r_k^- m_k^+$) *glnV44 thi-1 recA1 gyrA96 relA1 Δ(lacIZYA-argF)U169 deoR (*Φ80d*lacΔ (lacZ)M15)*	Invitrogen Life Technologies
KS1000	F′ *lacI*^{q} *lac*$^+$ *pro*$^+$ */ ara Δ(lac-pro) Δ(tsp)*::Kan^R *eda51::Tn10(*Tet^R*) gyrA(*Nal^R*) rpoB thi-1 argI(am)*	(370)
LMC10	A *lac* deletion restriction-negative derivative of the O6:K15:H16 ETEC derived strain C921b-2, which was cured of the plasmid encoding Rns and CS3	(49)
LMC10Δ*hns*	LMC10 *hns*::Ap^R	This study
MC4100	F$^-$ *araD139 Δ(argF-lac)U169 rpsL150 relA1 flb-5301 ptsF25 deoC1 rbsR*	(52)
NEB 5-alpha	F′ *proA*$^+$*B*$^+$ *lacI*^{q} *Δ(lacZ)M15 zzf::Tn10(*Tet^R*)/fhuA2Δ(argF-lacZ)U169* (Φ80 *Δ(lacZ)M15) glnV44 gyrA96 recA1 endA1 thi-1 hsdR17*	New England Biolabs
PD32	MC4100 *hns-206*::Ap^R	(80)
SU101	*E. coli* JL 1434 *lexA71*::Tn5 (Def) *sulA211 Δ(lacIPOZYA)169/*F′ *lacI*^{q} *lacZ*ΔM15::Tn9 (op$^+$/op$^+$)	(82)
T7 Express *I*^{q}	F′ *proA*$^+$*B*$^+$ *lacI*^{q}*zzf::Tn10(*Tet^R*)/fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10—*Tet^S*)2 [dcm] R(zgb-210::Tn10—*Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^R)]	Stratagene

2.1.1.3 Bacterial culture conditions

Bacteria were routinely grown on L agar plates and in shaken aerobic liquid cultures at 37°C, except where otherwise stated. Liquid cultures were inoculated by transferring single colonies from agar plate cultures into an appropriate volume of L broth and grown overnight. Where mid-logarithmic cultures were required, overnight cultures were diluted 1:50 in fresh media and grown to the appropriate optical density at 600 nm.

For thermoregulation studies it was critical that the growth temperature of certain cultures was maintained at 20°C or below. To achieve this, a water bath set to 20°C was placed in a cold room. Liquid cultures were then grown by incubation without shaking in the water baths maintained at 20°C or 37°C. Agar plates were grown as usual in a 37°C incubator or were placed in waterproof plastic bags and submerged in the 20°C water bath.

2.1.1.4 Antibiotics and media additions

All stock antibiotics and media additives were filter sterilised through 0.2 μm Millex filters (Millipore) and stored at –20°C. Carbenicillin, kanamycin and spectinomycin were prepared as 50 mg/ml stocks in ddH_20 and used at a working concentration of 50 μg/ml. Tetracycline and chloramphenicol were prepared as 10 mg/ml and 30 mg/ml stocks respectively in 70% (v/v) ethanol and used at working concentrations of 10 μg/ml and 30 μg/ml respectively. The *lac* operon inducer IPTG was prepared as a 100 mM stock in ddH_20 and used at concentrations of 0.1-1 mM as appropriate. X-Gal, a chromogenic substrate for β-galactosidase, was prepared as a 25 mg/ml solution in *N, N*-dimethyl formamide and used in agar plates at a final concentration of 50 μg/ml.

2.1.2 Plasmids and oligonucleotides

2.1.2.1 Plasmids

All plasmids used in this study are listed in Table 2.2 with relevant descriptions and sources. For plasmids constructed in this study the details of construction are described in the relevant chapters.

TABLE 2.2. Plasmids used in this study

Plasmid	Relevant details	Reference or source
p660AraC	Tc^r pSR660 derivative carrying the *lexA*$_{DBD}$-*araC* fusion	(193)
pACYC177	Ap^r Kan^r p15A replicon, cloning vector	(55)
pASK75	Ap^r ColE1 replicon, expression vector with *tetA* promoter/operator and constitutively expressed *tetR*	(371)
pASK75Spec	Ap^r $Spec^r$ pASK75 with a spectinomycin resistance cassette inserted	This study
pASKrns	Ap^r pASK75 derivative containing anhydrotetracycline-inducible *rns*	This study
pASKrnsSpec	Ap^r $Spec^r$ pASKrns with a spectinomycin resistance cassette inserted	This study
pBSKII+	Ap^r ColE1 replicon, high-copy-number cloning vector	Stratagene
pC102	Spc^r pSS2192 derivative with 15 bp insertion encoding RCLNS after C102 of Rns	This study
pCL1920	Spc^r pSC101 replicon, low-copy-number cloning vector	(226)
pCooGFP	Ap^r Cm^r ColE1 replicon, reporter plasmid containing *coo* promoter region (from bp –393 to bp +62) fused to a promoterless *gfp*+ gene	S.G. Smith unpublished
pCoo-88	Cm^r Wild type *coo* promoter region (from bp –88 to bp + 62) inserted into pZep08	This study
pCoo-96	Cm^r Wild type *coo* promoter region (from bp –96 to bp + 62) inserted into pZep08	This study
pCoo-352	Cm^r Wild type *coo* promoter region (from bp –352 to bp + 62) inserted into pZep08	This study
pCoo-359	Cm^r Wild type *coo* promoter region (from bp –359 to bp + 62) inserted into pZep08	This study
pCoo-359-I	Cm^r *coo* promoter region (from bp –359 to bp + 62) containing CGC substituted for TAT motif at Rns binding site I inserted into pZep08	This study
pCoo-201-II	Cm^r *coo* promoter region (from bp –201 to bp + 62) containing CGC substituted for TAT motif at Rns binding site II inserted into pZep08	This study
pCoo-359-II	Cm^r *coo* promoter region (from bp –359 to bp + 62) containing CGC substituted for TAT motif at Rns binding site II inserted into pZep08	This study
pCoo-146-I&II	Cm^r *coo* promoter region (from bp –146 to bp + 62) containing CGC substituted for TAT motif at both Rns binding sites I and II inserted into pZep08	This study
pCoo-359-I&II	Cm^r *coo* promoter region (from bp –359 to bp + 62) containing CGC substituted for TAT motif at both Rns binding sites I and II inserted into pZep08	This study
pCooGFP-I	Ap^r Cm^r Derivative of pCooGFP with CGC substituted for TAT motif in Rns binding site I	This study
pCooGFP-II	Ap^r Cm^r Derivative of pCooGFP with CGC substituted for TAT motif in Rns binding site II	This study

pCooGFP-I&II	Ap^r Cm^r Derivative of pCooGFP with CGC substituted for TAT motif in both Rns binding sites I and II	This study
pCooGFP-2	Kan^r p15A replicon, reporter plasmid containing *coo* promoter region (from bp –393 to bp +62) fused to a promoterless *gfp*+ gene	This study
pCS1	Ap^r *coo* promoter region (from bp –393 to bp +62) inserted into pBSKII+	S.G. Smith unpublished
pCS1-SiteI	Ap^r Derivative of pCS1 with CGC substituted for TAT motif in Rns binding site I	This study
pCS1-SiteII	Ap^r Derivative of pCS1 with CGC substituted for TAT motif in Rns binding site II	This study
pCS1-SiteI&II	Ap^r Derivative of pCS1 with CGC substituted for TAT motifs in both Rns binding sites I and II	This study
pET-3b	Ap^r pMB1 replicon, cloning and expression vector	Novagen
pF205	Spc^r pSS2192 derivative with 15 bp insertion encoding MFKHF after F205 of Rns	This study
pGPS4	Cm^r Transprimer donor component of GPS-LS system	New England Biolabs
pGPS5	Kan^r Transprimer donor component of GPS-LS system	New England Biolabs
pHisΔNACRS	Ap^r Derivative of pHisRns minus the nucleotides encoding the NACRS amino acid sequence	This study
pHisC102	Ap^r Derivative of pHisRns containing the 15 bp insertion from pC102	This study
pHisC102A	Ap^r *rns* (C102A) ORF in pQE-30	This study
pHisC30	Ap^r pHisRns derivative with 15 bp insertion encoding CLNNC after C30 of Rns	This study
pHisF205	Ap^r Derivative of pHisRns containing the 15 bp insertion from pF205	This study
pHisL198	Ap^r Derivative of pHisRns containing the 15 bp insertion from pL198	This study
pHisL51	Ap^r pHisRns derivative with 15 bp insertion encoding FKQFL after L51 of Rns	This study
pHisQ227	Ap^r Derivative of pHisRns containing the 15 bp insertion from pQ227	This study
pHisR103A	Ap^r *rns* (R103A) ORF in pQE-30	This study
pHisRns	Ap^r *rns* ORF inserted into pQE-30	S.G. Smith unpublished
pHisS158F	Ap^r *rns* (S158F) ORF in pQE-30	This study
pHisT23	Ap^r pHisRns derivative with 15 bp insertion encoding CLNNT after T23 of Rns	This study
pHisV253Δ	Ap^r Derivative of pHisRns containing the 15 bp insertion from pV253Δ	This study
pHisY242	Ap^r Derivative of pHisRns containing the 15 bp insertion from pY242	This study
pHNSHIS	Ap^r *hns* ORF plus C-terminal 6 his codons inserted into pET-3b	This study
pK256	Spc^r pSS2192 derivative with 15 bp insertion encoding CLNTK after K256 of Rns	This study
pKOBEG	Cm^r pSC101 replicon, Temperature-sensitive plasmid carrying the *red* and *gam* genes of λ-phage under control P_{araBAD} promoter	(56)
pKRP13	Sp^r Sm^r Plasmid source of spectinomycin resistance cassette	(322)

pL198	Spc^r pSS2192 derivative with 15 bp insertion encoding FKQRL after L198 of Rns	This study
pL221	Spc^r pSS2192 derivative with 15 bp insertion encoding CLNIL after L221 of Rns	This study
pM-ΔNACRS	Ap^r Derivative of pMRns5 minus the nucleotides encoding the NACRS amino acid sequence	This study
pMAL-c2	Ap^r ColE1 replicon, MBP-protein fusion expression vector	New England Biolabs
pM-C102	Ap^r Derivative of pMRns5 containing the 15 bp insertion from pC102	This study
pM-Q227	Ap^r Derivative of pMRns5 containing the 15 bp insertion from pQ227	This study
pM-RhaSLinker	Ap^r Derivative of pMRns5 with the DNA sequence corresponding to the NACRS amino acid sequence of Rns replaced with nucleotides encoding the LENSASR linker sequence of the RhaS protein	This study
pMRns5	Ap^r *rns* ORF inserted into pMAL-c2	(373)
pPrns2	Ap^r *rns* promoter region inserted into pBSKII+	This study
pQ227	Spc^r pSS2192 derivative with 15 bp insertion encoding CLNNQ after Q227 of Rns	This study
pQE-30	Ap^r ColE1 replicon, expression vector, T5 promoter/*lac* operator element; 5′ six-His tag coding sequence	Qiagen
pRare	Cm^r p15A replicon, encodes tRNAs for rare codons in *E. coli*	Novagen
pRns660	Tc^r pSR660 derivative carrying the $lexA_{DBD}$-*rns* fusion	This study
pRnsLacZ-2	Tc^r *rns* promoter region inserted into pRW50	This study
pRW50	Tc^r RK2 replicon, low-copy-number vector carrying a promoterless *lacZ* gene	(233)
pS239	Spc^r pSS2192 derivative with 15 bp insertion encoding CLNTS after S239 of Rns	This study
pS239Δ	Spc^r pSS2192 derivative with 15 bp insertion encoding VOchTSS after S239 of Rns (Och, ochre stop codon)	This study
pSR660	Tc^r ColE1 replicon, vector encoding the LexA DNA binding domain (DBD) sequence used for homodimerisation studies	(69)
pSS2192	Spc^r *rns* gene and upstream area inserted into pCL1920	(373)
pTrc99a	Ap^r Cloning vector with an IPTG-inducible P_{trc} promoter	(6)
pTrcRns	Ap^r *rns* ORF inserted into pTrc99a	S.G. Smith unpublished
pV253Δ	Spc^r pSS2192 derivative with 15 bp insertion encoding VOchTGV after V253 of Rns (Och, ochre stop codon)	This study
pY242	Spc^r pSS2192 derivative with 15 bp insertion encoding CLNTY after Y242 of Rns	This study
pZep08	Cm^r Ap^r ColE1 replicon, encodes promoterless *gfp*+ gene downstream of MCS	(172)

2.1.2.2 Oligonucleotides

The sequences of all oligonucleotides used in this study are listed in Table 2.3. All oligonucleotides were synthesised by MWG-Biotech, Ebersberg, Germany.

2.2 Nucleic acid methodologies

2.2.1 Transformation of *E. coli* strains

Two distinct methods were used for the introduction of foreign DNA into *E. coli* cells. Cells were made competent either by repeated washes with a cold calcium chloride solution, or cold Millipore grade water, and were then transformed by heat shock (242) or electroporation (97) respectively. Significantly greater transformation efficiencies were achieved using the electroporation method. Generally, the calcium chloride method was used for routine transformation of intact plasmids into *E. coli* K-12 strains. The electroporation method was used with *E. coli* K-12 strains when higher efficiencies were required, and was used routinely with ETEC strain LMC10.

2.2.1.1 Transformation of *E. coli* K-12 strains using calcium chloride method

A 3 ml culture of the strain to be made competent was inoculated from a single colony into L broth and incubated overnight with shaking. This culture was then diluted 1:100 in 100 ml of fresh L broth and grown to an OD_{600nm} of between 0.4 and 0.6. The cultures were then incubated on ice for 20 min and the bacteria were harvested by centrifugation at 6000 x *g* for 8 min at 4°C. The pellet was resuspended in 20 ml of cold $CaCl_2$ solution (60 mM $CaCl_2$, 15% (v/v) glycerol, 10 mM PIPES, pH 7) and harvested by centrifugation as before. This pellet was resuspended in 20 ml cold $CaCl_2$ solution and incubated on ice for 30 min before again harvesting the bacteria as before. The pellet was then resuspended in a final volume of 4 ml of cold $CaCl_2$ and incubated overnight on ice before aliquoting in 100 µl volumes and storage at –80°C.

TABLE 2.3. Oligonucleotide primers used in this study

Primer	Sequence
C102-F	5′ – GAT ACC AAT GCT GCT AGA AGC ATG TCA AG –3′
C102-R	5′ – CTT GAC ATG CTT CTA GCA GCA TTG GTA TC –3′
CS1-blunt	5′ – CAC TAA AGG GAA CAA AAG CTG GAG C –3′
CS1-XbaI	5′ – CGC TCT AGA TAC GAC TCA CTA TAG G –3′
CSBS1	5′ – GGC ATA ATA AAG CCG TCC AG –3′
CSBS2	5′ – CGC AAC AAA GGG AAT CAT AAG –3′
hns-flank_F	5′ – AAG CGG GTA AAT GAC TGC TG –3′
hns-flank_R	5′ – CGC AAT TGA TAT GCT GAT CG –3′
InvPCR-1	5′ – ATG TCA AGA AAA ATA ATG ACA ACA G –3′
InvPCR-2	5′ – GGT ATC TAT TTT TGA CAT GCC –3′
LexA-1	5′ – CGC CTC GAG AAT GGA CTT TAA ATA CAC TG –3′
Linker-1	5′ – GCA TCA CGT ATG TCA AGA AAA ATA ATG AC –3′
Linker-2	5′ – GCT GTT TTC CAG GGT ATC TAT TTT TGA C –3′
New_CS1-Blunt	5′ – CGT GAA TTC GGC CGC TGT ATG AAT C –3′
New_Prns-EcoRI	5′ – CAT GAA TTC GGC CGC TGT ATG AAT C –3′
Prns-HindIII	5′ – CGC AAG CTT CAG TGT ATT TAA AGT CC –3′
R103-F	5′ – CCA ATG CTT GTG CTA GCA TGT CAA G –3′
R103-R	5′ – CTT GAC ATG CTA GCA CAA GCA TTG G –3′
Ranmut3	5′ – CAC AGA ATT CAT TAA AGA GGA GAA ATT AAC TAT G –3′
Rns-19r	5′ – CGC GGG ATC CTT ATC CAC CTT TAA AAT AAG TG –3′
Rns-F	5′ – CGC GGA TCC ATG GAC TTT AAA TAC ACT G –3′
RnsF_Xba	5′ – CGC TCT AGA ATG GAC TTT AAA TAC ACT G –3′
RnsKpn	5′ – CGT GGT ACC TTA TCC ACC TTT AAA ATA AGT G –3′
RnsR_Hind	5′ – CGC AAG CTT TTA TCC ACC TTT AAA ATA AG –3′
Rns-UpF	5′ – AGC GGC ATA ACC TGA ATC TG –3′
Rns-UpR	5′ – AAA ACA TGT AAT TTT CTC TTC TTC C –3′
rt-gapF	5′ – CGT TCT GGG CTA CAC CGA AGA TGA CG –3′
rt-gapR	5′ – AAC CGG TTT CGT TGT CGT ACC AGG A –3′
rt-rnsF	5′ – AGC CAA CCT TGC AAA AAG TG –3′
rt-rnsR	5′ – TGC CTT AGC ATA TCT CCG TTC –3′
Site_I_F	5′ – CGC AGG AAT AAA TCG GCG TTT AAA TGT CAC CAA GG –3′
Site_I_R	5′ – CCT TGG TGA CAT TTA AAC GCC GAT TTA TTC CTG CG –3′

Site_II_F	5′ – GCT GTG TGT TAT TTA TTC GCT TGA TTG TTG ATT GC –3′
Site_II_R	5′ – GCA ATC AAC AAT CAA GCG AAT AAA TAA CAC ACA GC –3′
VM1	5′ – GCA GTT CAA AAT GGT TGT CG –3′
VM2	5′ – TGG CCA CTT TAA ACA AAC AGC –3′
VM3	5′ – GCG GTA TTA TGA TGT TTG TTG C –3′
VM4	5′ – AAT CAC TTT TTG CAA GGT TGG –3′

Competent bacterial cells (100 µl), prepared as described above, were mixed with between 10 ng and 1 µg of plasmid DNA and incubated on ice for 10 min. The cells were then transferred to a 42°C water bath and incubated for 90 s, thus inducing the uptake of the DNA. Cells were immediately transferred to ice. Following a 1 min incubation on ice, the cells were mixed with 1 ml of pre-warmed SOC broth and incubated at 37°C for 1 h to allow expression of plasmid-borne antibiotic resistance genes. Aliquots of 50 µl and 1 ml of the transformation mixture were then spread on selective L agar plates and incubated overnight at 37°C. Alternatively, competent cells (*E. coli* NEB 5-alpha and T7 Express I^q) were purchased from New England Biolabs (NEB) and transformed according to the manufacturer's instructions. Transformants were single colony purified on fresh L agar plates containing appropriate antibiotics.

2.2.1.2 Transformation of *E. coli* strains by electroporation

A 3 ml culture of the strain to be made electrocompetent was inoculated from a single colony into L broth and incubated overnight with shaking. This culture was then diluted 1:100 in 100 ml of fresh L broth and grown to an OD_{600nm} of between 0.4 and 0.6. The cultures were then incubated on ice for 20 min before the bacteria were harvested by centrifugation at 6000 x *g* at 4°C for 8 min. The pellet was resuspended in 100 ml of sterile ice-cold water and the bacteria were then harvested as before. The cell pellet was resuspended in 50 ml ice-cold sterile water and harvested again. The pellet was then resuspended in 4 ml sterile ice-cold 10% (v/v) glycerol, harvested and resuspended again in ice-cold 10% (v/v) glycerol to a final volume of 400 µl. 40 µl aliquots were stored at –80°C.

Electro-competent bacteria (40 µl) were mixed with between 10 ng and 2 µg of DNA (in a volume not exceeding 4 µl) and immediately transferred into a cold electroporation cuvette (BioSmith, 0.2 cm electrode width). The electroporation was carried out in a Bio-Rad gene pulser at 12.5 $kVcm^{-1}$, 25 µF and 200 Ω. The bacteria were then mixed with 1 ml of pre-warmed SOC broth and incubated at 37°C for 1 h to allow expression of plasmid-borne antibiotic resistance markers. Aliquots of 50 µl and 1 ml quantities of the transformation mixture were spread on selective L agar plates and transformants were single colony purified onto fresh L agar plates following growth overnight at 37°C.

2.2.2 λ Red-mediated allele replacement

The inactivation of chromosomal genes was achieved using an allele-replacement system based on the gene products of the λ phage red operon (75). This system is based on the ability of the *gam*, *bet* and *exo* gene products to inhibit the Exonuclease V activity of RecBCD, allowing the transformation of linear DNA fragments, and promoting homologous recombination at regions of homology between the chromosome and the transformed linear DNA fragment. The λ-phage *redγβα* operon was located on the pKOBEG plasmid (56) under the control of the L-arabinose-inducible P_{araBAD} promoter. This plasmid has a temperature sensitive pSC101 replicon, which permits plasmid replication at 30°C but not at higher temperatures. The plasmid also confers resistance to the antibiotic chloramphenicol. Use of the λ Red allele replacement system enabled the transfer of a mutated *hns* allele, amplified from the chromosome of *E. coli* strain PD32, to the genome of ETEC strain LMC10 (see Chapter 6).

2.2.2.1 λ Red mediated allele replacement in ETEC strain LMC10

The *E.coli* strain PD32 was used as the source of an *hns* gene interrupted by the insertion of an ampicillin resistance cassette (80). A linear DNA fragment containing the interrupted gene, plus ~400 bp of flanking DNA, was produced by PCR amplification of the desired region from a chromosomal preparation of PD32. The linear fragment was isolated and purified from an agarose gel (Section 2.2.4.2) and concentrated by Pellet Paint®-precipitation (Section 2.2.4.5).

ETEC strain LMC10 bacteria harbouring the pKOBEG plasmid were grown to mid-logarithmic phase at 30°C in the presence of 30 μg/ml chloramphenicol. Expression of the Red α, β and γ proteins was induced by addition of 0.2% L-arabinose. The induced bacteria were grown for a further 2 h and made electrocompetent (Section 2.2.1.2). These electrocompetent bacteria were then transformed with approximately 1 μg of the purified linear DNA containing the disrupted gene. Following the 37°C expression step of the transformation procedure, the bacteria were spread on agar plates containing ampicillin (to select for the interrupted gene) but not chloramphenicol. All subsequent steps were carried out at 37°C in the absence of chloramphenicol to ensure rapid loss of the pKOBEG plasmid. Putative mutants were single-colony purified twice and then

screened for the presence of the novel mutant allele by PCR and appearance on MacConkey salicin agar. Loss of the pKOBEG plasmid was confirmed by testing for chloramphenicol sensitivity on agar containing the antibiotic.

2.2.3 Purification of plasmid and chromosomal DNA

2.2.3.1 Small-scale purification of plasmid DNA

The HiYield Plasmid Mini Kit (Real Biotech Corporation) was used for the routine purification of plasmid DNA from overnight bacterial cultures. Purification was carried out according to the manufacturer's instructions using a modified alkaline lysis method. Briefly, depending on plasmid copy-number, bacteria from a 3-10 ml overnight culture were harvested and lysed in an alkaline solution containing SDS (to denature proteins) and RNase (to degrade RNA). The alkaline conditions also denatured both the chromosomal and plasmid DNA which were released upon cell lysis. The solution was then rapidly neutralised causing the plasmid DNA to re-anneal and the chromosomal DNA to precipitate. The lysate was then cleared by centrifugation to remove cellular debris and chromosomal DNA and the soluble plasmid DNA was purified using a mini cation-exchange column. The purified DNA was eluted from the column in a Tris/EDTA buffer following a desalting and washing step.

2.2.3.2 Large-scale purification of plasmid DNA

The Fast Ion Plasmid Midi Kit (Real Biotech Corporation) was used to purify plasmid DNA from 50-100 ml overnight bacterial cultures according to the manufacturer's instructions using a modified alkaline lysis method similar to that described above. Bacteria were lysed as before and the cleared lysate was passed through a cation-exchange column, which binds the re-natured plasmid DNA. The column with bound DNA was washed repeatedly and the DNA was eluted in a high-salt buffer. The DNA was then further purified and desalted by precipitation with isopropanol and resuspended in Tris/EDTA buffer.

2.2.3.3 Purification of total genomic DNA

Chromosomal DNA was purified using the Puregene genomic DNA purification kit (Gentra Systems). Briefly, bacteria from a 0.5 ml overnight culture were harvested by centrifugation, resuspended in a cell lysis solution, containing Tris [hydroxymethyl] aminomethane, EDTA and SDS and incubated at 80°C for 5 min to lyse the bacteria. Contaminating RNA was removed by treatment with RNase and cellular proteins were removed by protein precipitation using ammonium acetate. The remaining DNA was then precipitated with isopropanol and resuspended in a Tris/EDTA buffer.

2.2.4 *In vitro* manipulation of DNA

2.2.4.1 Restriction endonuclease digestion of DNA

All restriction digests were carried out using enzymes supplied by NEB according to the manufacturer's instructions. Briefly, 0.1-2 μg of purified DNA was incubated with 10-20 units of restriction enzyme in the appropriate NEB buffer for 2 h at the appropriate temperature. Digests with multiple enzymes were carried out in the recommended double digest buffer or in an appropriate buffer in which all enzymes had 100% activity. Where no suitable buffer was available sequential digests were performed.

2.2.4.2 Purification of DNA fragments

Following PCR amplification or digestion with restriction endonucleases, linear DNA fragments were purified directly from solution using the High Pure PCR Product Purification Kit (Roche) or from an agarose gel slice using the Geneclean II Kit (Q Bio gene). To purify DNA from solution, 5 volumes of Binding Buffer was added and mixed thoroughly. Binding Buffer contains the chaotropic salt guanidine thiocyanate, which denatures proteins. The solution was then added to a spin column where the DNA fragments bind to a glass fibre matrix. The bound DNA was then purified in a series of ethanol based wash steps to remove salts, enzymes and unincorporated nucleotides. The purified DNA was eluted from the column in Tris Buffer. Purification of DNA from an agarose gel slice was achieved by melting the gel slice at 55°C in the

presence of 4.5 volumes of 6 M sodium iodide and 0.5 volumes of TBE modifier, a concentrated salt solution. Once the gel slice had melted and the DNA was in solution, 5-10 μl of Glassmilk, an aqueous suspension of silica particles, was added and the mixture was incubated for 5 min at 55°C. The DNA bound to the silica particles was harvested by centrifugation at 15,800 x *g* for 10 s and resuspended in New Wash, an ethanol based wash solution. The wash step was performed twice more after which the pellet was air-dried and the bound DNA then eluted with Millipore grade water.

2.2.4.3 Ligation of DNA fragments

The T4 DNA Ligase was used to catalyse the formation of a phosphodiester bond between the exposed 5′ phosphate and 3′ hydroxyl groups of linear DNA fragments in an ATP-dependent fashion (114). These ligations were carried out using the Quick Ligation kit (NEB) and were routinely used for the cloning of DNA fragments into appropriate plasmid vectors. 50 ng of digested vector DNA was mixed with a three-fold molar excess of digested insert in a total volume of 10 μl and mixed with 10 μl of 2X Quick Ligation Reaction Buffer. 1 μl of Quick T4 DNA Ligase was added and the reaction was incubated at room temperature for 10 min. The ligated DNA molecules were then transformed into either calcium-chloride competent NEB 5-alpha cells or electrocompetent *E. coli* XL-1 Blue cells. Where calcium-chloride competent cells were used, 10 μl of the ligation reaction was directly transformed. For electrocompetent cells, the ligated DNA was first purified as previously described (Section 2.2.4.2) and then concentrated by pellet-paint precipitation (Section 2.2.4.5) prior to electroporation.

2.2.4.4 Agarose gel electrophoresis

DNA samples were separated on a 0.8-2% agarose gel, depending on the size of the DNA. Briefly, for a 1% gel, 1 g of agarose was added to 100 ml of 0.5X TBE buffer (44.5 mM Tris borate, pH 8.3, 1 mM EDTA) and microwaved to dissolve the agarose. Ethidium bromide was added to a final concentration of 1 μg/ml and the molten gel was poured into a gel mould and allowed to set. DNA samples were prepared by adding an appropriate volume of 5X sample loading buffer (25 mM Tris pH 7.6, 30% (v/v) glycerol, 0.125% (w/v) bromophenol blue) and these samples were subjected to

electrophoresis through the gel at 135 V for 45 min in 0.5X TBE buffer. The separated DNA fragments were photographed while illuminated under UV light.

2.2.4.5 Purification of DNA samples from solution by ethanol precipitation

Ethanol precipitation of DNA was routinely used for the preparation of desiccated DNA samples for DNA sequencing. Ethanol is added to the aqueous DNA sample to deplete the hydration shell surrounding the DNA molecules exposing the negatively charged phosphate groups. A cation is then added to mask this charge and allow a precipitate to form. Sodium acetate was routinely used as the source of the cation (Na^+). Briefly, the DNA sample was mixed with 2.5 volumes of ethanol followed by 0.1 volumes of 3 M sodium acetate (pH 5.2) and incubated at –80°C for 1 h to allow a precipitate to form. The precipitated DNA was then recovered by centrifugation at 21,000 x *g* for 15 min, washed with 500 µl of 70% ethanol to remove contaminating salt and centrifuged again as before. The precipitated DNA was then dried at 55°C for 10 min. If an aqueous DNA solution was required, the precipitated DNA was dissolved in Millipore grade water at 55°C for 10 min.

For concentrating dilute DNA samples, Pellet Paint® Co-Precipitant was routinely used. Pellet Paint® Co-Precipitant is a visible dye-labeled carrier formulated specifically for use in alcohol precipitation of nucleic acids. 2 µl of Pellet Paint® Co-Precipitant was added to the DNA solution followed by 0.1 volume of 3 M Sodium Acetate and 2 volumes of ethanol. The samples were vortexed briefly and incubated at room temperature for 2 min. Precipitated DNA was collected by centrifugation at 21,000 x *g* for 5 min. DNA was subsequently washed with 70% ethanol and air-dried. The dried DNA pellet was resuspended in an appropriate volume of Millipore grade water.

2.2.5 Polymerase chain reaction

The polymerase chain reaction (PCR) was used for the amplification of specific DNA fragments for use in cloning reactions and for the confirmation of constructed plasmids and mutant alleles. PCR is based on the ability of certain thermostable DNA polymerases to synthesise a new DNA strand complementary to a provided single-stranded, denatured DNA template when primed with specific complementary oligonucleotides (342). The procedure involves successive rounds of thermal

denaturation of a double-stranded DNA template, hybridization of two complementary oligonucleotides (primers) and synthesis of the new DNA strand by the DNA polymerase. The primers are designed to be complementary to opposite strands at either end of the fragment to be amplified and orientated such that their 3′ ends face each other. The new DNA strand is synthesised from provided dNTPs by the DNA polymerase in the presence of Mg^{2+}. Each new strand acts as a template in further rounds of amplification and the procedure thus results in exponential amplification of the desired DNA fragment.

2.2.5.1 Amplification of DNA by polymerase chain reaction

Two different DNA polymerases were used for the routine amplification of DNA fragments. *Taq* DNA polymerase (NEB), a recombinant purified thermostable polymerase from *Thermus aquaticus* YT-1, which is highly efficient but lacks the 3′-5′ exonuclease activity necessary for error correction (proof-reading), was used in PCR reactions where the sequence accuracy of the amplified DNA was not critical; for example, screening recombinant plasmids for inserts or screening for mutant alleles in purified chromosomal DNA. Where a high degree of sequence fidelity was required, for example to generate DNA fragments for cloning reactions, Phusion™ High Fidelity DNA Polymerase (Finnzymes) was used. Phusion™ is also highly processive but retains the 3′-5′ exonuclease activity necessary for proof-reading resulting in a significant decrease in the rate of nucleotide misincorporation.

Phusion™ PCR reactions were carried out according to the manufacturer's instructions in an MJ Research PTC-200 peltier thermal cycler. Briefly, 10 µl of 5X Phusion HF buffer (containing 1.5 mM $MgCl_2$) was mixed with 0.2 mM of each dNTP, 0.5 µM of each primer, 1 pg-10 ng of template DNA, 1 unit of Phusion™ and ddH_2O to a final volume of 50 µl. The PCR reactions were transferred to the thermal cycler and incubated as follows:

1	Denaturation: 98°C for 30 s
2	Denaturation: 98°C for 10 s
3	Oligonucleotide annealing: a temperature corresponding to the lowest melting temperature of the primer pair for 30 s
4	Extension: 72°C for 15 s per kilobase of expected DNA product

5 Repeat steps 2-4 for an additional 29 cycles

6 Final Extension: 72°C for 10 min

Taq PCR reactions were carried out as above with some exceptions. *Taq* buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3) was substituted for Phusion HF buffer in the reaction mixture and 1 unit of *Taq* polymerase was added to each reaction. In the thermal cycler, denaturation was carried out at 94°C for 2 min initially and for 30 s during the cycling steps and extension was conducted at 72°C for 1 min per kilobase of expected DNA product.

Purified plasmid DNA, genomic DNA or cell lysates were used as templates in PCR reactions. Cell lysates were prepared by transferring a single bacterial colony to a PCR tube containing 20 µl Millipore grade water using a sterile pipette tip and incubating at 100°C for 5 min. 1 µl of this crude lysate was used in each PCR reaction.

2.2.6 Mutagenesis

2.2.6.1 Pentapeptide scanning mutagenesis

The GPS-LS linker-scanning system (NEB) was used to introduce 15 bp insertions at random positions throughout the *rns* open reading frame (ORF). This system involves the random insertion of a transposon derivative (transprimer) into the target gene followed by removal of most of the transprimer by restriction digest. Subsequent religation of the target backbone results in a 15 bp insertion remaining in the gene of interest (30). In most cases the mutated genes now encode proteins with pentapeptide insertions. The activity of these mutant proteins can then be assessed to determine the effect, if any, of the pentapeptide insertion.

The mutagenesis procedure was carried out according to the manufacturer's instructions. The target *rns* encoding plasmid (pSS2192 or pHisRns) was mixed with the transprimer donor plasmid (pGPS5 or pGPS4), TnsABC* transposase, GPS buffer and start solution (magnesium acetate). After incubation at 37°C for 1 h followed by inactivation by heating at 75°C for 10 min, the mutagenised plasmids were transformed into *E. coli* XL-1. Colonies containing target plasmids with transprimer insertions were isolated by selecting for both the resistance marker of the target plasmid and that of the

transprimer. All of the several thousand resultant colonies were pooled and grown with relevant antibiotic selection overnight. Plasmid DNA was prepared from this overnight culture and subsequently digested with *Pme*I to remove the bulk of the transprimer. The transprimer-free backbone was religated and transformed into *E. coli* XL-1 to amplify the library of plasmids containing 15 bp insertions. Plasmid DNA was prepared from a pool of the resultant colonies and then transformed into *E. coli* XL-1 containing a reporter plasmid (pCooGFP or pCooGFP-2). Thus bacteria harbouring mutant *rns* genes could be easily discerned by observing their fluorescence under UV light. Colonies with altered fluorescence (relative to that seen with wild type Rns) were selected and the presence of a 15 bp insertion in their *rns* gene was verified by a PCR/restriction endonuclease screen. The precise sequence and location of the insertion was determined by DNA sequencing.

2.2.6.2 Random mutagenesis

To generate random mutations the *rns* gene was amplified from pHisRns by PCR using *Taq* DNA polymerase. The mutation frequency associated with using the low-fidelity polymerase *Taq* was increased by reducing the final concentration of dCTP present in the PCR reaction from 200 μM to 20 μM and adding 200 μM of dITP in place of the depleted dCTP (220). The products of several such PCR reactions were purified and cloned into pBSKII+ to produce a pool of plasmids containing random mutations within the *rns* gene. A 520 bp fragment corresponding to the N-terminal domain of Rns was removed by restriction enzyme digest from the plasmids and subcloned into the pHisRns background to produce a second pool of plasmids containing random mutations only within the first half of the *rns* gene. This pool of plasmids was transformed into *E. coli* XL-1/pCooGFP-2 cells to identify mutants unable to activate the *coo* promoter in pCooGFP-2 thus resulting in non-fluorescent cells.

2.2.6.3 Site-directed mutagenesis

Site-directed mutagenesis was performed using the QuikChange kit (Stratagene) according to the manufacturer's instructions. In this system a target plasmid containing the DNA region of interest is combined in a PCR reaction with two oligonucleotide

primers that are complementary to opposite strands of the target and contain the desired mutation. During the PCR reaction the primers are extended by the *Pfu* high fidelity DNA polymerase. Incorporation of the primers produces a mutated plasmid containing staggered nicks. After the PCR reaction, the resultant product is treated with *Dpn*I. As *Dpn*I is specific for methylated and hemimethylated DNA, it digests only the parental DNA and not the newly synthesised mutation-containing DNA. This mutated nicked vector DNA is then repaired *in vivo* when transformed into *E. coli* XL-1 cells.

2.2.6.4 Mutagenesis by inverse PCR

Inverse PCR was used to eliminate and to replace portions of the *rns* gene. In this technique a plasmid encoding the gene of interest was combined in a PCR reaction with specially designed primers. These primers flank the area that is to be deleted or replaced and are divergent, i.e. their 3′ ends face away from each other instead of towards each other. When a new portion of DNA was to be introduced, the nucleotides encoding the novel fragment were incorporated into the primers. When the PCR reaction was complete, it was treated with *Dpn*I to digest parental plasmid DNA. The PCR product was then gel purified and ligated as described in Section 2.2.4.3 before transformation into NEB 5-alpha cells.

2.2.7 Preparation and analysis of RNA

During isolation and handling of RNA, several precautions were taken to minimise the risk of contamination with exogenous RNases. These precautions included wearing gloves, working in a designated RNA handling area that had been cleaned with RNaseZAP® (Sigma) and working with designated tips, tubes, pipettes and electrophoresis tanks. Additionally, where possible, all steps in the handling of RNA were performed expediently and on ice.

2.2.7.1 Isolation of RNA

Total RNA was isolated from bacterial cultures using the RiboPure™-Bacteria kit (Ambion). Briefly, 6 ml of a mid-logarithmic phase bacterial culture was harvested by centrifugation. Where several RNA isolations were to be performed simultaneously, the

harvested cell pellets were resuspended in 1.5 ml RNA*later*™ (Ambion) and stored at 4°C until all samples were ready. Cell pellets, collected by centrifugation from a mid-logarithmic phase culture or from an RNA*later*™ suspension, were resuspended in RNA*wiz*, a combination of phenolic, detergent and chaotropic denaturants that inhibit RNases and stabilise RNA. The bacterial cell walls were disrupted by vortexing this suspension in the presence of Zirconia beads. The resultant bacterial lysate was mixed with chloroform and centrifuged to retrieve the RNA-containing upper aqueous phase. This crude RNA preparation was diluted with ethanol and bound to a silica filter. After several wash steps to remove contaminants, the RNA was eluted in a low ionic strength solution. Contaminating genomic DNA was removed by DNaseI treatment. RNA concentration (A_{260nm}) and purity (A_{260nm}:A_{280nnm} ratio) was assessed using a NanoDrop® ND-8000 spectrophotometer (NanoDrop Technologies). RNA samples were stored at –80°C.

2.2.7.2 Reverse transcriptase PCR

Reverse transcriptase PCR (RT-PCR) was performed using the Qiagen OneStep RT-PCR kit. Briefly, 20 ng of template total RNA was mixed with 5 µl of 5X Qiagen OneStep RT-PCR buffer, 0.4 mM of each dNTP, 0.6 µM of each primer, 1 µl of Qiagen OneStep RT-PCR enzyme mix (contains Omniscript™ reverse transcriptase, Sensiscript™ reverse transcriptase and HotStarTaq® DNA polymerase) and RNase-free water to a final volume of 25 µl. The reactions were transferred to an MJ Research PTC-200 peltier thermal cycler and incubated as follows:

1. Reverse transcription: 50°C for 30 min
2. Activation of HotStarTaq® DNA polymerase: 95°C for 15 min
3. Denaturation: 94°C for 30 s
4. Oligonucleotide annealing: a temperature corresponding to 5°C below the melting point of the primer pair for 30 s
5. Extension: 72°C for 1 min
6. Repeat steps 3-5 for an additional 34 cycles
7. Final extension: 72°C for 10 min

The PCR products were visualised on a 2% agarose gel (Section 2.2.4.4).

2.3 Analysis and manipulation of proteins

2.3.1 SDS-PAGE

Proteins were separated on discontinuous denaturing polyacrylamide gels by the method of Laemmli (222). Using this method, proteins are denatured in SDS and β-mercaptoethanol and separated on the basis of their size as they travel through a polyacrylamide gel towards the anode. SDS binds to most proteins in a constant weight ratio, masking their natural charge with its own negative charge and giving each protein a similar mass:charge ratio allowing separation on the basis of size rather than charge. The discontinuous gel is formed using buffers of differing composition and pH to firstly focus the separating proteins into narrow well-defined bands and then separate these focused proteins on the basis of their size.

2.3.1.1 Preparation of total cellular extract for SDS-PAGE analysis

The OD_{600nm} of an overnight culture was measured and 1 ml of the culture was then harvested by centrifugation at 15,800 x *g* for 1 min. The pelleted bacteria were resuspended in an appropriate volume of Laemmli buffer (75 mM Tris-HCl, pH 6.8, 20% (v/v) glycerol, 2% (w/v) SDS, 2% (v/v) β-mercaptoethanol, 10 μg/ml bromophenol blue) such that the final concentration was 10 OD_{600nm} units/ml. Samples were stored at –20°C. Prior to electrophoresis the samples were heated at 100°C for 5 min to denature all proteins.

2.3.1.2 Preparation of heat shock samples

Heat released extracts of ETEC strain LMC10 were prepared using the method of Smyth (374) with slight modifications. Bacteria, from solid or liquid culture, were resuspended in 0.15 M NaCl at a concentration of 50 OD_{600nm} units/ml. The suspension was incubated at 65°C for 15 min after which whole cells were removed by centrifugation at 10,000 x *g* for 10 min. The supernatant fraction was collected and stored at –20°C. For SDS-PAGE analysis, 1 volume of heat shock sample was combined with 1 volume Laemmli buffer, the mixture was heated at 100°C for 5 min and 20 μl was subjected to electrophoresis.

2.3.1.3 Electrophoresis of protein samples

Discontinuous gels were prepared using standard protocols (12), and 8, 10, 12 and 15% polyacrylamide gels were routinely used. Typically, 10 µl of each boiled Laemmli sample was loaded per well and broad range (NEB) or low range (Favorgen Biotech Corp) protein molecular weight markers were included on all gels. Electrophoresis was carried out in a Bio-Rad Mini-Protean III gel tank and was usually performed at 200 V for 55 min. Gels were stained using Coomassie brilliant blue R-250 (0.25% Coomassie brilliant blue R-250, 45% (v/v) methanol, 10% (v/v) acetic acid) and destained using Coomassie destain solution (45% (v/v) methanol, 10% (v/v) acetic acid) or transferred to nitrocellulose or PVDF membranes for use in Western immunoblots (see Section 2.3.3).

2.3.2 Purification of proteins and production of antiserum

The pMRns5 plasmid (373) had previously been constructed using the pMAL protein fusion and purification system (NEB). It contains the *rns* gene fused in-frame to the 3′ end of the *malE* gene, which encodes maltose-binding protein (MBP). The fusion gene is transcribed under the control of an IPTG-inducible P_{tac} promoter. The QIA*express* system (Qiagen) was used to construct a plasmid (pHisRns) in which the *rns* gene was tagged at the 5′ end with six histidine codons. This plasmid was used to produce Rns protein tagged with six histidine residues at the amino terminus. Mutant forms of *rns* were subcloned into pMRns5 and pHisRns from pSS2192 derivatives (Table 2.2). Alternatively, the plasmids themselves were subjected to mutagenesis (see Section 2.2.6) to generate plasmids that expressed mutant derivatives of Rns protein.

The pHNSHIS plasmid was constructed by cloning the *hns* gene plus nucleotides encoding six histidine residues at its 3′ end into the expression vector pET3b. This plasmid was used for overexpression of C-terminal His-tagged HNS using the T7 RNA polymerase system (395).

Each of the above described plasmids were used to produce large amounts of His-tagged or MBP fusion protein allowing generation of anti-Rns antiserum and purification of the tagged or fusion protein.

2.3.2.1 Large-scale purification of His-tagged Rns (or mutant derivatives) and His-tagged H-NS

His•Bind® Quick Cartridges (Novagen) were used for rapid purification of His-tagged protein by immobilised metal affinity chromatography. The His-Tag sequence (six consecutive histidine residues) binds to divalent cations (Ni^{2+}) immobilised on His•Bind® resins. After unbound proteins are washed away, the target protein is recovered by elution with imidazole. This system enables the purification of proteins under non-denaturing conditions.

A 10 ml overnight culture of *E. coli* KS1000 containing the plasmids pRare and pHisRns was grown in L-broth and then used to inoculate 500 ml fresh prewarmed L broth. The culture was grown to an OD_{600nm} of 0.6 at which point expression of His-tagged Rns was induced by addition of IPTG to a final concentration of 1 mM. The culture was incubated for a further 5 h and then harvested by centrifugation at 6,000 x *g* for 10 min. The cell pellet was resuspended in 6 ml binding buffer (0.5 M NaCl, 5 mM imidazole, 20 mM Tris-HCl, pH 7.9). Lysis of the bacterial cells was achieved by freezing the cell suspension at –20°C overnight followed by thawing and sonication. After sonication the insoluble material was removed by centrifugation at 21,000 x *g* at 4°C for 30 min. The soluble supernatant was made up to a volume of 50 ml with binding buffer. A His•Bind® Quick 900 cartridge (Novagen) was equilibrated with 6 ml binding buffer. Using a 20 ml syringe, the soluble cell extract was passed through the cartridge at a rate of 2 drops/s. Unbound protein and cellular debris were removed by washing with 20 ml binding buffer. The column was then washed with 10 ml wash buffer (binding buffer with the addition of 60 mM imidazole). Bound, His-tagged Rns was subsequently eluted in 2.5 ml elution buffer (binding buffer with the addition of 1 M imidazole).

Expression of His-tagged H-NS protein was conducted as described above with the following alterations. Several freshly transformed colonies of *E. coli* T7 Express I^q harbouring pHNSHIS were transferred to 3 ml Luria broth and grown for 30 min. This starter culture was then used to inoculate 500 ml fresh prewarmed Luria broth. The culture was grown at 30°C until an OD_{600nm} of 0.5 was reached. Expression of His-tagged H-NS was then induced by addition of IPTG to a final concentration of 0.5 mM followed by 2 further hours of growth at 30°C. The bacteria were then harvested and cell lysis and protein purification was carried out as described above.

Large-scale purified His-tagged proteins were transferred into storage buffer (50 mM sodium phosphate, 0.3 M NaCl, 50% (v/v) glycerol, pH 8 or 6) using PD10 buffer exchange columns (Amersham Biosciences) and stored in single use aliquots at –80°C.

2.3.2.2 Small-scale purification of His-tagged Rns (or mutant derivatives) and His-tagged H-NS

Small-scale rapid purification of His-tagged proteins was performed using HIS-Select™ Spin Columns (Sigma). The spin columns contain silica particles derivatized with a chelate charged with nickel. The columns are selective for recombinant proteins with histidine tags.

Expression of His-tagged Rns (or mutant derivatives) for small-scale purification was carried out two ways. Initially, a 5 ml culture of *E. coli* KS1000/pRare containing the plasmid encoding the protein of interest was grown overnight in L broth and then used to inoculate 100 ml fresh L broth. This sub-culture was grown to an OD_{600nm} of 0.6 and was then induced with 1 mM final concentration IPTG and grown for a further 5 h before harvesting. During later purifications the expression step was conducted using OnEx™ medium (see Section 2.1.1.2). A 10 ml overnight culture of *E. coli* KS1000/pRare containing the plasmid encoding the protein of interest was grown in OnEx™ medium and harvested the next morning. Expression of His-tagged H-NS for small-scale purification was achieved as described in Section 2.3.2.1, except it was carried out in 100 ml Luria broth. In all cases, once the cells were harvested they were resuspended in 3 ml equilibration buffer (50 mM sodium phosphate, 0.3 M NaCl, 10 mM imidazole, pH 8.0) and lysed by a freeze-thaw cycle followed by sonication. The cell lysate was then centrifuged at 21,000 x *g* at 4°C for 30 min to enable collection of the soluble fraction. The HIS-Select spin column was equilibrated with equilibration buffer. The soluble cell extract was then loaded onto the column. Unbound protein was washed from the column with two passages of wash buffer (50 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole, pH 8.0). Finally protein was eluted from the column in 150 µl elution buffer (50 mM sodium phosphate, 0.3 M NaCl, 250 mM imidazole, pH 8.0). Where required, small-scale purified His-tagged proteins were subjected to ultrafiltration using Vivaspin 500 columns (MW cut-off 10,000 and 100,000) as described by the manufacturers (Sartorius). Prior to cross-linking experiments (Section 2.3.5.2), protein samples were dialysed into storage buffer (50 mM sodium phosphate,

0.3 M NaCl, 50% (v/v) glycerol, pH 8) using Slide-A-Lyzer® mini dialysis units, 3,500 MWCO (Pierce). Small-scale purified His-tagged proteins were stored at 4°C.

2.3.2.3 Large-scale purification of MBP-Rns (or mutant derivatives)

Purification of MBP-fusion proteins involves preparing a cell extract in which the protein of interest has been over-expressed. This crude cell extract is then passed over a column containing an agarose resin derivatised with amylose, a polysaccharide composed of maltose subunits. Due to the affinity of MBP for amylose the fusion protein will bind to the column. Unbound proteins are removed from the column by washing after which the fusion protein is eluted from the column with free maltose. MBP has a higher affinity for maltose than amylose so it preferentially binds to the maltose as it passes through the column and in this way it is removed from the immobilised amylose.

A 20 ml overnight culture of *E. coli* KS1000/pRare harbouring the plasmid encoding the MBP fusion protein of interest was grown in OnEx™ medium. Bacteria were harvested and resuspended in 1 ml sonication buffer (50 mM Tris-HCl (pH 7.5), 10% (v/v) sucrose, 100 mM NaCl, 1 mM EDTA) and then frozen at –80°C for 2-3 h. The cell suspensions were thawed on ice and then lysed by sonication. The insoluble material was removed by centrifugation at 21,000 x *g* at 4°C for 30 min. The soluble supernatant was collected and made up to a volume of 50 ml in column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4). An MBP affinity column was prepared by pouring 1 ml of amylose resin into a 2.5 x 10 cm column. The column bed was washed with 10 ml column buffer. The soluble lysate was loaded onto the column at a rate of approximately 1 ml/min using a peristaltic pump. Unbound proteins were removed by washing the column with 20 ml column buffer. The bound MBP-fusion protein was eluted from the column in 5 ml column buffer supplemented with 10 mM maltose and collected in approximately 0.5 ml fractions. The elution fractions were analysed by SDS-PAGE to identify those that contained protein. Protein containing fractions were divided into single use aliquots and stored at –80°C.

2.3.2.4 Quantitation of protein

The concentration of proteins in solution was estimated by A_{280nm} on a NanoDrop® ND-8000 spectrophotometer (NanoDrop Technologies) or using the Bradford reagent (39). In the Bradford assay, proteins form a complex with Brilliant Blue G dye causing a shift in the maximum absorbance from 465 to 595 nm. The amount of absorbance at 595 nm is proportional to the amount of protein present. A protein concentration dilution series was made using bovine serum albumin (BSA) diluted in PBS. 5 µl of each dilution in the series, of PBS alone and of the protein sample of interest was added to 95 µl of Bradford reagent, in duplicate, in a 96 well plate. The absorbance at 595 nm was read and recorded. A standard protein concentration curve was constructed using the BSA dilution series. The concentration of the protein of interest was then extrapolated from the standard curve.

2.3.2.5 Rns antiserum

An anti-Rns antiserum was generated by immunising a specific pathogen free rabbit with purified His-tagged Rns. The initial immunisation consisted of 250 µg protein in Freund's complete adjuvant. This was followed by two further boosts of 125 µg protein in Freund's incomplete adjuvant administered 21 and 42 days after the first immunisation. A high antibody titre was detected in a blood sample taken from the rabbit on day 59. A final boost of 125 µg protein in Freund's incomplete adjuvant was administered on day 70 and the rabbit was exsanguinated on day 76.

The resultant antiserum was absorbed against a lysate of *E. coli* KS1000/pRare/pTrc99a to remove cross-reacting antibodies. The lysate was prepared from a 500 ml culture that was harvested by centrifugation, resuspended in 5 ml sonication buffer and subsequently lysed by sonication. The lysate was divided in two, and one half was heated at 100°C for 5 min. The heated and unheated lysates were combined, mixed with 7 ml antiserum and incubated overnight at 4°C with continuous mixing. Insoluble materials, including protein-antibody complexes, were removed by centrifugation at 21,000 x *g* for 30 min. The supernatant was collected and the remaining anti-Rns antibodies present in this supernatant were precipitated by adding an equal volume of saturated ammonium sulfate, pH 7.0. The precipitate was collected by centrifugation and resuspended in 0.3 volumes of PBS. The absorbed antiserum was

then dialysed against 6 L PBS at 4°C overnight. After dialysis the antiserum was centrifuged at 4,000 x *g* for 30 min and the supernatant was stored in aliquots at –20°C.

2.3.3 Western immunoblotting

2.3.3.1 Electro-transfer of separated proteins

Proteins were separated on SDS-polyacrylamide gels and transferred to nitrocellulose membranes using a Biometra Fastblot semi-dry transfer apparatus, or to PVDF membranes using a Bio-Rad Mini Trans-Blot wet transfer apparatus. For transfer to nitrocellulose membranes, the polyacrylamide gel and the membrane were equilibrated in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol) and blotting was carried out according to the manufacturer's instructions at 5 mA/cm^2 for 20 min. Where higher sensitivity of detection was required, Immobilon-P PVDF membrane (Millipore) was used. Prior to the wet transfer, the PVDF membrane was activated by soaking in 100% methanol. It was then rinsed in Millipore grade water before equilibration, along with the polyacrylamide gel, in transfer buffer (25 mM Tris, 192 mM glycine, 10% (v/v) methanol). The wet transfer was performed according to the manufacturer's instructions. Following transfer all membranes were stained with Ponceau S stain (0.5% (w/v) Ponceau S, 1% (v/v) acetic acid) to visualise the transferred proteins. Excess Ponceau S stain was removed by washes with Millipore grade water before subsequent manipulation of the membranes. Nitrocellulose membranes were blocked by incubation in blocking buffer (5% non-fat powdered milk in phosphate-buffered saline plus 0.05% Tween-20) at 4°C overnight or for 1 h at room temperature with shaking. Immobilon-P PVDF membranes were blocked by incubation in blocking buffer (3% non-fat powdered milk in phosphate-buffered saline plus 0.05% Tween-20) for 1 h at room temperature with shaking.

2.3.3.2 Detection of bound proteins

Blocked membranes were incubated with antiserum diluted in blocking buffer for 1 h at room temperature. Blocking buffer was composed of non-fat powdered milk, 5% (w/v) for nitrocellulose membranes or 3% (w/v) for PVDF membranes, in PBS

containing 0.05% (v/v) Tween-20. Typical antibody dilutions were 1:1,000 for anti-Rns antiserum, 1:400 for anti-CS1 antiserum and 1:15,000 for anti-MBP antiserum. The membrane was then washed three times for 15 min with PBS containing 0.05% Tween-20 and incubated with a secondary HRP-linked anti-rabbit antibody diluted 1:20,000 in blocking buffer for 1 h. The blot was then washed as before and bound antibody was detected using the SuperSignal West Pico chemiluminescent HRP substrate (Pierce). Chemiluminescence was detected using BioMax Light Film (Kodak) and films were developed manually using standard Kodak developer and fixer solutions according to the manufacturer's instructions. The length of exposure was varied to adapt for the signal strength. If a greater sensitivity of detection was necessary then the alternative chemiluminescent substrate SuperSignal® West Femto Maximum Sensitivity Substrate was used. In such cases it was necessary to reduce the antibody dilutions to 1:10,000 for anti-Rns antiserum and 1:50,000 for the secondary HRP-linked anti-rabbit antibody and to increase the number of washes after the antibody incubations from 3 to 6.

2.3.4 Electrophoretic mobility shift assay (EMSA)

The binding of purified MBP-Rns (or mutant derivatives) and His-tagged H-NS to the *coo* and *rns* promoter regions was analysed by the electrophoretic mobility shift assay (EMSA) using the LightShift® Chemiluminescent EMSA kit (Pierce). This assay is based on the fact that the migration of DNA-protein complexes in a native polyacrylamide gel is slower than that of non-bound DNA thus causing a "shift" in migration of the labelled DNA band. To generate DNA probes for EMSAs, a 195 bp or a 431 bp region of the *coo* promoter was amplified by PCR from pCooGFP-2 or from the mutant derivatives of pCS1 (see Table 2.2), and a 434 bp region of the *rns* promoter was amplified by PCR from pSS2192. The primers in these PCR reactions (VM1 & 2, CSBS1 & CSBS2 and VM3 & 4 respectively, see Table 2.3) had been biotinylated at their 5′ ends; therefore they produced biotin-end labelled PCR products. The PCR products were gel purified as described above (Section 2.2.4.2). Approximately 20 or 40 pg of biotinylated DNA probe was incubated with a range of protein concentrations at 37°C for 10 min in a reaction buffer. Unless stated otherwise, the reaction buffer consisted of 10 mM Tris (pH 7.5), 50 mM KCl, 1 mM dithiothreitol, 0.1 μg BSA/μl, 0.025% (v/v) Nonidet P-40, 6.5% (v/v) glycerol and, except where noted differently, 2

ng poly(dI.dC)/μl. For competition experiments, the first protein was incubated with the biotinylated DNA probe in reaction buffer at 37°C for 10 min after which time the binding reactions were challenged with the addition of the second protein and incubated at 37°C for a further 10 min. Binding reactions were separated on 5% non-denaturing polyacrylamide gels at 100 V for 65 min (for reactions using 195 bp DNA probes) or 90-120 min (for reactions using the 434 bp DNA probe) and then transferred to a nylon membrane at 80 V for 1 h. After the transfer, DNA was cross-linked to the membrane at 120 mJ/cm^2 using a UV-light cross-linker (Uvitec). The biotinylated DNA was detected with a streptavidin-horseradish peroxidase conjugate and a chemiluminescent substrate as directed by the LightShift® Chemiluminescent EMSA kit protocol (Pierce).

2.3.5 Protein-protein cross-linking

2.3.5.1 *In vivo* cross-linking

The method used was essentially that described by Adams *et al.* (2). Briefly, overnight cultures of *E. coli* containing a Rns expressing plasmid or an empty vector negative control were harvested by centrifugation. The cell pellets were washed with 0.9% (w/v) NaCl and resuspended in 125 mM HEPES (pH 7.3) at an OD_{600nm} of 1.0. 1 ml of cell suspension was treated with 1.0 or 0.1 mM dithiobis(succinimydylpropionate) (DSP) (Pierce) at room temperature for 15 min. The reaction was quenched with 50 μl of 1 M Tris-HCl (pH 8.0). The cells were then incubated on ice for 5 min before being collected by centrifugation and resuspended in Laemmli buffer (see Section 2.3.1.1) without β-mercaptoethanol. The samples were heated at 100°C for 10 min prior to separation on a 15% SDS polyacrylamide gel. Protein complexes were detected by Western immunoblotting.

2.3.5.2 *In vitro* cross-linking

Where *in vitro* cross-linking was carried out with purified His-tagged Rns, the zero-length chemical cross-linker 1-ethyl-3-(3′-dimethylaminopropyl) carbodiimide (EDC) and the catalyst *N*-hydroxy-succinimide (NHS) were added to 1 μg of protein. The reaction volume was made up to 20 μl with ddH_20. The final concentrations of EDC and NHS were 50 mM and 200 mM respectively (79, 431). The reaction mixture was

incubated at room temperature for 60 min. The reaction was stopped by the addition of an equal volume of Laemmli buffer. The samples were then incubated at 60-100°C for 5 min. Alternatively the reaction was stopped by the addition of Laemmli buffer and urea (final concentration 4 M) after which the samples were incubated at room temperature for a further 10 min. Portions of samples were then loaded onto an SDS polyacrylamide gel and protein complexes were detected by Western immunoblotting.

Where *in vitro* cross-linking was carried out with purified MBP-fusion proteins, the procedure was conducted as described above with some exceptions. 1.25 μg of protein was used and the reaction volume was made up to 20 μl with 7 mM MES buffer (pH 6). The reaction was stopped by treating the samples with urea using a modified form of the method described by Soulié *et al.* (381, 382). Briefly, a master mix was prepared in which 15 μl (per sample to be treated) of a mixture containing 6.7% (w/v) SDS and 4.6 M β-mercaptoethanol was added to 12.5 mg (per sample to be treated) of crystalline ultra-pure urea. The mixture was then vortexed for 2 min to solubilise the urea. 22.5 μl of this master mix was added to each cross-linking reaction and the samples were then heated at 100°C for 70 s. Following a brief centrifugation, 20 μl of each sample was loaded onto an 8% SDS polyacrylamide gel. The entire process from dissolving the urea until loading the gel took less than 10 min. Protein complexes were detected by Western immunoblotting.

2.3.6 Gel filtration chromatography

Gel filtration chromatography of purified MBP-Rns was carried out collaboratively by Dr. Robert Fagan at Imperial College London. MBP-Rns (0.5 ml at approximately 5mg/ml) was passed through a Superdex 200 10/30 column (GE Healthcare) and eluted with PBS at a flow rate of 0.4 ml/min. The protein content of each fraction was measured spectrophotometrically at a wavelength of 280 nm. The column was calibrated with six high and low molecular weight calibration standards (Sigma) ranging in size from 29 to 669 kDa to obtain a calibration graph for determination of the molecular weight of the eluted protein.

Eluted fractions were lyophilised and then analysed by this researcher to verify the presence of MBP-Rns by SDS-PAGE and Western immunoblotting.

2.4 Gene expression analysis

2.4.1 Assay of β-galactosidase activity

A DNA fragment containing the *rns* promoter was cloned upstream of the *lacZ* gene in the vector pRW-50 to create the transcriptional fusion reporter plasmid pRnsLacZ-2. Therefore, the *lacZ* gene in pRnsLacZ-2 is transcribed from the *rns* promoter but is translated from its own ribosome binding site. In bacterial cells the *lacZ* gene product, β-galactosidase, cleaves the β-galactoside linkage of lactose. The colourless synthetic substrate ONPG contains the β-galactoside linkage and therefore can also be cleaved by β-galactosidase. Cleavage of ONPG produces galactose and *o*-nitrophenol, which is yellow. Determining the levels of *o*-nitrophenol gives an indication of the levels of β-galactosidase present which in turn provides a measure of the level of activity of the promoter controlling the *lacZ* gene.

The β-galactosidase activity was measured using a standard Miller assay (259). Briefly, cultures were grown to mid-logarithmic phase in duplicate. 100 µl of each culture was added, in duplicate, to a 2 ml tube containing 900 µl of Z buffer (60 mM $Na_2HPO_4.7H_2O$, 40 mM $NaH_2PO_4.H_2O$, 10 mM KCl, 1 mM $MgSO_4.7H_2O$, 50 mM β-mercaptoethanol). After addition of 50 µl of chloroform and 50 µl of 0.1% (w/v) SDS the mixture was vortexed vigorously to permeabilise the bacterial cells and then placed at 28°C for 5 min. 200 µl of the chromogenic substrate ONPG (4 mg/ml in Z buffer) was added to start the reaction and the reaction was allowed to progress at 28°C for approximately 5 to 10 min. When sufficient yellow colour had developed the reaction was stopped by addition of 500 µl of 1 M Na_2CO_3 and the tubes were centrifuged at 21,000 x *g* for 10 min to pellet cellular debris. The optical density of the supernatant at 420 nm was then determined. The level of β-galactosidase activity was calculated using the following equation:

$$\text{β-galactosidase activity (Miller Units)} = \frac{1000 \times OD_{420nm}}{t \times v \times OD_{600nm}}$$

where t is the duration (in min) of the reaction and v is the volume (in ml) of culture used in the reaction.

2.4.2 Spectrofluorimetry

Areas of the *coo* promoter were cloned upstream of the promoterless *gfp*+ gene in the vector pZep08. This created transcriptional fusion reporter constructs in which the *gfp*+ gene is transcribed from the *coo* promoter fragment but retains its own ribosome binding site. The amount of green fluorescent protein (GFP, specifically the GFP+ allele) produced by cells harbouring such constructs was a measure of the activity of the *coo* promoter fragment that was present. GFP levels were measured as followed: *E. coli* cultures were grown in duplicate to mid-logarithmic phase in L broth and then placed on ice for 20 min. Alternatively, cultures were grown overnight on L agar. In the case of L agar grown cultures, bacterial growth was collected from the plates and emulsified in 1 ml of L broth. All cultures to be analysed were adjusted to a similar OD_{600nm}. Duplicate 200 µl aliquots of each culture were transferred to 1.5 ml tubes and the cells were pelleted by centrifugation. The cell pellets were washed by resuspension in 200 µl PBS and then pelleted again. The cells were lysed by resuspension in 500 µl 1X CelLytic B, a non-denaturing detergent (Sigma). Duplicate 100 µl aliquots of each cell lysate were dispensed to a black, flat-bottomed 96 well plate. Fluorescence levels were measured in a Wallac 1420 Victor^2 multilabel counter set at an excitation wavelength of 485 nm and an emission wavelength of 535 nm. The values obtained were standardised with respect to optical density and expressed per ml of culture. The final values were expressed as an average of all of the readings obtained for each strain.

Chapter 3 Mutational analysis of the Rns transcriptional regulator of ETEC

3.1 Introduction

Rns, like most other AraC family members, has two predicted HTH motifs in its C-terminal domain (267). The HTH motif is the most common DNA binding motif. It is approximately 20 residues long and is composed of two α-helices linked by a turn of about four residues. The second α-helix within the motif is called the recognition helix as it inserts into the major groove of the DNA where it engages in base-specific DNA contacts (210). Among AraC family members, there is a low level of sequence conservation at the first predicted HTH motif (HTH1). This may represent involvement of HTH1 in the recognition of the different target sites of each member (139). In contrast, the sequence of the second predicted HTH motif (HTH2) of the family was found to be much less variable, provoking the suggestion that this motif may be involved in a common role other than DNA binding, such as forming contacts with RNA polymerase (139). Indeed reports on the role of HTH2 in members of the AraC family have been somewhat varied. A study of the effects of mutations in the predicted second HTH of AraC itself was unable to demonstrate a specific HTH-like interaction occurring at this region (42). Analysis of the crystal structure of the AraC family protein Rob in complex with its *micF* promoter revealed that the protein only inserted its first HTH into the major groove of its binding site while the second HTH merely lay on the surface of the DNA helix making contacts with the DNA backbone (221). However, the crystal structure of the AraC family member MarA complexed with its DNA binding site revealed that the recognition helices of both HTH1 and 2 of the protein were inserted into adjacent major grooves of the DNA where they made sequence specific interactions (327). This supported the findings of a study that had concluded that the second HTH motif of AraC does contribute to DNA binding by the protein (280). More recently, mutational analysis of the AraC family members ToxT and XylS revealed that each of their HTHs are important for DNA binding, indeed the relative contribution of HTH2 of ToxT to DNA binding is even greater than that of its HTH1 (57, 87). Therefore some AraC proteins use both of their HTH motifs in DNA binding. While this is also thought to be the case for Rns it has not yet been confirmed. In fact, aside from its primary sequence, much is unknown about Rns such as which areas of the protein are required for DNA binding and for activation of the *coo* and *rns* promoters. The discovery of such information would benefit the study of Rns itself and that of other AraC family members.

Scanning linker mutagenesis (SLM) is a valuable method for studying the functional organisation of proteins (175). It has been used on a range of proteins from several different species to both correlate the known structure of certain proteins to their function, for example the β-lactamase of pBR322 (167) and the XerD recombinase of *Salmonella enterica* serovar Typhimurium (hereafter referred to as *S.* Typhimurium) (47), and to characterise proteins of largely unknown structure, for example the McrA endonuclease of *E. coli* (9). SLM involves the generation and analysis of proteins containing random peptide insertions. One form of SLM is based on the random insertion of a transprimer, a derivative of the bacterial transposon Tn7, into a cloned target gene (30). The reaction is performed *in vitro* using a mutant form of the transposase TnsABC (TnsABC*) that has lost its target site selectivity and thus promotes insertion at random locations (391). The transprimer contains rare 8 bp long cleavage sites for the restriction endonuclease *Pme*I 5 bp from each of its outer ends. During the transposition process, 5 bp of target site DNA are duplicated at the insertion point (Fig. 3.1). Therefore, after transposition, *Pme*I digestion followed by religation of the target DNA will remove the bulk of the transprimer and result in a 15 bp insertion remaining in the gene of interest. This 15 bp insertion comprises 5 bp from each end of the transprimer plus the 5 bp of duplicated target DNA (175). In most cases the 15 bp insertion produces a five amino acid, or pentapeptide, insertion in the protein encoded by the target gene. As the 15 bp insertion contains variable duplicated target site DNA, the composition of the pentapeptide insertion will not be constant (175). However, in two of the six possible frames of insertion the sequence of the *Pme*I site within the 15 bp insertion encodes a stop codon meaning that the resultant protein will be a truncate (30). The effects of the disruption caused by the insertions on protein activity will vary depending on the insertion location. Insertions that interrupt areas critical to protein function, for example areas of secondary structure, have deleterious effects on protein activity while insertions within surface exposed loops or unstructured regions are generally tolerated (174).

The various studies that used pentapeptide scanning mutagenesis to unravel protein structure-function relationships have found that the insertions introduced by the system can be small enough to enable the isolation of mutants with relatively subtle phenotypes (175). However, random mutagenesis is an even subtler tool used to gain structural and functional information about little understood proteins. Random mutagenesis of the gene of interest can be achieved by amplifying the gene using an error prone PCR

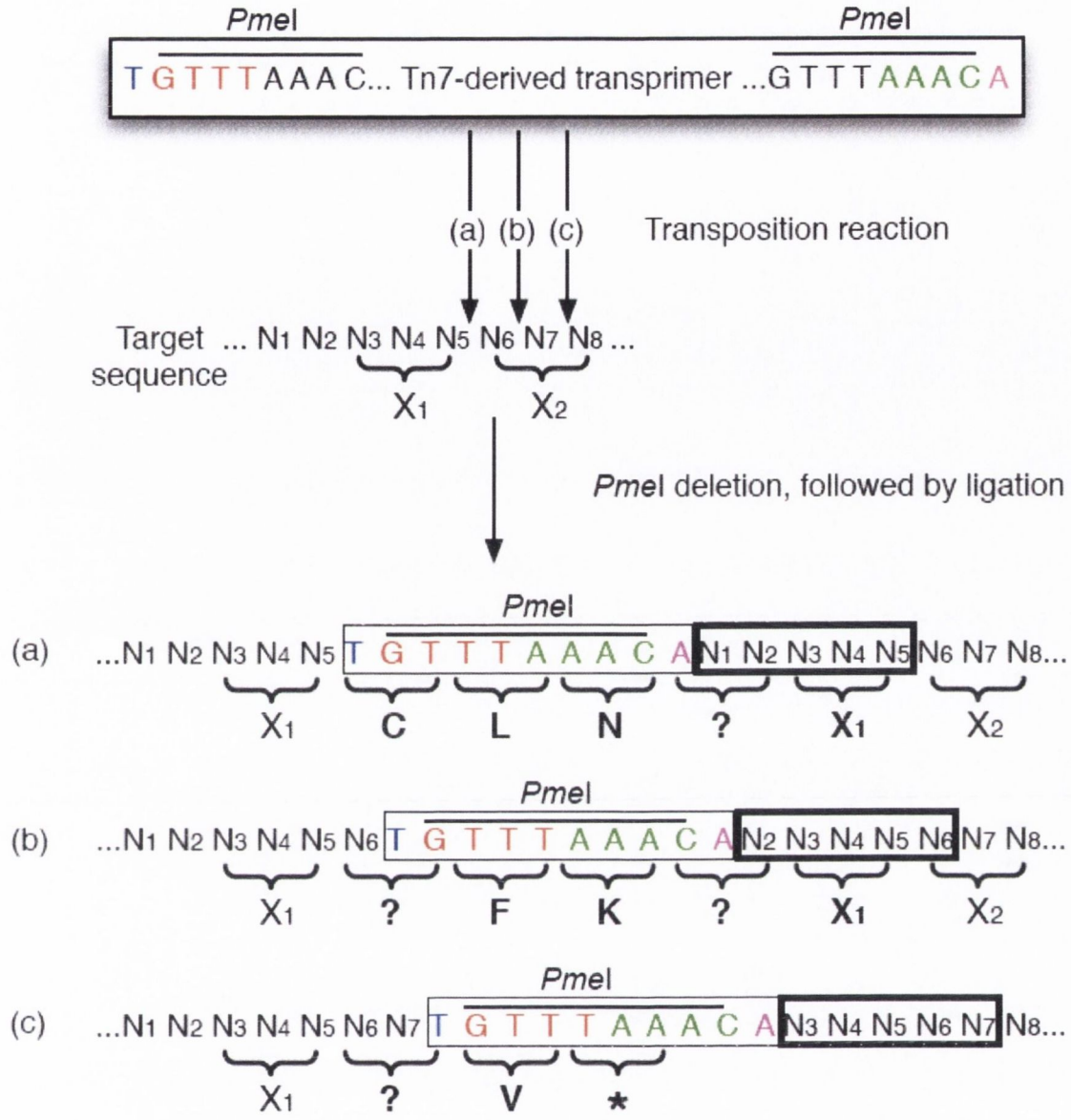

Fig. 3.1. Schematic representation of the pentapeptide scanning mutagenesis procedure. For simplicity, only the top strand of DNA of the transprimer and target are depicted. The Tn7-derived transprimer is shown at the top of the figure. Only the nine nucleotides present at the ends of the transprimer are shown. *Pme*I sites are indicated and over-lined. Coloured residues (blue, red, green and pink) denote transprimer-derived sequences that form part of the 15 bp insertion. The target sequence nucleotides are represented as N_1 N_2 N_3 etc. Insertion can occur in three possible reading frames depicted as (a), (b) and (c). The five nucleotides duplicated during insertion are boxed in a thick black line. The 15 bp insertions after N_5, N_6 and N_7 encode different sets of amino acids represented in bold. Some of the inserted amino acids are determined by the transprimer-derived sequence while others (denoted as "?") vary depending on the target sequence. The insertion in frame (c) introduces a stop codon (represented by an asterisk) and therefore results in a premature truncation of the protein. Adapted from Hallet *et al.* (167).

method. This method is based on both the intrinsic error frequency of *Taq* DNA polymerase, and the principle that when the concentration of one of the four nucleotides in a PCR reaction is reduced, incorporation of any of the other three nucleotides in its place is preferred (220). Both the low efficiency of such a PCR amplification, and the mutation frequency, can be increased by adding dITP to the reaction in place of the limiting nucleotide (384). During amplification, dITP will be incorporated instead of the depleted nucleotide and in the following PCR cycle one of the other three nucleotides (or dITP itself) will be incorporated as a complement to dITP, thus creating mutations, most commonly point mutations but occasionally insertions and deletions (220).

The aim of the study presented in this chapter was to use the techniques of scanning linker mutagenesis and random mutagenesis to gain insights into the structure-function relationship of the Rns protein. The activities of mutated proteins were tested to identify areas of Rns that can withstand mutation and remain functional, and areas that are critical for the functioning of Rns and thus cannot tolerate any disruption.

3.2 Results

3.2.1 Generation and screening of mutant forms of Rns

3.2.1.1 Generation of a library of 15 bp insertions in pSS2192

The first step in the characterisation of the Rns protein was to use the GPS-LS linker scanning system (NEB) to mutagenise the *rns* encoding plasmid pSS2192 (Table 2.2 and Fig. 3.2 (A)). This mutagenesis system is based on that described by Biery *et al.* (30). The pSS2192 plasmid was chosen as the mutagenesis target because like the wild type *rns* encoding plasmid, it has a low copy number (approximately four copies per cell). As described in Sections 2.2.6.1 and 3.1, pSS2192 was initially mutagenised by the random insertion of a kanamycin-resistant derivative of the bacterial transposon Tn7 (transprimer) from the donor plasmid pGPS5. The mutagenised plasmids were digested with the restriction endonuclease *Pme*I to remove the majority of the transprimer and then religated to produce a collection of pSS2192 derivatives containing 15 bp insertions at random positions. Insertions within the spectinomycin resistance marker and critical regions of the origin of replication of pSS2192 are selected against; therefore a library of plasmids harbouring 15 bp insertions predominantly throughout the *rns* ORF and the area upstream of the ORF had been created.

3.2.1.2 Isolation of mutants with insertions affecting the activity, or expression, of Rns

The transcriptional fusion reporter plasmid pCooGFP contains the *coo* promoter fused to a promoterless *gfp*+ gene, which encodes green fluorescent protein (GFP) (Table 2.2 and Fig. 3.2 (B)). Wild type Rns can activate expression of GFP from the *coo* promoter on this plasmid. The library of pSS2192 plasmids containing 15 bp insertions was transformed into *E. coli* XL-1/pCooGFP. This allowed the isolation of pSS2192 derivatives with insertions affecting the expression or activity of Rns as colonies containing these plasmids displayed increased or decreased GFP expression (and therefore fluorescence under UV light) relative to those containing unmutagenised

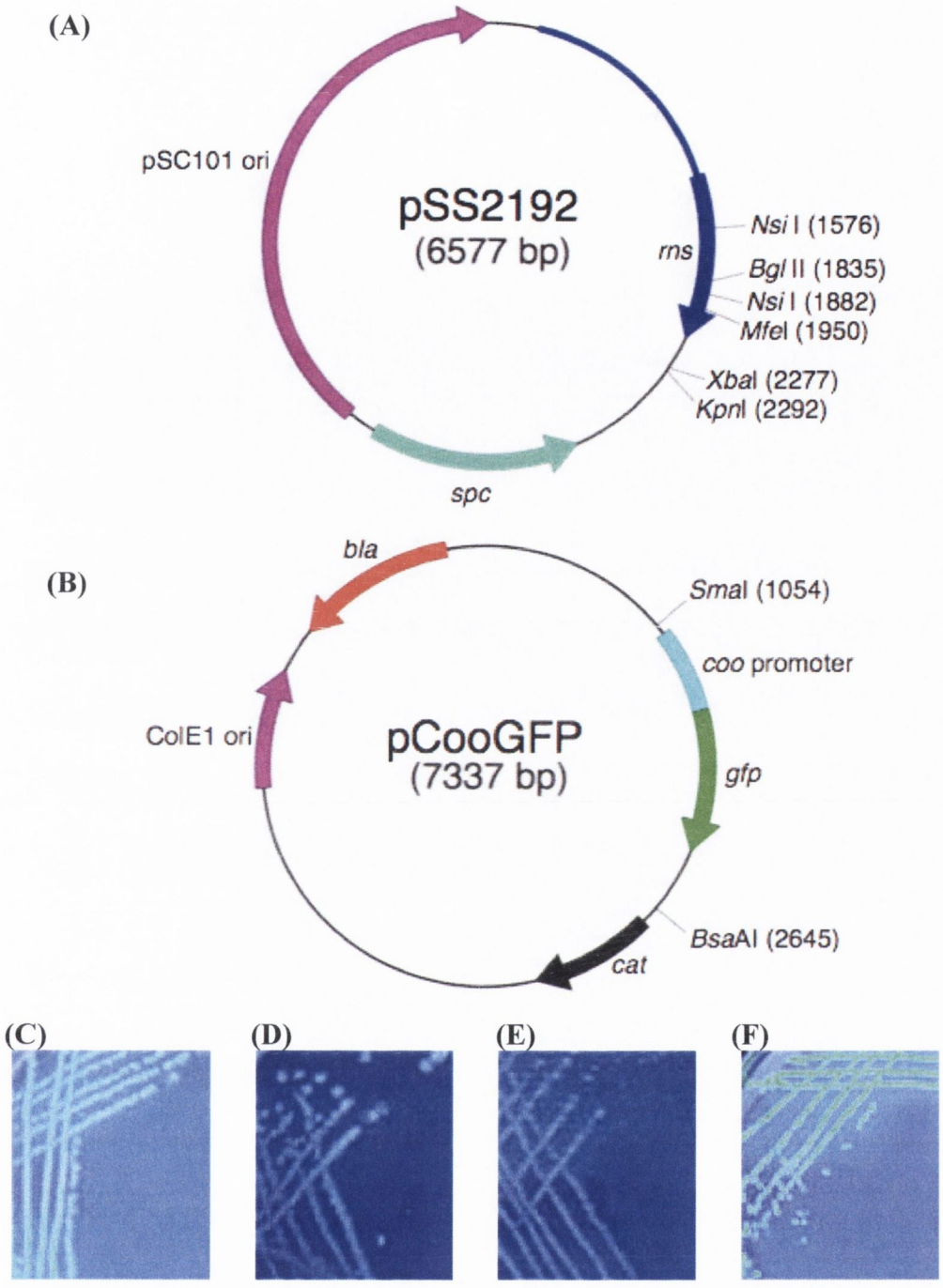

Fig. 3.2. Structures of plasmids pSS2192 and pCooGFP and appearance of bacterial colonies irradiated with UV light. **(A)** Map of the plasmid pSS2192. Important features such as the *rns* gene and pSC101 origin are indicated. Relevant restriction endonuclease sites are marked. **(B)** Map of the plasmid pCooGFP. The location of the *coo* promoter fused to the promoterless *gfp* (encodes green fluorescent protein, GFP) gene is shown. Relevant restriction endonuclease sites are marked. **(C-F)** Images of typical fluorescent and non-fluorescent colonies. Shown are cultures of *E. coli* XL-1/pCooGFP harbouring pSS2192 (*rns* encoding plasmid, positive control) **(C)**, pCL1920 (empty vector plasmid, negative control) **(D)**, or pSS2192 derivatives encoding the Rns mutants provisionally titled A3 **(E)** and E3 **(F)**.

pSS2192 (Fig. 3.2 (C-F)). Of approximately 5,000 colonies, the vast majority (>75%) were non-fluorescent under UV light while a small proportion (~1%) were highly fluorescent under UV light. A selection of these non-fluorescent and highly fluorescent colonies found to be sensitive to kanamycin (i.e. free of the transposon) were chosen for further analysis. At this time the insertion mutants were given provisional names such as A1-14, B1-10 and E1-15.

3.2.1.3 Screening pSS2192 derivatives to identify 15 bp insertions in the *rns* ORF

The *rns* ORF contains three *Dra*I sites (TTT/AAA) located 7 bp, 249 bp and 784 bp downstream from the translation start site (Fig. 3.3 (A)). Therefore, *Dra*I digestion of the 798 bp wild type *rns* ORF produces two large fragments (easily detectable after agarose gel electrophoresis) of 242 bp and 535 bp, and two small fragments of 9 bp and 12 bp. The 15 bp insertion generated by the GPS-LS linker scanning system includes a rare *Pme*I site (GTTT/AAAC), which also contains a *Dra*I site. Therefore, *Dra*I digestion of a mutant *rns* ORF harbouring a 15 bp insertion will produce a different set of fragments to that found for digestion of wild type *rns,* as the enzyme will also cut at the insertion site. Thus, to confirm that the altered fluorescence phenotypes described above were due to insertions in the *rns* ORF (and not elsewhere in pSS2192) and to determine the approximate location of the insertions, the *rns* ORF of the mutated pSS2192 plasmids within the chosen 48 non-fluorescent colonies was amplified by PCR using the primers Rns-F and Rns-19r (Table 2.3). The PCR products were subjected to *Dra*I digestion and the resulting fragments were then separated by agarose gel electrophoresis. A representative gel is shown in Fig. 3.3 (B). As expected, *Dra*I digestion of wild type *rns* produced a band of 798 bp corresponding to uncut DNA and two other bands of 535 bp and 242 bp. There were also bands of 251 bp and 547 bp due to incomplete digestion at the sites at the ends of the PCR products. Four mutants had the same pattern of digestion fragments as wild type *rns* and therefore did not appear to have an insertion within the *rns* ORF. DNA fragments were not detected for some mutants, A2 and A9 for example. This could have been due to the 15 bp insertion occurring within the binding sites of the primers thus preventing PCR amplification of these mutants. However, examination of the *Dra*I digestion products of most mutants revealed the presence of additional fragments not seen in the digest of wild type *rns*. This confirmed that these mutants had a 15 bp insertion within the *rns* ORF. In many

cases an extra band of between 250 bp and 500 bp could be seen, indicating that the insertion had occurred within the 535 bp fragment of *rns*. This was not surprising as this fragment encodes the C-terminal domain of the Rns protein that contains the two predicted HTH motifs, so insertions here would be expected to disrupt the activity of Rns thus resulting in decreased GFP expression within the *E. coli* XL-1/pCooGFP background. Based on the results of the *Dra*I digests, a selection of the mutant *rns* ORFs were sequenced to identify the precise location and sequence of the insertion. In addition, the secondary structure of Rns was predicted to gain a better understanding of the regions where the insertions had occurred and to relate this to the effects of the insertions on Rns activity. At the time of writing, a tertiary structure of Rns or a close homologue was not available, therefore the jpred and PROF programmes were used to obtain a predicted secondary structure (64, 338). Both programmes predicted a consensus of features within Rns (Fig. 3.4) as they have previously done for the AraC family member ToxT (57). Of the mutant genes sequenced, seven encoded Rns proteins with pentapeptide insertions within or in the vicinity of one of the protein's predicted HTH motifs, one encoded a Rns protein with a pentapeptide insertion in the N-terminal domain of the protein, and six encoded truncated forms of Rns (Table 3.1). Some of the independently isolated mutants contained identical insertions. Once the mutant proteins had been characterised they (and the pSS2192 derivatives encoding them) were formally named by the amino acid residue to the N-terminal side of the insertion position (Table 3.1). Truncates were indicated by a delta symbol (Δ).

3.2.1.4 Screening pSS2192 derivatives to identify 15 bp insertions in the region upstream of the *rns* ORF

The pSS2192 plasmids from the selected 15 highly fluorescent colonies were also screened for 15 bp insertions within the *rns* ORF as described above. However the fragments produced by the *Dra*I digests of the *rns* ORF of these mutants exactly matched that of wild type *rns* (results not shown). Therefore the 15 bp insertions within these mutants must be located outside of the *rns* ORF, possibly within the upstream area of the gene that is also present in pSS2192. To test this possibility, the primers Rns-UpF and Rns-UpR (Table 2.3) were used to amplify by PCR a 773 bp fragment of pSS2192. This fragment consists of the first 132 bp of the *rns* ORF (including the first *Dra*I site) in addition to 641 bp of DNA found immediately upstream of the ORF (Fig.

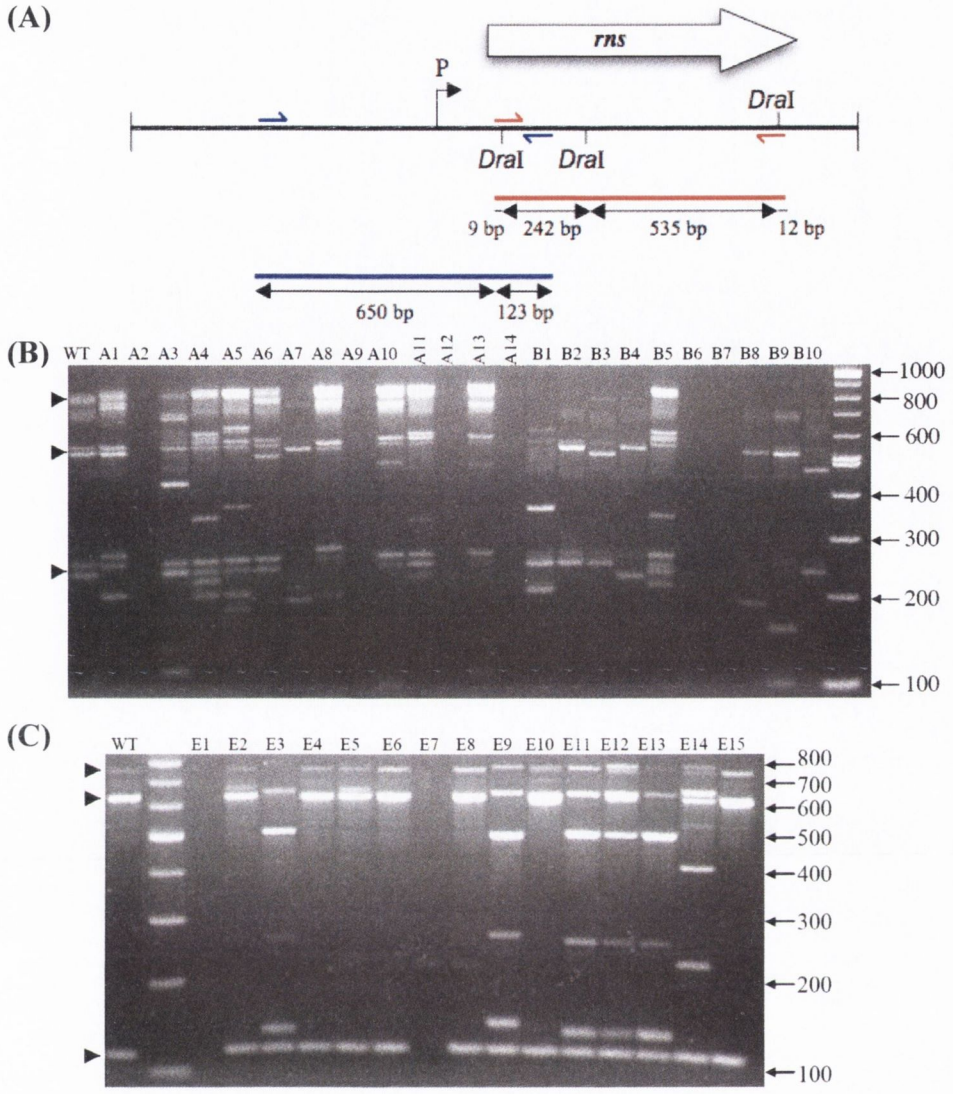

Fig. 3.3. *Dra*I digestion of *rns* and mutant *rns* genes. (A) Diagram of the 2.0 kb *rns*-containing insert of pSS2192. The *rns* ORF is represented by an open arrow. The location of the *rns* promoter (P) and the *Dra*I sites are shown. The primers used to amplify the *rns* ORF and the upstream area of *rns* are represented as red and blue arrows (respectively), while the PCR products resulting from such amplifications are represented as red and blue lines (respectively). The lengths of the fragments produced by *Dra*I digestion of the PCR products are indicated. Screening pSS2192 derivatives for 15 bp insertions within the *rns* ORF **(B)** and within the upstream area of *rns* **(C).** Wild type pSS2192 and selected pSS2192 derivatives were subjected to PCR amplification of the 798 bp *rns* ORF **(B)** or a 773 bp fragment of *rns* consisting of 641 bp upstream of the *rns* ORF and the first 132 bp of the ORF **(C)**. The PCR products were digested with *Dra*I and then separated on a 2% (wt/vol) agarose gel. The provisional names of the mutants encoded by the pSS2192 derivatives are indicated above each lane. The bands expected from *Dra*I digestion of wild type *rns* DNA are indicated by arrowheads on the left. The presence of any other bands indicates that the mutant contains a 15-bp insertion within the screened area. DNA molecular size standards (in base pairs) are indicated on the right.

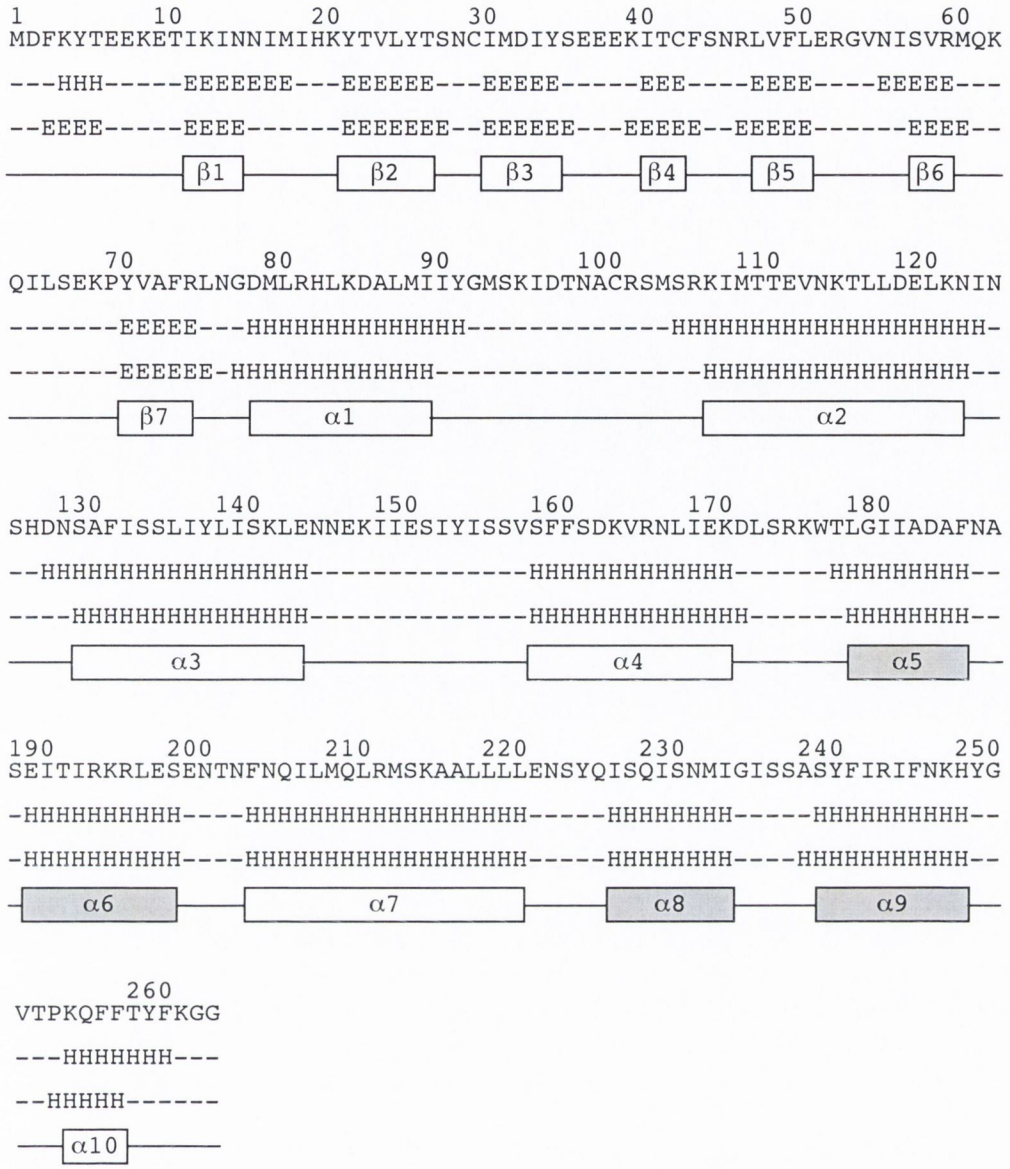

Fig. 3.4. Predicted secondary structure of the Rns protein. The numbered amino acid sequence of Rns is shown on the top line. Secondary structure predictions from jpred and PROF are shown in the second and third lines, respectively, (H = α-helix, E = β-strand). The consensus prediction is depicted on the bottom line (α = α-helix, β = β-strand). Helices predicted to be involved in the helix-turn-helix motifs are shown in grey. There was no consensus prediction for the extreme N-terminus of Rns but subsequent analysis has revealed that this area is predicted to be unstructured or disordered (see Chapter 4).

TABLE 3.1. Rns derivatives resulting from mutagenesis of pSS2192

Formal name[a]	Provisional name	Pentapeptide insertion	Predicted structure at insertion site[b]
RnsI17Δ[c]	A1	IVOch[d]TI	Unstructured
RnsH20Δ	A8	KYVOchT	Unstructured
RnsC102	C12	RCLNS	Unstructured
RnsI192Δ	A11	VOchTEI	Recognition helix of HTH1 (α6)
RnsR195Δ	B5	NVOchTR	Recognition helix of HTH1 (α6)
RnsL198	A4, C3	FKQRL	Recognition helix of HTH1 (α6)
RnsF205	A5	MFKHF	Helix between HTH1 and 2 (α7)
RnsL221	C2	CLNIL	Helix between HTH1 and 2 (α7)
RnsQ227	A3	CLNNQ	First helix of HTH2 (α8)
RnsS239	A13	CLNTS	Turn of HTH2
RnsS239Δ	D4	VOchTSS	Turn of HTH2
RnsY242	A10	CLNTY	Recognition helix of HTH2 (α9)
RnsV253Δ	A6, D9	VOchTGV	Helix downstream of HTH2 (α10)
RnsK256	D10	CLNTK	Helix downstream of HTH2 (α10)

[a]Mutant Rns proteins are named by the amino acid residue to the N-terminal side of the insertion position

[b]Stucture predictions from jpred and PROF, see Fig. 3.4. Areas not predicted to be of either α or β structure are designated as unstructured

[c]Truncates are indicated by a delta symbol (Δ)

[d]Och, ochre stop codon

3.3 (A)). The PCR products amplified from each of the mutants were digested with *Dra*I and the resulting fragments were analysed by agarose gel electrophoresis as before (Fig. 3.3 (C)). The set of digestion products for several of the mutants (E2, E4, E5, E6, E8, E10, E15) was identical to that obtained for wild type *rns*, i.e. fragments of 650 bp and 123 bp. Therefore, these mutants do not harbour the 15 bp insertion in the *rns* ORF or the 641 bp upstream of the ORF. The insertion may be located elsewhere in pSS2192 and may have led to an increase in the plasmid copy number such that the levels of Rns protein were increased thus resulting in elevated GFP expression in the *E. coli* XL-1/pCooGFP background. Previous studies have found that mutations in the pSC101 replicon, which is present in pSS2192, result in plasmids with copy numbers four to eight-fold greater than wild type. These mutations map to the *rep* gene that

encodes a protein involved in the regulation of plasmid replication (420, 436).

For two mutants (E1 and E7) no DNA fragments were detected. As described earlier this was probably due to a lack of PCR amplification. However, the digests of the remaining six mutants produced fragments not seen in the digest of the wild type *rns* PCR product. Sequencing of four of these mutants confirmed that they contained the 15 bp insertion in the 641 bp of DNA upstream of the *rns* ORF. For three independently isolated mutants (E3, E11 and E13) the 15 bp insertion had occurred 45 bp downstream of the *rns* transcription start site (named *rns*45). In one case, mutant E9, the 15 bp insertion was 36 bp downstream of the *rns* transcription start site (named *rns*36).

3.2.1.5 Screening for expression of Rns insertion mutants

It was essential to establish that the 15 bp insertions within the *rns* ORF were not mitigating against expression of the encoded Rns protein. This would rule out the possibility that the lack of activation of the *coo* promoter in the GFP based screen described above was due to an absence of Rns protein rather than the inability of a mutant form of Rns to function. The simplest method to monitor protein expression would be to perform Western immunoblot analysis, using anti-Rns antiserum, on strains harbouring the mutant constructs. However this was complicated by the discovery that even expression of wild type Rns from pSS2192 could not be detected using standard Western immunoblot methods, possibly due to the low copy number of the plasmid. The presence of Rns in whole-cell lysates of *E. coli* XL-1/pSS2192 could be detected when an extremely sensitive enhanced chemiluminescent substrate (SuperSignal® West Femto Maximum Sensitivity Substrate, Pierce) was used to probe immunoblots. This substrate enables the detection of femtogram amounts of antigen. However, when the same immunoblot procedure was performed on whole-cell lysates of overnight cultures of *E. coli* XL-1 harbouring a selection of the mutant derivatives of pSS2192, mutant Rns proteins were not detected (Fig. 3.5 (A)). This does not necessarily indicate that Rns protein is not expressed from the mutant pSS2192 derivatives. Rns produced from pSS2192 is expressed from its native promoter, which is positively autoregulated. Insertions that may adversely affect Rns function at the *coo* promoter, as was seen in the GFP screen, may also adversely affect its function at the *rns* promoter. Therefore it is possible that pSS2192 derivatives with 15 bp insertions within the *rns* ORF produce Rns derivatives with pentapeptide insertions, but unlike wild type Rns these mutant

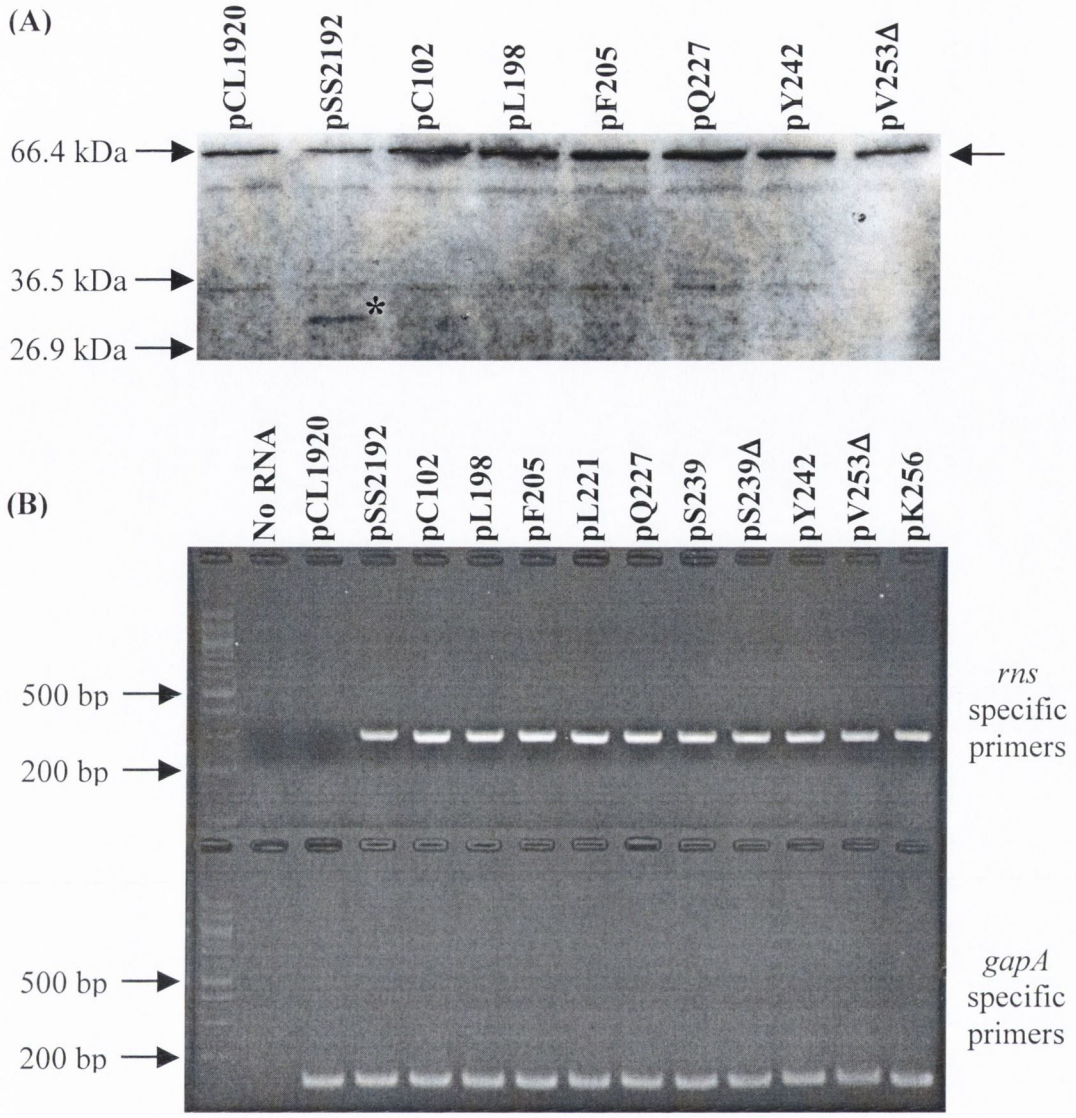

Fig. 3.5. Expression of *rns* and mutant *rns* genes. (A) Western immunoblot analysis of the presence of Rns in whole-cell lysates of *E. coli* XL-1 harbouring the indicated plasmids. The positions of protein molecular mass markers are shown on the left. An asterisk indicates the position of Rns. A non-specific cross-reactive band migrating alongside the 66.4 kDa marker (indicated by an arrow on the right) serves as a loading control. **(B)** RT-PCR analysis performed using total RNA isolated from cultures of *E. coli* XL-1 (harbouring the indicated plasmids) and primers specific for the *rns* transcript or the *gapA* transcript as indicated. Lane 2 contains an RNA-free negative control. The positions of selected DNA molecular size standards are indicated on the left.

proteins may not then be able to activate the *rns* promoter to produce the even greater levels of protein that can be detected by Western immunoblotting.

While it was not possible at this time to show that the insertion mutations had not prevented protein expression, it was possible to investigate whether the mutants were at least transcribed. This was done by isolating total RNA from mid-logarithmic phase cultures of *E. coli* XL-1 harbouring the empty vector control pCL1920, the positive control pSS2192 or a selection of the pSS2192 derivatives containing 15 bp insertions within the *rns* ORF. The presence of *rns* transcripts in the cultures was detected by performing a one-step end-point RT-PCR assay using the isolated total RNA as template. This assay involves the reverse transcription of the RNA to produce cDNA, which then serves as the template in a standard PCR amplification that was conducted with the primers rt-rnsF and rt-rnsR (Table 2.3). These primers are designed to amplify a 311 bp area present in the *rns* transcript. Control RT-PCR assays were simultaneously carried out using primers rt-gapF and rt-gapR (Table 2.3) that are specific for the constitutively expressed housekeeping gene *gapA* (261). An equal amount of product was obtained from each of the RNA containing samples when RT-PCR was conducted with the rt-gapF and rt-gapR primers (Fig. 3.5 (B)). This confirms that there was minimal variance between the samples, i.e. the RNA preparations were of good quality and an equal amount of RNA was used in each reaction. When RT-PCR was conducted with the rt-rnsF and rt-rnsR primers no product was detected for the pCL1920 sample as this is the empty vector negative control and does not contain the *rns* gene. An equal amount of product was detected for the pSS2192 sample and the selection of pSS2192 insertion containing derivatives however. This confirms that the presence of the 15 bp insertions does not prevent transcription of the mutant *rns* genes.

3.2.1.6 Generation and screening of a library of 15 bp insertions in pHisRns

While the initial round of mutagenesis on pSS2192 had been successful, it was evident that this low copy number plasmid system was challenging to work with. For example it was necessary to grow large volumes of bacterial cultures harbouring the pSS2192 derivatives in order to purify sufficient amounts of plasmid DNA. In addition, it would not be possible to purify the mutant proteins expressed by the pSS2192 derivatives. Therefore it was decided to conduct another round of mutagenesis with the higher copy number (an estimated 20 copies per cell) plasmid pHisRns (Table 2.2, Fig.

3.6 (A)). This plasmid allows high-level expression of Rns under the control of an inducible T5 promoter and, as it encodes Rns protein with six histidine residues at its amino terminus, any insertion mutants created in this plasmid could then be easily purified. The mutagenesis procedure was carried out as described above, this time using a chloramphenicol-resistant derivative of the bacterial transposon Tn7 (from the donor plasmid pGPS4). The random insertion of this transprimer into pHisRns was followed by *Pme*I digestion and then ligation to obtain a library of pHisRns plasmids containing 15 bp insertions.

As both the pHisRns plasmid and the pCooGFP reporter plasmid carry ampicillin resistance and have ColE1 based origins of replication they are not compatible and consequently cannot be present simultaneously in the same bacterial cell. Therefore it was necessary to construct a GFP reporter plasmid that would be compatible with pHisRns to enable screening of the 15 bp insertion containing pHisRns plasmids. A 1.6 kb region of DNA that contains the *coo* promoter fused to the promoterless *gfp+* gene was excised from pCooGFP by digestion with *Sma*I and *Bsa*AI and this blunt-ended fragment was cloned into the plasmid pACYC177 that had been digested with *Hin*CII. The resulting transcriptional fusion reporter plasmid, named pCooGFP-2 (Table 2.2, Fig. 3.6 (B)), carries kanamycin resistance and has a p15A-based origin of replication. The library of pHisRns plasmids containing 15 bp insertions was transformed into *E. coli* XL-1/pCooGFP-2. Out of approximately 2,000 colonies, 65 displaying reduced or no fluorescence were screened for the presence of a 15 bp insertion within the *rns* ORF of their pHisRns derivative by the PCR and *Dra*I digestion screen as described above. To determine the areas of Rns that can tolerate insertions and remain functional at the *coo* promoter, 95 colonies displaying wild-type levels of fluorescence were also selected and screened for the presence of 15 bp insertions in the *rns* ORF of the pHisRns derivative they contained. Of the pHisRns derivatives within the non-fluorescent colonies, 13 found to contain insertions in the *rns* ORF were then screened for expression of full length Rns protein by Western immunoblotting. Three were found to produce full length Rns (results not shown) and were sequenced to determine the precise location of their insertions. One was found to encode a pentapeptide insertion after the threonine residue at position 23 of Rns, the second encoded a pentapeptide insertion after the cysteine at position 30 of Rns and the third encoded a pentapeptide insertion after the lysine residue at position 51 of Rns. The mutant plasmids and proteins were named accordingly (i.e. pHisRnsT23 and RnsT23 etc).

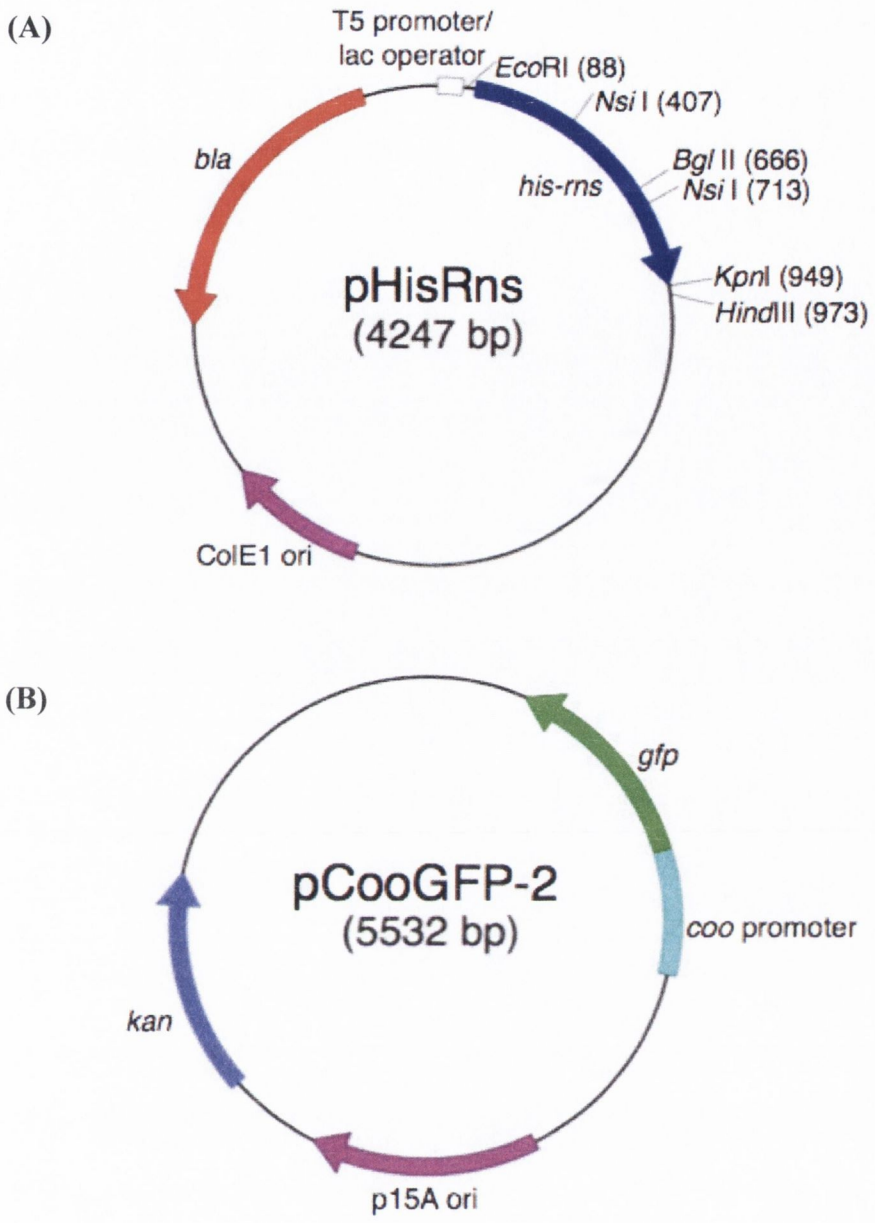

Fig. 3.6. Structures of the plasmids pHisRns and pCooGFP-2. **(A)** Map of the plasmid pHisRns. The position of the *his-rns* gene fusion is noted along with other relevant features and restriction endonuclease sites important for cloning. **(B)** Map of the plasmid pCooGFP-2. The location of the *coo* promoter fused to the promoterless *gfp* (encodes GFP) gene is shown.

Of the pHisRns derivatives within the colonies displaying wild-type levels of fluorescence, 13 were found to contain 15 bp insertions within the *rns* ORF. Sequencing revealed the exact sequence and location of the insertions (Table 3.2) and that only ten were unique as three insertions had been independently isolated twice.

The three-dimensional structure of Rns is not known but both the jpred and PROF secondary structure prediction programs predict that the N-terminal region is of primarily β-structure while the C-terminal domain is mainly α-helical (Fig. 3.4). Combining the results of the mutagenesis on both pSS2192 and pHisRns it could be surmised that pentapeptide insertions that cause Rns to be non-functional at the *coo* promoter mainly occurred in the α-helical C-terminal half of the protein, while tolerated insertions clustered in the β-strand rich N-terminal half of the protein.

TABLE 3.2. Functional insertion mutants resulting from mutagenesis of pHisRns

Formal name[a]	**Provisional name**	**Pentapeptide insertion**	**Predicted structure at insertion site**[b]
RnsE10	b35	CLNKE	Unstructured
RnsT11	a42	CLNKT	Unstructured
RnsK13	a20	CLNIK	β-strand (β1)
RnsI17	a4	CLNNI	Unstructured
RnsI34	a19, b24	CLNNI	β-strand (β3)
RnsK108	b36	CLNRK	α-helix (α2)
RnsL118	b27	CLNTL	α-helix (α2)
RnsL119	a9	FKHLL	α-helix (α2)
RnsS127	a23, b11	LFKHS	Unstructured
RnsN130	a46, b43	CLNNN	Unstructured

[a]Mutant Rns proteins are named by the amino acid residue to the N-terminal side of the insertion position

[b]Stucture predictions from jpred and PROF, see Fig. 3.4. Areas not predicted to be of either α or β structure are designated as unstructured

3.2.1.7 Random mutagenesis of *rns*

The *rns* ORF was subjected to random mutagenesis as a follow up to the pentapeptide scanning mutagenesis and in an attempt to isolate more non-functional mutations in the N-terminal half of Rns. The *rns* ORF was amplified from pHisRns by PCR using *Taq*

DNA polymerase and the primers Ranmut3 and RnsKpn (Table 2.3). The primers Ranmut3 and RnsKpn have *Eco*RI and *Kpn*I sites incorporated into them respectively. The concentration of dCTP in the PCR reaction was reduced to limiting amounts while 0.2 M of dITP was added in its place to increase the frequency of mutations arising in the PCR products. Several PCR reactions were performed. The PCR products were purified, digested with *Eco*RI and *Kpn*I and then ligated into the pBSKII+ vector that had also been digested with *Eco*RI and *Kpn*I. This produced a pool of plasmids containing random mutations within the *rns* ORF. The plasmid library was digested with *Eco*RI and *Bgl*II to isolate a 520 bp fragment encoding the N-terminal half of Rns. This fragment was ligated into the pHisRns backbone from which the equivalent fragment had been excised by digestion with *Eco*RI and *Bgl*II. This resulted in a library of plasmids containing random mutations only within the first half of the *rns* ORF. These plasmids were transformed into *E. coli* XL-1/pCooGFP-2 to identify mutants unable to activate the *coo* promoter. Out of approximately 2,000 colonies screened, 131 displaying reduced fluorescence were selected. Overnight cultures of each were set up and were induced with 1 mM IPTG for two hours prior to harvesting. Whole cell lysates of each of the cultures were screened for the presence of full length Rns protein by Western immunoblotting (Fig. 3.7). Of the 131 potential mutants, four were found to express full length Rns and the *rns* gene present in each was amplified by PCR and sequenced to determine what mutation was present. Three of the mutations were not random point mutations but 15 bp insertions thought to be the remnants of a transposition event. On a previous occasion, Insertion Element 10 (IS10) had been found to transpose into pHisRns. IS10 forms part of the composite tetracycline resistance transposon Tn10, which is present in the F plasmid of *E. coli* XL-1. UV light, used to identify fluorescent and non-fluorescent colonies containing mutated pHisRns, has been shown to induce IS10 transposition in *E. coli* (106). These mutants were not studied further. The remaining mutation was identified as the point mutation C473T, which results in the amino acid substitution S158F.

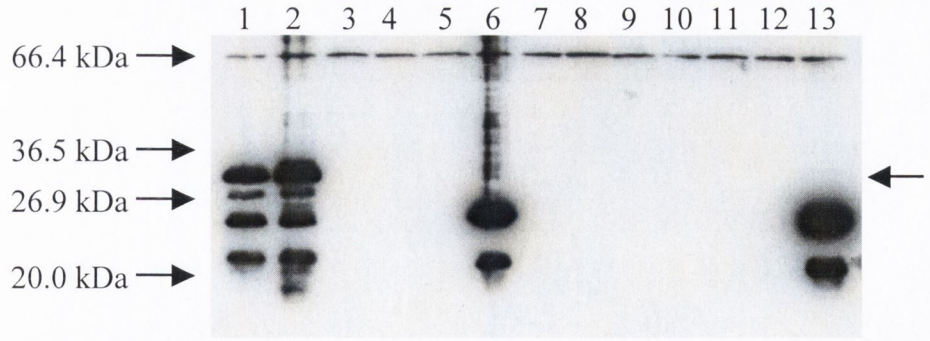

Fig. 3.7. Expression of Rns and putative random mutants. Representative Western immunoblot analysis (using anti-Rns antiserum) of whole-cell lysates of *E. coli* XL-1/pCooGFP-2 harbouring pHisRns (lane 1) or pHisRns derivatives containing *rns* inserts that had been subjected to random mutagenesis. The positions of protein molecular mass markers are shown on the left. A non-specific cross-reactive band migrating alongside the 66.4 kDa marker serves as a loading control. An arrow on the right indicates the position of full-length Rns. Other bands are most likely truncates (lanes 6 and 13) or break down products. The mutant in lane 2 (RnsS158F) is expressed as full-length protein.

3.2.2 Analysis of the activity of mutant Rns proteins

3.2.2.1 Transactivation of the *coo* promoter by Rns and mutant derivatives

As detailed above, the mutant derivatives of pSS2192 and pHisRns had been transformed into *E. coli* XL-1 containing a reporter plasmid (pCooGFP or pCooGFP-2) to provide an indication of the ability of the mutant Rns proteins to activate the *coo* promoter. It was now necessary to quantify the amount of GFP expression in such cultures and thus accurately determine the level of activity of the Rns mutants. Cultures were subjected to fluorimetric analysis as described in Section 2.4.2. The fluorescence values obtained in the analyses were expressed as a percentage of the fluorescence detected in cultures containing wild type Rns. Mutations were considered to be tolerated, i.e. not significant, if the fluorescence levels in cultures containing the mutants were within two-fold of the fluorescence levels in cultures containing wild type Rns. Cultures encoding Rns proteins containing pentapeptide insertions within, or in the vicinity of, either of the protein's predicted HTH motifs (i.e. RnsL198, RnsF205, RnsL221, RnsQ227, RnsS239, RnsY242, and RnsK256) had extremely low levels of fluorescence, less than 5% of that detected in the presence of wild type Rns (Fig. 3.8 (A)). This was also true for the two truncates tested, RnsS239Δ and RnsV253Δ. The fluorescence levels of cultures containing the pentapeptide insertion mutants HisRnsT23, HisRnsC30 and HisRnsL51 and the point mutant HisRnsS158F were equally low (Fig. 3.8 (B)). These fluorescence levels were similar to those in cultures harbouring the *rns* free controls pCL1920 or pQE-30. This indicates that there were negligible levels of GFP in each of these cultures and therefore these Rns mutants were unable to activate the *coo* promoter.

The mutant protein RnsC102 retained an intermediate level of activity at the *coo* promoter. Bacteria expressing this protein were 12% as fluorescent as those expressing wild type Rns (Fig. 3.8 (A)). Therefore the pentapeptide insertion after C102 seems to have a less dramatic effect on Rns activity than insertions in the HTH regions.

The fluorimetric analysis confirmed that the insertion mutants isolated from fluorescent colonies of *E. coli* XL-1/pCooGFP-2 harbouring pHisRns derivatives (Table 3.2) were functional at the *coo* promoter (Fig. 3.8 (C)). The fluorescence levels in cultures containing these mutants ranged from 76% to 178% of the level of fluorescence in the presence of wild type HisRns. These pentapeptide insertions were not considered

to have had a significant effect on protein activity and therefore had not occurred in critical regions of Rns. The resultant mutants could tolerate the insertions and largely retain their ability to activate the *coo* promoter.

The fluorescence levels in cultures containing the mutants *rns*36 and *rns*45 were 1.6-fold and 3.7-fold higher (respectively) than that in cultures possessing wild type *rns* (results not shown). As the fluorescence detected for *rns*36 was less than two-fold greater than wild type levels, only the insertion in *rns*45 was considered to have had a significant effect. This insertion 45 bp downstream of the *rns* transcription start site may have resulted in increased levels of Rns, which in turn drove the expression of increased levels of GFP from the *coo* promoter on the reporter plasmid.

3.2.2.2 Purification of His-tagged Rns and subcloning of pSS2192 based mutations into pHisRns

Further characterisation of Rns mutants would require purified protein. However, the mutant Rns proteins encoded by the pSS2192 derivatives (Table 3.1) could not be over-expressed or readily purified. It is possible to over-express and purify His-tagged Rns protein from cultures of *E. coli* KS1000 harbouring the plasmids pRare and pHisRns. The *E. coli* strain KS1000 is deficient in the production of Tsp protease (370) and this has been shown to improve expression of the Rns protein (373). The pRare plasmid encodes tRNAs for codons that rarely occur in *E. coli*. Such codons occur frequently in the *rns* gene and this can reduce the translation of Rns in the cell. The combined use of the KS1000 strain and the pRare plasmid enables the expression of sufficient amounts of soluble Rns protein for purification. Over-expression of His-tagged Rns was induced with the addition of IPTG to logarithmic-phase cultures of KS100/pRare/pHisRns, or by growing the cultures overnight in Overnight Express™ Instant TB Medium (Novagen). Whole cell lysates of induced cultures were separated into soluble and insoluble fractions by centrifugation. His-tagged Rns was purified by passing the soluble fraction through a nickel column and then eluting the protein from the column. An example of purified His-tagged Rns analysed by SDS-PAGE, along with samples of the uninduced and induced cultures, and the insoluble and soluble fractions, is shown in Fig. 3.9 (A).

Several of the insertion mutations present in the pSS2192 derivatives (Table 3.1) were chosen to be subcloned into the pHisRns background. In each of the plasmids pL198, pF205, pY242 and pV253Δ, the 15 bp insertions are located between *Bgl*II and *Kpn*I

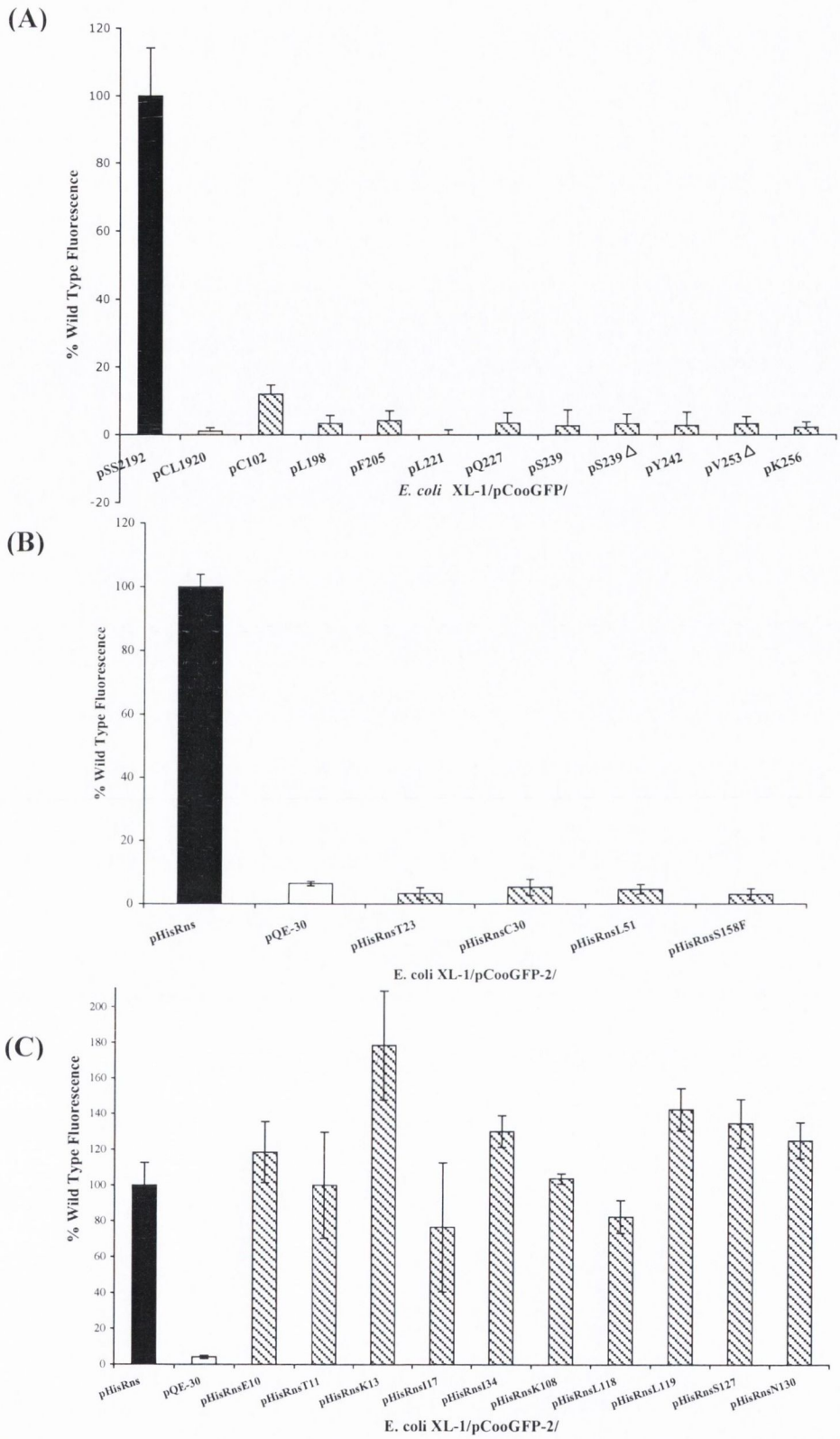

Fig. 3.8. Transactivation of the *coo* promoter by Rns and mutant derivatives. Fluorescence levels of the indicated bacterial cultures (see horizontal axes) were expressed as a percentage of the level of fluorescence in cultures containing wild type Rns. Data represents averages of triplicate measurements of duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Sets of bacterial cultures analysed on different days are shown in separate panels.

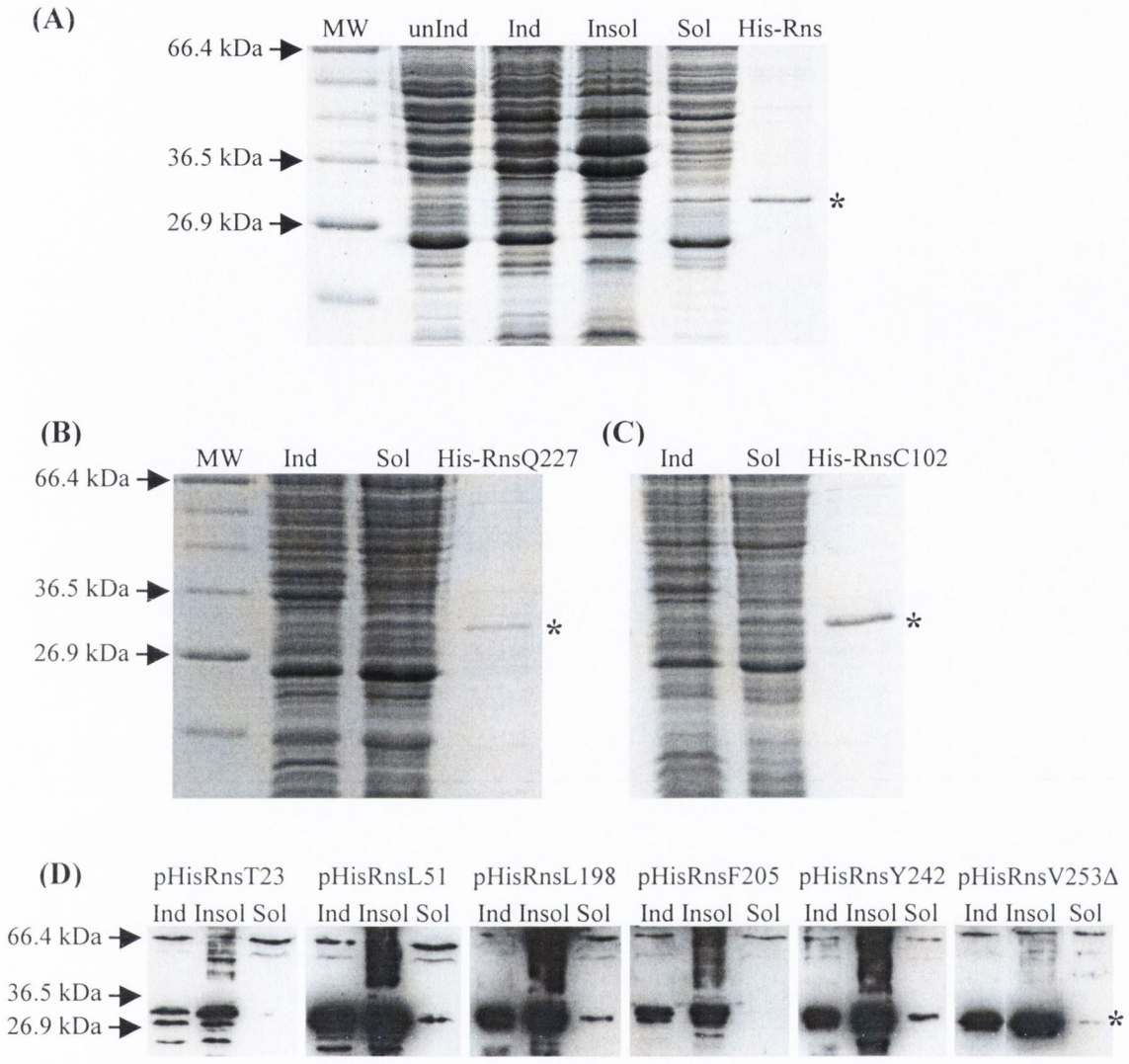

Fig. 3.9. Expression and purification of His-tagged Rns and mutant derivatives. SDS-PAGE analysis of uninduced (unInd, grown in L broth) and induced (Ind, grown in Overnight Express medium) cultures, as well as insoluble (Insol) and soluble (Sol) fractions of the induced cultures of *E. coli* KS1000/pRare harbouring pHisRns **(A)**, pHisQ227 **(B)** or pHisC102 **(C).** Also included are the purified proteins themselves (indicated by an asterisk on the right) and selected protein molecular mass markers (MW, sizes indicated on the left). **(D)** Western immunoblot analysis of induced cultures of *E. coli* KS1000/pRare harbouring the indicated plasmids, and the insoluble and soluble fractions of the induced cultures. An asterisk on the right indicates the position of Rns.

sites. These plasmids were digested with *Bgl*II and *Kpn*I to isolate the insertion containing fragments, which were then ligated into a pHisRns backbone that had also been digested with *Bgl*II and *Kpn*I to remove the equivalent wild type *rns* fragment. The resulting plasmids were named pHisL198, pHisF205, pHisY242 and pHisV253Δ (Table 2.2). The 15 bp insertion in pC102 is located between two *Nsi*I sites. The 302 bp fragment resulting from *Nsi*I digestion was excised from pHisRns and replaced with the equivalent fragment derived from *Nsi*I digestion of pC102 to create pHisC102 (Table 2.2). A diagnostic restriction digest was carried out to confirm that the newly inserted fragment was in the correct orientation. Finally, a *Bgl*II-*Hin*dIII DNA fragment was excised from pHisRns and replaced with a *Bgl*II-*Hin*dIII fragment containing the 15 bp insertion encoding the Q227 pentapeptide insertion to create pHisQ227 (Table 2.2).

All of these pHisRns derivatives, along with the plasmids pHisT23, pHisL51 and pHisS158F that had been isolated by performing mutagenesis directly on pHisRns, were transformed into *E. coli* KS1000/pRare. Using the procedure described above and in Section 2.3.2.2 it was possible to purify His-tagged RnsQ227 and His-tagged RnsC102 protein (Fig. 3.9 (B) and (C)). This confirmed that the 15 bp insertions in these mutants had not disrupted protein expression. However, when the procedure was conducted using pHisL198, pHisF205, pHisY242, pHisV253Δ, pHisT23, pHisL51 and pHisS158F purified protein was not recovered in the material eluted from the nickel columns. The expression of His-RnsT23, His-RnsC30, His-RnsL51 and His-RnsS158F had already been confirmed by Western immunoblotting while these mutants were being screened. Samples of induced *E. coli* KS1000/pRare cultures harbouring pHisT23, pHisL51, pHisL198, pHisF205, pHisY242 and pHisV253Δ, and of the insoluble and soluble fractions prepared from these cultures during the attempted purification procedures were analysed by Western immunoblotting to determine if the mutant proteins had been expressed (Fig. 3.9 (D)). The Western immunoblots confirmed that on induction the proteins were expressed. Therefore, while it had not been possible to detect these proteins when they were expressed from the native *rns* promoter in pSS2192, they could be easily detected when they were expressed from the T5 promoter in a medium copy number plasmid. However the mutant proteins were predominantly present in the insoluble fraction of the whole cell lysates. The severely reduced amount of each of these proteins in the soluble fraction is the likely reason why they could not be purified.

3.2.2.3 Further quantification of the ability of Rns mutants to activate transcription at the *coo* promoter

It was not possible to obtain purified His-tagged Rns insertion mutants L198, F205, Y242 or the truncate V253Δ for use in *in vitro* assays of protein function. However, now that they were encoded by pHisRns based plasmids it had been demonstrated that each of these mutant proteins, and His-tagged RnsC102 and RnsQ227, were expressed *in vivo*. Therefore the fluorimetric analysis of these mutants was repeated using their pHisRns based plasmids and the pCooGFP-2 reporter plasmid in *E. coli* XL-1 (Fig. 3.10). Again the mutant containing an insertion after position C102 exhibited partial activity at the *coo* promoter, cultures containing pHisC102 were 27% as fluorescent as those containing pHisRns. This was a higher level of activity than that detected when the protein was encoded by the pSS2192 derivative pC102. This could be due to the greater levels of mutant Rns protein in cells containing the medium copy number pHisC102 rather than the low copy number pC102 where protein expression is driven by the native *rns* promoter.

As before, the fluorescence levels of cultures expressing the truncate RnsV253Δ or mutants with insertions in or near either of the predicted HTH regions of Rns were similar to the levels found in cultures containing the *rns* free control pQE-30. However it is now known that the lack of activation of the *coo* promoter in cultures containing these mutants is not simply caused by the Rns mutants not being expressed. Therefore the lack of expression of GFP in cultures containing these mutants is due to the insertions causing a defect in the activity of Rns at the *coo* promoter.

3.2.2.4 Activation of CS1 expression by Rns and mutant derivatives in ETEC

LMC10 is a strain of enterotoxigenic *E. coli* that carries the plasmid containing the CS1 fimbrial genes but lacks the plasmid harbouring the *rns* gene. To test the ability of a selection of the mutant Rns proteins to activate the *coo* promoter in its natural environment (i.e. as the promoter of CS1 fimbrial expression), the pSS2192 and pHisRns derivatives of interest were transformed into LMC10. Transformants were restreaked and then grown overnight in CFA broth as this has been shown to lead to optimal CS1 expression. Heat released extracts (heat shocks) of the overnight cultures were prepared as described in Section 2.3.1.2 and analysed by SDS-PAGE to test for the presence of CooA, the 15.2 kDa major subunit of CS1 fimbriae.

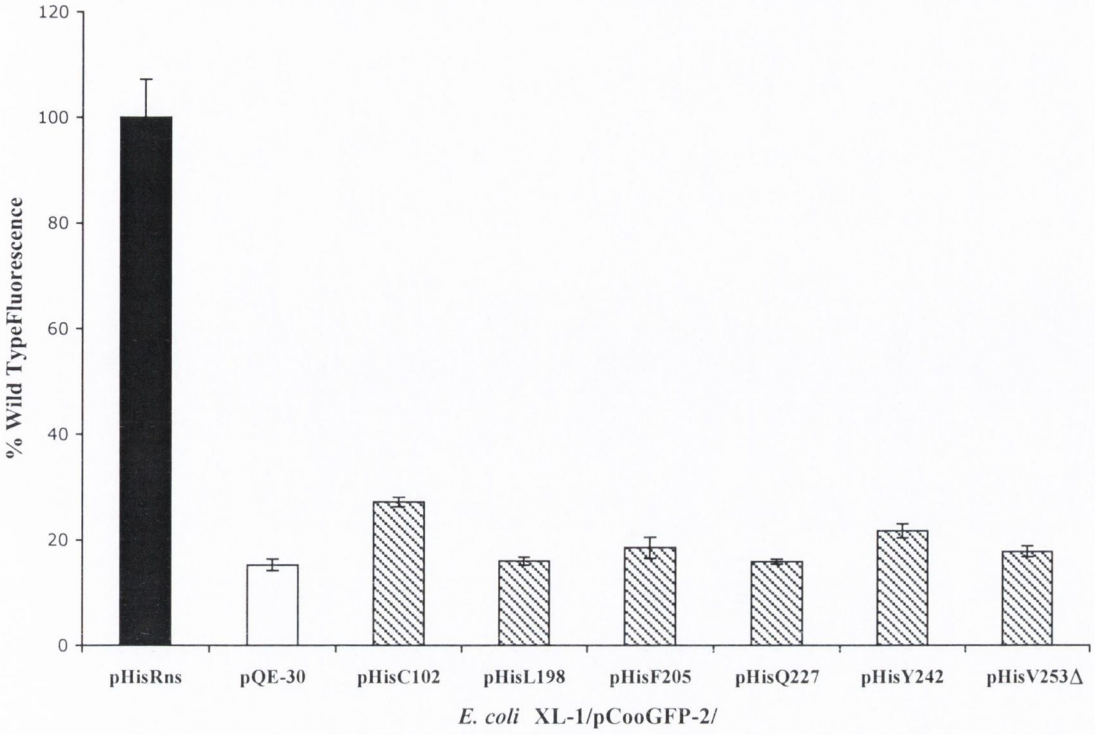

Fig. 3.10. Transactivation of the *coo* promoter by His-tagged Rns and mutant derivatives. Fluorescence levels of the indicated bacterial cultures (see horizontal axis) were expressed as a percentage of the level of fluorescence in cultures containing wild type Rns. Data represents averages of triplicate measurements of duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown.

The results obtained for cultures containing the pHisRns derivatives were unusual. The transformation efficiency had been low, the resultant colonies had been smaller than usual, and on SDS-PAGE analysis CooA was not even detected in the samples prepared from cultures containing the positive control pHisRns (results not shown). The reason for these unusual findings is likely due to the powerful T5 promoter present in pHisRns plasmids and their derivatives. The extremely high transcription rate initiated at this promoter is only repressed and regulated by the presence of high levels of the *lac* repressor encoded by *lacI*. The *E. coli* XL-1 strain carries the $lacI^q$ mutation that results in overproduction of the *lac* repressor, yet even in that strain the level of basal transcription that occurred in the absence of induction with IPTG was sufficient to perform fluorimetric analysis using the pHisRns based constructs. Although the exact genotype of the ETEC strain LMC10 is not known, it is described as a *lac* deletion strain therefore it is likely that it does not contain the *lacI* gene. If the T5 promoter is not repressed during transformation or cell growth, even low levels of expression from the promoter can reduce the transformation efficiency and select against the desired transformants. The plasmids in transformants that do grow often contain deletions and mutations. Therefore it was not possible to work with the pHisRns derivatives in LMC10.

The pSS2192 derivatives were tolerated in the LMC10 strain; therefore it was possible to reliably assess the ability of the proteins encoded by these derivatives to activate CS1 expression (Fig. 3.11). As expected, no CooA was present in the heat shock of the LMC10 culture containing the negative vector control pCL1920 while a large amount of CooA had been produced in the culture containing the positive control pSS2192. CS1 expression in the presence of the RnsC102 mutant was much lower than in the presence of wild type Rns. As determined by densitometry, the amount of CooA in the heat shock of the culture expressing RnsC102 was five-fold less than that in the sample expressing wild type Rns. This confirms the earlier result that the pentapeptide insertion at C102 reduced, but did not eliminate, the activity of Rns at the *coo* promoter.

The SDS-PAGE analysis demonstrated that the two truncates tested, RnsS239Δ and RnsV253Δ, and all of the mutant proteins with insertions in or near to either of the predicted HTH motifs of Rns failed to activate CS1 expression. A band corresponding to CooA was not detectable in the heat shocks of the LMC10 cultures in which they were contained. This verifies the fluorimetric analysis that found these mutants to be inactive at the *coo* promoter.

The level of CooA expression in *E. coli* LMC10 containing the *rns*45 mutant was determined by analysing whole-cell lysates of overnight cultures by SDS-PAGE (results not shown). This mutant had been found in the fluorimetric analysis to cause highly elevated levels of expression from the *coo* promoter. While CooA protein was present in the lysate of the LMC10 culture containing *rns*45, it was found by densitometry to be equivalent to the amount present in the lysate of the culture containing wild type Rns. The likely explanation for this is that the LMC10 cells had reached saturation levels of CS1 expression.

3.2.2.5 Binding of His-tagged Rns to *coo* promoter DNA

While the fluorimetric analyses conducted with the reporter plasmids pCooGFP and pCooGFP-2 had been useful to study the activity of Rns and its mutant derivatives at the *coo* promoter *in vivo*, it was also desirable to study the binding of Rns to the *coo* promoter *in vitro*. This was done by performing electrophoretic mobility shift assays (EMSAs) with purified Rns protein and a purified fragment of *coo* promoter DNA. In EMSAs, the protein and a labelled DNA probe of interest are incubated together and then subjected to non-denaturing PAGE. If the protein has bound to the DNA probe, it will cause the probe to migrate more slowly in the gel than unbound DNA. This reduced electrophoretic mobility is observed as a "shift" of the DNA band when the labelled DNA probes are detected. To construct the probe for use in EMSAs, a 195 bp region of DNA, comprising the area of the *coo* promoter from bp -196 to bp -2 (relative to the transcription start site) and including the two Rns binding sites previously identified by DNase I footprinting (Fig. 1.6) (268), was amplified by PCR using the 5′-biotinylated primers VM1 and VM2 (Table 2.3) and pCooGFP as a template. In initial EMSAs, the incubation of His-tagged Rns with the *coo* promoter DNA probe was performed in a basic reaction buffer consisting of 10 mM Tris (pH 7.5), 50 mM KCl, 1 mM dithiothreitol and 3 mM magnesium chloride. Using these conditions, specific complexes resulting from protein binding to DNA were occasionally detected, but non-specific protein-DNA complexes that had not migrated beyond the wells of the gel were frequently observed (Fig. 3.12 (A)). The composition of the reaction buffer was then modified in order to increase the specificity of complex formation. The presence of magnesium chloride was found to have no effect on complex formation (results not shown) so it was omitted, while poly(dI.dC) and BSA were added to the reaction buffer.

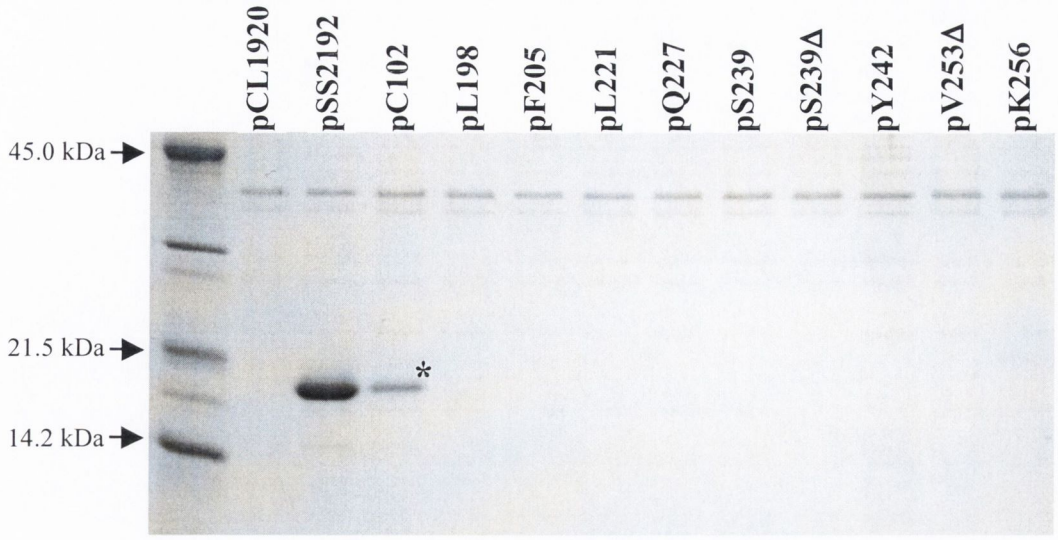

Fig. 3.11. Expression of the CooA fimbrial subunit in the presence of Rns insertion mutants. SDS-PAGE analysis of heat-released extracts of cultures of ETEC strain LMC10 harbouring the indicated plasmids. The positions of selected protein molecular mass markers are indicated on the left. An asterisk indicates the position of bands corresponding to the CooA fimbrial subunit.

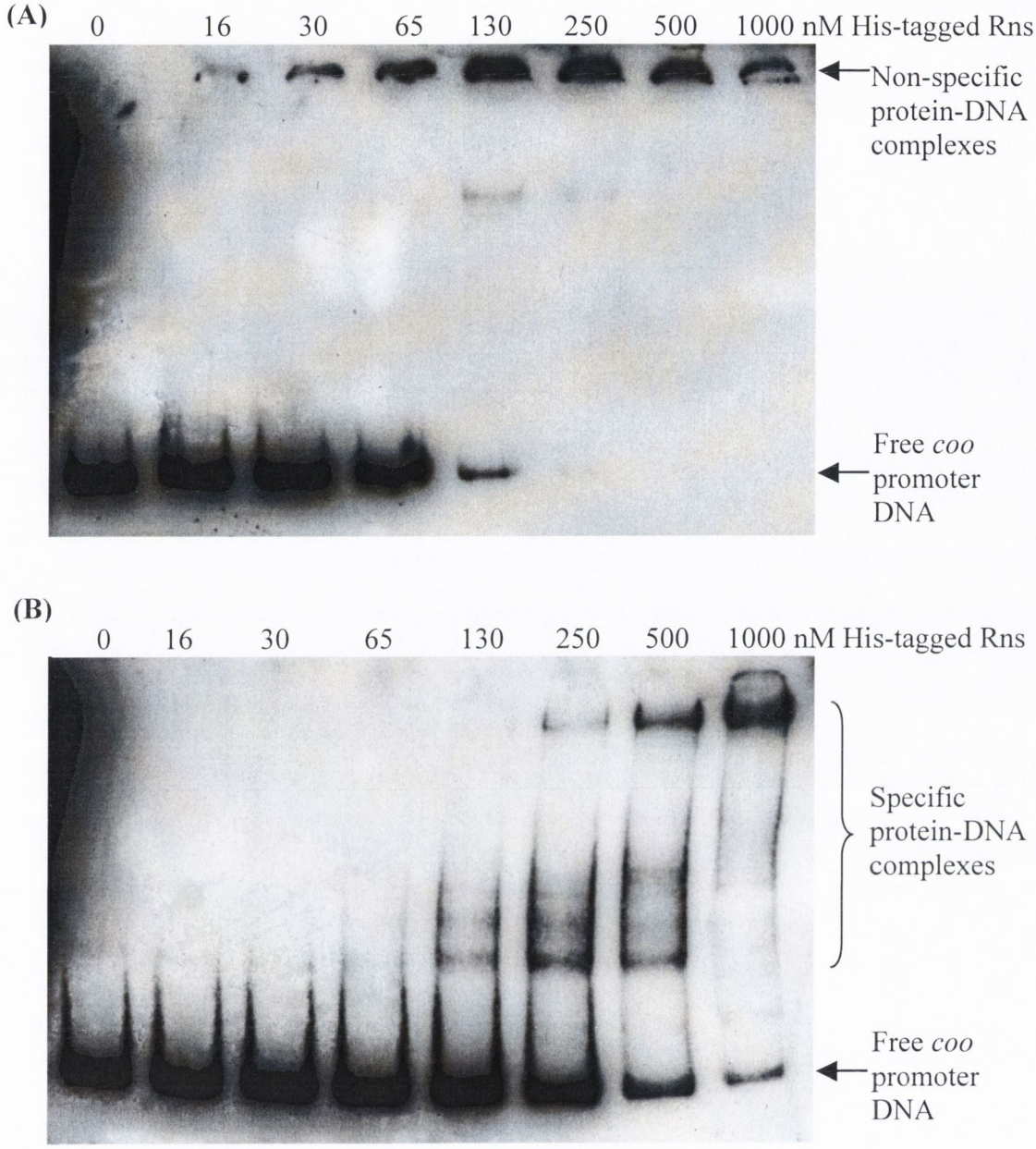

Fig. 3.12. Interaction of His-tagged Rns with *coo* promoter DNA. (A) EMSAs were carried out with approximately 40 pg of a biotinylated *coo* promoter DNA probe and increasing concentrations of His-tagged Rns, as shown across the top of the figure, in the presence of the basic reaction reaction buffer consisting of 10 mM Tris (pH 7.5), 50 mM KCl, 1 mM dithiothreitol and 3 mM magnesium chloride. The unbound DNA fragments and non-specific protein-DNA complexes (that have not migrated beyond the wells of the gel) are indicated on the right. **(B)** EMSAs were carried out with approximately 40 pg of a biotinylated *coo* promoter DNA probe and increasing concentrations of His-tagged Rns, as shown across the top of the figure, in the presence of the basic reaction reaction buffer plus poly (dI.dC) and BSA and without magnesium chloride. The unbound DNA fragments and specific protein-DNA complexes are indicated on the right.

Poly(dI.dC) is a form of non-specific competitor DNA that has been found to reduce non-specific interactions in EMSAs and thus improve the detection of complexes of interest. Similarly, non-specific proteins such as BSA have been found to enhance the formation and detection of specific protein-DNA complexes in EMSAs (216). The resolution of protein-DNA complexes was much improved using these optimised conditions. Several distinct and specific protein-DNA complexes were detected when EMSAs were performed with His-tagged Rns and the *coo* promoter DNA fragment, confirming that Rns binds to this region (Fig. 3.12 (B)). The complexes detected when protein concentrations of 130 nM, and higher, were incubated with 40 pg of DNA could be due to His-tagged Rns occupying each of its binding sites individually at first (i.e. Rns bound to site I alone and Rns bound to site II alone), and then binding to both sites I and II, and then perhaps even greater amounts of Rns binding co-operatively to both sites. In the presence of 1 μM His-tagged Rns, less than 10% of the *coo* promoter DNA probe remained unbound when compared with the amount of free DNA observed in the absence of His-tagged Rns.

Although the result shown in Fig. 3.12 was reproducible, it was found that even using the optimised reaction conditions it was not possible to achieve reliable EMSAs on every occasion. This seemed to be due to lot-to-lot variation in the protein used in the assays. In addition, the DNA binding activity of His-tagged Rns appeared to diminish over time on storage. This would complicate the routine comparison of *coo* promoter binding by His-tagged Rns and the mutant proteins. At this point it was decided to investigate the use of a maltose binding protein (MBP) Rns fusion in EMSAs as an alternative to His-tagged protein.

3.2.2.6 Purification of MBP-Rns and subcloning of pSS2192 based mutations into pMRns5

As an alternative to utilising His-tagged protein, the fusion protein MBP-Rns was employed. This protein was expressed from pMRns5 (Table 2.2 and Fig. 3.13 (A)) (373). The maltose binding protein, MBP, has been found to improve the solubility of target proteins when fused to their amino-terminus (203). This is extremely useful for analysing AraC family proteins, which are notoriously insoluble (139). The fusion proteins can then be conveniently purified on an amylose resin affinity column (205). Additionally, it has been shown that the fusion of MBP to the amino terminus of Rns

does not interfere with the activity of the protein *in vivo* or *in vitro* (268). SDS-PAGE analysis of induced cultures of *E. coli* KS1000/pRare containing the pMRns5 plasmid revealed one major inducible protein corresponding to MBP-Rns (Fig. 3.13 (B)). This protein was not present in cultures of *E. coli* KS1000/pRare containing the vector control pMALc-2. MBP-Rns was purified on an amylose resin column and eluted in several fractions that were also analysed by SDS-PAGE. Purified MBP-Rns could be reliably stored long term in single use aliquots at –80°C.

At this point it was decided to focus on two of the pentapeptide insertion mutants. The RnsQ227 mutant was chosen as a representative of the mutants with insertions in the vicinity of the predicted HTH regions, and the RnsC102 mutant was chosen as both its insertion site and its status of retaining a partial level of activity were unique. DNA fragments containing the 15 bp insertions encoding these mutations were isolated by restriction digest from their pSS2192 derivatives and inserted into the pMRns5 background. This was done by excising from pMRns5 the 302 bp fragment that results from *Nsi*I digestion and replacing it with the equivalent fragment produced by *Nsi*I digestion of pC102 to create pM-C102 (Table 2.2). A diagnostic restriction digest was conducted on pM-C102 to verify that the newly inserted fragment was in the correct orientation. The insertion-containing fragment of pQ227 was isolated by digesting the plasmid with *Mfe*I and *Xba*I, and was then ligated into a pMRns5 backbone from which the equivalent fragment had also been removed by digestion with *Mfe*I and *Xba*I. The resulting plasmid was named pM-Q227 (Table 2.2).

The plasmids pM-C102 and pM-Q227 were transformed into *E. coli* KS1000/pRare and it was confirmed that it was possible to over-express and purify MBP fusions of both mutant proteins. Further analysis of RnsQ227 is discussed in the following sections of this chapter, while further study of RnsC102 is discussed in Chapter 4.

3.2.2.7 Investigation of the interaction of MBP-Rns and MBP-RnsQ227 with the *coo* promoter

Binding of Rns, and the mutant RnsQ227, to the *coo* promoter was again analysed by performing EMSAs using the biotinylated 195 bp fragment of *coo* promoter DNA described earlier as the target. On this occasion, however, the protein used was purified MBP-Rns and MBP-RnsQ227. In addition, the reaction conditions were further optimised by including the non-ionic detergent Nonidet P-40 (NP-40) in the reaction

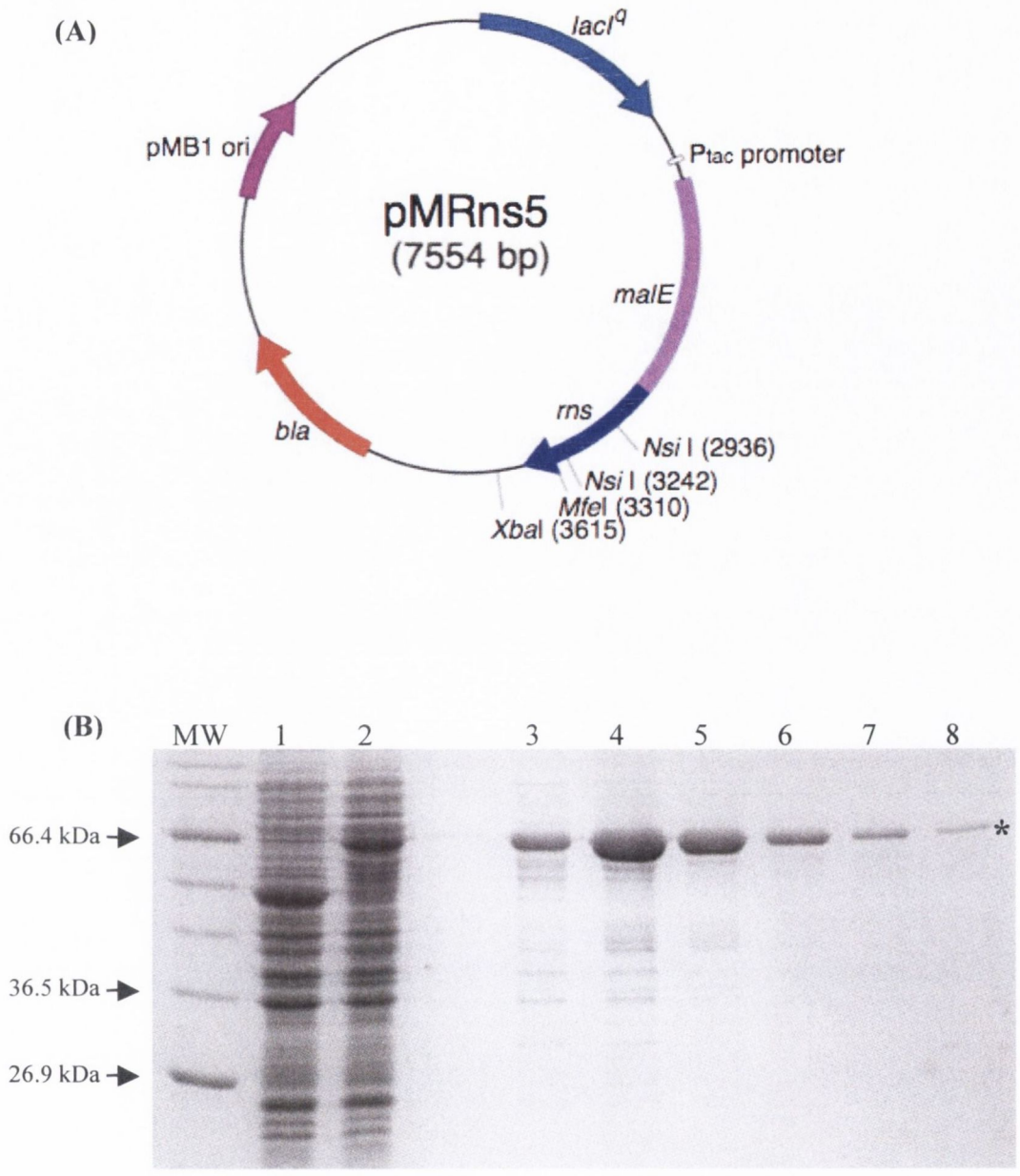

Fig. 3.13. Structure of pMRns5 and expression of MBP-Rns fusion protein. (A) Map of plasmid pMRns5. The position of the *malE-rns* gene fusion is marked. Other relevant features and restriction endonuclease sites are indicated. **(B)** SDS-PAGE analysis of cultures of *E. coli* KS1000/pRare harbouring pMALc2 (lane 1) or pMRns5 (lane 2) grown in Overnight Express medium, and of fractions of purified MBP-Rns eluted from an amylose resin column (lanes 3-8). The positions of selected protein molecular mass markers (MW) are indicated on the left. An asterisk on the right indicates the position of MBP-Rns.

buffer at a final concentration of 0.025% (v/v). The inclusion of non-ionic detergents has been found to increase the affinity of protein-DNA binding in EMSA reactions (171). When EMSAs were performed with approximately 20 pg of biotinylated *coo* promoter DNA and a range of concentrations of MBP-Rns, three major protein-DNA complexes were detected (Fig. 3.14). Formation of the first complex was apparent even in the presence of a concentration of MBP-Rns as low as 3 nM. The amount of unbound DNA probe was vastly reduced in the presence of 34 nM MBP-Rns and at this protein concentration the second, higher molecular weight complex was also formed. These complexes may result from Rns occupying binding sites I and II at the *coo* promoter. This phenomenon is examined further in Chapter 5. The DNA probe was completely bound in the presence of 171 nM MBP-Rns. At the highest protein concentration tested, 514 nM, the first complex was no longer present and a third complex of very low mobility appeared. This low mobility complex was also detected when EMSAs were performed with 514 nM MBP-RnsQ227 (Fig. 3.14). However no other complexes were formed by MBP-RnsQ227. The amount of unbound DNA probe that remained in the presence of 171 nM MBP-RnsQ227 was equal to the amount of free probe in the protein free control (as determined by densitometry). Therefore the mutant RnsQ227 does not appear to bind to *coo* promoter DNA. This implies that the inactivity of RnsQ227 at the *coo* promoter is caused by the inability of the mutant to bind there. This lack of binding could result from the pentapeptide insertion at position Q227 disrupting the second HTH motif of Rns. It is likely that the low mobility complex formed in the presence of 514 nM MBP-Rns and MBP-RnsQ227 is due to non-specific interactions of the probe when proteins are present in high concentrations.

3.2.2.8 Measurement of RnsQ227 activity at the *rns* promoter

In addition to activating expression from the *coo* promoter, Rns also positively regulates its own promoter (134). To assess the ability of the mutant RnsQ227 to activate the *rns* promoter, a plasmid containing a transcriptional fusion of the *rns* promoter to the *lacZ* reporter gene (encodes β-galactosidase) was constructed. This was accomplished using the vector pRW50 (Table 2.2), a 16.9 kb low-copy-number plasmid that carries a promoterless *lacZ* gene downstream from a multiple cloning site. The reporter plasmid was constructed in a two-step process to reduce the complications involved in working with such a large vector. A 728 bp region of DNA, containing the

rns promoter and comprising the 709 bp located immediately upstream of the *rns* ORF, in addition to the first 19 bp of the ORF, was amplified by PCR using the primers New_Prns-EcoRI and Prns-HindIII (Table 2.3) and the plasmid pSS2192 as a template. The primers New_Prns-EcoRI and Prns-HindIII have *Eco*RI and *Hin*dIII sites incorporated into them, respectively. The PCR product was digested with *Eco*RI and *Hin*dIII and ligated into the pBSKII+ vector also digested with the same enzymes. The resulting plasmid, named pPrns2 (Table 2.2), was sequenced to confirm the DNA sequence of the insert. The *rns* promoter-containing insert of pPrns2 was isolated by digesting the plasmid with *Eco*RI and *Hin*dIII and was then ligated into the pRW50 vector digested with the same enzymes to create the reporter plasmid pRnsLacZ-2 (Table 2.2 and Fig. 3.15 (A)). In pRnsLacZ-2, the *rns* promoter controls transcription of the *lacZ* gene, therefore the level of β-galactosidase activity in cells containing this reporter plasmid is a measure of the activity of the *rns* promoter.

The pRnsLacZ-2 plasmid, along with pMALc2, or pMRns5, or pM-Q227, was transformed into the ETEC *lac* deletion strain LMC10. The levels of β-galactosidase activity in each of these cultures were measured as described in Section 2.4.1 (Fig. 3.15 (B)). The levels of β-galactosidase activity in cells containing the positive control pMRns5 were consistently 1.6-fold higher than the levels in cells containing the negative empty vector control pMALc2. This is due to positive regulation of the *rns* promoter by the MBP-Rns protein encoded by pMRns5. However, the β-galactosidase activity levels in cells containing pM-Q227 were similar to those in cells containing pMALc2. Therefore, in addition to its inactivity at the *coo* promoter, the mutant RnsQ227 appears to be unable to activate transcription from the *rns* promoter.

3.2.2.9 Interaction of MBP-Rns and MBP-RnsQ227 with the *rns* promoter

To determine if the lack of activation of the *rns* promoter by RnsQ227 is due to an inability of the mutant protein to bind there, EMSAs were performed using a DNA probe containing the *rns* promoter region. The probe was produced by PCR amplification of a 434 bp area of *rns* promoter DNA using the 5′-biotinylated primers VM3 and VM4 (Table 2.3) and the plasmid pSS2192 as a template. This 434 bp fragment of DNA (bp -305 to bp +129 with respect to the transcription start site of *rns*) includes the three sites shown, by DNase I footprinting, to be bound by Rns (Fig 1.7) (269). Three protein-DNA complexes were observed when EMSAs were performed

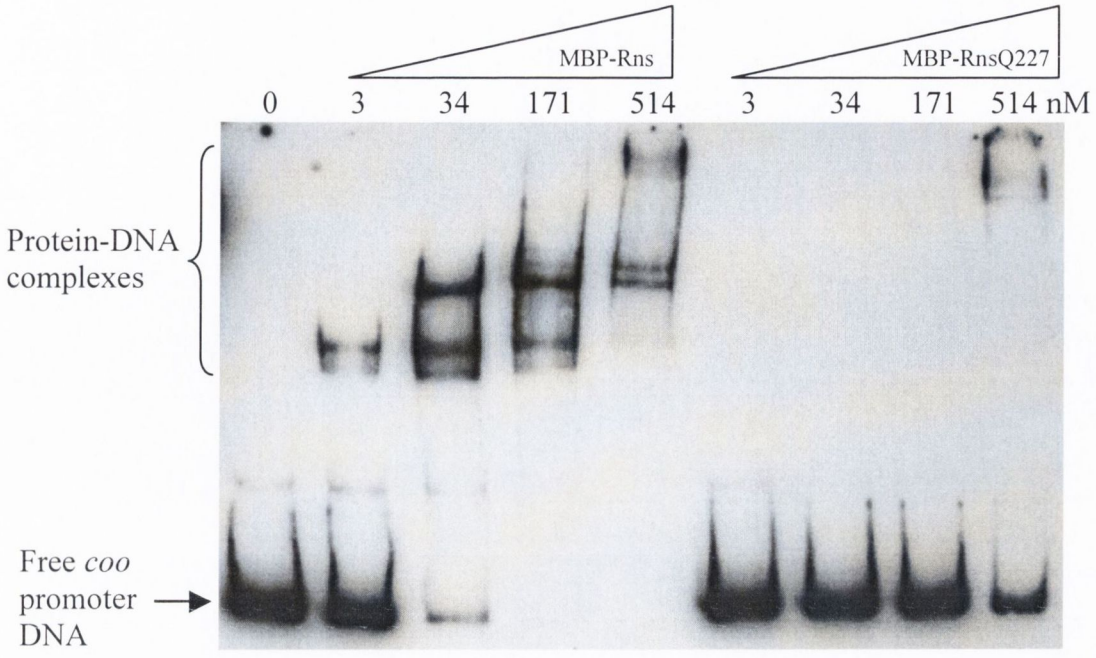

Fig. 3.14. Interaction of MBP-Rns and MBP-RnsQ227 with *coo* promoter DNA. EMSAs were carried out with approximately 20 pg of a biotinylated *coo* promoter DNA probe and a range of concentrations of MBP-Rns or MBP-RnsQ227 as indicated at the top of the figure. The unbound DNA fragments and protein-DNA complexes are indicated on the left.

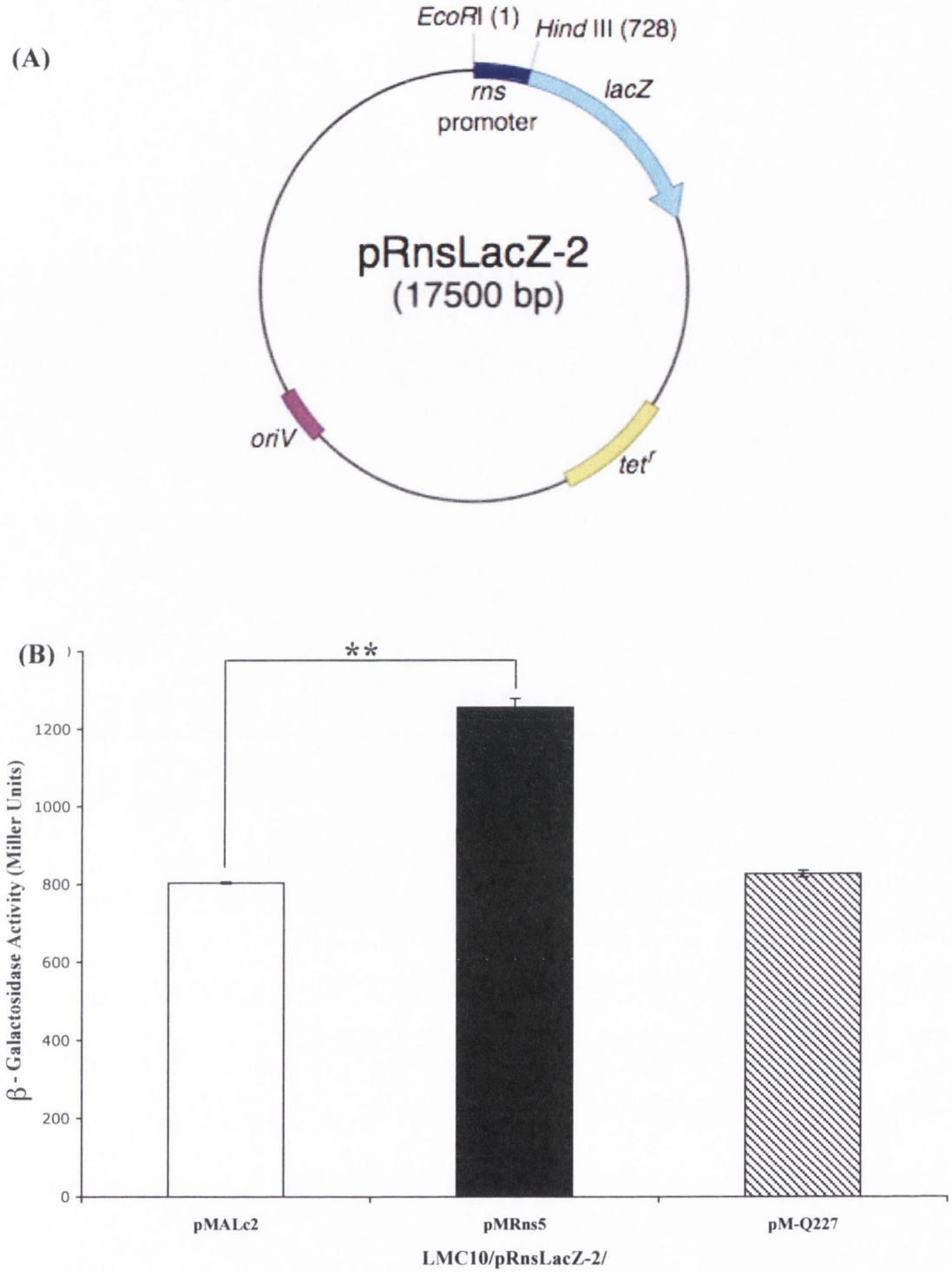

Fig. 3.15. Structure of pRnsLacZ-2 and regulation of the *rns* promoter. (A) Map of plasmid pRnsLacZ-2. The position of the *rns* promoter-*lacZ* fusion is marked. Other relevant features and restriction endonuclease sites are indicated. **(B)** The β-galactosidase activities of the indicated bacterial cultures (see horizontal axis) were measured. Data represents averages of duplicate assays on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

with this probe and a range of concentrations of MBP-Rns (Fig. 3.16 (A)). Formation of the first, highest mobility, complex was detected when only 3 nM of MBP-Rns was present. When the concentration of MBP-Rns in the binding reaction was increased to 17 nM even more of this first complex was formed in addition to the formation of a second lower mobility complex. When 68 nM of MBP-Rns was added no DNA probe remained unbound and a third complex of even lower mobility was formed while the presence of the first complex had reduced dramatically. In the presence of 171 nM MBP-Rns only the third, lowest mobility, complex was detected. These results confirm that Rns binds to three sites in the area of DNA that was used. It can be postulated that the complexes are a result of Rns binding to one site initially, then at higher protein concentrations to two sites and finally binding to all three sites at even higher protein concentrations.

Protein-DNA complexes were not detected when the EMSAs were repeated with MBP-RnsQ227 (Fig. 3.16 (B)). The DNA probe remained unbound even in the presence of 171 nM MBP-RnsQ227. Thus the pentapeptide insertion after position Q227 of Rns results in a mutant protein that has lost the ability to bind to DNA at the *rns* promoter region and therefore it cannot activate transcription from the promoter.

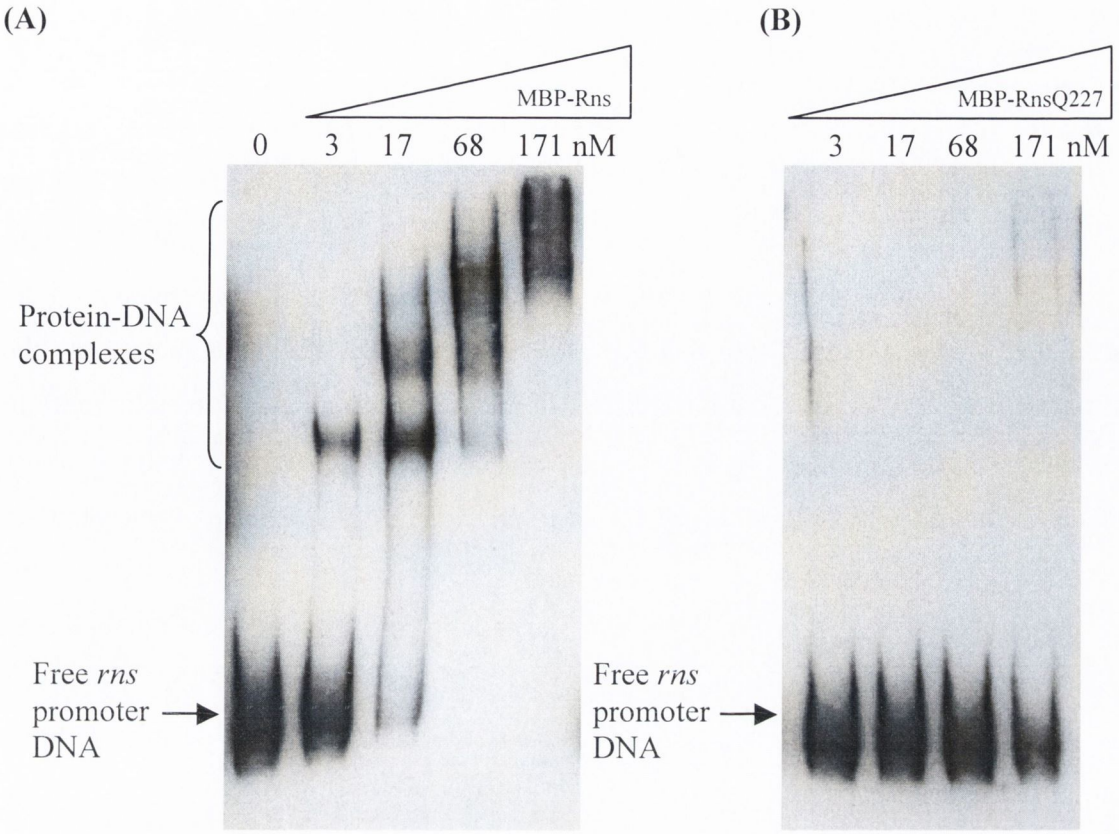

Fig. 3.16. Interaction of MBP-Rns and MBP-RnsQ227 with *rns* promoter DNA. EMSAs were carried out with approximately 20 pg of a biotinylated *rns* promoter DNA probe and a range of concentrations of MBP-Rns **(A)** or MBP-RnsQ227 **(B)** as indicated at the top of the figures. The unbound DNA fragments and protein-DNA complexes are indicated on the left.

binding to three sites near the *rns* promoter (269). The 15 bp insertion in *rns*45 occurs within the second of these sites. This raises the possibility that Rns binds there during autoactivation to displace a repressor. Therefore the 15 bp insertion may have disrupted the repressor binding site, eliminating the need for Rns binding and causing the apparent upregulation of Rns expression seen for *rns*45. An issue with this explanation is that the authors of the report that identified these binding sites suggest that the region downstream of the *rns* promoter does not appear to contain any repressor binding sites (269). However, they also state that the role of binding site 2 is unclear and additional investigation is needed to better understand it. The effect of the 15 bp insertion in this area adds a further element of complexity to this already unusual regulatory system.

Of the 14 Rns mutants containing pentapeptide insertions in the N-terminal domain of the protein, 10 were found by the fluorimetric analysis to tolerate the insertion as they retained 76-178% of wild type activity at the *coo* promoter. A similar result was found for the Rns-related AraC family member PerA of EPEC; several mutations in the N-terminal region of this protein also had little effect on its activation function (311). The most striking similarity of the PerA analysis to this study is that the insertion after position K108 of Rns did not affect protein activity, while mutation of the corresponding conserved residue K108 of PerA was also found to have no effect on protein function (311). Therefore this area of Rns, and related proteins, appears to be tolerant of alteration.

The insertions after positions E10, T11, I17, S127 and N130 of Rns are in regions of the protein that are not predicted to be either α-helices or β-strands. These mutants remained functional at the *coo* promoter, thus supporting previous reports that insertions in unstructured areas of proteins are less deleterious than those in regions of secondary structure (175). However, the insertions in the functional mutants RnsK13 and RnsI34 occurred in the first and second predicted β-strands, and the insertions after positions L118 and L119 of Rns are in the second predicted α-helix of Rns. Therefore it appears that these disrupted secondary structure elements may not be required for Rns to activate expression from the *coo* promoter. Mutations in an equivalent helical region of PerA were also found not to affect protein function (311).

As so many insertions in the β-strand rich N-terminal region of Rns were tolerated, it is tempting to speculate that this area of the protein is insensitive to disruption. However, while the activity of most N-terminal mutants was essentially unaffected, four

mutants were isolated with insertions in this region that adversely affected their function. RnsC102 retained a reduced ability to activate expression from the *coo* promoter in the GFP reporter plasmids and in ETEC strain LMC10 and is discussed further in Chapter 4. A more serious effect was caused by the insertions after T23, C30 and L51 of Rns. Although these mutant proteins were expressed, they did not activate the *coo* promoter. These insertions may have affected essential secondary structural elements of Rns. Rns is predicted to have β-strands in the areas around residues T23 and L51 (Fig. 3.4). Several mutations in predicted β-strand regions in the N-terminus of ToxT of *V. cholerae* also rendered the protein inactive (57) while mutations within the β-strand rich N-terminal region of PerA severely compromised protein function (311). It is possible that the area of Rns disrupted by the insertions at T23, C30 and L51 forms a structure that is critical for some aspect of transcriptional activation, possibly even forming contacts with RNA polymerase. Unfortunately additional analysis was restricted by an inability to purify the proteins. While little is known about the structure or function of the N-terminal domain of this class of AraC-like proteins, a recent study found it to be crucial for Rns activity. Deletion analysis of Rns revealed that in the absence of its N-terminal 60 amino acids the protein was unable to activate or repress a selection of Rns-regulated promoters *in vivo*, and removal of the first 80 amino acids of Rns prevented it from binding DNA *in vitro* (19). The study also isolated two N-terminal point mutants, RnsI14T and RnsN16D that were unable to activate the *rns* promoter. The authors suggest that these residues may be involved in an interaction between the N- and C-terminal domains of Rns and that the DNA binding ability of Rns may depend on such an interaction (19). It is possible that the pentapeptide insertions in the mutants RnsT23, RnsC30 and RnsL51 could have disrupted this proposed interaction. The inability of these mutants to activate the *coo* promoter provides further evidence that regions of the N-terminal domain of Rns are critical for its function.

A single mutant, RnsS158F, was isolated following random mutagenesis. This could be due to the fact that the area of the *rns* gene encoding the N-terminal half of the protein alone was randomly mutagenised and only non-functional mutants were screened for. As described above, the N-terminal domain of Rns was found to be largely tolerant of pentapeptide insertions so it is possible that very few point mutations in this region would cause the protein to become inactive. Also, only mutants encoding full length Rns protein were selected for further analysis. The *rns* gene contains numerous stretches of A and T residues so it is likely that many point mutations could

have introduced a premature stop codon. Indeed, during screening it was found that a large number of the mutants were truncates. Nevertheless, the RnsS158F mutant was confirmed to be expressed as a full-length protein and to be unable to activate expression from the *coo* promoter. The serine at position 158 of Rns is very well conserved in closely related AraC family proteins including CfaD and CsvR of ETEC, AggR of EAEC, VirF of *S. flexneri* and PerA (Fig. 1.8). Additionally it lies in one of the well-conserved regions of the N-terminal domain. Therefore it is likely that this area is important in the functioning of these proteins and the effect of the mutation at S158 of Rns on protein activity confirms this. The mutation is a non-conservative substitution from a serine residue with an uncharged polar side chain to a phenylalanine residue with a non-polar bulky aromatic side chain. A similar mutation in the same area of VirF, S150A, also abolished activity of that protein (309). It is possible that the S158F mutation in Rns eliminated an interaction made by the wild type residue or that this area is key to the structural integrity of Rns. It was not possible to purify the mutant protein to examine it further.

In all seven incidences of pentapeptide insertions within the C-terminal domain of Rns, the insertion had a severe negative effect on the ability of the protein to activate expression from the *coo* promoter in the GFP reporter plasmids and in the ETEC strain LMC10. It could be confirmed that each of these mutants were transcribed but expression of the mutant proteins from their pSS2192-derived plasmids could not be demonstrated. As this lack of demonstrable expression could have been caused by the *rns* autoactivation mechanism enhancing the negative effects of the mutations when the proteins were expressed from their native promoter, several were re-cloned downstream of the T5 promoter in the pHisRns plasmid background. It was then possible to show that the mutant proteins were expressed and a reassessment of their ability to activate a *coo* promoter-*gfp* fusion confirmed that they were non-functional. Most of the proteins had reduced solubility compared to wild-type Rns and therefore could not be purified.

One of the C-terminal insertions had occurred in the first of the predicted HTH motifs of Rns. The insertion after position L198 is likely to have disrupted the recognition helix of this potential HTH domain and therefore eliminated its function, i.e. it may have prevented the DNA binding that this area of Rns is predicted to be involved in. Mutation of the corresponding neighbouring residue, E196, in VirF also reduced the activity of this protein (309). The importance of HTH1 in the function of AraC family proteins has been well documented. Mutations of several residues within the first HTH

of AraC itself reduced the ability of the protein to bind to, and activate transcription from, the *araBAD* promoter (42, 53, 133). Mutations in HTH1 of the Rns-related regulators VirF and PerA also reduced or abolished the activity of these proteins (309, 311) while DNA binding by the AraC family members MarA and SoxS was especially sensitive to alterations in the first HTH motif (143, 159). Thus, the loss of function of RnsL198 is unsurprising and provides further confirmation of the importance of HTH1.

The pentapeptide insertions in the mutants RnsF205 and RnsL221 had occurred in a predicted long α-helix situated between HTH1 and 2 of Rns. An equivalent helix was observed in the crystal structure of the AraC family protein Rob where it was found to fix the relative orientations of its two HTH motifs (221). Therefore it is possible that insertions in this potential α-helix of Rns could alter DNA binding indirectly by disrupting the conformation of the protein such that the potential DNA-binding helices are no longer in the proper alignment with the DNA. The mutant proteins would then no longer be able to bind to and activate the *coo* promoter. Mutations in the corresponding predicted α-helix of PerA also led to an inactive protein (311). An alternative explanation for the inactivating effect of the insertions at F205 and L221 is that a coiled coil structure is predicted between residues 195 and 222 of Rns by the COILS program. Coiled coil domains are a common structural element involved in protein oligomerisation (240). If this area of Rns is involved in oligomerisation, which Rns may require for its activity, then disruption of this potential coiled coil could cause the protein to be non-functional. The possible oligomerisation of Rns is discussed further in Chapter 4.

The inactivity of Rns proteins with insertions in the predicted second HTH motif implies that this region has an important function. Three insertions in the area of HTH2 were characterised. The insertion in RnsS239 occurs at the end of the turn of the second HTH, an area that typically does not tolerate distortions (10). The insertion in RnsY242 occurs within the recognition helix of the HTH motif. Both of these pentapeptide insertions are likely to have greatly disrupted these critical elements of HTH2 and thus prevent the DNA interaction this region of Rns is predicted to be involved in. In each case the residue after which the insertion occurs is well conserved among the Rns-related regulators. Mutation of the residue corresponding to S239 in VirF reduced protein function while mutation of the residues equivalent to Y242 in both VirF and PerA also completely abolished protein activity (309, 311). The insertion in RnsQ227 was immediately upstream of HTH2. As it was possible to purify this mutant protein it

was chosen for further analysis as a representative of the HTH insertion mutants. RnsQ227 was shown to be incapable of activating the *coo* promoter in the GFP reporter plasmids and in ETEC strain LMC10, and to also have no activity at the *rns* promoter, therefore the insertion had affected a vital aspect of Rns function. When EMSAs were performed, wild type Rns was shown to bind both *coo* and *rns* promoter DNA fragments, forming several complexes reflecting its multiple binding sites at each. RnsQ227 did not bind at either promoter thus confirming the prediction that HTH2 of Rns is required for DNA binding and the inactivity of this HTH mutant at least, is due to the insertion disrupting this binding.

The insertion in the mutant RnsK256 had occurred in the final predicted α-helix of the protein, downstream of HTH2. Mutations in the equivalent region of the AraC family member XylS also rendered the protein inactive (243). This suggests an important function for this region at the C-terminus of the protein. It is possible that, like the α-helix between the two HTH motifs, it is required to stabilise or correctly orientate these predicted DNA binding domains and is therefore intolerant of disruption.

All truncates isolated during this study were unable to activate the *coo* promoter. This is unsurprising for RnsI17Δ and RnsH20Δ as over 90% of the protein is missing in each, and for RnsI192Δ and RnsI195Δ as most of the predicted DNA binding domain has been lost. These truncates were not analysed further. The truncation in RnsS239Δ occurred just after the turn in HTH2. That this mutant was unable to activate expression from the *coo* promoter is further proof that HTH2 is critical for Rns activity. However it must be considered that the loss of 26 amino acids in this truncate may disrupt the folding of the entire C-terminal domain of Rns, thus causing the loss of function. The fact that RnsV253Δ was inactive is intriguing as only the last 12 residues of Rns are missing in this truncate. This further supports the theory that the final predicted α-helix downstream of HTH2 plays an essential role in the functioning of Rns.

In summary this mutational analysis has defined regions critical for the activity of the Rns protein. The predominantly α-helical C-terminal domain was found to be vital for Rns function and disruption of the predicted HTH motifs was demonstrated to prevent DNA binding. In contrast, the β-strand rich N-terminal domain of Rns was largely tolerant of insertions. These results have both provided novel information on the structure-function relationship of Rns and supported the findings of similar studies on related proteins.

Chapter 4 Rns is a dimer and requires a central flexible region for full activity

4.1 Introduction

It is a well established concept that a folded, three-dimensional structure is a prerequisite for protein function. However, there is increasing recognition that this view may need to be re-assessed as many proteins contain areas that are predicted to be disordered yet carry out important functional roles (434). There is no standard definition of protein disorder, but disordered areas can be described as non-globular regions lacking regular secondary structure and displaying a high level of flexibility in the polypeptide chain (434). Many proteins containing such regions are involved in molecular recognition and their disordered segments have been found to undergo a transition to an ordered state upon binding their ligand or DNA target (102). For other proteins the function of these regions is dependent on their disorder, for example they act as flexible linkers (99). Flexible linkers are important for connecting separate functional domains of proteins, and may play an active role in the positioning of the domains relative to each other. Alternatively these linkers can simply act as passive spacers, tethering adjacent domains within a certain distance of one another. Biochemical and genetic techniques were used to determine that amino acids 171-177 within an unstructured area of AraC act as a linker between the protein's N-terminal domain (required for dimerisation and ligand binding) and its C-terminal domain (involved in DNA binding and transcription activation) (119). Recently, similar linkers were also identified in the AraC family members RhaS and RhaR (212). The linkers of all three proteins were subjected to mutational analysis (120, 212). The results revealed that the identity of the individual amino acids within the linkers is not critical and that the linkers do not play a direct role in the activity of the proteins but merely serve to flexibly connect their two domains (120, 212).

The subfamily of AraC-like proteins involved in the regulation of stress response, such as MarA and Rob, are monomers while several of the members of the AraC family that are involved in regulating the metabolism of sugars are active as dimers (248). The AraC protein itself functions as a dimer in solution in both the presence and absence of arabinose (430), and it also binds to DNA as a dimer (176). The crystal structure of the dimerisation/arabinose-binding domain of AraC revealed that the protein dimerises via an antiparallel coiled coil formed between an α-helix located at the end of the N-terminal domain of each monomer (378). The ends of this coiled coil are anchored by three critical leucine residues that pack together in a "knobs-into-holes" manner (378).

This mode of packing is the defining characteristic of coiled coil domains (240). Mutagenesis revealed that corresponding conserved leucines in UreR and XylS are also critical for the dimerisation of these AraC family members (306, 339).

It is not yet known whether the AraC family members that regulate virulence act primarily as monomers or oligomers (248). Amongst the Rns-related regulators it has been suggested, based on genetic evidence, that VirF of *S. flexneri* is a dimer (309) and PerA is a monomer (193) while a recent report suggested that it was not possible to detect dimerisation of the Rns N-terminal domain *in vivo* or of full-length Rns *in vitro* (19).

Many different approaches, including the use of gene fusion systems, chemical cross-linkers and gel filtration chromatography, are available to monitor protein-protein interactions and determine whether a protein of interest oligomerises. Gene fusion systems typically take advantage of the domain structure of transcriptional regulators such as the LexA repressor or the λ phage repressor, cI (69, 82, 184). Both of these repressors contain dimerisation and DNA binding domains and are only functional as dimers. Their native dimerisation domain can be removed and replaced with the protein of interest. If this protein is capable of dimerising it will restore function to the DNA binding domain of the repressor. This is assessed by monitoring the transcription of a reporter gene located downstream of a promoter regulated by the repressor. Another method for assessing the oligomeric state of proteins involves the use of chemical cross-linkers; reagents that react with functional groups on closely associated proteins to form a covalent linkage between them. Cross-linking treatments can be carried out on purified proteins *in vitro* or, if the cross-linker is membrane-permeable, on intact cells expressing the protein of interest. The cross-linking reaction physically links the protein complexes such that they survive separation by SDS-PAGE and can then be detected by Coomassie staining or Western immunoblotting. Thus it can be determined if the protein of interest migrated as a monomer or higher order oligomer. Gel filtration chromatography is a method used to measure the molecular weight of proteins. The elution profile of proteins passed through a gel filtration column can be compared to profiles of protein molecular weight standards to determine if the size of the protein of interest corresponds to the size expected for a monomer or an oligomer.

The aims of the study presented in this chapter were to assess whether there is a role for protein disorder in Rns and to use several different methods to establish if Rns forms oligomers.

4.2 Results

4.2.1 Analysis of the C102 containing region of Rns

4.2.1.1 Site-directed mutagenesis of amino acids C102 and R103 of Rns

As described in Chapter 3, the insertion mutant RnsC102 was found to have retained a partial level of activity at the *coo* promoter. The insertion had occurred in an uncharacterised area of Rns that was not predicted to be α-helical or of β-structure (Fig. 3.4). In addition, the pentapeptide insertion within RnsC102 was not predicted to have introduced an α-helix or β-sheet in the protein. To gain a better understanding of why the insertion had reduced but not abolished protein activity, it was decided to create more subtle point mutations in the vicinity of the C102 residue and assess their effect on Rns function. Alanine-scanning mutagenesis, a technique in which the individual amino acids of a protein are changed to alanine, is an approach that is frequently used to identify residues critical for protein function. Therefore the C102 residue itself and the neighbouring residue R103 were individually changed to alanine (there is already a naturally occurring alanine residue at position 101 of Rns). This was done by performing site-directed mutagenesis on the plasmid pHisRns using the QuikChange kit (Stratagene) and the primer pair C102-F and C102-R or the primer pair R103-F and R103-R (Table 2.3). The *rns* ORF of the resulting plasmids pHisC102A and pHisR103A was sequenced to verify each mutation. Since it was possible to purify the mutant proteins (results not shown) it was confirmed that they were expressed and soluble. The pHisC102A and pHisR103A plasmids were transformed into *E. coli* XL-1/pCooGFP-2 and the resultant colonies were found to be at least as fluorescent as those containing pHisRns (see Section 4.2.1.4 for further analysis). Therefore a pentapeptide insertion after position C102 of Rns resulted in a mutant with reduced levels of activity at the *coo* promoter, while the C102A and R103A point mutations appeared to have no effect on the ability of Rns to activate expression from the *coo* promoter.

4.2.1.2 A predicted area of disorder around position C102 of Rns

Secondary structure predictions could not provide a reason for the effects of the mutations in the area around the C102 residue of Rns. Therefore alternative prediction programmes were employed. GlobPlot (http://globplot.embl.de) is a web-based tool that plots the tendency of protein sequences to be ordered (globular) or disordered (non-globular). It has been found to successfully identify disordered regions in well characterised proteins (230). The primary sequence of Rns was analysed by the GlobPlot programme. In addition to correctly detecting the potential twin HTH motifs in the C-terminal domain of Rns (the defining characteristic of AraC family members), the GlobPlot analysis revealed that the extreme N- and C-termini of Rns and the region encompassing residues 100 to 104 are predicted to be disordered (Fig. 4.1). Spritz (http://distill.ucd.ie/spritz) and DRIP-PRED (http://www.sbc.su.se/~maccallr/disorder), two other computational methods for protein disorder prediction, also predicted that an area of disorder is present around residues 100-104 of Rns (419). When the GlobPlot analysis was performed on the sequence of Rns including the pentapeptide insertion after position C102 however, the disorder in this region was no longer predicted to be present. In contrast, the single amino acid changes of C102A and R103A were not predicted to disrupt the disorder in this area (results not shown). It is possible that for full activity, the Rns protein may have a requirement for disorder in the area including residues 100 to 104. This would explain why the disorder eliminating pentapeptide insertion after C102 caused a reduction in Rns activity while the alanine substitutions at positions C102 and R103, that had no affect on this disorder, were tolerated.

4.2.1.3 Deletion of the NACRS sequence of Rns

To determine the importance of predicted protein disorder for Rns activity a mutant protein was produced in which the amino acid sequence NACRS, that comprises the predicted disordered region between positions 100 to 104, was deleted. A deletion mutant *rns* ORF encoding such a protein was created in both the pHisRns and pMRns5 backgrounds. This was achieved by performing inverse PCR using the divergent primers InvPCR-1 and InvPCR-2 (Table 2.3) that flank the region to be deleted, and the pHisRns or pMRns5 plasmid as a template. The inverse PCR amplification generated linear DNA fragments that lack the nucleotides encoding the NACRS sequence. These linear fragments were ligated to create the plasmids pHisΔNACRS and pM-ΔNACRS

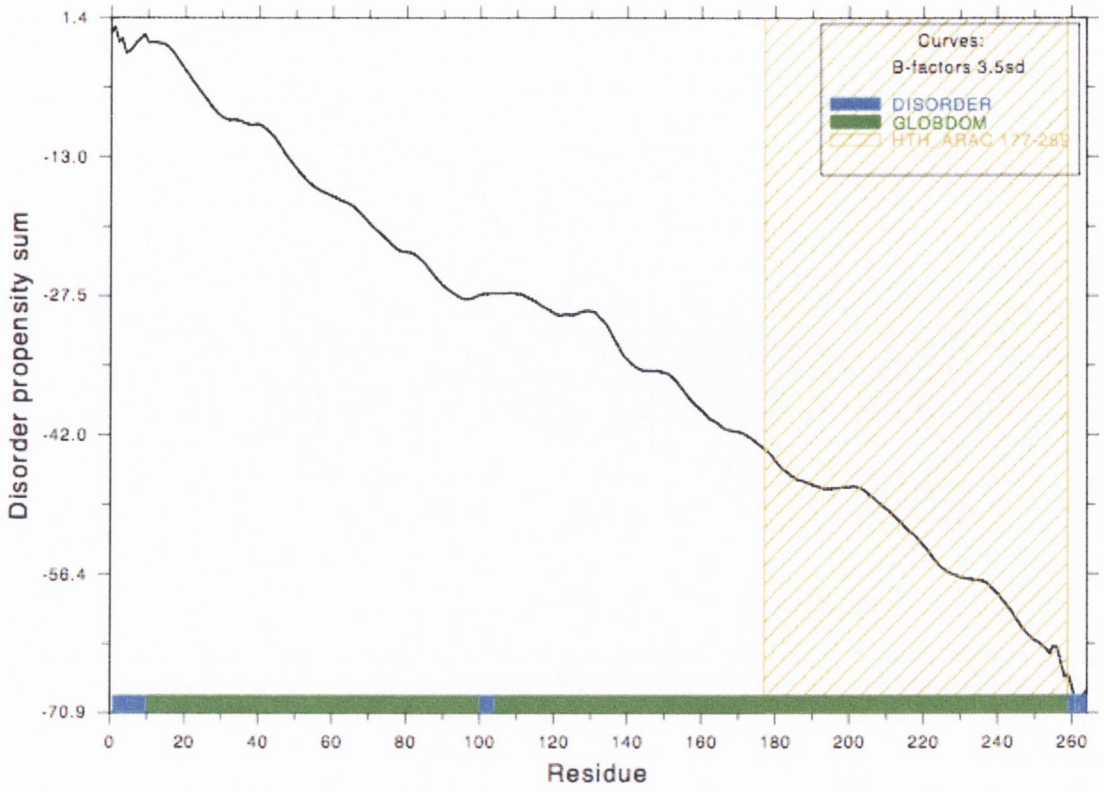

Fig. 4.1. GlobPlot analysis of the amino acid sequence of Rns. The primary sequence of Rns was analysed by GlobPlot, a web service that plots the tendency within a protein sequence for order/globularity and disorder. Putative globular domains (GlobDoms) are shown in green. Disorder propensity is indicated by an upward slope of the curve and putative disordered regions are shown in blue. The characteristic AraC family C-terminal twin HTH motifs are predicted to be present between amino acids 177 and 259 of Rns and are indicated by yellow hatching. GlobPlot predicts three regions of disorder within Rns: amino acids 1-10, 100-104 and 259-265.

(Table 2.2). The *rns* ORF within both plasmids was sequenced to verify that the deletion had occurred. Since it was possible to purify both the His-tagged and MBP fusion forms of the Rns deletion mutants this confirmed that these proteins were expressed and soluble.

4.2.1.4 Activity of the NACRS sequence mutants at the *coo* promoter

Fluorimetric analysis was performed to quantify the ability of the Rns mutants to activate transcription from the *coo* promoter. The fluorescence levels in cultures of *E. coli* XL-1 harbouring both the reporter plasmid pCooGFP-2 and a pHisRns based plasmid encoding one of the mutants were measured as described in Section 2.4.2 (Fig. 4.2). As before, the RnsC102 pentapeptide insertion mutant displayed partial activity. Cultures containing pHisC102 were ~26% as fluorescent as those containing the positive control pHisRns. Cultures containing the alanine substitution mutant R103A were as fluorescent as cultures containing wild type Rns, while the fluorescence levels in cultures containing the mutant C102A were 1.3-fold greater than those seen for wild type Rns, which was not considered significant. These results confirm what had been observed when simply viewing colonies of *E. coli* XL-1/pCooGFP-2 harbouring pHisC102A or pHisR103A. However, the fluorescence levels in cultures containing the deletion mutant encoded by pHisΔNACRS were similar to those in cultures harbouring the *rns* free control pQE-30. Thus single alanine substitutions in the NACRS sequence of Rns did not have an adverse effect on the ability of the protein to activate the *coo* promoter. Conversely, a pentapeptide insertion in this putative disordered area reduced protein activity and deletion of the area eliminated Rns activation of the *coo* promoter.

The reduced activities of the RnsC102 insertion mutant and RnsΔNACRS deletion mutant could be due to deficiencies in their ability to bind to DNA at the *coo* promoter. To determine if this was the case, EMSAs were performed with the *coo* promoter DNA probe and purified MBP fusions of each of the mutants (Fig. 4.3 (A)). MBP-RnsC102 was found to be proficient for binding to the *coo* promoter. The insertion mutant formed the same set of complexes as wild type Rns, however it did so with less affinity (Fig. 4.3 (B)). At a protein concentration of 3 nM there was noticeably less of the first, highest mobility, protein-DNA complex formed by MBP-RnsC102 compared with the amount formed by MBP-Rns. Greater concentrations of MBP-RnsC102 were required to completely bind the DNA probe. Approximately 20% (as determined by

densitometry) of probe remained unbound in the presence of 171 nM MBP-RnsC102 while at the same concentration of MBP-Rns no free probe remained. Finally, MBP-RnsC102 displayed a greater tendency than wild type Rns to form the low mobility non-specific protein-DNA complex. In the presence of 514 nM MBP-RnsC102 only the non-specific complex was formed but at the same concentration of MBP-Rns, in addition to the non-specific complex, one of the specific protein-DNA complexes was still present. In contrast to MBP-RnsC102, MBP-RnsΔNACRS did not specifically bind the *coo* promoter DNA probe. Even at the highest concentration of the deletion mutant tested, the majority of the probe remained unbound and only a small amount of the non-specific complex had been formed. Therefore the EMSA results help to explain the activity levels of the RnsC102 and RnsΔNACRS mutants at the *coo* promoter. The pentapeptide insertion after position C102 of Rns caused the mutant protein to bind to the promoter with less affinity, thus contributing to its reduced activity there. Elimination of the NACRS sequence of Rns has a more deleterious effect; it results in a mutant protein that is incapable of binding to the *coo* promoter and consequently cannot activate transcription.

4.2.1.5 Activity of MBP-RnsC102 and MBP-RnsΔNACRS at the *rns* promoter

The ability of the RnsC102 and RnsΔNACRS mutants to activate the *rns* promoter was also assessed. β-Galactosidase levels in ETEC strain LMC10 cells carrying the pRnsLacZ-2 reporter plasmid along with pM-C102 or pM-ΔNACRS (or control plasmids) were measured (Fig. 4.4). The RnsC102 insertion mutant activated transcription from the *rns* promoter, but to a level lower than wild type Rns. The β-galactosidase expression levels in cells containing MBP-RnsC102 were approximately 1.3-fold less than those in cells containing MBP-Rns ($P < 0.005$). The RnsΔNACRS deletion mutant was completely defective in its ability to activate transcription at the *rns* promoter. The level of β-galactosidase expression in cells harbouring pM-ΔNACRS was equivalent to that in cells containing the negative vector control pMALc2.

To follow up these measurements of *rns* promoter activity *in vivo*, EMSAs were performed to determine the ability of MBP-RnsC102 and MBP-RnsΔNACRS to bind to the *rns* promoter DNA probe *in vitro* (Fig. 4.5 (A)). The results were similar to those of the EMSAs performed with the *coo* promoter DNA probe. MBP-RnsC102 is capable of

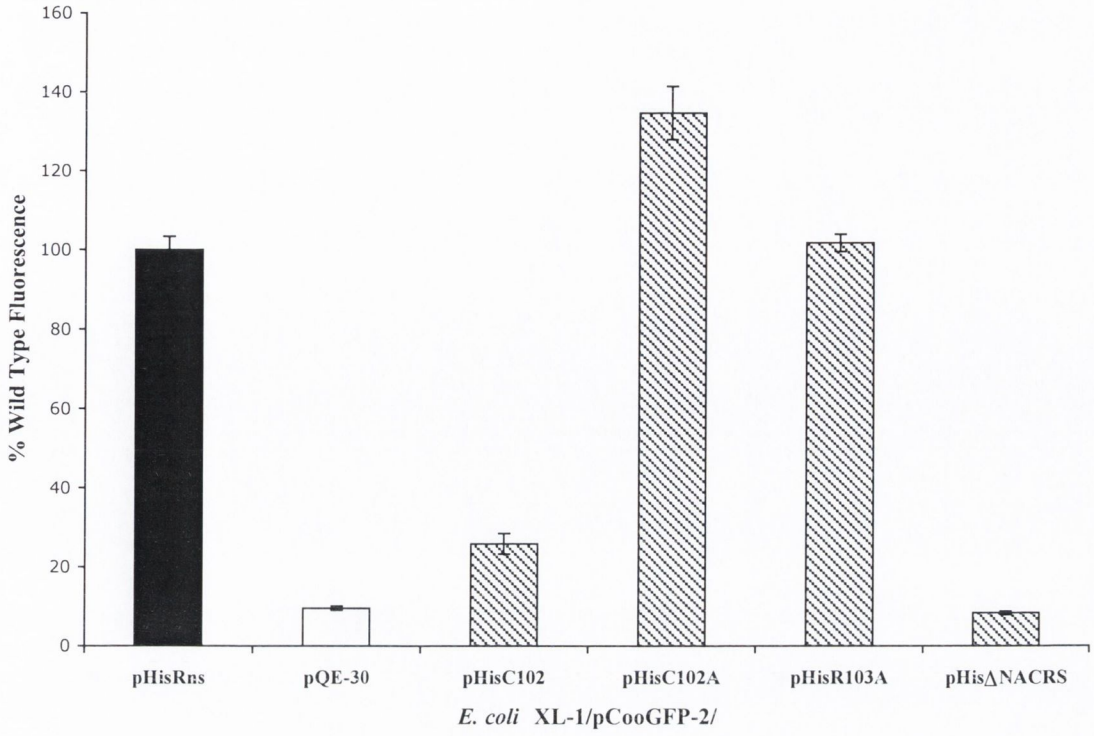

Fig. 4.2. Transactivation of the *coo* promoter by His-tagged Rns and NACRS sequence derivatives. Fluorescence levels of the indicated bacterial cultures (see horizontal axis) were expressed as a percentage of the level of fluorescence in cultures containing wild type Rns. Data represents averages of triplicate measurements of duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown.

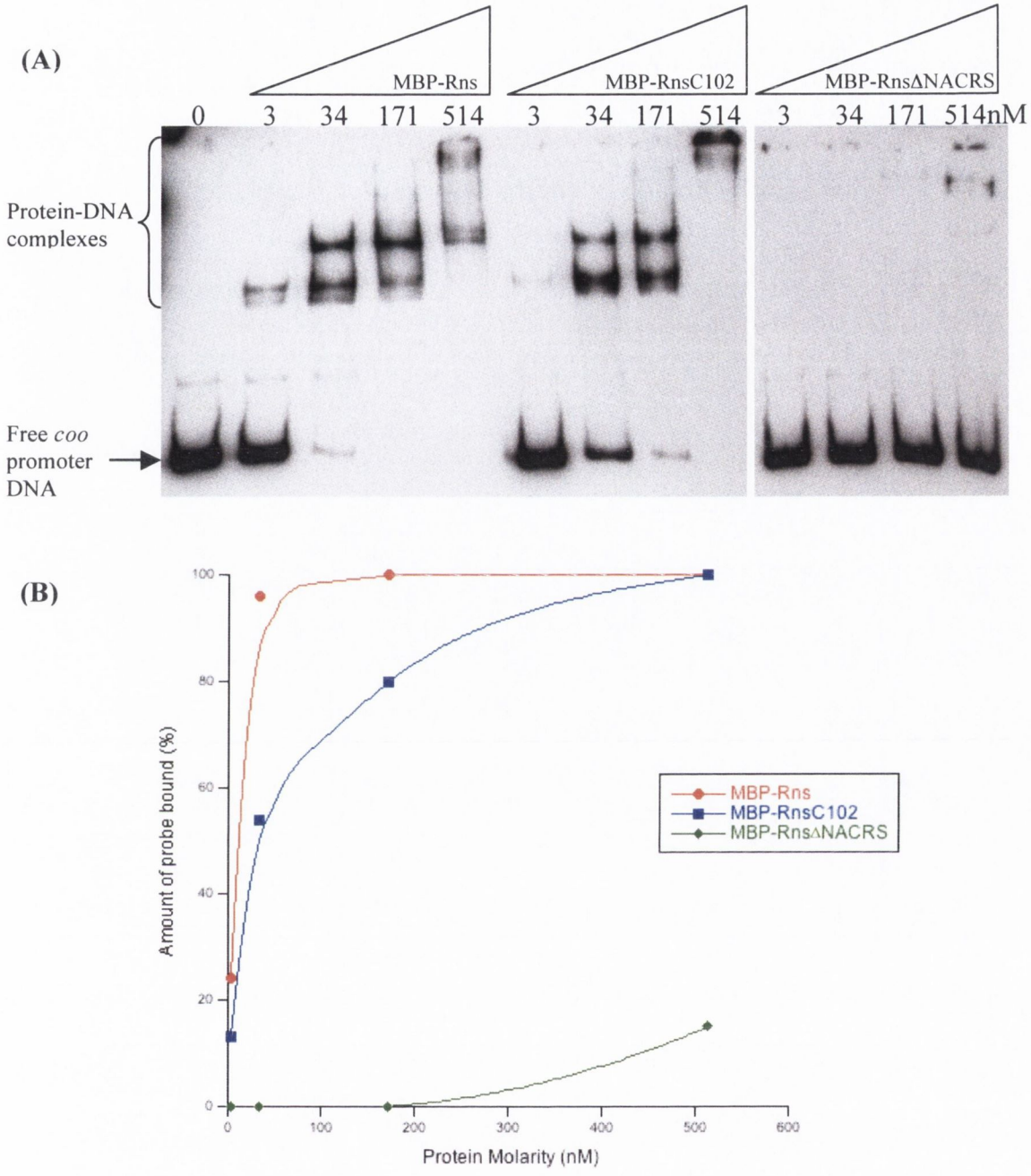

Fig. 4.3. Interaction of MBP-Rns and mutant derivatives with *coo* promoter DNA. **(A)** EMSAs were carried out with approximately 20 pg of a biotinylated *coo* promoter DNA probe and a range of concentrations of MBP-Rns, MBP-RnsC102 or MBP-RnsΔNACRS as indicated at the top of the figure. The unbound DNA fragments and protein-DNA complexes are indicated on the left. **(B)** The percentage of bound *coo* promoter DNA in the EMSA experiments is plotted as a function of the amount of protein added.

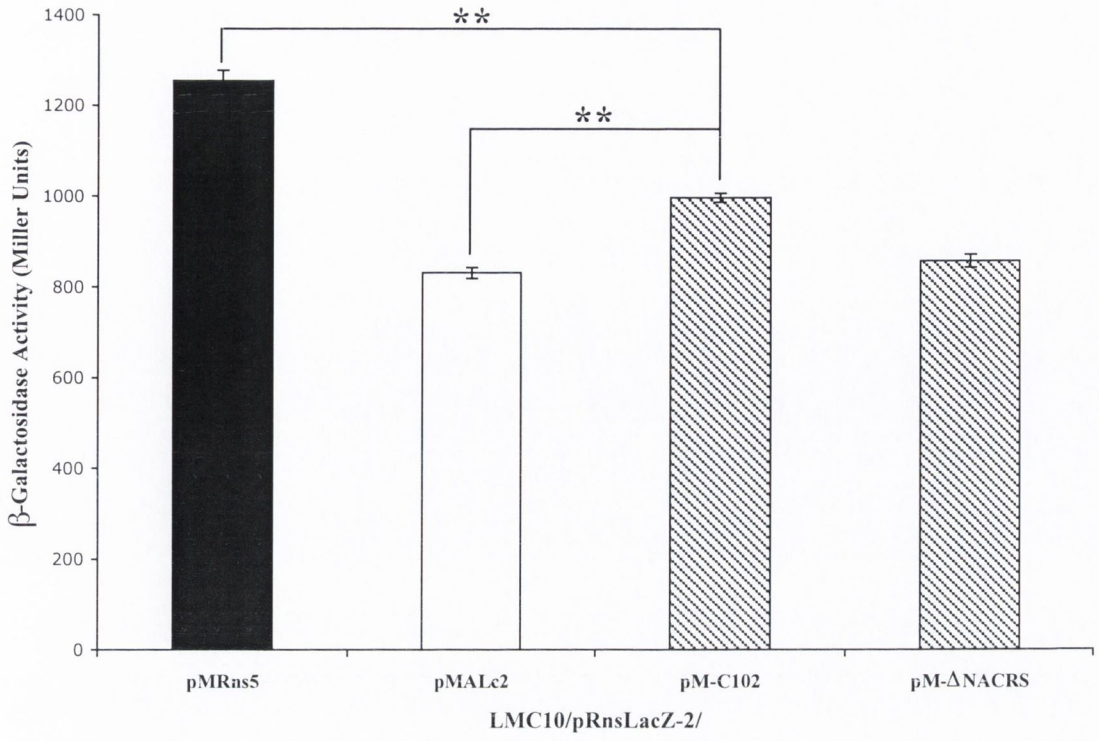

Fig. 4.4. Transactivation of the *rns* promoter by MBP-Rns and mutant derivatives. The β-galactosidase activities of the indicated bacterial cultures (see horizontal axis) were measured. Data represents averages of duplicate assays on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

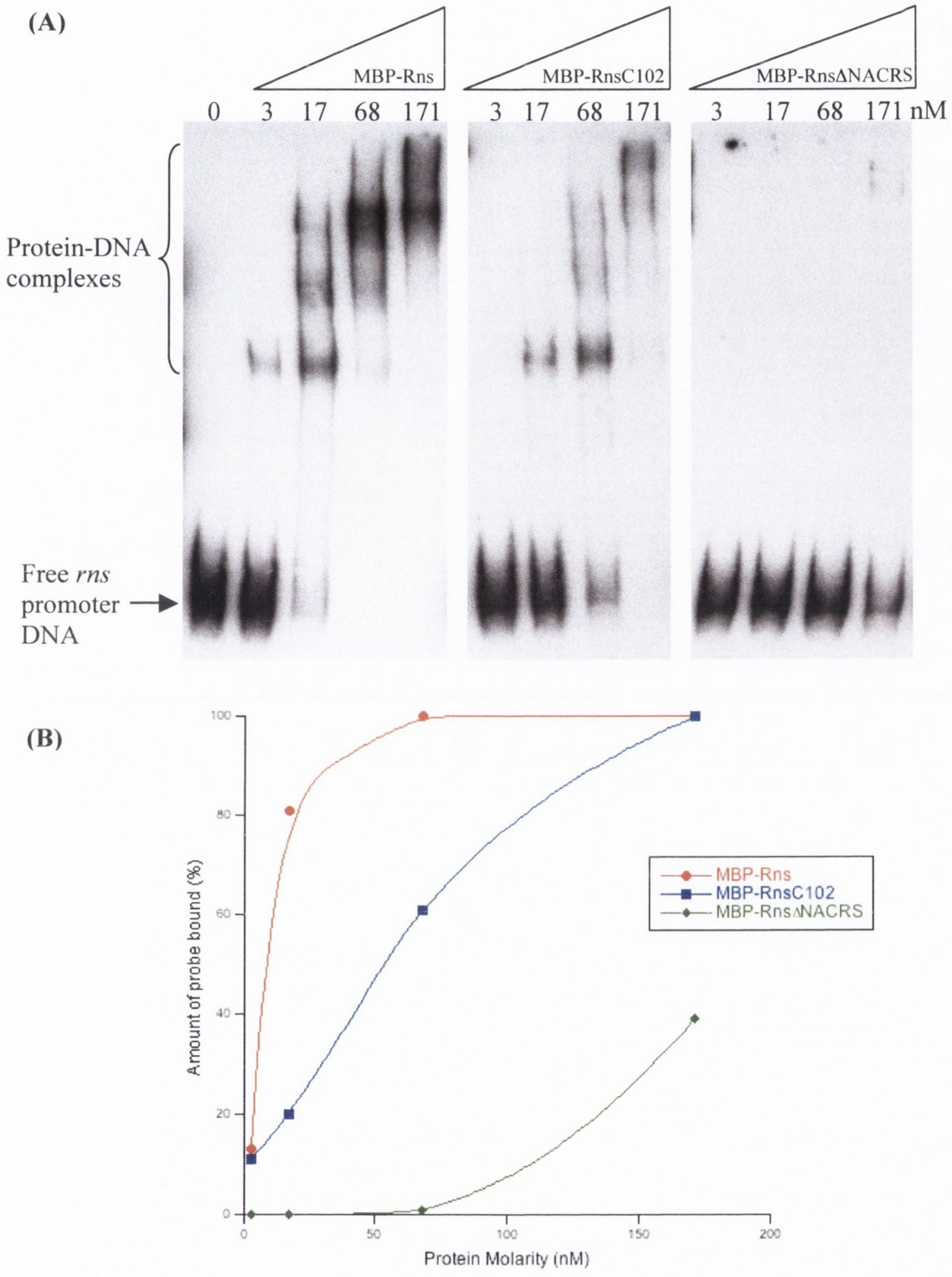

Fig. 4.5. Interaction of MBP-Rns and mutant derivatives with *rns* promoter DNA. **(A)** EMSAs were carried out with approximately 20 pg of a biotinylated *rns* promoter DNA probe and a range of concentrations of MBP-Rns, MBP-RnsC102 or MBP-RnsΔNACRS as indicated at the top of the figure. The unbound DNA fragments and protein-DNA complexes are indicated on the left. **(B)** The percentage of bound *rns* promoter DNA in the EMSA experiments is plotted as a function of the amount of protein added.

binding to the *rns* promoter probe but does so with less affinity than wild type MBP-Rns (Fig. 4.5 (B)). In the presence of 17 nM MBP-Rns all three protein-DNA complexes were formed and at 68 nM MBP-Rns no DNA probe remained unbound. However, in the presence of 17 nM MBP-RnsC102 only the first protein-DNA complex was formed and at 68 nM MBP-RnsC102 approximately 40% (as determined by densitometry) of the *rns* probe remained unbound. Protein-DNA complexes were not detected when EMSAs were performed with MBP-ΔNACRS, thus the deletion mutant also appears to be unable to bind to *rns* promoter DNA. Therefore disruption of the NACRS sequence of Rns impairs the functioning of the protein at both the *coo* and *rns* promoters.

4.2.1.6 Replacement of the NACRS sequence with an alternative disordered peptide

The results of the previous experiments suggested the possibility that in order for Rns to be fully functional there is a requirement for a region of disorder around residues 100-104. In addition, it appeared that as long as this disorder is present, the identity of the amino acids in the region is not critical. To test these possibilities further a mutant derivative of Rns was created in which LENSASR, the seven-residue linker sequence of the AraC family member RhaS of *E. coli*, was substituted for Rns residues 100-104 (NACRS). The RhaS linker was selected for several reasons. It its native setting the LENSASR sequence has been suggested to function only to flexibly link the two functional domains of RhaS. Furthermore, a derivative of the RhaR protein containing the RhaS linker in place of its own linker sequence was functional with only small defects in its activity (212). Finally, GlobPlot analysis revealed that within the Rns amino acid sequence, replacing the deleted NACRS residues with the RhaS linker LENSASR restored the predicted disorder in this region (results not shown).

A mutant *rns* ORF encoding the Rns derivative with the RhaS linker substitution was created in the pMRns5 background by performing inverse PCR using the divergent primers Linker-1 and Linker-2 (Table 2.3). These primers flank the region corresponding to NACRS and each have 5′ extensions comprising nucleotides encoding a portion of the LENSASR sequence. Thus the inverse PCR amplification produced linear DNA fragments containing the nucleotides that code for the LENSASR sequence in place of those encoding NACRS. These linear fragments were ligated to create the

plasmid pM-RhaSLinker (Table 2.2). The *rns* ORF within the plasmid was sequenced to verify that the correct substitution had occurred. It was possible to purify the MBP fusion form of the mutant protein. This confirmed that it was expressed and soluble.

4.2.1.7 Activity of MBP-(Rns with RhaS linker) at the *coo* and *rns* promoters

The ability of the Rns derivative carrying the RhaS linker to activate transcription from the *coo* promoter in pCooGFP-2 was assessed using fluorimetric analysis. Cultures of *E. coli* XL-1/pCooGFP-2 harbouring pM-RhaSLinker were found to be as fluorescent as those containing pMRns5, (the relative fluorescence levels of the cultures were 108% ± 3 and 100% ± 3, respectively). Therefore the Rns derivative containing the RhaS linker appeared to be fully functional at the *coo* promoter. To confirm this, EMSAs were performed to study the binding of the Rns derivative to the promoter. MBP-(Rns with RhaS linker) was found to bind the *coo* promoter DNA probe and form the specific protein-DNA complexes at a similar rate to wild type MBP-Rns (Fig. 4.6 (A)). To establish whether the RhaS linker Rns derivative was also as proficient as wild type Rns at binding to the *rns* promoter, a second set of EMSAs were conducted. Once again the MBP fusion of the derivative was found to complex with the *rns* promoter DNA probe in a manner similar to that observed for MBP-Rns (Fig. 4.6 (B)).

Therefore, replacement of the NACRS sequence with an alternative putative disordered region produced a form of Rns that displayed wild type activity at both the *coo* and *rns* promoters. This supports the theory that Rns activity is dependant on the presence of an area of disorder in the vicinity of residues 100-104.

4.2.2 Ascertaining the oligomeric state of Rns

4.2.2.1 A LexA-based genetic system to analyse Rns dimerisation *in vivo*

To gain an improved understanding of the Rns protein it is important to establish if it exists as a monomer or perhaps as a dimer like some other members of the AraC family of proteins. A LexA-based genetic system was employed as a first step towards ascertaining the oligomeric state of Rns. LexA, a protein involved in the control of SOS response genes, regulates expression of the *sulA* gene by binding to the *sulA*

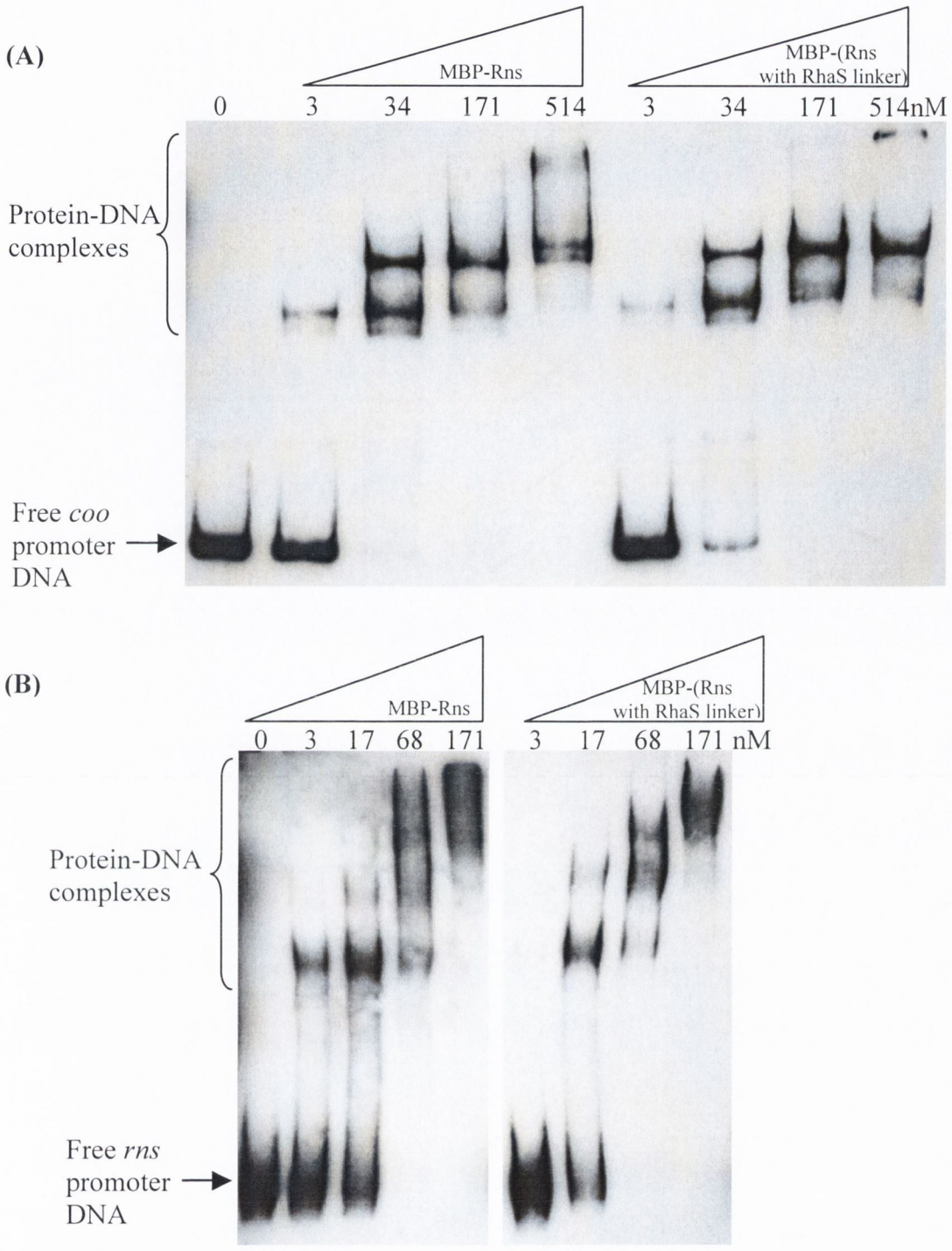

Fig. 4.6. Interaction of MBP-Rns and MBP-(Rns with RhaS linker) with *coo* and *rns* promoter DNA. EMSAs were carried out with approximately 20 pg of a biotinylated *coo* promoter DNA probe **(A)** or approximately 20 pg of a biotinylated *rns* promoter DNA probe **(B)** and a range of concentrations of MBP-Rns or MBP-(Rns with RhaS linker) as indicated at the top of each panel. The unbound DNA fragments and protein-DNA complexes are indicated on the left

promoter region and repressing transcription. Although the LexA DNA-binding domain (DBD) alone can recognise the *sulA* operator, the repressor is functional only as a dimer. Thus the ability of a protein to dimerise can be evaluated by determining whether a fusion of this protein and the LexA DBD is capable of repressing transcription of a chromosomal *sulA*::*lacZ* fusion in the *E. coli* reporter strain SU101 (Fig. 4.7 (A)). The vector pSR660, in which a *lac* promoter controls expression of a portion of the *lexA* gene encoding the DNA-binding domain of LexA alone, was used to construct a plasmid encoding a fusion of Rns and the LexA DBD. The *rns* ORF was amplified by PCR using the primer set LexA-1 and RnsKpn (Table 2.3) and the pSS2192 plasmid as a template. The LexA-1 and RnsKpn primers have an *Xho*I and a *Kpn*I site incorporated into them, respectively, resulting in a PCR product with an *Xho*I site at its 5′ end and a *Kpn*I site at its 3′ end. The PCR product was digested with *Xho*I and *Kpn*I and ligated in frame with the LexA DBD in the pSR660 vector that had also been digested with *Xho*I and *Kpn*I. The resulting recombinant plasmid, named pRns660 (Fig. 4.7 (B)), was sequenced to confirm that the insert was correct and in frame with the LexA DBD-coding sequence.

Initially it was verified that pRns660 expressed the LexA DBD-Rns fusion when induced. Whole cell lysates of *E. coli* DH5α, harbouring pSR660 or pRns660 and grown to logarithmic phase in the presence of IPTG were analysed by Western immunoblotting using anti-LexA DBD antibody (Santa Cruz Biotechnology). Proteins corresponding to the LexA DBD alone and the LexA DBD-Rns fusion were detected in the samples obtained from *E. coli* DH5α/pSR660 and *E. coli* DH5α/pRns660 respectively (Fig. 4.8 (A)). Next it was necessary to confirm that Rns remained functional when fused to the LexA DBD. The pSR660 and pRns660 plasmids were transformed into *E. coli* MC4100 cells harbouring the reporter plasmid pCooGFP-2. While cultures containing pSR660 were non-fluorescent, the fluorescence levels in cultures of *E. coli* MC4100/pCooGFP-2 containing pRns660 were three-fold greater than those in cultures containing pMRns5 (Fig. 4.8 (B)). Therefore the LexA DBD-Rns fusion was functional.

As the LexA DBD-Rns fusion had been shown to be both expressed and active it was now possible to assess whether it formed homodimers. This was performed by measuring the repression of β-galactosidase activity in induced cells of *E. coli* strain SU101 harbouring pSR660 and comparing it to the repression levels in cells containing

pSR660 or p660AraC (Fig. 4.8 (C)). The p660AraC plasmid encodes a LexA fusion to the known homodimer AraC and therefore serves as a positive control. The reporter strain SU101 constitutively transcribes *lacZ* from the strong *sulA* promoter. Consequently, prior to measuring β-galactosidase activity, it was necessary to grow both the starter overnight cultures and the subsequent subcultures of each strain in the presence of 1 mM IPTG. This ensured that the plasmids expressing the LexA DBD fusions had been induced for several hours and any pre-existing β-galactosidase had been degraded. The LexA DBD-AraC fusion and the LexA DBD-Rns fusion were found to repress β-galactosidase activity by 97%, and 67% respectively, compared to *E. coli* SU101 with the LexA DBD alone. Percent repression relative to the control cultures of SU101/pSR660 was calculated as [1.0 – (Miller units of sample/Miller units of control)] × 100 (69). Therefore, these results imply that the Rns protein appears to homodimerise *in vivo*.

4.2.2.2 Cross-linking of His-tagged Rns *in vivo*

A cross-linking method was employed to further assess the possibility that Rns may form dimers *in vivo*. *E. coli* KS1000 harbouring the plasmids pRare and pHisRns were grown overnight in Overnight Express™ Instant TB Medium (Novagen) to induce expression of His-tagged Rns. Samples of the bacteria were then treated with 0, 0.1 or 1 mM of the membrane-permeable, thiol-cleavable, amine-reactive cross-linker dithiobis(succinimidylpropionate) (DSP) as described in Section 2.3.5.1. Following DSP treatment, the samples were denatured by heating at 100°C for 10 min in the presence of Laemmli buffer lacking the thiol β-mercaptoethanol (see Section 2.3.1.1). The samples were then analysed by SDS-PAGE and Western immunoblotting using an anti-His tag probe (Pierce). A band corresponding to monomeric His-tagged Rns (~30 kDa) was detected following treatment with 0 or 0.1 mM DSP (Fig. 4.9). This band was also present following treatment with 1 mM DSP but it was significantly reduced. The reduction of monomeric Rns protein in the presence of an increased concentration of the cross-linking reagent suggested that Rns could be an oligomer. However, oligomeric cross-linked complexes of His-tagged Rns were not detected. Additional cross-linking experiments were necessary to investigate this inconsistency.

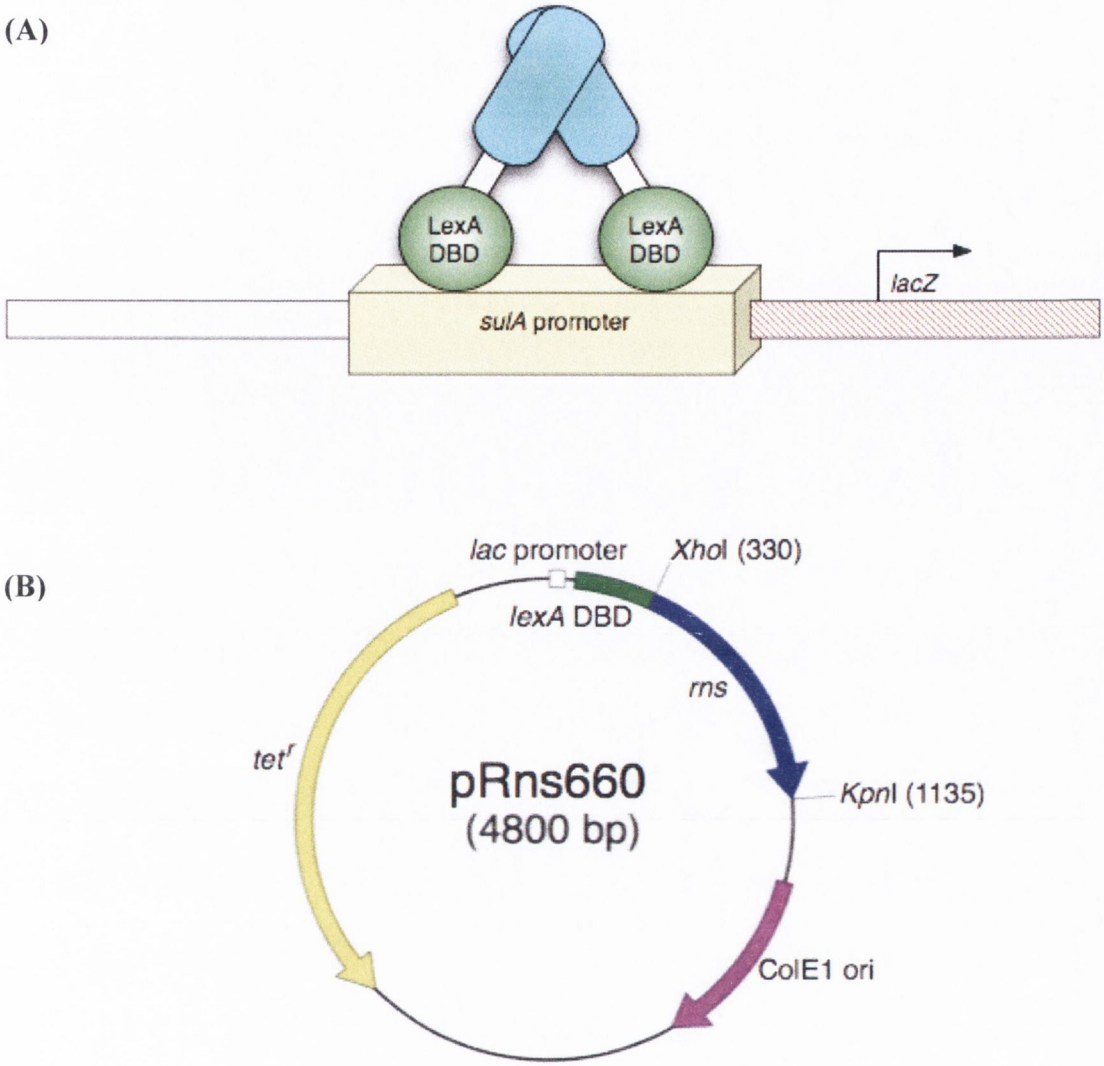

Fig. 4.7. LexA-based system for analysis of homodimerisation. (A) Overview of the LexA system. The protein of interest (depicted in blue) is fused to the LexA DNA-binding domain (LexA DBD, shown in green). In the absence of dimerisation the LexA DBD is unable to bind to the *sulA* promoter and repress transcription. If the protein of interest is capable of dimerising it will restore the ability of LexA to function, resulting in repression of a *sulA*::*lacZ* fusion. Adapted from Dmitrova *et al.* (82). **(B)** Map of the LexA DBD-Rns fusion expressing plasmid pRns660. Relevant features and restriction endonuclease sites are indicated.

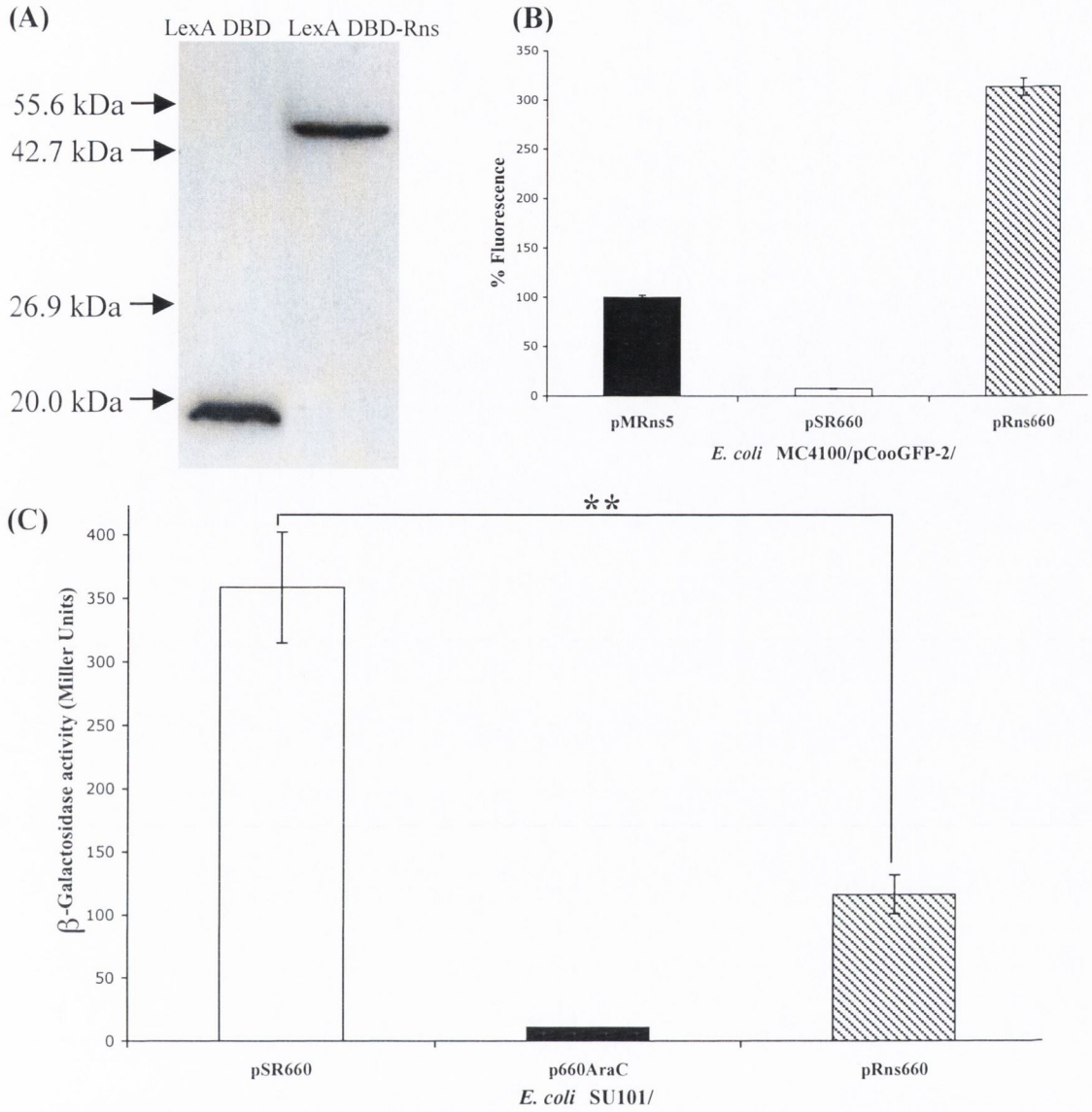

Fig. 4.8. Analysis of the LexA DBD-Rns fusion. (A) Western immunoblot analysis (performed with an anti-LexA DBD antibody) of whole cell lysates of induced cultures of *E. coli* DH5α/pSR660 (lane 1) and DH5α/pRns660 (lane 2) demonstrating that the LexA DBD and LexA DBD-Rns fusion (respectively) are expressed. The positions of molecular mass markers are indicated on the left **(B)** Transactivation of the *coo* promoter by the LexA DBD-Rns fusion. Fluorescence levels of the indicated cultures (see horizontal axis) were expressed as a percentage of the level of fluorescence in cultures containing MBP-Rns. Data represents averages of triplicate measurements on duplicate cultures. Error bars indicate standard deviation values. **(C)** Regulation of the *sulA* promoter by LexA DBD or LexA DBD fusions. β-galactosidase activities of *E. coli* SU101 (contains a chromosomal *sulA*::*lacZ* transcriptional fusion) carrying pSR660 (encodes LexA DBD), p660AraC (encodes LexA DBD-AraC) or pRns660 (encodes LexA DBD-Rns) were measured. Data represents averages of duplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently three times; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

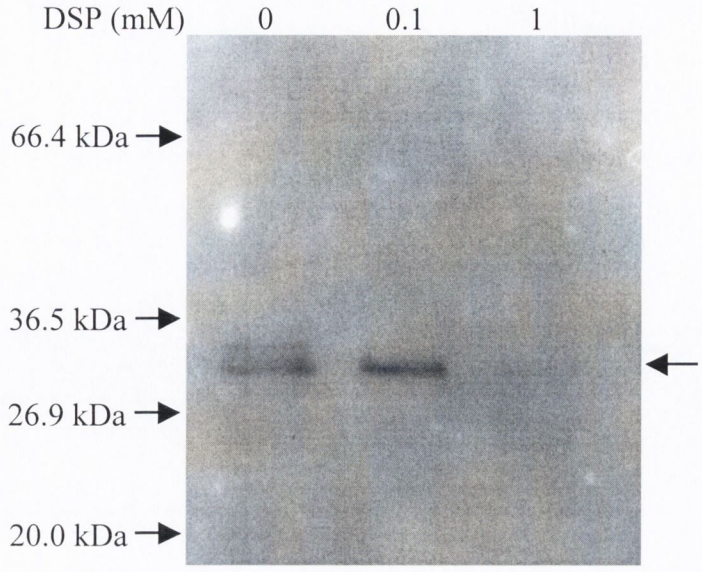

Fig. 4.9. Cross-linking of His-tagged Rns *in vivo*. Cells of *E. coli* KS1000/pRare/pHisRns were grown overnight in Overnight Express™ Instant TB Medium (Novagen) to induce expression of His-tagged Rns. Samples of the cells were treated with the cross-linker DSP (the concentration of DSP used is shown above each lane) and were then denatured by heating at 100°C for 10 min in the presence of Laemmli buffer (lacking β-mercaptoethanol) and analysed by Western immunoblotting using an anti-His tag probe. The positions of molecular mass markers are indicated on the left. An arrow on the right indicates the position of monomeric His-tagged Rns.

4.2.2.3 Cross-linking of His-tagged Rns *in vitro*

Subsequent cross-linking studies were conducted *in vitro* with purified Rns protein. Experiments were initially performed with His-tagged Rns. The *E. coli* nucleoid-associated protein, H-NS, served as a positive control as it has previously been demonstrated to exist as a dimer (127). His-tagged Rns and His-tagged H-NS were purified as described in Section 2.3.2.2. To eliminate the imidazole used in the purification procedure and to stabilise the proteins during storage, the purified proteins were dialysed into a sodium phosphate buffer containing 50% (v/v) glycerol (see Section 2.3.2.2). The His-tagged proteins were subjected to an *N*-hydroxy-succinimide (NHS) catalysed 1-ethyl-3-(3′-dimethylaminopropyl) carbodimide (EDC) cross-linking reaction. EDC is a carboxyl and amine-reactive zero-length cross-linker. The products of the cross-linking reactions were denatured by heating at 100°C for 5 min in the presence of Laemmli buffer and then separated by SDS-PAGE (Fig. 4.10 (A)). After treatment with EDC and NHS, the amount of monomeric His-tagged H-NS present was reduced and a band corresponding to the approximate expected size of an H-NS dimer (~32 kDa) was detected. The monomeric form of His-tagged Rns was also significantly reduced following cross-linking treatment. As was found with the *in vivo* assay, however, a high molecular weight complex corresponding to a possible oligomeric form of Rns was not detected. It was evident that the cross-linking procedure was having an effect on Rns, but what exactly was occurring was unclear. To examine this further both a control sample and a sample of EDC-NHS cross-linked His-tagged Rns were analysed by Western immunoblotting with anti-Rns antiserum (Fig. 4.10 (B)). Despite the enhanced sensitivity of detection provided by the Western immunoblot procedure, there was no evidence of the presence of Rns oligomers in the cross-linked sample. There was, however, the first indication of a potential explanation of why bands corresponding to Rns monomers were much reduced or no longer detected following cross-linking treatment. The anti-Rns antiserum had reacted with a mass at the top of the Western immunoblot, i.e. at the location of the interface between the stacking and resolving gel components of the SDS polyacrylamide gel. It is possible that this mass corresponded to a large aggregate of Rns protein that had not migrated beyond the stacking gel. If so, instead of revealing the presence of oligomeric forms of Rns, cross-linking could cause any Rns protein that had not aggregated to associate with the aggregate and remain in the stacking gel. This would explain the apparent

disappearance of Rns after cross-linking. As a standard procedure, the stacking gel was routinely removed after SDS-PAGE analysis and only the resolving gel was used in Coomassie staining or Western immunoblotting. Therefore the potential existence of this aggregate had not been noticed previously.

4.2.2.4 Ultrafiltration, and subsequent cross-linking, of His-tagged Rns *in vitro*

To explore the possibility that an aggregate may be present in samples of His-tagged Rns, the purified protein was subjected to ultrafiltration in a Vivaspin column with a molecular weight cut-off (MWCO) of 100 kDa. The MWCO was chosen such that soluble protein would pass through the filter (into the filtrate) while potential aggregate forms would be retained (in the retentate). Samples of retentate and filtrate were denatured and analysed by SDS-PAGE (after which the stacking gel was not removed) and Western immunoblotting using anti-Rns antiserum (Fig. 4.11 (A)). In addition to a band corresponding to monomeric Rns, a high molecular weight mass that had not migrated beyond the stacking gel was detected in the retentate sample. Monomeric Rns was also present in the filtrate, however the large mass was significantly reduced in this sample. Even less of the mass was detected when the filtrate was not heated at 100°C before loading onto the gel. Heating of samples in Laemmli buffer prior to SDS-PAGE analysis can increase protein aggregation (341). Thus, the findings that the mass had not passed through the 100 kDa MWCO ultrafiltration column, and that its formation was reduced when the sample was not heated, indicate that it is likely to be a large aggregate of Rns. However, it appeared that ultrafiltration of His-tagged Rns resulted in a filtrate comprising soluble protein that was largely free of the potential aggregate. Therefore the His-tagged Rns present in such a filtrate sample was used in EDC-NHS-catalysed cross-linking reactions. In addition, after cross-linking the samples were denatured either by heating (at 60, 80 or 100°C) for 5 min in the presence of Laemmli buffer or by incubation at room temperature for 5 min in Laemmli buffer plus 4 M urea. The samples were then analysed by SDS-PAGE (after which the stacking gel was not removed) and Western immunoblotting using anti-Rns antiserum (Fig. 4.11 (B)). It was evident that the method of denaturation used had an effect on His-tagged Rns. Less Rns was detected as the denaturation temperature increased and a mass was present in the stacking gel for several of the samples denatured by heating. Regardless of denaturation method however, for each set of samples the amount of monomeric Rns

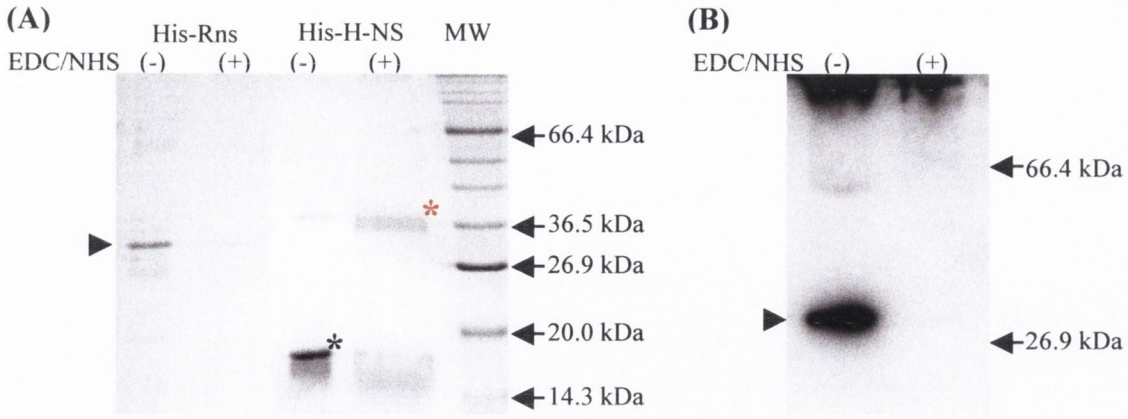

Fig. 4.10. Cross-linking of His-tagged Rns *in vitro*. **(A)** SDS-PAGE analysis of purified His-tagged Rns and His-tagged H-NS incubated with (+) or without (-) EDC-NHS cross-linking reagents. Prior to electrophoresis the protein samples were denatured by heating at 100°C for 5 min in the presence of Laemmli buffer. The positions of protein molecular mass markers (MW) are indicated on the right. Monomeric His-tagged Rns is indicated by an arrowhead on the left. Monomeric and dimeric His-tagged H-NS are indicated by black and red asterisks, respectively. **(B)** Western immunoblot analysis (using anti-Rns antiserum) of purified His-tagged Rns incubated with (+) or without (-) EDC-NHS cross-linking reagents. Prior to electrophoresis the protein samples were denatured by heating at 100°C for 5 min in the presence of Laemmli buffer. The positions of protein molecular mass markers are indicated on the right. Monomeric His-tagged Rns is indicated by an arrowhead on the left.

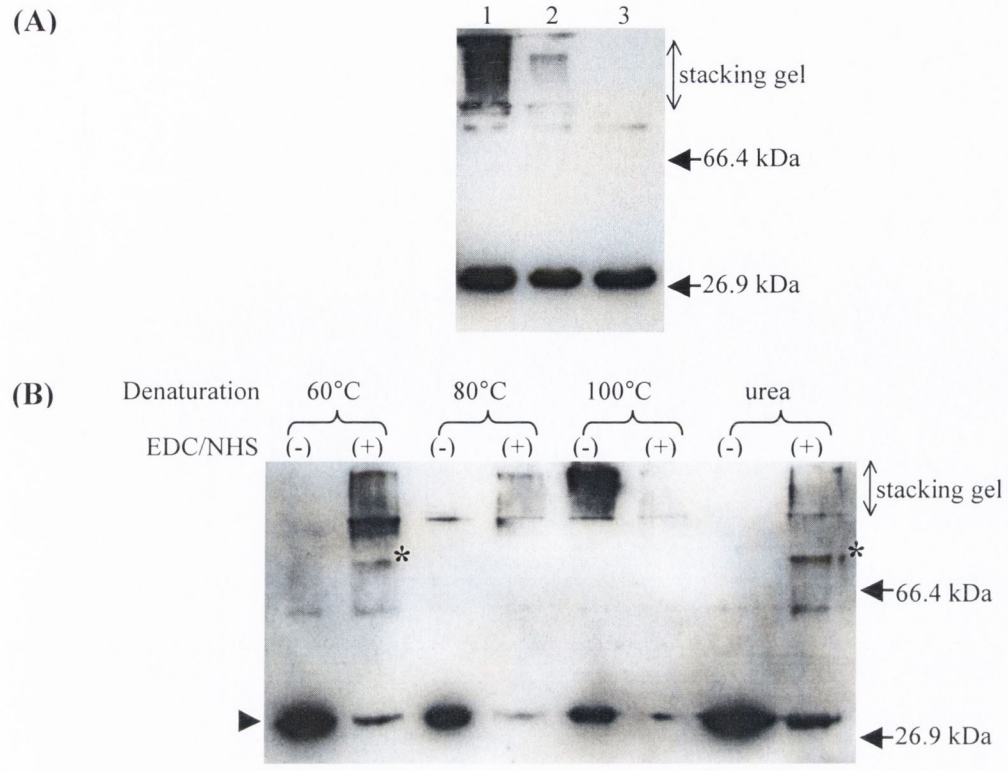

Fig. 4.11. Ultrafiltration and cross-linking of His-tagged Rns *in vitro*. **(A)** Western immunoblot analysis (using anti-Rns antiserum) of samples of the retentate (lane 1) and the filtrate (lanes 2 and 3) produced following ultrafiltration of purified His-tagged Rns in a column with a molecular weight cut-off of 100 kDa. Lanes 1 and 2 contain retentate and filtrate, respectively, mixed with an equal volume of Laemmli buffer and heated at 100°C for 5 min prior to electrophoresis. Lane 3 contains filtrate mixed with an equal volume of Laemmli buffer and electrophoresed without heating. The portion of the Western immunoblot that corresponded to the stacking gel is indicated on the right, as are the positions of protein molecular mass markers. **(B)** Western immunoblot analysis (using anti-Rns antiserum) of samples of purified His-tagged Rns incubated with (+) or without (-) EDC-NHS cross-linking reagents. Prior to electrophoresis the protein samples were denatured by heating at the indicated temperatures for 5 min in the presence of Laemmli buffer or by incubation at room temperature for 5 min in Laemmli buffer plus 4 M urea. The portion of the Western immunoblot that corresponded to the stacking gel is indicated on the right, as are the positions of protein molecular mass markers. The position of monomeric His-tagged Rns is indicated by an arrowhead on the left. Asterisks indicate the positions of bands potentially corresponding to dimeric His-tagged Rns.

detected was reduced following cross-linking. Most importantly, in the samples denatured by heating at 60°C or by treatment with urea, a novel band that had migrated above the 66.4 kDa molecular marker was present after cross-linking. This band may correspond to a dimer of Rns and it was observed on several occasions when cross-linking of His-tagged Rns was repeated. However, detection of this putative Rns dimer was not consistently reproducible. As mentioned in Chapter 3, the DNA-binding ability of His-tagged Rns was found to be variable while MBP-Rns performed more reliably. Therefore ensuing cross-linking experiments were performed with MBP fusion proteins.

4.2.2.5 Cross-linking of MBP-Rns, and mutant derivatives, *in vitro*

MBP is a monomer (387), and did not interfere with cross-linking of an MBP fusion of XylS, an AraC family member (339). Therefore purified MBP-Rns protein was used in NHS-catalysed EDC cross-linking reactions *in vitro*. The purified proteins MBP-paramyosinΔSal (a fusion of MBP to a fragment of the multimeric protein paramyosin from *Dirofilaria immitis*) and MBP (NEB) were included as positive and negative controls, respectively. After cross-linking, one set of protein samples was denatured using the standard method of heating at 100°C for 5 min in the presence of Laemmli buffer. Another set of samples was denatured using a modified urea-based method (see Section 2.3.5.2) demonstrated to reduce protein aggregation (381, 382). All samples were then analysed by SDS-PAGE and Western immunoblotting using anti-MBP antiserum. When the samples were denatured by heating in Laemmli buffer, cross-linking was found to result in: a reduced amount of the MBP monomer, the detection of an oligomeric form of MBP-paramyosinΔSal and a lack of detection of any form of MBP-Rns (Fig. 4.12 (A)). Using the urea-based denaturation procedure, however, detection of each of the proteins after cross-linking was much improved. Therefore this denaturation method was used in all subsequent cross-linking experiments and provided consistently reproducible results, an example of which is shown in Fig. 4.12 (B). Under these conditions the EDC-NHS treatment was found to have no effect on monomeric MBP but to result in the appearance of bands corresponding to dimers and trimers of MBP-paramyosinΔSal. For cross-linked samples of MBP-Rns, in addition to the band corresponding to monomeric protein, a species migrating with an apparent molecular weight more than two-fold greater than that of the monomer was present. This species was reliably detected after EDC-NHS cross-linking of MBP-Rns and may be a dimer of

the fusion protein. Its migration in the gel did not precisely correspond to the mobility expected for a dimer of MBP-Rns (~146 kDa) but this could be explained by the finding that cross-links can alter the migration rate of proteins in SDS-PAGE (157). A reduced amount of a similar species was detected in samples of MBP-Rns that had not been cross-linked. This species may be due to dimers of MBP-Rns that had survived the urea-based denaturation procedure. Therefore these cross-linking studies revealed that MBP-Rns appeared to be capable of dimerising.

The ability of the mutant proteins MBP-RnsC102, MBP-RnsQ227 and MBP-RnsΔNACRS to dimerise *in vitro* was also analysed in EDC-NHS-catalysed cross-linking experiments (Fig. 4.12 (C)). Following cross-linking treatment the putative dimer band was present in samples of both MBP-RnsC102 and MBP-RnsQ227. When MBP-RnsΔNACRS was cross-linked this band was also present but to a far lesser extent. In addition, the amount of monomeric MBP-RnsΔNACRS detected after cross-linking was much reduced. Therefore, it appeared that the mutations in each of these derivatives of Rns had not prevented the proteins from dimerising. However, the deletion mutant MBP-RnsΔNACRS may have a reduced ability to form dimers and may also have a tendency to aggregate or precipitate when cross-linked.

4.2.2.6 Gel filtration chromatography of MBP-Rns

To further characterise the ability of Rns to form dimers, the oligomeric state of MBP-Rns in solution was examined by gel filtration chromatography. Two major peaks were observed following gel filtration analysis of purified MBP-Rns (Fig. 4.13 (A)). The first peak was in the column void volume and was therefore outside of the range of calibration. The second peak corresponded to a protein with a molecular mass of 134 kDa (by comparison with a series of standard proteins). This does not correlate exactly with the molecular mass predicted for a dimer of MBP-Rns (~146 kDa). However, as only perfectly globular proteins migrate precisely according to size during gel filtration it was still likely that this peak represented a dimeric form of MBP-Rns. Aliquots of the elution fractions comprising each of the two peaks were denatured by incubation at room temperature in Laemmli buffer plus 8 M urea. They were then examined by SDS-PAGE and Western immunoblotting with anti-Rns antiserum to determine if MBP-Rns was present. MBP-Rns was not detected in the fractions corresponding to the first peak (results not shown). This peak may have been due to non-specific aggregation in the

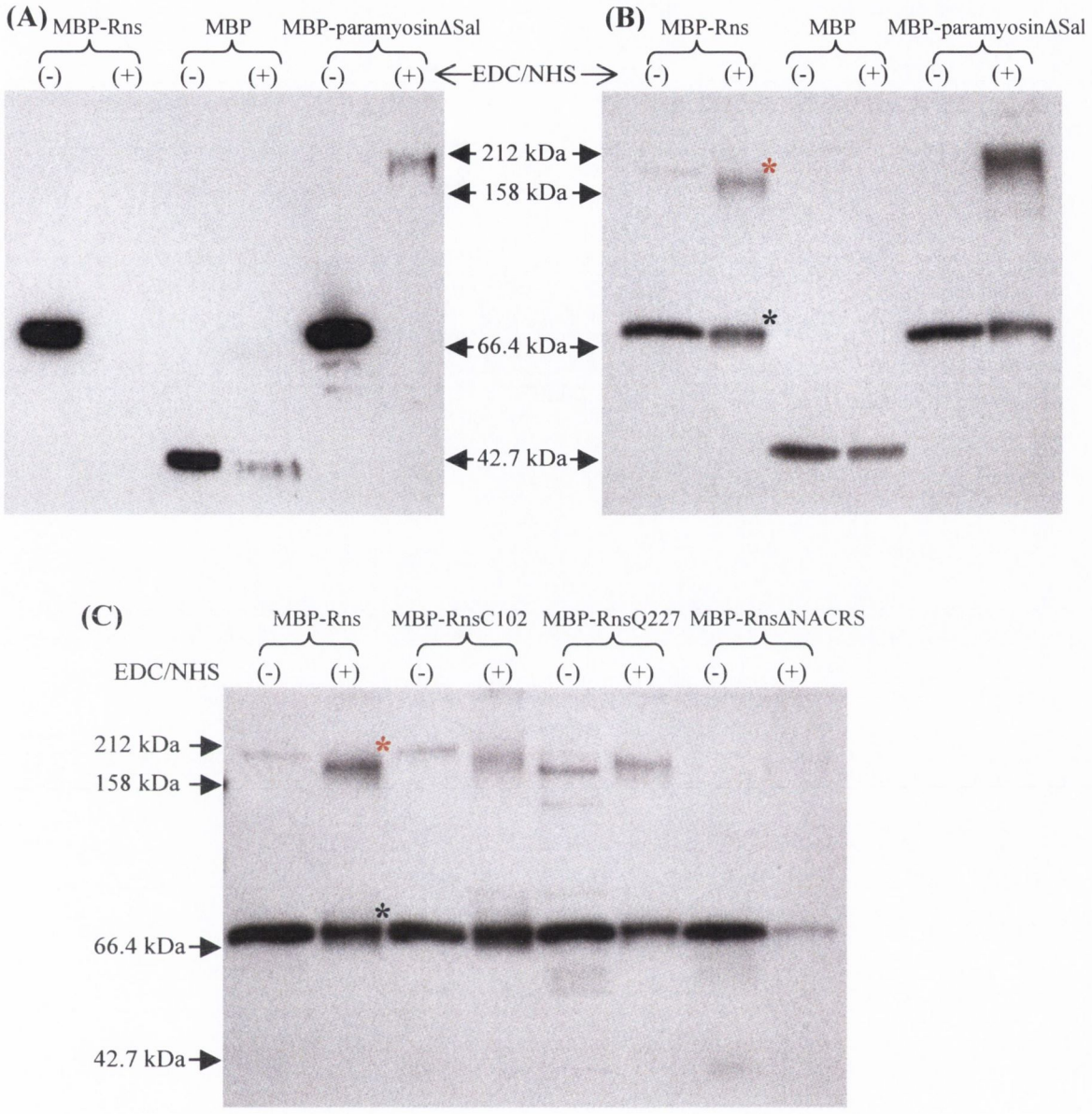

Fig. 4.12. Cross-linking of MBP and MBP fusions *in vitro*. Western immunoblot analysis (using anti-MBP antiserum) of MBP or MBP fusions, as indicated at the top of each panel, incubated with (+) or without (-) EDC-NHS cross-linking reagents. Prior to electrophoresis the protein samples were denatured either by heating at 100°C for 5 min in the presence of Laemmli buffer **(A)** or by using a urea-based treatment demonstrated to reduce protein aggregation **(B)** and **(C)**, see Section 2.3.5.2. The positions of protein molecular mass markers are indicated. The positions of bands corresponding to monomeric and potentially dimeric MBP-Rns are indicated by black and red asterisks, respectively.

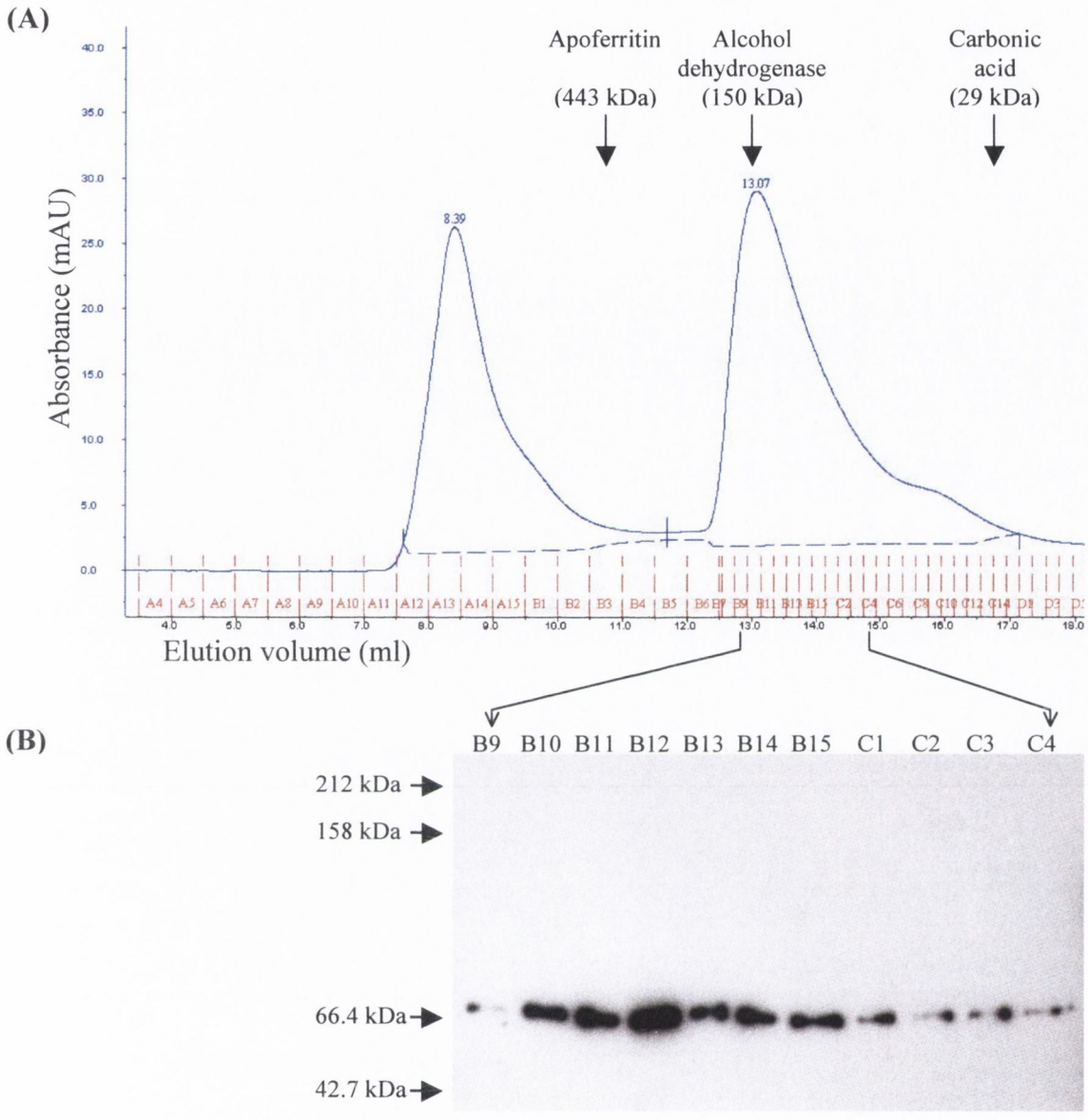

Fig. 4.13. Gel filtration chromatography of MBP-Rns. (A) Elution profile of MBP-Rns following gel filtration analysis. Protein elution was monitored at 280 nm. Fractions of 0.5 ml were collected and designated as A1-15, B1-15, C1-15 etc. as indicated in red. The elution volumes are indicated above the peaks. Arrows at the top of the figure indicate the volumes at which the protein standards apoferritin, alcohol dehydrogenase and carbonic anhydrase eluted. **(B)** Western immunoblot analysis (using anti-Rns antiserum) of aliquots of the indicated gel filtration elution fractions. Prior to electrophoresis the samples were denatured by incubation at room temperature in Laemmli buffer plus 8 M urea. The positions of protein molecular mass markers are indicated on the left.

protein sample. However, after denaturation, monomeric MBP-Rns (~73 kDa) was detected in the fractions that had comprised the peak corresponding to a size of 134 kDa (Fig. 4.13 (B)). Therefore gel filtration analysis of MBP-Rns demonstrated that the protein exists mainly as a dimer in solution.

4.3 Discussion

In recent years there has been increasing awareness that many proteins contain functionally important disordered regions (99). While studying the RnsC102 pentapeptide insertion mutant, it became evident that this may also be true of Rns. In order to be fully active, the Rns protein appears to depend on the existence of an area of disorder between its N- and C-terminal domains.

The pentapeptide insertion after position C102 of Rns had occurred in a part of the protein that was predicted to be unstructured. In general such sequences are tolerant of insertions (175). Further analysis revealed that within this unstructured region of Rns, the amino acids N100 to S104 were predicted to form a disordered area. However, following the introduction of the pentapeptide insertion after residue C102 this area of disorder was no longer predicted to be present. The RnsC102 insertion mutant displayed a reduced ability to activate transcription at both the *coo* and *rns* promoters. This partial level of activity may be due, at least in part, to the insertion affecting DNA binding by the protein. Although the mutant was capable of binding to *coo* and *rns* promoter DNA, it was found to do so with less affinity than wild type Rns. A pentapeptide insertion within the putative disordered region of Rns reduced the activity of the protein, but elimination of this region was even more deleterious to protein function. A Rns derivative lacking the NACRS residues that comprise the potential area of disorder did not bind to DNA and was unable to activate expression from the *coo* and *rns* promoters. Therefore the presence of this disordered region was clearly important for Rns activity. However, analysis of additional Rns mutant derivatives revealed that the identity of the amino acids composing the required area of disorder was not crucial. Individual alanine substitutions at positions C102 and R103 of Rns maintained the predicted disorder and were found to have no adverse effects on Rns-dependent expression from the *coo* promoter; RnsC102A and RnsR103A activated the *coo* promoter as efficiently as wild type Rns. Indeed all five of the residues comprising the area of predicted disorder in Rns were found to be replaceable. A derivative of Rns in which the seven amino acids comprising the flexible linker of RhaS, LENSASR (212), were substituted for the NACRS amino acid sequence was fully functional. The substitution recreated the necessary predicted disordered region and the Rns derivative with the RhaS linker sequence activated the *coo* promoter and bound to DNA as well as wild type Rns. Thus even the length of the disordered area in Rns is not critical.

What function might this essential region of predicted disorder in the Rns protein serve? Deletion and disruption of the region abolished and reduced DNA binding by Rns respectively. However, amino acids 100-104 of Rns are not in the predicted DNA-binding domain of the protein and are unlikely to make direct contact with DNA. Therefore the Rns sequence NACRS is not likely to have a direct involvement in the DNA binding activity of the protein. It is possible that, like the LENSASR sequence of RhaS, the NACRS residues of Rns form a flexible linker that connects the two domains of the protein. While also being disorder-promoting residues, the alanine and serine amino acids within the NACRS sequence are known to permit protein flexibility. In addition, flexibility is a characteristic associated with disordered regions of proteins (100). Hence, it may be the flexibility that is provided by the disorder in this area of Rns that is essential for protein function. A flexible interdomain linker may be necessary to allow the C-terminal domain of Rns to bind to DNA while the N-terminal domain performs its as yet undiscovered function.

The NACRS sequence of Rns shares several similarities with the known flexible linkers of AraC itself and the AraC family members RhaS and RhaR. It was demonstrated that the amino acid sequence of the linkers of each of these proteins did not have to be maintained precisely, as the linker residues could be individually altered without significantly reducing protein activity (120, 212). Specifically, single alanine substitutions in the linker regions of both RhaS and RhaR had minor effects, if any, on the ability of the proteins to regulate transcription (212), as was found for the NACRS sequence of Rns. Furthermore, like the Rns derivative in which the NACRS residues had been replaced by the RhaS linker LENSASR, a RhaR derivative carrying the RhaS linker in place of its own was functional (212). Only alterations that were thought to reduce the flexibility of the linkers of the AraC, RhaS and RhaR proteins resulted in mutant derivatives with defects in their activity (120, 212). Therefore it appears that like Rns, these proteins require an area of flexibility between their N- and C-terminal domains.

Flexible linker regions have generally been demonstrated to be tolerant of amino acid insertions. For example, pentapeptide insertions within the linker sequences of the McrA endonuclease of *E. coli*, the XerD recombinase of *S.* Typhimurium and the CI repressor of coliphage 186 had no discernable effect on the activities of these proteins (9, 47, 366). Indeed the linker region of AraC was first identified by screening the protein for an area that was tolerant of insertions (119). In contrast, the Rns protein was

not found to be fully tolerant of the pentapeptide insertion after position C102 within the NACRS sequence. However, as discussed above, this was most likely due to the insertion disrupting the essential disorder and flexibility in the region. In addition, as detailed in Chapter 3, the area immediately downstream of the NACRS residues (amino acids K108 to N130) is tolerant of insertions. A recent study included a suggestion, based on unpublished scanning linker mutagenesis results, that residues 100 through 131 of Rns probably functions as a flexible linker (19). The findings of the work described in this, and the preceding chapter support this suggestion. Furthermore, this work has specifically identified the NACRS sequence of Rns that may comprise the minimal linker region of the protein or at least represents a small segment within a larger linker in which disorder or flexibility is vital.

In general, AraC-like proteins that regulate carbon metabolism are dimeric while those that regulate genes involved in stress response function as monomers (248). The oligomeric state of the AraC-like proteins involved in virulence gene regulation remains mostly unknown. In this study, cross-linking assays, a LexA-based reporter system, and gel filtration chromatography indicated that the Rns protein is able to dimerise. All three techniques employed His-tagged Rns or a Rns fusion. However these forms of the protein have been demonstrated to be functional and, due to the high insolubility of Rns, it is necessary to use His-tagged Rns or MBP-Rns for *in vitro* studies.

The ability of Rns to homodimerise *in vivo* was initially examined with a LexA-based gene fusion system. This system has previously been used to assess the dimerisation of several AraC family members including AraC itself, XylS and the virulence regulators UreR, PerA and ToxT (46, 193, 306, 317, 339). LexA dimers repress expression of the *sulA* gene, however the LexA DBD alone is unable to dimerise and thus cannot repress a chromosomal transcriptional fusion of the *sulA* promoter to *lacZ*. Therefore the finding that transcription from the *sulA* promoter was repressed by a fusion of the LexA DBD to full length Rns revealed that Rns is capable of dimerising. In contrast, it was recently reported that dimerisation of Rns could not be demonstrated using an alternative gene fusion system based on the λ phage repressor, cI (19). In this report by Basturea *et al.* only a portion of Rns (residues 1-154) was used and found to be unable to restore repressor function to the cI DBD (19). The findings that a fusion of full length Rns to the LexA DBD is functional, and therefore a dimer, while a fusion of Rns(1-154) to the cI DBD is not, may be due to the dimerisation domain of Rns residing within its C-terminal amino acids (residues 155-265).

Cross-linking assays were conducted to further assess the possibility that the Rns protein dimerises. MBP-Rns was subjected to NHS-catalysed EDC cross-linking reactions after which it was denatured using a urea-based treatment and then analysed by SDS-PAGE and Western immunoblotting. Under these conditions protein complexes corresponding in size to putative dimers of MBP-Rns were consistently detected. Again this differs from the findings of Basturea *et al.* who reported that MBP-Rns dimers were not observed following SDS-PAGE analysis of denatured samples of MBP-Rns that had been exposed to the chemical cross-linker glutaraldehyde (19). The report does not specify the method used to denature the protein samples prior to SDS-PAGE analysis. However, as the work performed in this chapter illustrates, the denaturation method is significant, as it can affect the results of cross-linking experiments. Basturea *et al.* do state that the mobility of MBP-Rns was reduced following cross-linking. Indeed the amount of monomeric MBP-Rns present decreased significantly as the concentration of cross-linker increased, however this is attributed to the binding of various amounts of glutaraldehyde to the MBP-Rns monomers (19). In addition, glutaraldehyde cross-linking of Rns was revealed to result in the presence of large complexes near the top of the gel (19). Therefore it is possible that the cross-linking of Basturea *et al.* was hampered by the aggregation of Rns protein.

While conducting the experiments included in this chapter, it became apparent that the Rns protein is prone to aggregation. This is the likely reason for the difficulties encountered during the initial cross-linking experiments using His-tagged Rns. Aggregation is a general characteristic of many AraC-like proteins. AraC itself, RhaS, ToxT and XylS have been reported to aggregate and this has impeded the analysis of each of these proteins (201, 317, 354, 429). There has been conjecture that AraC may be incompletely folded prior to binding to DNA; thus it could have an excessive number of hydrophobic residues exposed, which would explain its tendency to aggregate (354). The tendency of Rns to aggregate appeared to increase when the protein was heated in the presence of Laemmli buffer, the standard method of denaturation used prior to SDS-PAGE analysis. Thermal denaturation can cause proteins to unfold and expose previously buried hydrophobic patches. Aggregation then occurs as a consequence of intermolecular interactions between these hydrophobic protein surfaces (160). Hydrophobic interactions are abolished in the presence of high concentrations of urea. Therefore, use of urea during the denaturation of MBP-Rns samples enabled the observation of the effects of cross-linking on the protein in the absence of aggregation.

This may explain why dimers of MBP-Rns were detected following the cross-linking assays described in this study but not after the cross-linking of Basturea *et al.* (19).

The final method used to assess the oligomeric state of Rns was gel filtration. As judged by SDS-PAGE and Western immunoblot analysis, following gel filtration chromatography MBP-Rns eluted as a peak with an apparent molecular mass of 134 kDa, implying that MBP-Rns exists as a dimer in solution.

The location of the dimerisation domain of Rns is not known. Pentapeptide insertions in the putative linker and HTH regions of Rns did not prevent dimerisation of MBP fusions of the RnsC102 and RnsQ227 mutants, respectively. Dimerisation of MBP-RnsΔNACRS appeared to be reduced relative to wild type protein but loss of the potential flexible linker may have indirectly affected the ability of Rns to form protein-protein interactions. Dimerisation of AraC is mediated by an antiparallel coiled coil formed between an α-helix at the end of the N-terminal domain of each monomer (378). UreR and XylS are proposed to use a similar mechanism to dimerise (306, 339). A potential α-helix formed by residues 131-146 of Rns, was identified by Basturea *et al.* and reported to have limited homology to the dimerisation helices of AraC, UreR and XylS (19). As described above, however, a portion of Rns (residues 1-154) containing this potential α-helix did not homodimerise (19). Although the coiled coil prediction program COILS (241) predicts a coiled coil structure in the vicinity of residues 131-146 of Rns (α-helix 3, Fig. 3.4), another coiled coil is predicted with even greater probability between residues 195-222 (α-helix 7) (Fig. 4.14). Therefore this sequence, situated between the two putative HTH motifs of Rns, may be involved in dimerisation.

The means by which a dimer of Rns would bind to the previously identified DNA binding sites at the *coo* and *rns* promoters is not clear. As the Rns binding sites are asymmetric it has been assumed that an asymmetric Rns monomer binds at each site, placing a recognition helix from each of its HTH motifs in adjacent major groove regions of the DNA (268, 269). It is possible, however, that Rns dimerises via an antiparallel coiled coil formed by the α-helix located between its HTH motifs. One of the Rns monomers comprising the dimer could then provide HTH1 for binding in one major groove of the DNA while the second monomer could provide HTH2 for DNA binding in the other major groove. Additional studies are necessary to elucidate the mechanism by which Rns dimerises and binds to DNA and the possible involvement of a flexible linker region in the activity of the protein.

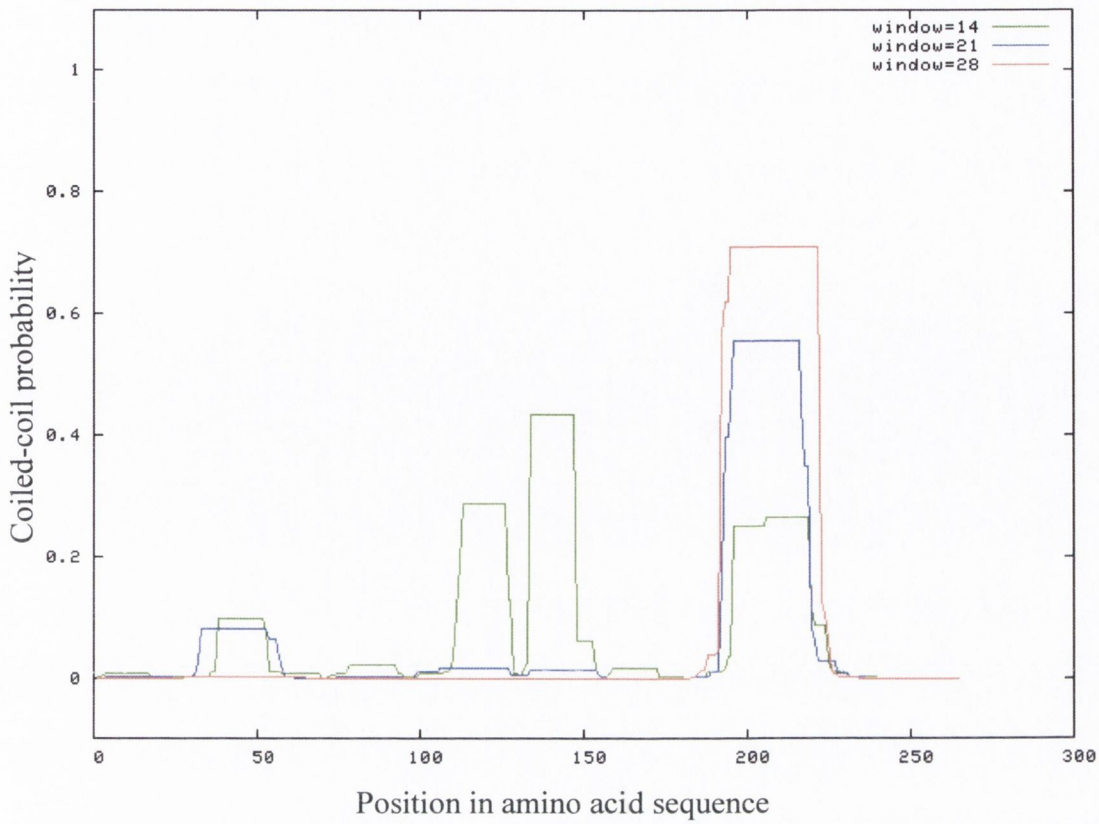

Fig. 4.14. COILS output for Rns. The amino acid sequence of Rns was analysed by COILS, a programme that calculates the probability that a protein sequence will adopt a coiled coil conformation. The probability is obtained in scanning windows of 14, 21 or 28 residues. Using the 14 residue scanning window the programme identified potential coiled coil structures between residues 100 and 150 of Rns and between residues 195 and 222. Using the 21 and 28 residue scanning windows the probability of a coiled coil structure between residues 195 and 222 of Rns is even greater.

Chapter 5 Regulation of the *coo* promoter

5.1 Introduction

The Rns transcriptional regulator is required for the expression of CS1 fimbriae in enterotoxigenic *E. coli* (49). Rns activates CS1 expression directly by binding to two sites upstream of the *coo* promoter. Both of these sites are required for full Rns-dependent transcription from the *coo* promoter. Site I is centred at position -112 and site II is centred at position -44 (numbering relative to the transcription start site) (268). Three thymines present within both sites were found to be critical for Rns binding (268). Subsequent analysis revealed that these thymine residues are also present in the three Rns binding sites at the *rns* promoter (269). Indeed all five of these experimentally determined Rns binding sites have a conserved arrangement of thymines. At each site two thymines are separated by an intervening adenosine and the third thymine is seven nucleotides upstream of this conserved TAT motif, but on the opposite strand (Fig. 1.6) (269).

In *E. coli*, specific activators (like Rns) and specific repressors (such as the *lac* repressor) control the expression of certain genes by binding to their promoters and up- or down-regulating transcription respectively (41). However, in addition to these dedicated transcription factors, a class of DNA-binding proteins including IHF, Fis, HU and H-NS act as global regulators of a wide variety of unrelated genes (93, 416). H-NS is a small (~16 kDa) abundant nucleoid-associated protein that plays a dual role in compacting and structuring DNA and in regulating transcription (92). Upon binding, H-NS affects the topology of DNA. H-NS has been demonstrated to condense DNA, constrain DNA supercoils and induce DNA bending (70, 386, 408). It is via these changes in DNA topology that H-NS influences nucleoid structure and transcription. H-NS mainly acts as a negative regulator, repressing the transcription of a large number of genes. Many of the genes modulated by H-NS are responsive to environmental signals such as temperature, osmolarity and pH (11). H-NS has also been found to silence horizontally acquired genes and is therefore proposed to be involved in protecting the cell from the inappropriate expression of such foreign DNA (91). H-NS contains an N-terminal dimerisation domain, a C-terminal DNA-binding domain and a flexible linker that connects the two domains and is required for higher order oligomerisation of the protein (329). In addition to H-NS many gram negative bacteria express StpA, a paralogue of H-NS (445). A third H-NS-like protein, Sfh, is encoded by a gene on a R27-like plasmid in *S. flexneri* (22). StpA and Sfh are structurally and

functionally homologous to H-NS and all three proteins can form heteromeric complexes (79). Different combinations of these proteins may have distinct roles in the cell.

H-NS was initially reported to bind to DNA with little sequence specificity, although it was found to preferentially bind to intrinsically curved and AT rich DNA (71). It was recently demonstrated, however, that H-NS binds sequence specifically to two identical high-affinity binding sites in the *E. coli proU* operon (38). It is thought that H-NS first binds (as a dimer at least) to such high-affinity or nucleation sites from where it then polymerises along the DNA, co-operatively binding to further low-affinity sites, leading eventually to the formation of extended nucleoprotein complexes (38, 330). It is these higher order oligomeric nucleoprotein complexes that are responsible for transcription repression by H-NS either by acting to exclude the binding of RNA polymerase or by trapping RNA polymerase at the promoter. Trapping of RNA polymerase, at the *rrnB* ribosomal RNA gene promoter for example, is brought about when H-NS complexes bound to a region upstream of the promoter interact with H-NS bound downstream of the promoter, resulting in the looping of the intervening DNA (72). RNA polymerase is physically trapped within this loop, thus transcript elongation is prevented. In *E. coli*, the position of H-NS binding correlates strongly with that of RNA polymerase binding (154, 291). In contrast, H-NS in *Salmonella* was not demonstrated to co-localise with RNA polymerase (237). Thus it appears that in *Salmonella* H-NS silencing occurs via the occlusion mechanism, while in *E. coli* the trapping mechanism is employed.

H-NS negatively regulates transcription from the *coo* promoter (270). H-NS was reported to predominantly exert its inhibitory effect in a region downstream of the promoter (+7 to +929) although it was suggested that H-NS may also act upstream of this region (270). However, the H-NS protein was not demonstrated to bind directly at the *coo* locus. H-NS has also been found to repress transcription at several other promoters that are activated by AraC family members including the VirF-regulated *virB* promoter, the CfaD-regulated *cfa* promoter, the UreR-regulated *ureR* promoter and the ToxT-regulated *ctx* promoter (58, 199, 402, 444). This has prompted a suggestion that in general, one of the functions of AraC family members may be to antagonise repression by H-NS at their target genes, i.e. to act as anti-repressors. Indeed, Rns was demonstrated to be involved in circumventing H-NS-mediated silencing of the *coo* promoter (270). The aims of the study presented in this chapter were to further analyse the binding of Rns to the *coo* promoter and to investigate the regulation of the promoter.

5.2 Results

5.2.1 Effects of mutations in the *coo* promoter region on Rns binding and promoter activity

5.2.1.1 Interaction of MBP-Rns with mutant *coo* promoter derivatives

As detailed in Chapters 3 and 4, when EMSAs were performed to study the binding of wild type MBP-Rns to *coo* promoter DNA, three main sets of protein-DNA complexes were observed (see Fig. 3.14, for example). To gain an improved understanding of the interaction of Rns with the *coo* promoter region and to learn more about these protein-DNA complexes, EMSAs were conducted with *coo* promoter DNA probes in which residues known to be important for Rns binding were mutated. All of the Rns binding sites at the *coo* and *rns* promoters contain a conserved TAT motif (268, 269). It has been demonstrated that at both of its binding sites located upstream of the *coo* promoter, Rns makes critical interactions with the thymine nucleotides of this TAT sequence (268). Therefore alteration of the TAT motif within *coo* promoter binding sites I and II would be expected to interfere with Rns binding there. The insert present in the pBSKII+ derived plasmid pCS1 contains bases -393 to +62 of *coo* DNA (numbering relative to the transcription start site) and therefore includes the *coo* promoter region. Site-directed mutagenesis was performed on pCS1 using the QuikChange kit (Stratagene) and the primer pair Site_I_F and Site_I_R or the primer pair Site_II_F and Site_II_R (Table 2.3). The *coo* promoter region within the resulting plasmids contains a CGC substitution for the TAT sequence within Rns binding site I (pCS1-SiteI) or within binding site II (pCS1-SiteII). Further site-directed mutagenesis conducted using pCS1-SiteII and the primers Site_I_F and Site_I_R produced the plasmid pCS1-SiteI&II in which the TAT sequence within both binding sites had been altered to CGC. All of the plasmids were sequenced to verify that the substitutions had occurred.

To construct the mutant probes for use in EMSAs, a 195 bp DNA fragment comprising the *coo* promoter region from -196 to -2 (relative to the transcription start site) was amplified by PCR using the 5′-biotinylated primers VM1 and VM2 (Table 2.3) and pCS1-SiteI, pCS1-SiteII or pCS1-SiteI&II as a template. EMSAs were then

performed to compare the interaction of MBP-Rns with a wild type *coo* promoter probe to the interactions of the protein with the mutated probes (Fig. 5.1). MBP-Rns bound less well to the mutated probes than it did to the wild type probe, thus confirming the importance of the TAT motif within sites I and II for Rns binding at the *coo* promoter. In the presence of 68 nM MBP-Rns, no free wild type probe remained. As determined by densitometry, at that protein concentration approximately 16% of both the site I mutant and the site II mutant probes was unbound while 97% of the site I&II mutant probe had not been bound. Despite the reduced binding affinity, MBP-Rns could still form protein-DNA complexes with the single site mutant *coo* promoter probes. Therefore disruption of one of the binding sites did not appear to prevent Rns from binding to the remaining unaltered site. In the presence of 68 nM MBP-Rns the wild type, site I mutant and site II mutant probes had formed two sets of complexes. The higher molecular weight complex was formed to a greater extent by the wild type probe while the mutant probes formed significantly more of the lower molecular weight complex. In addition, the lower molecular weight complexes formed by each probe had different mobilities. When formed by MBP-Rns and wild type *coo* promoter DNA, the lower molecular weight complex could appear as a doublet (see Fig. 4.3 for example). Therefore the EMSAs performed with the mutated *coo* probes revealed that this complex appears to result from MBP-Rns occupying binding site I alone and binding site II alone and that the mobility of the former is slightly less than that of the latter. It appears that the higher molecular weight complex is due to MBP-Rns occupying both binding sites I and II. The single site mutants formed a limited amount of this complex, however it was not formed when MBP-Rns was incubated with the site I&II mutant probe. It is possible that MBP-Rns bound to one wild type *coo* promoter site may facilitate at least some binding of protein to the other mutated site. This is suggestive of co-operative binding. Finally, the uppermost complex of least mobility was formed by all of the probes. That this complex was formed even by the site I&II mutant *coo* promoter probe supports the assumption that it is caused by non-specific binding in the presence of excess protein. Thus the use of mutated probes in EMSAs with MBP-Rns provided a clearer understanding of the interaction of Rns with *coo* promoter DNA.

5.2.1.2 Construction of reporter plasmids with variant *coo* promoter regions

pZep08-based reporter plasmids with *coo* promoter fragments containing the CGC for TAT substitution in Rns binding sites I and/or II fused upstream of a promoterless *gfp* gene were constructed. The *coo* promoter region (-393 to +62) containing the CGC substitution at Rns binding site I was amplified by PCR using the primers CS1-Blunt and CS1-XbaI (Table 2.3) and the plasmid pCS1-SiteI as a template. The PCR product was digested with *Xba*I and ligated into the pZep08 vector digested with *Sma*I and *Xba*I. DNA sequencing of the resulting plasmid revealed that the *coo* promoter insert present (-359 to +62) was shorter than expected. This plasmid was named pCoo-359-I (Table 2.2 and Fig. 5.2) where the number (-359) refers to the upstream end point of the *coo* promoter region present in the insert, and the roman numeral indicates that the insert contains a CGC for TAT substitution in Rns binding site I. To produce reporter plasmids with inserts that were equivalent in length to the insert in pCoo-359-I, the *coo* promoter regions within pCooGFP, pCS1-SiteII and pCS1-SiteI&II were amplified by PCR using the primers New_CS1-Blunt (Table 2.3) and CS1-XbaI. The PCR products were digested with *Xba*I and cloned into pZep08 digested with *Sma*I and *Xba*I. The inserts present in the resulting plasmids (pCoo-88, pCoo-96, pCoo-352, pCoo-359, pCoo-201-II, pCoo-359-II, pCoo-146-I&II and pCoo-359-I&II, Table 2.2 and Fig. 5.2) were sequenced and although each was found to have the expected downstream end-point of +62 (relative to the *coo* transcription start site), the upstream end-points of the inserts ranged from -359 (as expected) to -88. These plasmids were named according to the system described above for pCoo-359-I. To generate reporter plasmids containing *coo* promoter region inserts of the length (-393 to +62) present in the reporter plasmids (pCooGFP and pCooGFP-2) used in all previous experiments, it was necessary to perform site-directed mutagenesis as described in Section 5.2.1.1 on pCooGFP. The resulting plasmids, named pCooGFP-I, pCooGFP-II and pCooGP-I&II (Table 2.2 and Fig. 5.2), were sequenced to verify that the inserts were of the correct length and contained the CGC for TAT substitutions.

5.2.1.3 Activity of the *coo* promoter variants *in vivo*

It was noticed that *E. coli* XL-1 cells containing only the reporter plasmids with mutated and/or reduced length *coo* promoter inserts were fluorescent. This was unexpected as it indicated that the *coo* promoter variants were displaying transcriptional

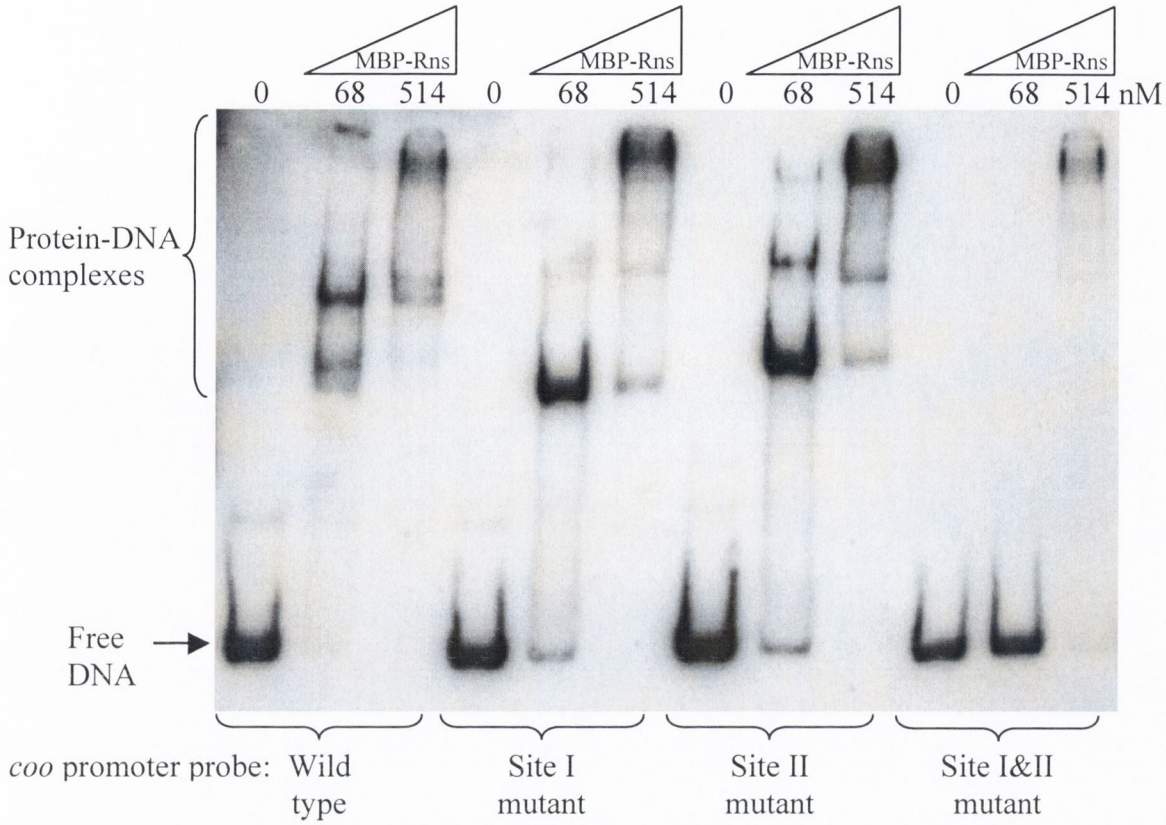

Fig. 5.1. Interaction of MBP-Rns with wild type and mutant derivatives of *coo* promoter DNA. EMSAs were carried out with approximately 20 pg of a wild type or mutated *coo* promoter DNA probe (as indicated at the bottom of the figure) and a range of concentrations of MBP-Rns (as indicated at the top of the figure). The unbound DNA fragments and protein-DNA complexes are indicated on the left.

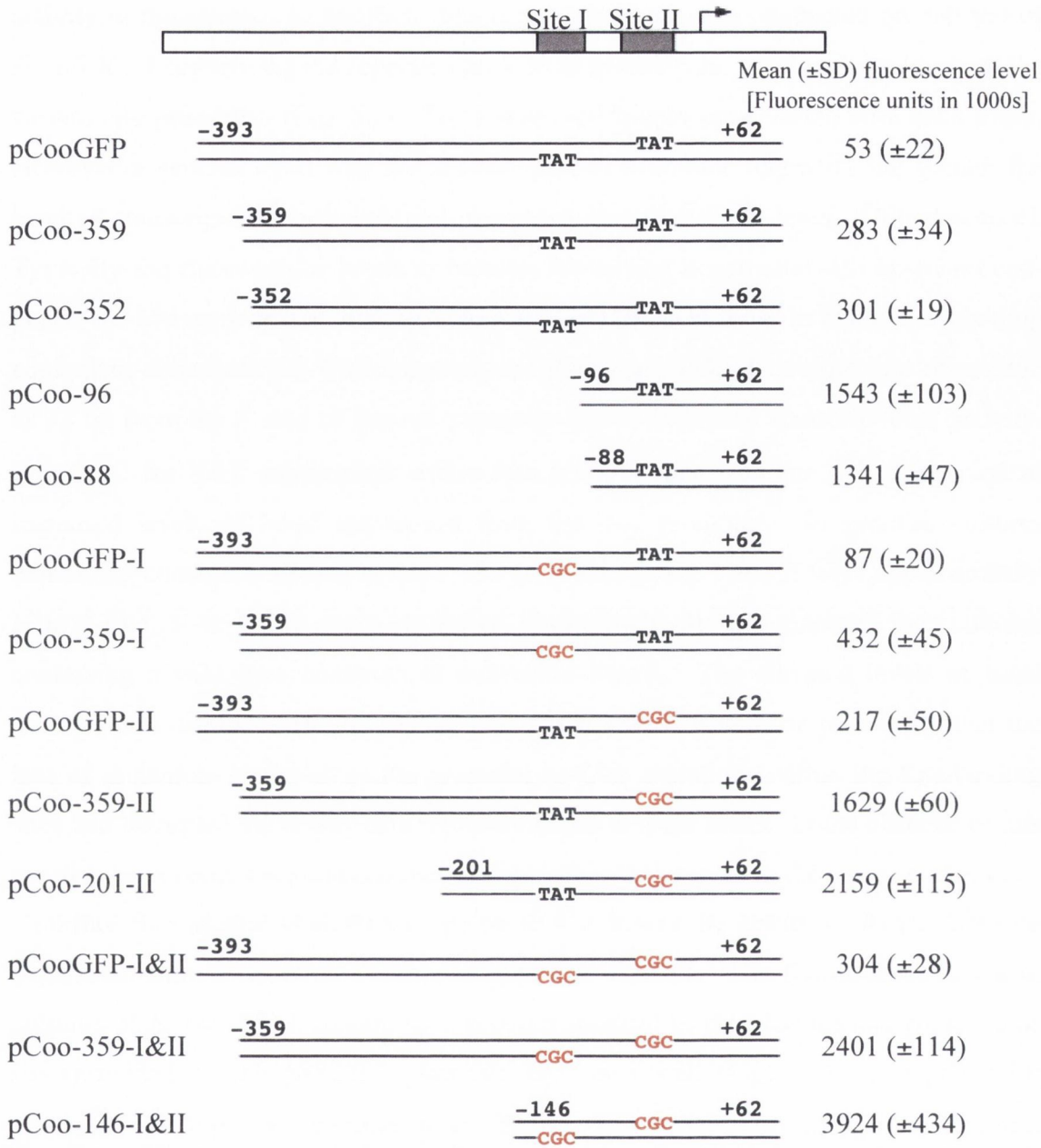

Fig. 5.2. Variant *coo* promoter fragments present in reporter plasmids and associated levels of basal promoter activity. At the top of the figure a schematic diagram represents the wild type *coo* promoter region. Rns binding sites I and II are shown as grey boxes and the transcription start site is represented as an arrow. Underneath and in relationship to this schematic diagram, different fragments of the *coo* promoter region cloned upstream of a promoterless *gfp* gene are represented as lines next to the reporter plasmid name. The end points of the *coo* promoter fragments are shown above the lines. Within Rns binding sites I and II each *coo* promoter fragment contained either the wild type TAT motif (coloured black) or a CGC substitution (coloured red). The TAT and CGC nucleotides are not to scale. The fluorescence level in cultures of *E. coli* XL-1 harbouring each plasmid is shown on the right in fluorescence units (in 1000s). Standard deviation (SD) values are indicated. This data represents averages of triplicate measurements on duplicate cultures. Measurements were performed independently at least three times; a representative data set is shown.

variable. Aside from this aberrant result, the loss of the upstream region (-393 to -359) did not appear to affect the ability of Rns to activate the *coo* promoter. However, the TAT motifs present in binding sites I and II are important for full Rns-dependent activation of the *coo* promoter *in vivo*.

5.2.2 Effect of H-NS on *coo* promoter activity

5.2.2.1 Does H-NS bind directly to the *coo* promoter region?

The transcriptional activity of the variant *coo* promoters in the absence of activation by Rns was postulated to be due to a reduced ability of a repressor to act at these promoters. The H-NS protein is a repressor of *coo* transcription (270). Therefore the possibility that it was H-NS-mediated repression that had been relieved by the loss of upstream sequences of the *coo* promoter and the CGC for TAT substitutions within the Rns binding sites was investigated. H-NS-dependent repression was reported to mainly occur in a region downstream of the *coo* transcription start site (+7 to +929) (270). Thus it was necessary to determine if H-NS could also repress the portion of the *coo* promoter present in the pCooGFP reporter plasmid used in this study. *E. coli* MC4100 and its *hns*-negative isogenic derivative, PD32, were transformed with pCooGFP and subjected to fluorimetric analysis (Fig. 5.4). Fluorescence levels in the *hns*$^-$ strain were 4-fold higher than those in the *hns*$^+$ strain ($P < 0.005$), demonstrating that the H-NS protein represses the *coo* promoter fragment comprising bp -393 to bp +62.

H-NS preferentially binds to curved DNA sequences (439). Previously, two-dimensional electrophoresis was used to demonstrate the presence of curved DNA in a region upstream of the *coo* operon (between bp -884 and bp -120) (373). In addition, analysis of the *coo* promoter region from -393 to +62 using the web-based bend.it DNA curvature prediction programme (http://www.icgeb.trieste.it/dna) revealed the presence of predicted areas of curvature (Fig. 5.5). Thus it is possible that H-NS binds directly to the *coo* promoter region. To examine this possibility, EMSAs were performed to monitor the interaction of H-NS with *coo* promoter DNA. The His-tagged H-NS expression vector pHNSHIS (Table 2.2) was constructed. His-tagged H-NS was overexpressed and purified as described in Section 2.3.2.1 (Fig. 5.6). Increasing concentrations of purified His-tagged H-NS protein were incubated with the 195 bp

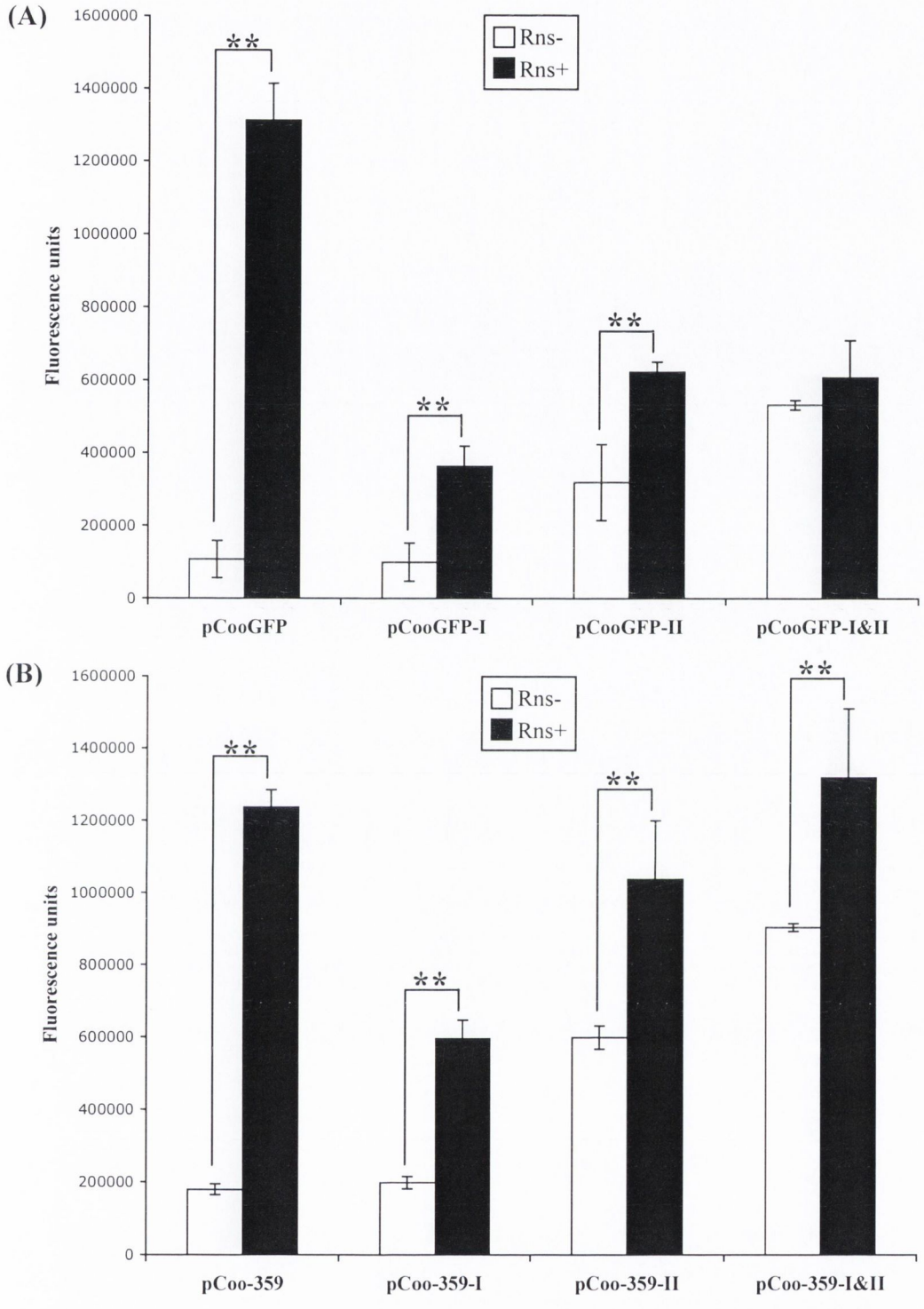

Fig. 5.3. Transactivation of *coo* promoter variants by Rns. The fluorescence levels in cultures of *E. coli* XL-1 harbouring the indicated reporter plasmids alone (Rns-) or in addition to the Rns expressing plasmid pSS2192 (Rns+) were measured. Data represents averages of triplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Sets of bacterial cultures analysed on different days are shown in separate panels. Statistical significance is indicated by **, where $P < 0.005$.

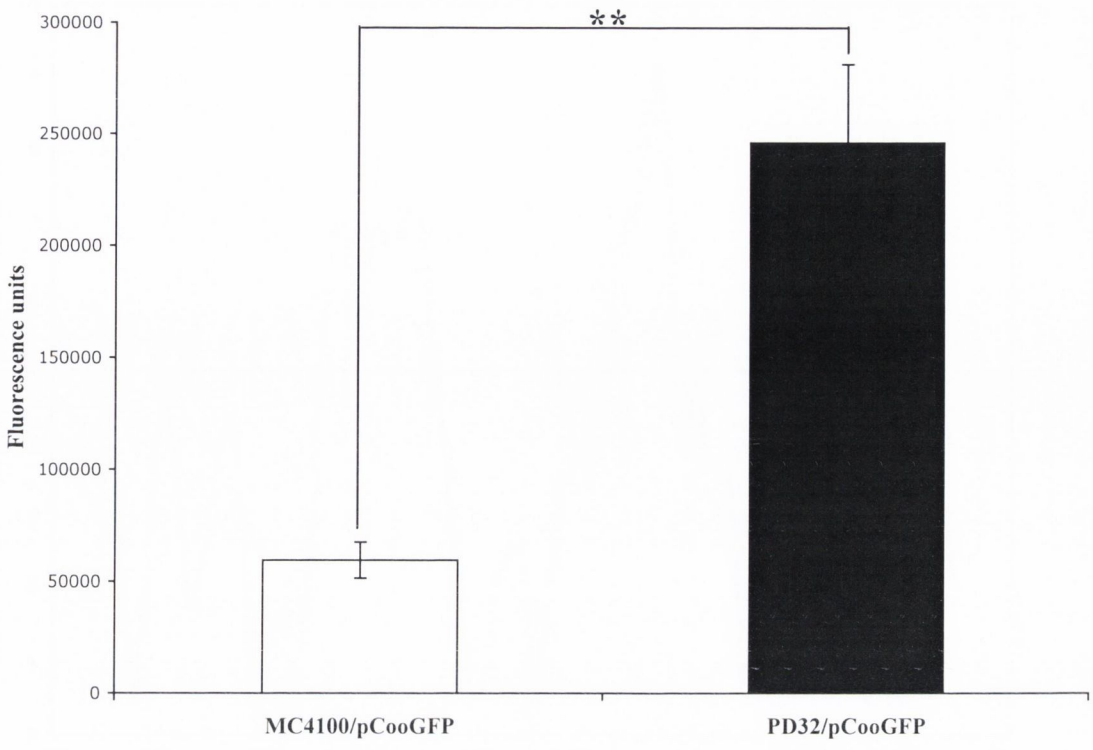

Fig. 5.4. Derepression of the *coo* promoter in the absence of H-NS. The fluorescence levels in cultures of *E. coli* MC4100 (*hns*$^{+}$) or PD32 (*hns*$^{-}$) harbouring the reporter plasmid pCooGFP were measured. Data represents averages of triplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least three times; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

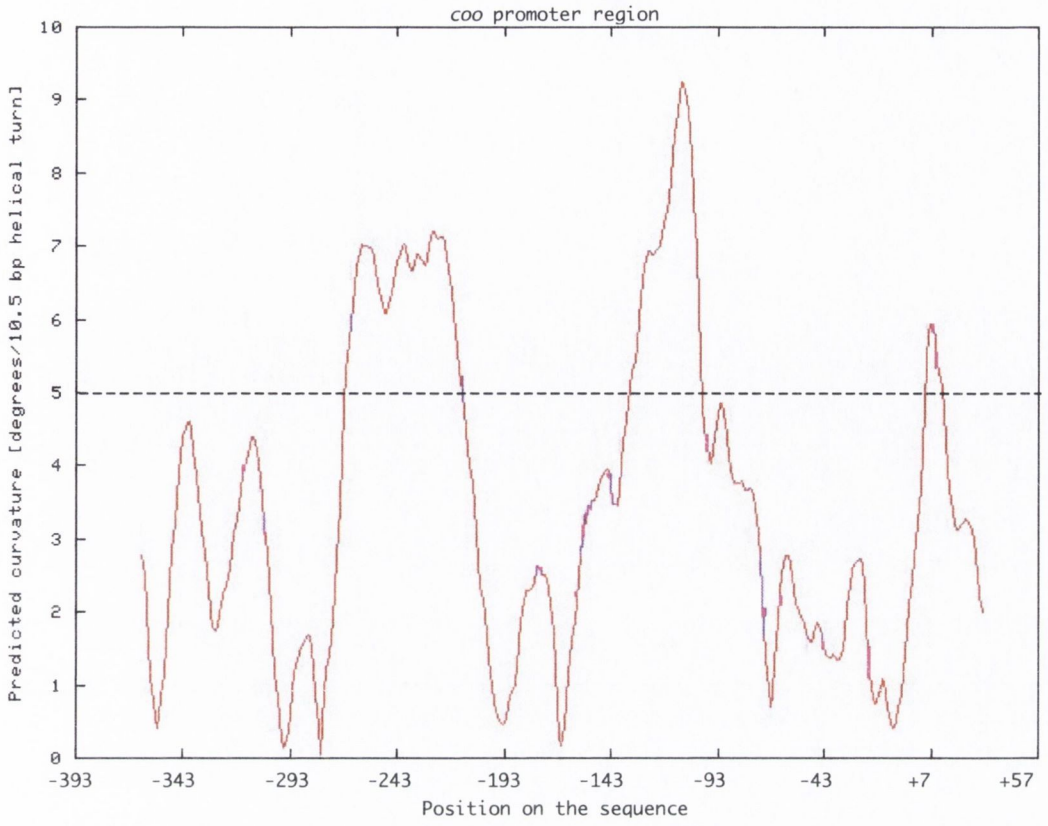

Fig. 5.5. *In silico* prediction of intrinsic curvature of the *coo* promoter region. The DNA sequence of the *coo* promoter region was analysed using the DNA curvature prediction programme, bend.it. Predicted curvature is expressed as degrees/helical turn and plotted against the position on the sequence (numbering relative to the *coo* transcription start site). Regions with > 5 degrees per helical turn of DNA (dashed line) represent predicted curved sequences.

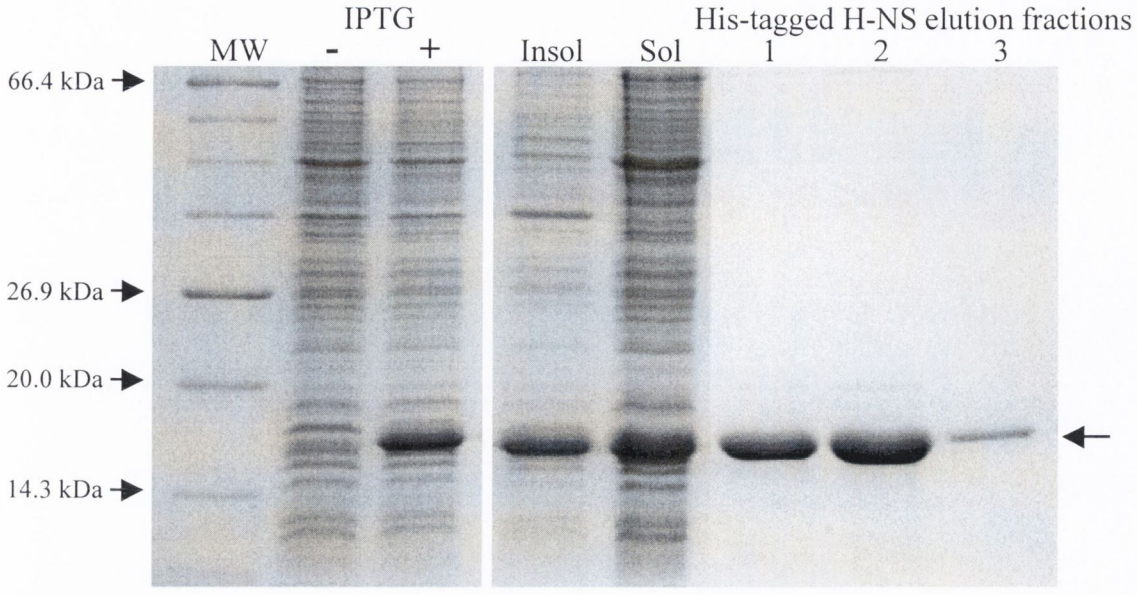

Fig. 5.6. Expression and purification of His-tagged H-NS. SDS-PAGE analysis of uninduced (IPTG -) and induced (IPTG +) cultures of *E.coli* T7 Express I^q harbouring pHNSHIS, insoluble (Insol) and soluble (sol) fractions of the induced cultures, and fractions of purified His-tagged H-NS. The positions of selected protein molecular mass markers (MW) are indicated on the left. An arrow on the right indicates the position of His-tagged H-NS.

wild type *coo* promoter probe (-196 to -2) in the presence and absence of an excess of non-specific competitor DNA (poly(dI.dC)) (Fig. 5.7 (A)). In the absence of poly(dI.dC), the *coo* promoter probe was completely bound when 0.56 μM His-tagged H-NS was added and a low mobility protein-DNA complex was detected. Increasing concentrations of H-NS led to the formation of protein-DNA complexes of even lower mobility. This suggested that H-NS interacts with several sites within the tested *coo* promoter region or that multiple H-NS molecules were binding. H-NS also bound the *coo* promoter probe when an excess of poly(dI.dC) was present. The binding affinity was less than was observed in the absence of non-specific competitor DNA, as 1.13 μM His-tagged H-NS was required to completely bind the probe. However, the EMSA results demonstrated that H-NS binds directly to the *coo* promoter. To examine if there were further H-NS binding sites upstream of position -196 of the *coo* promoter region, i.e. within the areas absent in some of the variant *coo* promoter reporter plasmids, a more extensive EMSA probe was constructed. A 431 bp fragment of DNA comprising the *coo* promoter region from -371 to +60 was amplified by PCR using the 5′-biotinylated primers CSBS1 and CSBS2 (Table 2.3) and pCooGFP as a template. The EMSAs described above were repeated with this extended probe (Fig. 5.7 (B)). As determined by densitometry, in the presence of 0.56 μM His-tagged H-NS approximately 90% of the 195 bp probe remained unbound while 80% of the 431 bp probe was not bound. Aside from this small difference, however, the affinity of H-NS for the 431 bp *coo* promoter probe did not appear to be significantly greater than that for the shorter probe. Addition of increasing concentrations of His-tagged H-NS to the 431 bp probe produced mobility shifts similar to those observed when H-NS was incubated with the 195 bp probe. Therefore, based on this assay, it was not possible to attribute the elevated basal transcriptional activity of the reduced length *coo* promoters to a loss of H-NS binding to linear DNA.

5.2.2.2 Effect of site I and II mutations on H-NS repression at the *coo* promoter

It remained possible that the CGC for TAT substitutions within Rns binding sites I and II had diminished H-NS-dependent repression of the *coo* promoter, perhaps by reducing H-NS binding there. The *coo* promoter sequences containing the site I and/or II mutations were examined using the bend.it programme. The CGC substitutions did not reduce the levels of predicted curvature present in the *coo* promoter (results not

shown). In addition to binding to curved DNA, H-NS was shown to specifically bind to two identical high-affinity binding sites (AATATATCGA) within a regulatory element of the *E. coli proU* gene (38). The DNA sequence from position -985 to +729, relative to the *coo* transcription start site, was searched for close matches to this 10 bp motif. A sequence with only one mismatch (AATAAATCGA) is present between positions -124 and -115 (Fig. 1.6). This sequence is within the area of the *coo* promoter protected from DNase I by MBP-Rns binding at site I (268). On the reverse strand, the final nucleotide of this sequence corresponds to the first thymine of the TAT motif of Rns binding site I. Therefore this nucleotide was altered in the site I and site I&II *coo* promoter mutants, creating a second mismatch to the *proU* sequence and potentially affecting H-NS binding to the *coo* promoter. EMSAs were performed to assess whether H-NS binding had indeed been reduced by the CGC for TAT substitutions at sites I and/or II. Increasing concentrations of His-tagged H-NS were incubated with the 195 bp wild type, or site mutated, *coo* promoter probes in the presence and absence of an excess of poly(dI.dC). In the absence of non-specific competitor DNA, H-NS appeared to bind to the wild type and mutated *coo* promoter probes in a similar manner (Fig. 5.8 (A)). When poly(dI.dC) was included, H-NS appeared to bind less well to the mutation containing probes than it did to the wild type *coo* promoter probe. As determined by densitometry, approximately 15% of the wild type probe remained unbound in the presence of 1.13 μM His-tagged H-NS (Fig. 5.8 (B)). In contrast, when 1.13 μM His-tagged H-NS was incubated with the mutant probes densitometric analysis revealed that approximately 40% of the site II mutant probe remained unbound while approximately 60% of both the site I and the site I&II mutant probes were not bound. However, when 2.27 μM His-tagged H-NS was added each of the *coo* promoter probes were bound equally. Therefore, when conducted in the presence of non-specific competitor DNA, the EMSA results appeared to indicate that the CGC for TAT substitutions at Rns binding sites I and/or II had decreased the affinity of H-NS binding to the *coo* promoter.

To examine if the site mutations had decreased H-NS-dependent repression of *coo* transcription, the GFP expression levels of a selection of the site I and/or II mutated *coo* promoter reporter plasmids were measured in a wild type (*E. coli* MC4100) and *hns* mutant (PD32) background (Fig. 5.9). If the ability of H-NS to repress the mutated *coo* promoters were reduced, the expression levels from these promoters would not be expected to increase significantly in the absence of H-NS. However, like the H-NS repressed wild type *coo* promoter, expression from each of the CGC substitution

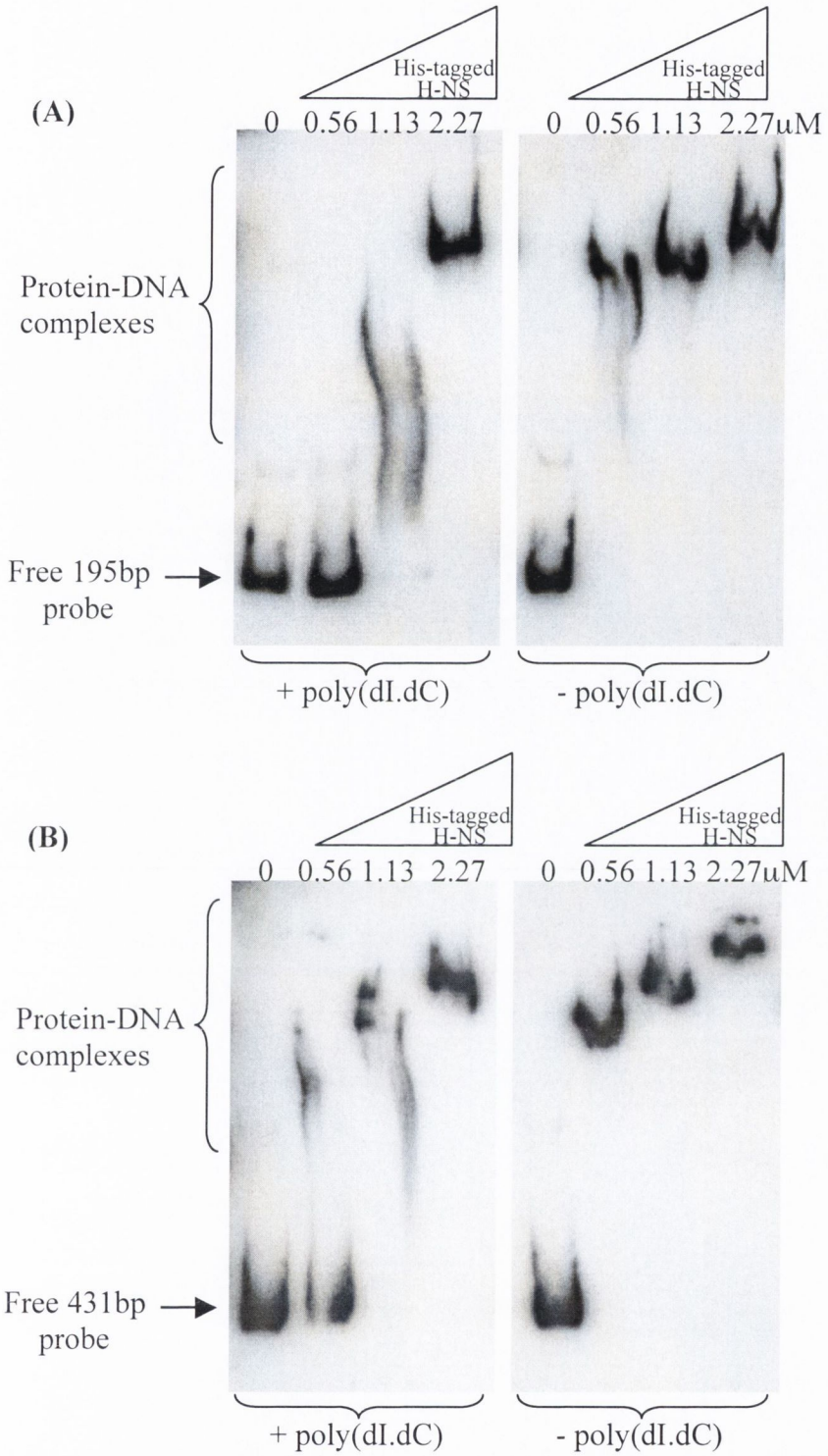

Fig. 5.7. Interaction of His-tagged H-NS with *coo* promoter DNA. EMSAs were carried out with approximately 20 pg of **(A)** a 195 bp biotinylated *coo* promoter DNA probe (-196 to -2) or **(B)** a 431 bp biotinylated *coo* promoter DNA probe (-371 to +60) and a range of concentrations of His-tagged H-NS as indicated at the top of each panel. EMSAs were conducted in the presence or absence of non-specific competitor DNA (poly(dI.dC)) as indicated underneath each panel. The unbound DNA fragments and protein-DNA complexes are indicated on the left.

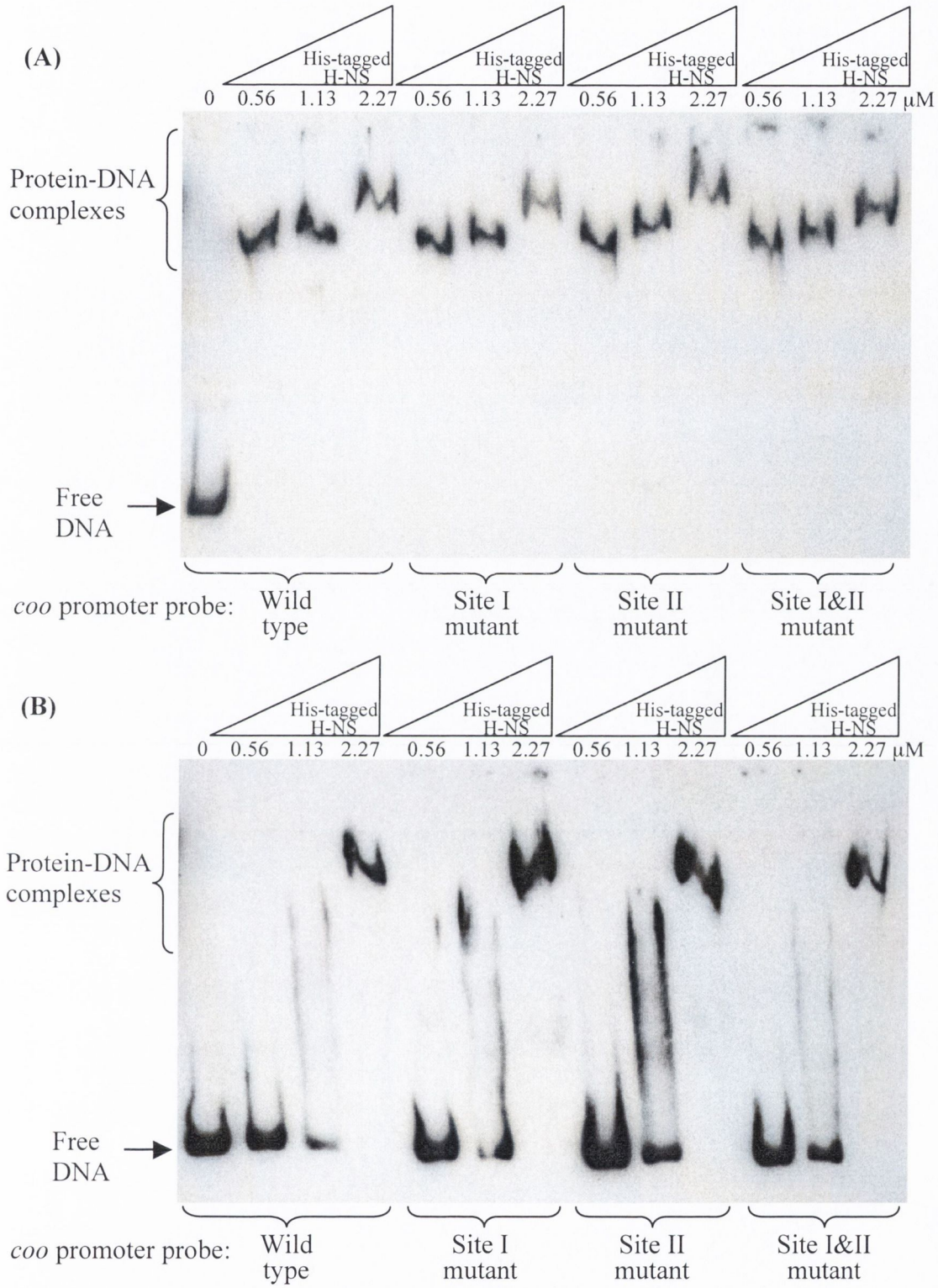

Fig. 5.8. Interaction of His-tagged H-NS with wild type and mutant derivatives of *coo* promoter DNA. EMSAs were carried out with approximately 20 pg of a wild type or mutated *coo* promoter DNA probe (as indicated at the bottom of the panels) and a range of concentrations of His-tagged H-NS (as indicated at the top of the panels). EMSAs were conducted in the absence **(A)** or presence **(B)** of non-specific competitor DNA (poly(dI.dC)). The unbound DNA fragments and protein-DNA complexes are indicated on the left.

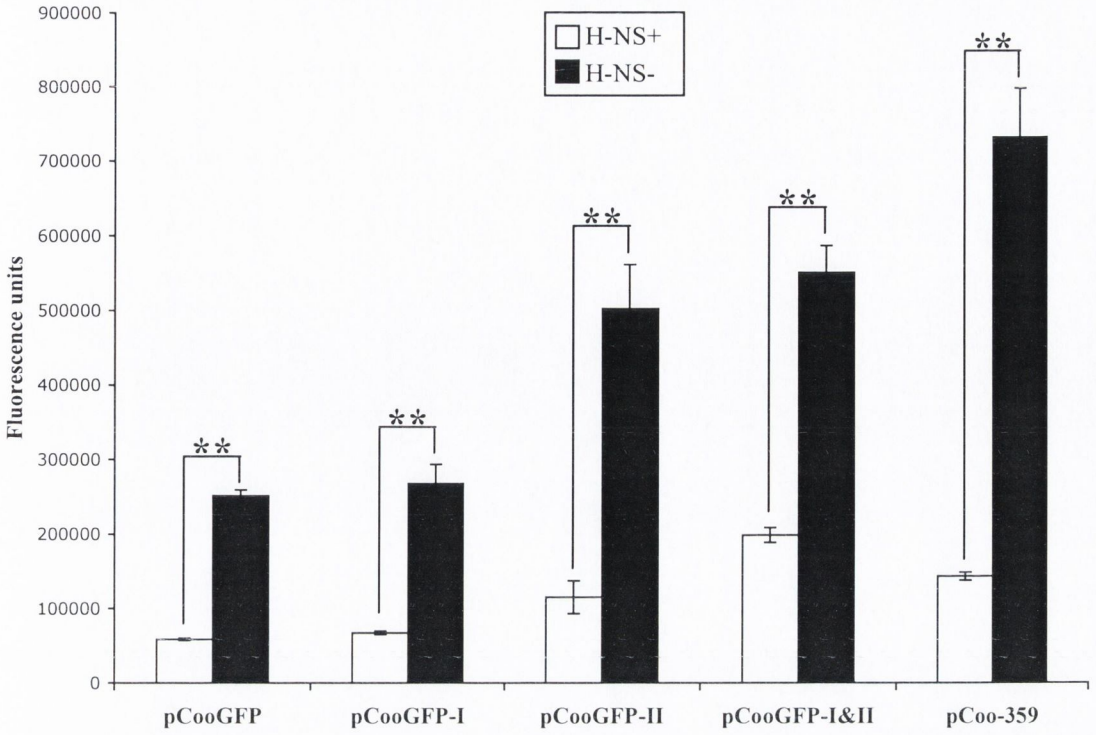

Fig. 5.9. Expression levels of *coo* promoter variants in the presence and absence of H-NS. The fluorescence levels in cultures of *E. coli* MC4100 (*hns*$^{+}$, white bars) or PD32 (*hns*$^{-}$, black bars) harbouring the indicated reporter plasmids were measured. Data represents averages of triplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

containing *coo* promoters was 3- to 4-fold higher in the absence of H-NS than it was in the presence of H-NS ($P < 0.005$). A similar trend was observed for the reduced length wild type *coo* promoter in the reporter plasmid pCoo-359. Thus, the results of these experiments indicate that the site mutations within the Rns binding sites at the *coo* promoter had not led to a reduced level of H-NS-mediated repression there.

5.2.3 Differential regulation of the *coo* promoter by H-NS and Rns

5.2.3.1 Effect of Rns on the *coo* promoter in the presence and absence of H-NS

As described above, *coo* promoter expression is derepressed in the absence of *hns*. The Rns expressing plasmid, pSS2192, was transformed into *E. coli* MC4100/pCooGFP and PD32/pCooGFP to compare the effects of Rns on *coo* promoter activity in wild type and *hns* mutant strains, respectively. The resulting transformants produced widely variable levels of fluorescence (results not shown). It appeared that the *E. coli* strains MC4100 and PD32 could not tolerate the presence of both pCooGFP and pSS2192. Therefore a different strategy was pursued. The alternative reporter plasmid pCooGFP-2 was used in place of pCooGFP. In addition, the pASK75 vector (Table 2.2) was used to construct a plasmid in which incremental levels of Rns could be inducibly expressed from a promoter other than its native *rns* promoter. The *rns* ORF was amplified by PCR using the primers RnsF_Xba and RnsR_Hind (Table 2.3) and pSS2192 as a template. The PCR product was digested with *Xba*I and *Hin*dIII and ligated into pASK75 that had previously been digested with the same enzymes to remove a segment of the vector encoding the OmpA signal peptide and a strep-tag. The resulting plasmid, pASKrns, was sequenced to confirm that the insert was correct. This plasmid expresses Rns from the *tetA* promoter/operator, which is tightly controlled by the TetR repressor. Expression from the *tet* promoter is induced by the addition of anhydrotetracycline (AHT). As both pASK75 and pASKrns carry only ampicillin resistance, it was necessary to modify these plasmids for use in the ampicillin resistant *E. coli* strain PD32. This was achieved by ligating a 2 kb spectinomycin resistance cassette, excised from plasmid pKRP13 (Table 2.2) by *Hin*dIII digestion, into *Hin*dIII digested pASKrns and pASK75. The structures of the resulting plasmids, named pASKrnsSpec (Fig. 5.10) and pASK75Spec (Table 2.2), were confirmed by restriction endonuclease mapping.

Both *E. coli* MC4100/pCooGFP-2 and PD32/pCooGFP-2 were transformed with either pASKrnsSpec or pASK75Spec. The resultant transformants were grown (in duplicate) to mid-logarithmic phase. The pASKrnsSpec containing cultures were divided into several aliquots to which increasing levels of the AHT inducer were added. All cultures were then grown for a further two hours before they were subjected to fluorimetric analysis. In the absence of Rns, *coo* promoter activity was approximately 4-fold higher ($P < 0.005$) in the H-NS$^-$ host background (PD32) than in the H-NS$^+$ background (MC4100) (Fig. 5.11). The fluorescence levels in these pASK75Spec-harbouring cultures were unaffected by the addition of AHT (results not shown). However, growth of the strains containing pASKrnsSpec in the presence of increasing amounts of AHT resulted in increasing levels of fluorescence (Fig. 5.12 (A)). For wild type *E. coli* MC4100/pCooGFP-2/pASKrnsSpec, *coo* promoter expression was activated over 12-fold ($P < 0.005$) when Rns was fully induced (200 μg/L AHT) compared to in the absence of AHT induction. In the *hns* mutant PD32, full Rns induction led to a 3-fold increase in *coo* expression. Thus Rns does not solely act to circumvent H-NS-mediated repression as even in the absence of H-NS, Rns can elevate *coo* promoter expression.

Western immunoblotting of samples of the cultures used in the fluorimetric analysis confirmed that increasing amounts of the AHT inducer resulted in increasing levels of Rns expression from pASKrnsSpec (Fig. 5.12 (B)). Rns expression from pASKrnsSpec was not detected in the absence of AHT induction. The western immunoblot analysis revealed that in the wild type MC4100 background, 20 μg/L AHT was sufficient to induce Rns expression while in the *hns* mutant PD32, Rns was only detected following induction with 50 μg/L AHT. In addition, potential Rns breakdown products were more prevalent in the *hns* mutant strain.

5.2.3.2 Competition between H-NS and Rns for the *coo* promoter region

Competition EMSAs were performed to assess whether the H-NS and Rns proteins compete with each other to interact with *coo* promoter region DNA. His-tagged H-NS and MBP-Rns were incubated separately with the 195 bp wild type *coo* promoter DNA probe to allow protein-DNA complexes to form. In an attempt to replace the pre-bound protein, a second incubation was then conducted during which increasing amounts of MBP-Rns were added to the H-NS binding reactions and increasing amounts of His-

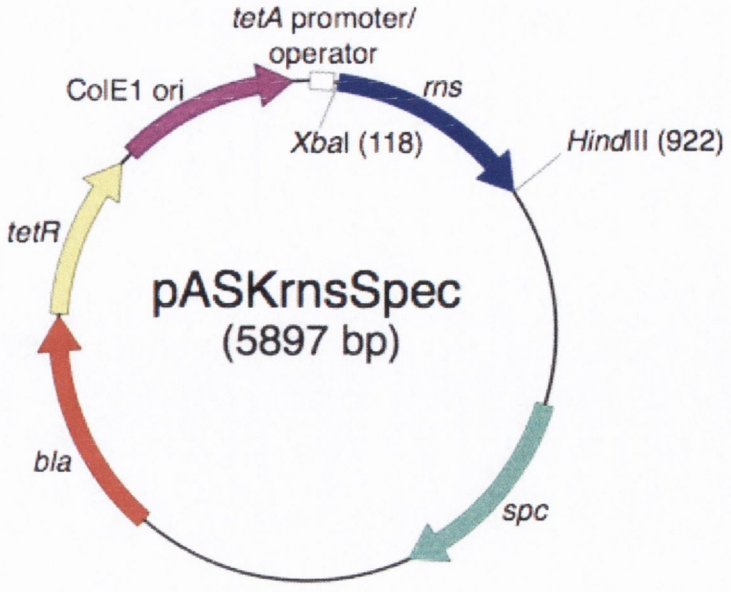

Fig. 5.10. Structure of the plasmid pASKrnsSpec. The plasmid pASKrnsSpec carries *rns* under the tight control of TetR (encoded by the *tetR* gene), the repressor of the *tetA* promoter/operator. The *tet* promoter is inducible by anhydrotetracycline. Relevant features and restriction endonuclease sites are indicated on the plasmid map.

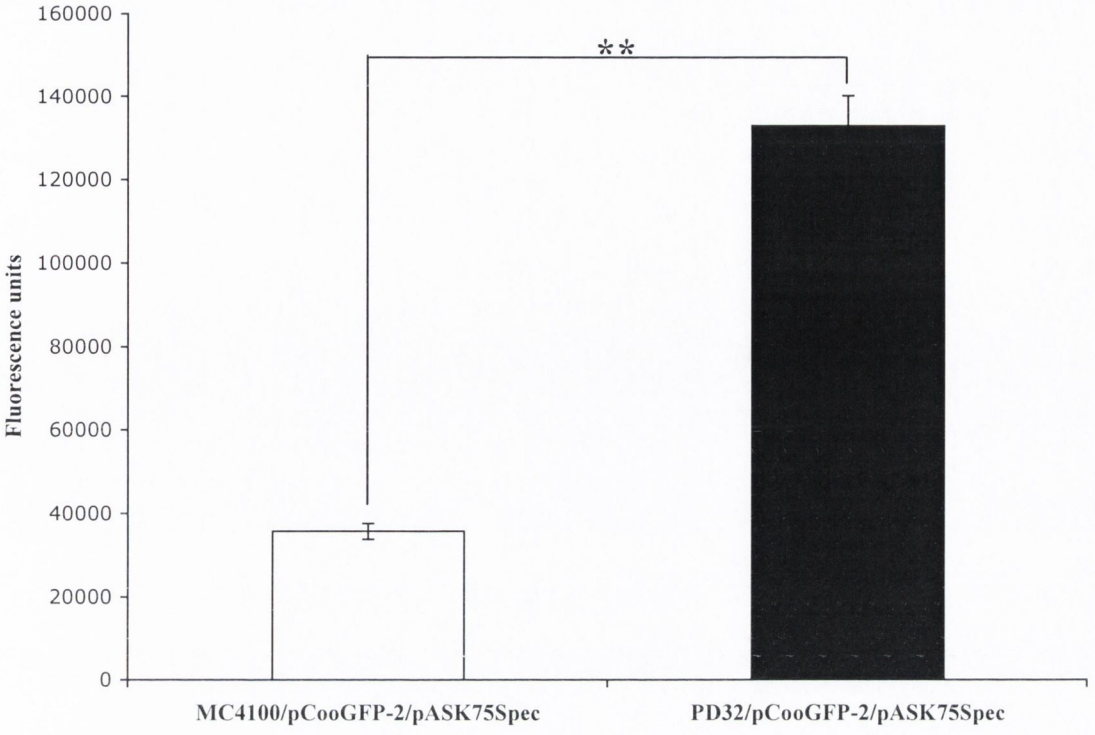

Fig. 5.11. Relative expression from the *coo* promoter in the presence and absence of H-NS. The fluorescence levels in cultures of *E. coli* MC4100 (*hns*$^+$) or PD32 (*hns*$^-$) harbouring the reporter plasmid pCooGFP-2 and the vector control pASK75Spec were measured. Data represents averages of duplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

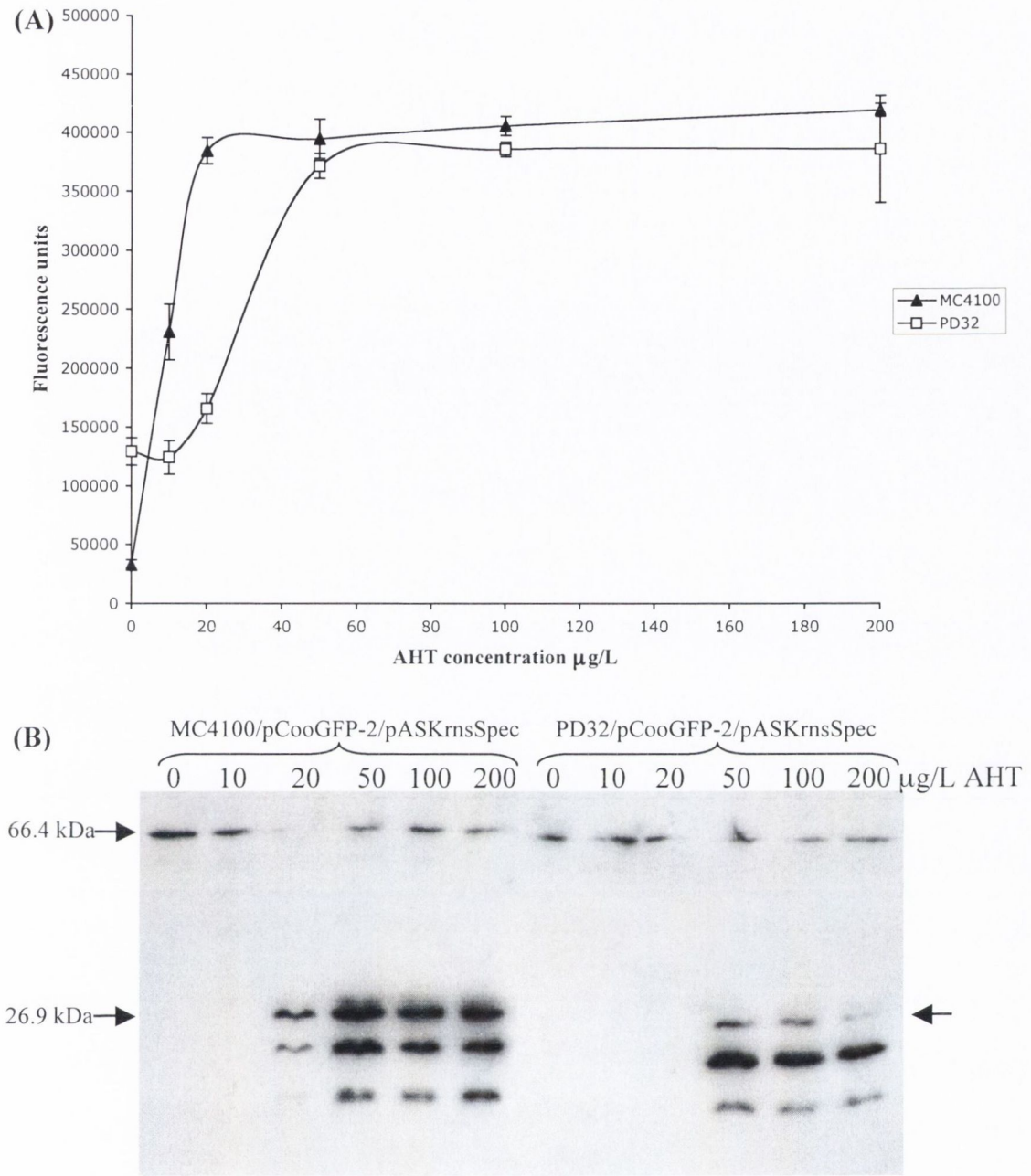

Fig. 5.12. Effect of Rns on *coo* promoter activity and induction of Rns expression in the presence and absence of H-NS. (A) Fluorescence levels of cultures of *E. coli* MC4100 (*hns*$^{+}$, solid triangles) or PD32 (*hns*$^{-}$, open squares), harbouring the reporter plasmid pCooGFP-2 and the Rns expressing plasmid pASKrnsSpec, measured as a function of anhydrotetracycline (AHT) concentration. Data represents averages of triplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. **(B)** Western immunoblot analysis (using anti-Rns antiserum) of whole-cell lysates of the cultures used in the fluorimetric analysis (as indicated above each lane). The positions of protein molecular mass markers are shown on the left. A non-specific cross-reactive band migrating alongside the 66.4 kDa marker serves as a loading control. An arrow on the right indicates the position of full-length Rns.

tagged H-NS were added to the Rns binding reactions. The protein-DNA complex formed by 2.27 μM His-tagged H-NS and the *coo* promoter probe alone can be observed in lane 2 of Fig. 5.13. The addition of increasing concentrations of MBP-Rns did not disrupt this H-NS-*coo* promoter complex (Fig. 5.13 lanes 3-5). Thus the H-NS-*coo* promoter complex was not replaced by the protein-DNA complexes formed in the presence of MBP-Rns and *coo* promoter DNA alone (Fig. 5.13, lanes 6-8). However, when 34 nM MBP-Rns was added to the H-NS binding reaction (lane 4), a low mobility complex appeared in addition to the H-NS-*coo* promoter complex. This low mobility complex may be due to co-occupation of the *coo* promoter probe by both Rns and H-NS. A complex that had not migrated beyond the wells of the gel, and was most likely due to non-specific binding in the presence of excess protein, was observed when 514 nM MBP-Rns was added to the H-NS binding reaction (lane 5). The protein-DNA complexes formed by 34 nM MBP-Rns and the *coo* promoter probe alone can be observed in lane 9 of Fig. 5.13. While MBP-Rns did not appear to be capable of displacing pre-bound H-NS from the *coo* promoter, addition of increasing concentrations of His-tagged H-NS led to the disruption of the Rns-*coo* promoter complexes (lanes 10-12). The addition of 0.56 and 1.13 μM His-tagged H-NS resulted in a reduced amount of Rns-*coo* promoter complexes and the reappearance of unbound DNA (lanes 10 and 11). When the highest concentration of H-NS was added, the Rns-*coo* promoter complexes were eliminated (lane 12). They were replaced by an H-NS-*coo* promoter complex equivalent to that formed in the presence of His-tagged H-NS and the *coo* promoter probe alone (lane 13). Thus it seems that H-NS initially removes Rns from the *coo* promoter and then binds the resultant free DNA. However, the low mobility complex postulated to result from the simultaneous binding of both Rns and H-NS to the *coo* promoter was again detected when 2.27 μM His-tagged H-NS was added to the Rns-*coo* promoter complexes (lane 12). Therefore it appears that although H-NS may displace the majority of pre-bound Rns from the *coo* promoter, a small amount of Rns protein can remain, even in the presence of the newly bound H-NS.

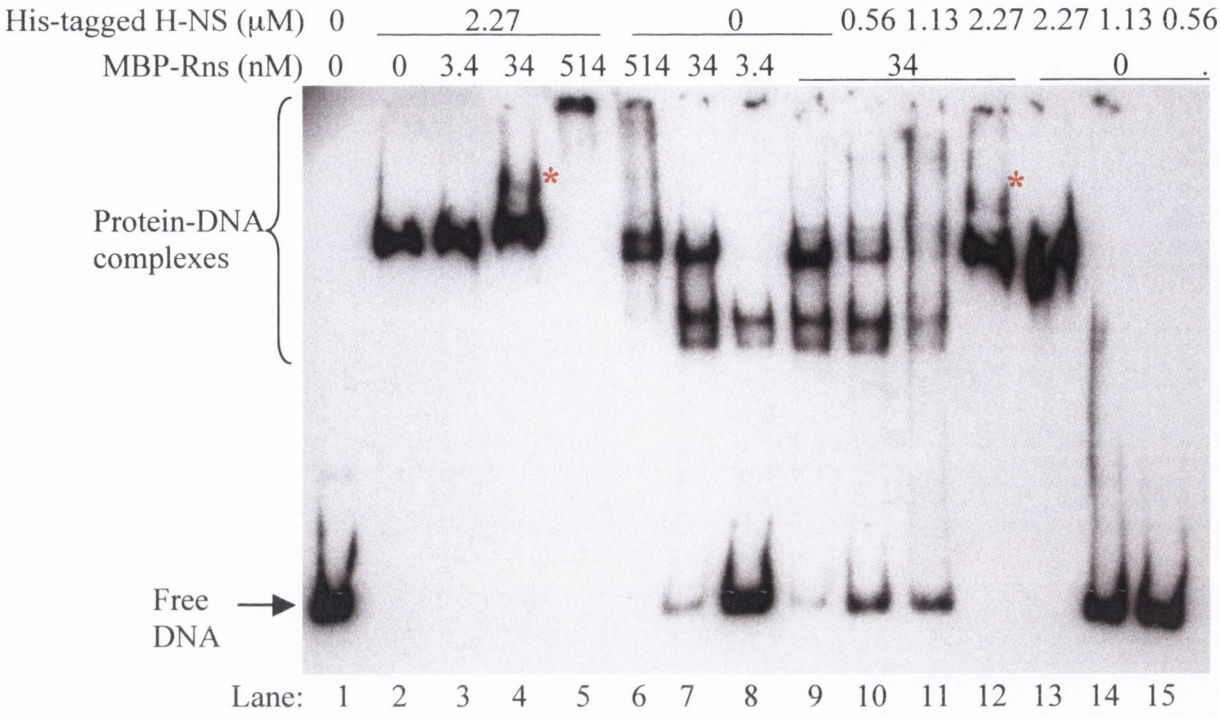

Fig. 5.13. Competition between His-tagged H-NS and MBP-Rns for interaction with *coo* promoter DNA. EMSAs were carried out with approximately 20 pg of a biotinylated *coo* promoter DNA probe. The *coo* promoter probe was incubated with various concentrations of MBP-Rns alone (lanes 6-8) or various concentrations of His-tagged H-NS alone (lanes 13-15), as indicated above each lane. Alternatively the *coo* promoter probe was initially incubated with 2.27 μM His-tagged H-NS and then challenged with increasing concentrations of MBP-Rns (lanes 2-5) or with 34 nM MBP-Rns and then challenged with increasing concentrations of His-tagged H-NS (lanes 9-12), as indicated above each lane. The unbound DNA fragments and protein-DNA complexes are indicated on the left. The positions of supershifted bands potentially corresponding to the simultaneous binding of MBP-Rns and His-tagged H-NS to the *coo* promoter DNA probe are indicated by red asterisks.

5.3 Discussion

The initial aim of the study described in this chapter was to examine the interaction of Rns with mutated forms of the *coo* promoter to gain a better understanding of Rns binding to *coo* DNA and of the protein-DNA complexes formed in EMSA experiments. This aim was achieved, however the unexpected increase in basal activity of the mutated *coo* promoters subsequently led to an investigation of the regulation of *coo* expression by both the specific transcription factor Rns and the global regulator H-NS.

Each of the Rns binding sites at both the *coo* and *rns* promoters contain a conserved TAT motif that is seven nucleotides downstream of, and on the opposite strand to, another conserved thymine residue (268, 269). Rns forms hydrophobic interactions with the C5-methyl groups of the three conserved thymines within binding sites I and II upstream of the *coo* promoter (268). Replacement of the individual thymine nucleotides with uracil, which lacks the C5-methyl group, interferes with Rns binding (268). Thus, in an attempt to disrupt the interaction of Rns with its binding sites at the *coo* promoter the nucleotide sequence CGC was substituted for the conserved TAT motif within sites I and/or II. The protein-DNA complexes formed by MBP-Rns and *coo* promoter fragments containing these substitutions were analysed by performing EMSA experiments. As expected, the CGC for TAT substitutions reduced Rns binding to *coo* promoter DNA. MBP-Rns did not bind specifically to a *coo* promoter probe in which both sites I and II were mutated. Therefore the conserved TAT motif was verified to be critical for Rns binding. However, MBP-Rns did form specific protein-DNA complexes with the probes containing CGC substitutions at site I or site II alone. The predominant complex formed was of high mobility and thought to be due to Rns occupying a single binding site. This confirms that Rns can bind to site I or site II individually. Another protein-DNA complex thought to be due to Rns occupying both sites I and II simultaneously was formed by MBP-Rns and the wild type *coo* promoter probe. This complex was formed to a reduced extent by the single site mutant probes. This reveals that co-operative binding of Rns to sites I and II may occur, i.e. Rns bound at a wild type site may facilitate a small amount of Rns binding to a mutated site. The double site mutant coo promoter probe did not form this complex.

The CGC for TAT substitutions within Rns binding sites I and/or II were then created in reporter plasmids containing the *coo* promoter (-393 to +62) fused to a *gfp* gene. The ability of Rns to activate expression from these *coo* promoter variants *in vivo* could then

be assessed. In the presence of Rns, expression from the site I mutated *coo* promoter increased 3.6-fold compared to the 12-fold increase in wild type *coo* promoter expression when Rns was supplied. Mutation of site II had an even greater effect on Rns-dependent activation as the presence of Rns led to only a 2-fold increase in expression from the site II mutated *coo* promoter. The Rns-dependent increase in expression from the *coo* promoter containing mutations in both sites I and II was even less again and was not always statistically significant. The more serious effect of the mutation in Rns binding site II could result from this being the more promoter-proximal site. Therefore, at the *coo* promoter, Rns binding to site II may have a greater impact of transcription activation than Rns binding to site I. In a previous analysis of the *coo* promoter Rns binding sites *in vivo*, the thymine found 7 nucleotides upstream and on the opposite strand to the TAT motif in Rns binding sites was replaced by a cytosine at sites I and/or II (268). The effect of such mutations on Rns-dependent expression from the *coo* promoter was similar to that described above for the CGC substitutions (268).

The reduced ability of Rns to activate expression from the mutation containing *coo* promoters was expected as it had been demonstrated that the mutations had decreased Rns binding to *coo* promoter DNA. However, the increase in basal expression levels from the *coo* promoter variants containing the site mutations was unexpected. In addition, *coo* promoter fragments in which upstream sequences (-393 to -88) had been lost were also found to display elevated levels of basal expression. The general trend was the less upstream sequence present, the greater the levels of *coo* expression. A progressive increase in basal transcription caused by the removal of upstream DNA sequences has also been reported to occur at the *ctx* promoter of *V. cholerae*, the *hilA* promoter of *S.* Typhimurium and the *yghJ* promoter of *E. coli* (353, 442, 444). In each of these cases it was postulated that a negative regulator might be operating at the upstream promoter sequences. Thus it was possible that a repressor may act both upstream of the *coo* promoter and at the Rns binding sites, and the loss or mutation of these sequences had alleviated this putative repression. A likely candidate for such a repressor was the global regulator H-NS as it was previously demonstrated to repress *coo* promoter expression, although this repression was suggested to largely occur downstream of the transcription start site (+7 to +929) (270). H-NS was found to repress the portion of the *coo* promoter used in this study (-393 to +62). The *coo* expression levels increased 4-fold in an *hns* mutant strain containing the reporter plasmid pCooGFP compared to the levels in a wild type strain harbouring the same

plasmid. Derepressed expression in the absence of H-NS has been found to occur at several other promoters regulated by AraC family members including the *ctx* promoter, the *ureR* promoter of *Proteus mirabilis* and the *cfa* promoter of ETEC (58, 199, 284).

H-NS had not been previously demonstrated to act directly at the *coo* locus, although it was predicted to do so (270). The *coo* promoter region displays characteristics correlated with H-NS binding; it is AT rich and *in silico* analysis revealed that it is predicted to contain areas of DNA curvature. EMSA experiments confirmed that H-NS binds directly to *coo* promoter DNA (-196 to -2), even in the presence of non-specific competitor DNA. Therefore H-NS was found to act upstream of the *coo* promoter in addition to its previously reported activity downstream of the promoter (270). H-NS also acts upstream and downstream of the *bgl* and *proU* promoters (85, 238). For both of these genes, it is postulated that H-NS bound to sequences flanking the promoters may form an extended nucleoprotein complex that represses transcription from the promoters by impeding the action of RNA polymerase (238, 359). H-NS-mediated repression of the *coo* promoter may occur by a similar mechanism.

EMSAs were performed to assess whether the Rns binding site mutations and lack of upstream *coo* promoter sequences had alleviated H-NS-mediated repression of the *coo* promoter by reducing H-NS binding to the region. Upstream of the *coo* promoter there is an area of predicted DNA curvature (~ -270 to -220) and a 10 bp sequence (-350 to -341, AAATTATCCA) with only 3 mismatches to the high-affinity H-NS binding site present in a regulatory element of the *proU* gene (38). However, His-tagged Rns did not appear to have a considerably greater affinity for a *coo* promoter DNA probe that contained an extended area of upstream sequence (-371 to +60) compared to the affinity it displayed for a shorter *coo* promoter probe (-196 to -2). Therefore, based on the EMSA experiments, it appeared that H-NS binding to the *coo* promoter region would be unaffected by the loss of upstream sequences. In contrast, under certain conditions H-NS binding to *coo* promoter DNA appeared to be reduced by the CGC for TAT substitutions within Rns binding sites I and/or II. The mutation in site II had affected another close match to the high-affinity H-NS binding site. An *in vivo* analysis was then conducted to compare the expression levels from a selection of the *coo* promoter variants in wild type *E. coli* and *hns* mutant strains. The increase in the level of expression, in the absence of H-NS, from a reduced length *coo* promoter and from the site I and/or II mutated *coo* promoters was comparable to that from the wild type *coo* promoter. This suggests that H-NS exerts a similar degree of repression on the variant

coo promoters as it does on the wild type *coo* promoter. Thus, despite the apparent reduction in H-NS binding to the site mutated *coo* promoters observed during the EMSA experiments, it was not possible to conclusively attribute the increase in basal activity of the *coo* promoter variants to a reduction in H-NS-mediated repression. This may be due to the experiments used to assess H-NS activity at the *coo* promoters. EMSAs were conducted with linear fragments of DNA and the *in vivo* analysis employed *coo-gfp* fusion reporter plasmids. These were the best tools available but they are only an approximation of the natural substrate for H-NS. However, alternative possibilities for the basal activity of the *coo* promoter variants must also be considered.

The negative effect of the upstream sequences (-393 to -88) on the *coo* promoter may result from these sequences forming repressive structures that prevent RNA polymerase from activating transcription from the promoter. Alternatively, these sequences may contain binding site(s) for a repressor other than H-NS. This hypothetical additional repressor may also bind at Rns binding sites I and II. Thus the elevated levels of basal activity of the *coo* promoter variants may be due to a reduction in the repression caused by a negative regulator other than H-NS. The involvement of an additional negative regulator in the control of the *coo* promoter has been postulated previously (270). A potential candidate for the additional repressor is StpA, the H-NS paralogue present in *E. coli*. StpA may work in concert with H-NS to repress *coo* transcription. Another possible explanation for the activity of the site mutation containing *coo* promoters in the absence of Rns is that instead of reducing the binding of a repressor to the region, the CGC for TAT substitutions may have facilitated the binding of a positive regulator of transcription. Putative candidates may include the nucleoid-associated protein Fis, which can act as an activator (93). Fis often antagonises H-NS-imposed repression and was reported to activate the Rns-regulated *rns* promoter (G. Munson, Abstr. 2004 ASM General Meeting). Transposon mutagenesis could be employed to isolate mutations in additional potential negative and/or positive regulators of the *coo* promoter. A further possibility, however, is that the Rns binding site mutations may have enhanced the *coo* promoter, i.e. they could have improved the binding of RNA polymerase. UP elements are ~20 bp AT rich sequences located immediately upstream of the promoter -35 hexamer that stimulate transcription by providing a point of contact for the C-terminal domain of the RNA polymerase α-subunit (337). UP elements can increase promoter activity in the absence of any protein factors other than RNA polymerase (321). The DNA sequence upstream of the -35 hexamer of the *coo* promoter is AT rich but is not a

strong match to the UP element consensus sequence proposed by Estrem *et al.* (118). However, the CGC for TAT substitution within Rns binding site II causes the region to more closely resemble the well characterised *rrnB* P1 and P2 UP elements, which respectively contain CTC and TGC at the equivalent position (337). Therefore the mutation at site II may have increased the strength of the *coo* promoter. At present none of these potential explanations for the basal activity of the *coo* promoter variants can be ruled out. Indeed a combination of the models proposed above may be occurring.

A strategy was devised to measure the activity of the *coo* promoter in an *hns* mutant when Rns was present. Even in the absence of H-NS, Rns increased *coo* promoter expression. Therefore, at the *coo* promoter, Rns does not solely counteract H-NS-mediated repression. This issue was addressed previously by Murphree *et al.* who monitored the effect of Rns on the activity of *coo-lacZ* fusions in an *hns* mutant strain (270). However, in that study the accurate measurement of *coo* promoter activity was hindered by the unstable *lac* phenotype of some of the strains used (270). UreR, CfaD and ToxT were demonstrated to increase the level of activity of the H-NS-repressed *ureR*, *cfa* and *ctx* promoters, respectively, in the absence of H-NS (58, 199, 444). Thus, like Rns, these AraC family members do not act only as anti-repressors of H-NS. In addition to overcoming the repressive effect of H-NS, Rns appears to increase transcription from the *coo* promoter via another mechanism. Rns may activate the promoter by facilitating the binding of RNA polymerase, as was found for ToxT (444), or it may act to diminish repression caused by a second negative regulator.

Competition EMSAs were performed to study the interplay between Rns and H-NS at the *coo* promoter. While the concentration of proteins used may not correspond to the actual conditions within the bacterial cell, the EMSAs enabled an analysis of the ability of Rns and H-NS to compete with each other for the *coo* promoter. Rns did not appear to be able to displace H-NS from the promoter, possibly as a result of H-NS-induced conformational changes in the DNA. In contrast, H-NS could displace Rns from the *coo* promoter. The addition of H-NS to *coo* promoter DNA pre-bound by Rns led to the reappearance of unbound DNA along with the loss of Rns-DNA complexes without the formation of H-NS-DNA complexes. A similar phenomenon was observed when the competition between H-NS and UreR for the *ureR-ureD* intergenic region was investigated (307). It suggests that an interaction between Rns and H-NS may be necessary to facilitate the removal of Rns by H-NS. Both the addition of H-NS to Rns-*coo* promoter complexes and the addition of Rns to H-NS-*coo* promoter complexes led

to the formation of a supershifted band of lower mobility than the specific bands resulting from the interaction of the *coo* promoter with H-NS or Rns alone. This supershifted band may be due to the simultaneous binding of Rns and H-NS to *coo* promoter DNA. The co-occupation of a promoter by H-NS and its anti-repressor protein has been demonstrated, for example at the *rrnB* P1 promoter which is differentially regulated by H-NS and Fis (4). Therefore it appears that, at least to some extent, Rns can bind to the *coo* promoter even in the presence of H-NS. Once bound, Rns may enable transcription to occur by disrupting the formation of an H-NS-based repressive nucleoprotein complex. Anti-repression of the *bgl* promoter was also proposed to involve the binding of proteins that interfere with the formation of a repressing nucleoprotein complex largely composed of H-NS (48). As discussed above, in addition to this anti-repression Rns may also directly activate *coo* transcription.

The unforeseen consequences of altering the *coo* promoter DNA sequence illustrates the potential complexity of *coo* promoter regulation. Further studies are necessary to identify all of the factors involved and to elucidate how these factors interact to influence *coo* expression.

Chapter 6 Thermoregulation of CS1 fimbrial expression

6.1 Introduction

Virulence gene expression in pathogenic bacteria is responsive to many environmental factors including osmolarity, pH and ion concentration (253). This enables bacteria to conserve energy by ensuring that virulence genes are only expressed under appropriate conditions, i.e. in the host environment. For human pathogens an increase from ambient temperature to 37°C is an important signal indicating their presence in a host, thus expression of many virulence factors is thermally regulated (213). This adaptation of pathogens to elevated temperature is distinct from the heat shock response, a protective mechanism triggered by shifting *E. coli* cells grown at 30°C to 42°C.

The mechanisms involved in thermoregulation of virulence gene expression include temperature-induced changes in protein conformation, RNA conformation and DNA topology (190). TlpA, an autoregulatory repressor encoded by an *S.* Typhimurium virulence plasmid, is one of the few known temperature-sensing proteins. An increase in temperature causes TlpA to undergo a reversible structural transition that renders the protein inactive (189). Another protein postulated to participate directly in the thermoregulation of gene expression is H-NS. Over 200 *Salmonella* genes were found to display H-NS-dependent up-regulation when the growth temperature was increased from 25°C to 37°C (290). The levels of *hns* mRNA and H-NS protein do not vary significantly in this temperature range (110, 279, 379). Therefore it has been suggested that the H-NS protein may act as a thermosensor by undergoing structural and functional alterations in response to temperature. There is some biochemical evidence for this suggestion. Although the affinity of the C-terminal domain of H-NS for DNA is largely unaffected by temperature changes, binding of full length H-NS to the *Salmonella hilC* promoter was decreased at 37°C (290). In addition, analysis of the N-terminal region of H-NS revealed that an increase in temperature led to a reduction of high order oligomerisation of the protein and an increase in the levels of H-NS dimers (290). A model was proposed in which the reduction in the oligomeric state of H-NS, as a result of an increase in temperature, leads to a loss in co-operative binding of H-NS to DNA. This facilitates the dissociation of H-NS from DNA at 37°C, thus enabling the transcription of H-NS-repressed thermoregulated genes (290). Previous findings that H-NS effects on DNA diminish as the temperature increases support this model (7). However, other studies appear to contradict the model as they have found that H-NS activity and tetramerisation actually increase as the temperature is elevated (54, 390).

Thermoregulation of some H-NS-repressed genes has been attributed to temperature-induced alterations in DNA topology rather than in the H-NS protein. Expression of the *virF* gene of *S. flexneri* is repressed below 32°C by a co-operative interaction between H-NS molecules bound to two sites, separated by a region of DNA curvature, at the *virF* promoter (126). In this system the curved DNA tract is proposed to be the thermosensor, as it undergoes a temperature-dependent conformational transition resulting in the disruption of the repression complex and activation of transcription at 37°C (315). An aspect of DNA topology known to be affected by temperature, and to alter gene expression, is supercoiling (90, 190). Thermally induced changes in DNA supercoiling can modulate the binding and activity of regulatory proteins and thus influence promoter activity. Induction of the *S. flexneri virB* promoter is proposed to involve a temperature-mediated alteration in local DNA supercoiling that enables the VirF regulator to both overcome repression by H-NS, and activate transcription (403).

RNA structure can also vary in response to temperature (190). Certain mRNA species are termed RNA thermometers as their 5′-untranslated regions are thermosensitive (271). The temperature sensing mechanism involves sequestration of the translation initiation sequences by an mRNA secondary structure at low temperatures (271). At elevated temperatures this structure melts, unmasking the ribosome binding site and enabling translation to occur (271). Temperature-dependent production of the *Yersinia pestis* AraC-like virulence regulator, LcrF, is thought to occur by this mechanism (180).

In *E. coli*, the expression of many fimbriae including 987P, bundle-forming pili and P fimbriae is thermoregulated (103, 153, 318). Fimbrial production is generally optimal at 37°C and, as is the case for CFA/I, CS1 and CS2 fimbriae of ETEC, non-existent at temperatures of 20°C and below (199, 374). The mechanisms involved in the temperature-regulated expression of fimbriae have been the subject of much study. Analysis of CFA/I fimbrial expression revealed that thermoregulation occurs at the transcriptional level of both the fimbrial subunit gene *cfaB* and the positive regulator *cfaD* (199). H-NS was found to be involved in the thermoregulatory response as CFA/I fimbriae were expressed at 20°C in an *hns* mutant strain (199). A study of the thermal control of transcription of the Pap operon was the first report of a role for H-NS in the thermoregulation of fimbrial expression (152). The aims of the study presented in this chapter were to determine the thermally sensitive step in the pathway of CS1 fimbrial expression and to establish if H-NS was involved in this thermoregulation.

6.2 Results

6.2.1 Analysis of the thermoregulation of CS1 fimbrial expression

6.2.1.1 Effect of temperature on the production of CS1 fimbriae in ETEC

The effect of temperature on CS1 fimbrial expression by ETEC strain LMC10 was assessed. The *rns*-containing plasmid pSS2192 or the empty vector control pCL1920 were transformed into LMC10. The transformants were grown at 20°C and 37°C on CFA agar and in CFA broth. Heat-released extracts of the bacterial cultures were prepared as described in Section 2.3.1.2 and analysed by SDS-PAGE and Western immunoblotting using anti-CS1 antiserum to test for the presence of the 15.2 kDa CooA major subunit of CS1 fimbriae. CooA was not detected in the heat-released extracts of LMC10/pCL1920 grown on CFA agar at either 20°C or 37°C (Fig. 6.1). ETEC strain LMC10 harbouring the Rns-expressing plasmid pSS2192 produced CooA following growth on CFA agar at 37°C, however CooA was not detected when the bacteria were grown at 20°C. Thus the thermoregulation of CS1 fimbrial production was confirmed. Analysis of the broth-grown cultures revealed the same result, i.e. CS1 fimbriae are produced in Rns-expressing cultures at 37°C but not at 20°C, however the level of fimbrial expression at 37°C was determined by densitometry to be approximately 50% less than that of the agar-grown cultures. Enhanced production of fimbriae by ETEC grown on solid CFA agar rather than liquid CFA broth is a known phenomenon (122, 147).

6.2.1.2 Construction of an ETEC *hns* mutant strain

To determine the contribution of H-NS to the thermoregulation of CS1 fimbrial expression in ETEC, an LMC10 *hns* mutant was constructed using the λ Red allele replacement system as described in Section 2.2.2. The *hns* gene of *E. coli* strain PD32 is disrupted by the presence of an ampicillin resistance cassette (80). This interrupted gene, plus ~400 bp of flanking DNA, was amplified by PCR using the primers hns-flank_F and hns-flank_R and PD32 genomic DNA as a template. The PCR product was

gel-purified, concentrated and then transformed into ETEC strain LMC10 harbouring the pKOBEG plasmid (Table 2.2). LMC10 had elevated levels of the λ Red proteins due to the induction of these genes from pKOBEG by the addition of L-arabinose, thus facilitating transfer of the interrupted *hns* gene to the bacterial chromosome. Putative mutants were initially confirmed by PCR analysis (Fig. 6.2 (A)). Further verification that the *hns* gene of the mutants was disrupted was obtained by comparing the appearance of the mutants grown on MacConkey agar containing 1% salicin to that of LMC10 and PD32 strains that had also been grown on this indicator media. As described in Section 2.1.1.2, *hns* mutants (such as *E. coli* PD32) display a β-glucoside-fermenting phenotype and therefore appear red on MacConkey salicin agar. In contrast, colonies of hns^+ bacteria (such as the ETEC strain LMC10) appear white on MacConkey salicin agar, as they are not capable of fermenting β-glucosides. Following growth on the indicator media, colonies of the putative LMC10 *hns* mutants were red, as were the colonies of *E. coli* PD32 (Fig. 6.2 (B)). The confirmed *hns* mutant strain was designated LMC10Δ*hns* (Table 2.1).

6.2.1.3 Effect of temperature on the production of CS1 fimbriae in an ETEC *hns* mutant

The effect of temperature on CS1 fimbrial expression by LMC10Δ*hns* was assessed as described above for the wild type ETEC strain. Heat-released extracts of LMC10Δ*hns* harbouring pSS2192 or pCL1920, grown at 20°C and 37°C on CFA agar and CFA broth, were analysed by Western immunoblotting using anti-CS1 antiserum. CooA was not detected in the heat-released extracts of LMC10Δ*hns*/pCL1920 grown on CFA agar at either temperature (Fig. 6.3). Thus it appears that even in an *hns* mutant strain, Rns is required for CS1 expression. Analysis of the extracts prepared from agar-grown LMC10Δ*hns*/pSS2192 revealed that for these Rns-expressing cultures CooA was present following growth at 37°C and, to a much reduced extent, following growth at 20°C. The same trends were observed in the broth-grown cultures (results not shown). Therefore, although production of CS1 fimbriae is abrogated at 20°C in the wild type ETEC strain, in the *hns* mutant a degree of derepressed CS1 fimbrial expression occurs at this temperature. Thus it appears that H-NS is involved in the repression of CS1 expression at 20°C. In addition to this H-NS-mediated repression there is another

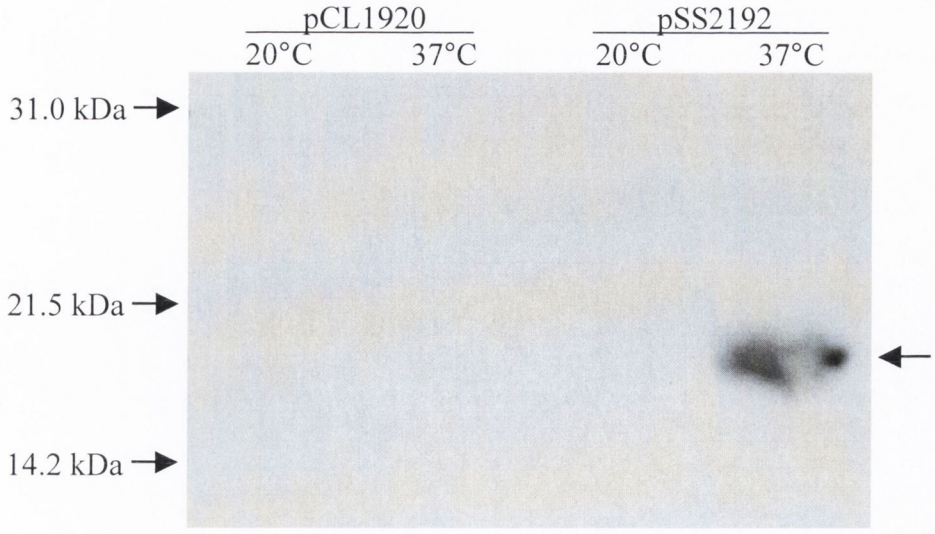

Fig. 6.1. Thermoregulation of CS1 fimbrial expression in wild type ETEC. Western immunoblot analysis of heat-released extracts of cultures of ETEC strain LMC10 harbouring pCL1920 or pSS2192, grown on CFA agar at 20°C or 37°C as indicated above each lane. The CooA fimbrial subunit protein (indicated by an arrow on the right) was detected with anti-CS1 antiserum. The positions of protein molecular mass markers are indicated on the left.

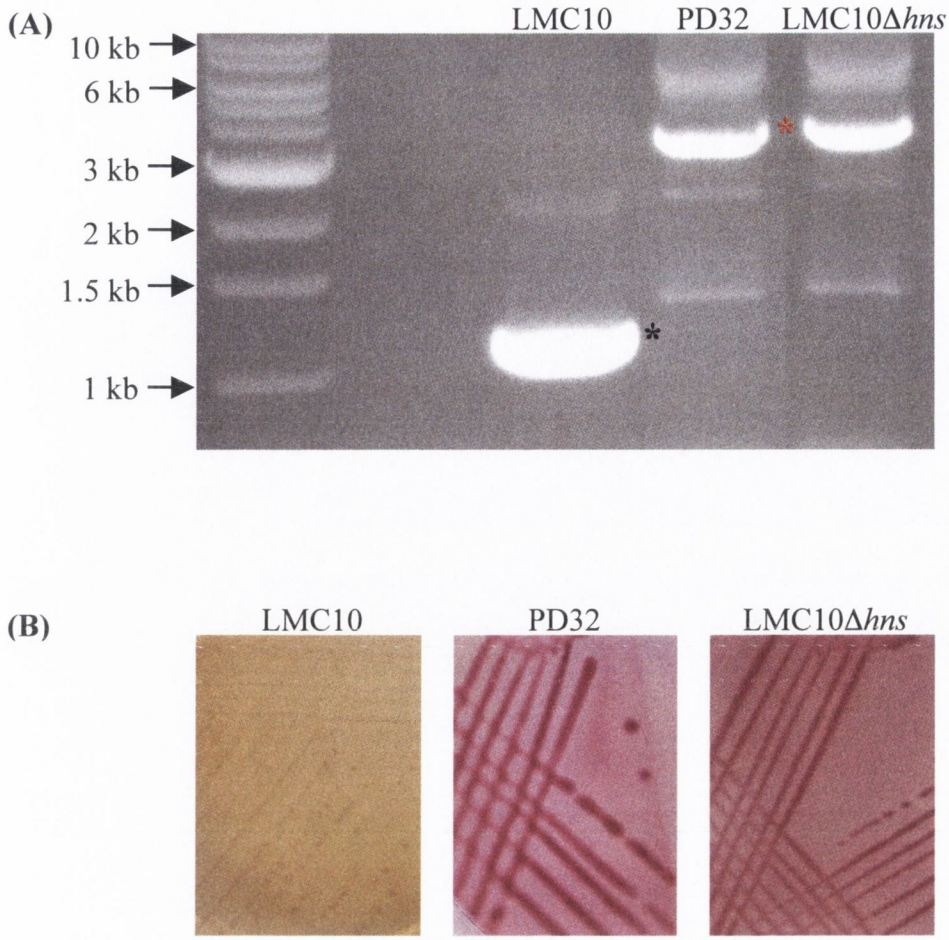

Fig. 6.2. Verification of the *hns* mutation in ETEC strain LMC10Δ*hns*. (A) Genomic DNA preparations from wild type (ETEC strain LMC10) or *hns* mutant strains (*E. coli* PD32 and ETEC strain LMC10Δ*hns*) were used as a template for PCR amplification of the *hns* gene plus ~400 bp of flanking DNA. The positions of selected DNA molecular size standards are indicated on the left. Lane 2 contains a DNA-free negative control. A black asterisk indicates the position of a band corresponding to an uninterrupted *hns* gene (plus ~400 bp of flanking DNA). A red asterisk indicates the position of bands corresponding to an *hns* gene disrupted by the presence of an ampicillin cassette (plus ~400 bp of flanking DNA). **(B)** Appearance of bacterial colonies grown on MacConkey agar containing 1% salicin. The identity of the colonies is indicated above each image. Fermentative bacteria (*hns* mutants) appear red on MacConkey agar containing salicin while non-fermenters (wild type) appear white or cream.

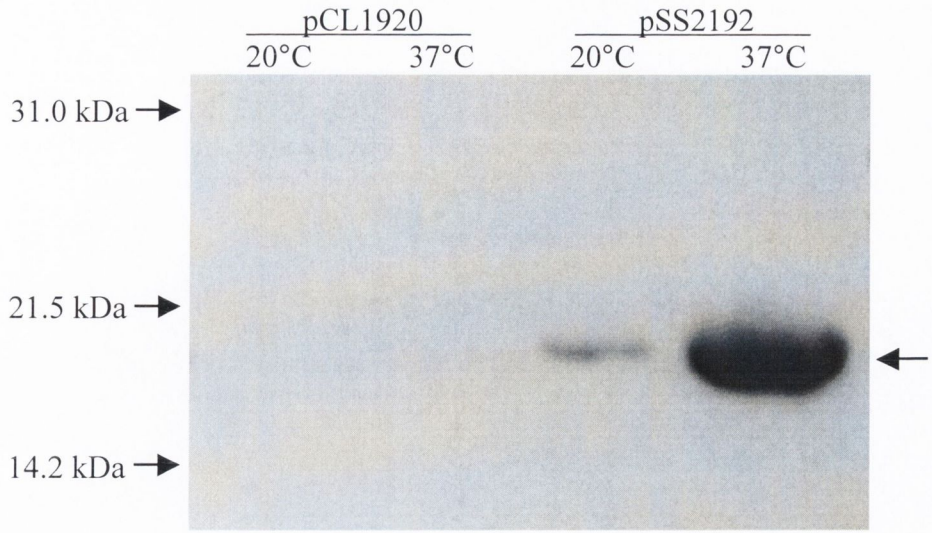

Fig. 6.3. Thermoregulation of CS1 fimbrial expression in an ETEC *hns* mutant. Western immunoblot analysis of heat-released extracts of cultures of ETEC strain LMC10Δ*hns* harbouring pCL1920 or pSS2192, grown on CFA agar at 20°C or 37°C as indicated above each lane. The CooA fimbrial subunit protein (indicated by an arrow on the right) was detected with anti-CS1 antiserum. The positions of protein molecular mass markers are indicated on the left.

aspect to the thermoregulation of fimbrial production, however, as even in the absence of H-NS the levels of CooA produced by bacteria grown at 20°C were considerably lower than the levels produced by bacteria grown at 37°C.

6.2.2 Determination of the level at which thermoregulation occurs

6.2.2.1 Effect of temperature on *coo* transcription

The temperature dependence of CS1 fimbrial production had been confirmed. To determine if *coo* transcription is affected by temperature, the level of *coo* promoter activity in bacterial cultures grown at 20°C and 37°C was assessed. Fluorimetric analysis was conducted on *E. coli* MC4100 that contained the reporter plasmid pCooGFP-2 in addition to pCL1920 or pSS2192, and had been grown at 20°C or at 37°C. The analysis was also performed on corresponding cultures of the *hns* mutant *E. coli* PD32 to measure *coo* promoter activity at the two temperatures in the absence of H-NS. At both 20°C and 37°C, *coo* transcription was derepressed in the absence of H-NS and activated by the presence of Rns (Fig. 6.4 (A)). Expression from the *coo* promoter in the wild type strain MC4100 harbouring the *rns*-containing plasmid pSS2192 was 3.2-fold higher ($P < 0.005$) at 37°C than it was at 20°C. Therefore *coo* transcription is influenced by growth temperature. The levels of Rns-dependent *coo* promoter expression in the *hns* mutant (PD32/pCooGFP-2/pSS2192) at 20°C and in the wild type strain (MC4100/pCooGFP-2/pSS2192) at 37°C were similar. Thus in the presence of Rns and the absence of H-NS, a significant level of *coo* transcription can occur at the generally non-permissive temperature of 20°C. However, the *hns* mutation did not eliminate thermoregulation of *coo* transcription. Expression from the *coo* promoter in the *hns* mutant strain PD32 containing pSS2192 was 2.1-fold higher ($P < 0.005$) at 37°C than at 20°C. The fluorimetric analysis was repeated using the clinical ETEC isolate LMC10 and its *hns*-negative isogenic derivative LMC10Δ*hns*, in place of the *E. coli* K-12 strains MC4100 and PD32 respectively. The results obtained with the ETEC strains were essentially similar to those obtained with the *E. coli* K-12 strains (Fig. 6.4 (B)). Therefore, in both *E. coli* K-12 and ETEC, H-NS-mediated repression of *coo* promoter expression occurs at 20°C and at 37°C and thermoregulation of CS1 fimbrial production

operates at the level of *coo* transcription at least. In addition, it appears that temperature regulation of *coo* transcription does not involve any ETEC-specific factors.

6.2.2.2 Effect of temperature on *rns* transcription

It had been established that a temperature effect is exerted at the level of *coo* transcription. However, the possibility that this thermoregulation was due to a thermally sensitive process occurring prior to *coo* transcription could not be ruled out. It was possible that the cellular level of Rns might be temperature dependent and thus the abundance of this transcriptional activator may be mediating the thermoregulatory response. To address this possibility, it was initially determined whether the *rns* gene is at least transcribed at 20°C. Total RNA was extracted after growth at 20°C and 37°C of *E. coli* strains MC4100 and PD32 harbouring pCL1920 or pSS2192. The isolated total RNA was then used as the template in one-step end-point RT-PCR assays performed to detect the presence of *rns* transcripts. During the assay the RNA is reverse transcribed to produce cDNA, which then serves as the template in a standard PCR amplification. Assays were conducted with the primers rt-rnsF and rt-rnsR (Table 2.3) that are designed to amplify a 311 bp area present in the *rns* transcript. Control reactions were simultaneously carried out using primers specific for the constitutively expressed housekeeping gene *gapA* (rt-gapF and rt-gapR, Table 2.3) (261). A uniform amount of PCR product was detected following each of the control reactions (Fig. 6.5) thus confirming that an equal amount of RNA was used in each sample, and that this RNA was of good quality. A product was not present after RT-PCR was performed using the rt-rnsF and rt-rnsR primers on samples isolated from pCL1920-harbouring cultures, as this empty vector control does not contain the *rns* gene. However, PCR products were detected after assays were conducted with the *rns* transcript specific primers and RNA extracted from cultures containing pSS2192, regardless of the temperature at which these cultures had been grown. Therefore, the *rns* gene is transcribed at 20°C (and 37°C) in both the wild type and *hns* mutant backgrounds.

To determine if temperature affects the levels of *rns* transcription, the activity of the *rns* promoter at 20°C and 37°C in the presence and absence of H-NS was then assessed. The *E. coli* K-12 strains MC4100 and PD32 were transformed with either the reporter plasmid pRnsLacZ-2, or the vector control pRW50. The levels of β-galactosidase expressed by the resultant transformants were measured following growth at 20°C and

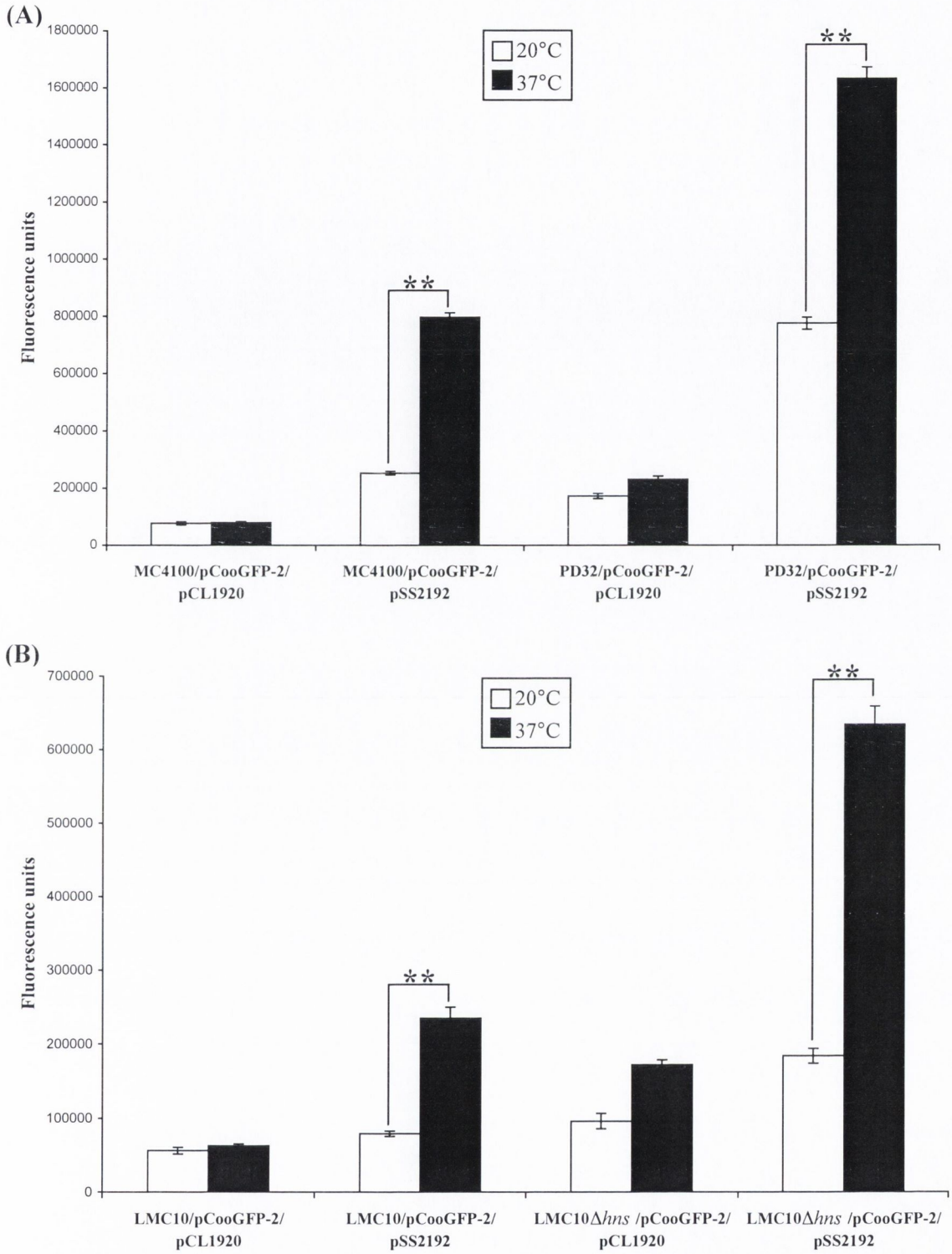

Fig. 6.4. Effect of temperature on *coo* promoter activity. The fluorescence levels of the indicated bacterial cultures (see horizontal axes) were measured following growth at 20°C or 37°C as indicated. Data represents averages of triplicate measurements on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown. Statistical significance is indicated by **, where $P < 0.005$.

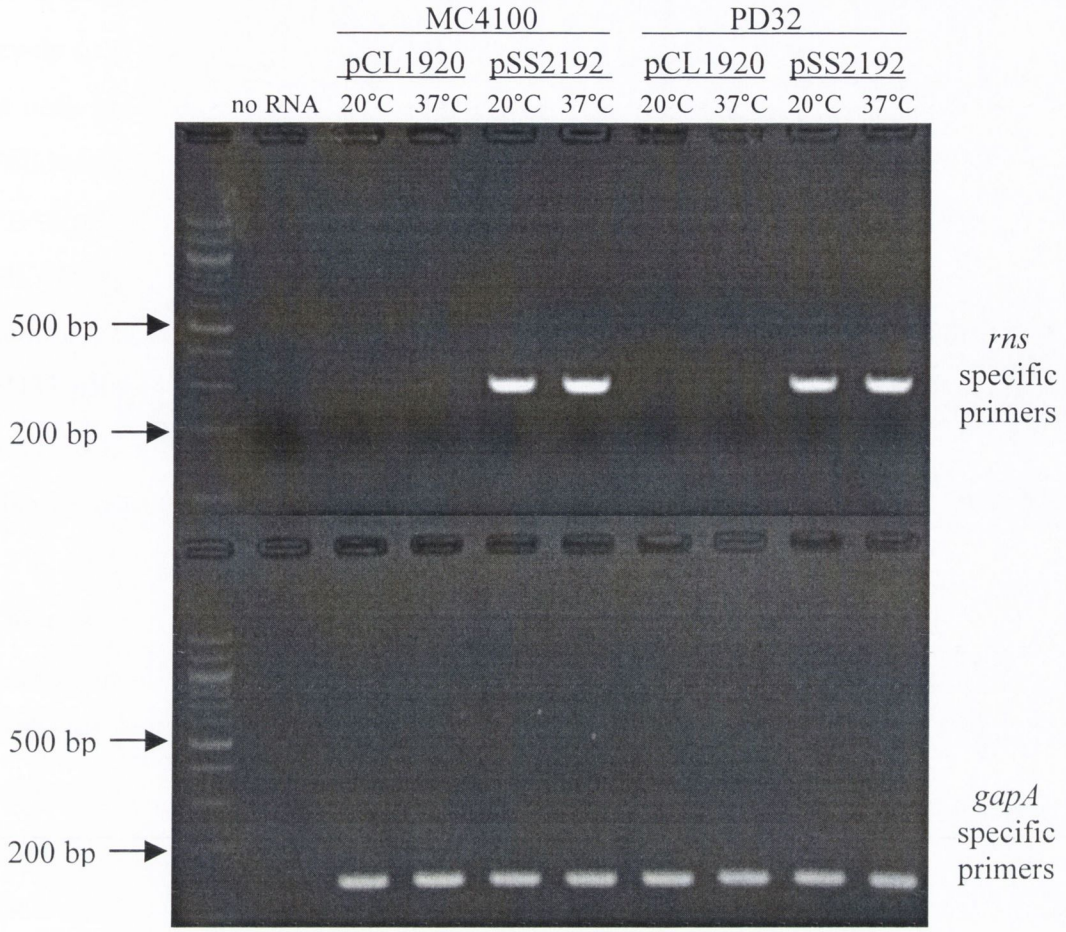

Fig. 6.5. The *rns* gene is transcribed at 20°C and 37°C. RT-PCR analysis was performed using primers specific for the *rns* transcript or the *gapA* transcript (as indicated on the right), and total RNA isolated from cultures of *E. coli* MC4100 (wild type) and PD32 (*hns* mutant), harbouring either pCL1920 (empty vector control) or pSS2192 (*rns*-containing plasmid), that had been grown at 20°C or 37°C as indicated above each lane. Lane 2 contains a RNA-free negative control. The positions of selected DNA molecular size standards are indicated on the left.

37°C (Fig. 6.6 (A)). Bacteria harbouring the vector plasmid pRW50 produced minimal levels of β-galactosidase. Analysis of the strains containing pRnsLacZ-2 revealed that at both growth temperatures, the *rns* promoter is derepressed in the absence of H-NS. Following growth at 37°C, the β-galactosidase expression levels in the *hns* mutant PD32/pRnsLacZ-2 were 5-fold greater ($P < 0.005$) than those in wild type MC4100/pRnsLacZ-2. Thus H-NS influences the activity of the *rns* promoter. In contrast, it appears that *rns* promoter activity is largely unaffected by temperature as PD32/pRnsLacZ-2 and MC4100/pRnsLacZ-2 produced similar β-galactosidase levels at both 20°C and 37°C. The effects of temperature and H-NS on *rns* transcription were also assessed in ETEC strains by measuring β-galactosidase production at 20°C and 37°C in LMC10 and LMC10Δ*hns* harbouring pRW50 or pRnsLacZ-2 (Fig. 6.6 (B)). The results were broadly similar to those observed for the *E. coli* K-12 strains, with the exception of the relative levels of *rns* promoter activity in the *hns* mutant at the two different growth temperatures. LMC10Δ*hns*/pRnsLacZ-2 expressed 1.4-fold more β-galactosidase at 37°C than it did at 20°C. Therefore, in an ETEC *hns* mutant background the *rns* promoter exhibits a reduced level of activity at 20°C compared to the level of activity at 37°C. However, *rns* transcription was not found to be thermally regulated in *E. coli* K-12 or wild type ETEC strains. In addition, transcription from the *rns* promoter was demonstrated to be repressed by the H-NS protein.

6.2.2.3 Interaction of His-tagged H-NS with the *rns* promoter

H-NS was found to negatively regulate the *rns* promoter. To determine whether H-NS binds directly to the *rns* promoter region, EMSAs were carried out to examine the interaction of H-NS with *rns* promoter DNA. Increasing concentrations of purified His-tagged H-NS were incubated with a 434 bp *rns* promoter probe in the presence of an excess of non-specific competitor DNA (poly(dI.dC)). The *rns* promoter probe was completely bound following incubation with 1.13 μM His-tagged H-NS and a low mobility protein-DNA complex was formed (Fig. 6.7). A complex of even lower mobility was detected when 2.27 μM His-tagged H-NS was incubated with the probe, indicating that there is more than one binding site for H-NS at the *rns* promoter or that numerous H-NS molecules were interacting with the DNA probe. Therefore, the EMSA results demonstrated that H-NS binds directly to the *rns* promoter.

6.2.2.4 Effect of temperature on Rns protein levels

The levels of *rns* promoter expression at 20°C and 37°C had been analysed and it appeared that, in general, temperature does not have a considerable effect on *rns* transcription. However, it remained possible that a thermally sensitive process may occur subsequent to *rns* transcription resulting in the thermoregulation of Rns protein levels. ETEC strains LMC10 and LMC10Δ*hns* harbouring pCL1920 or pSS2192 were grown at 20°C and 37°C and then analysed by Western immunoblotting using anti-Rns antiserum to assess the influence of temperature on the level of Rns protein in the cell. Rns protein was not detected in strains containing the empty vector control pCL1920 (Fig. 6.8). In LMC10/pSS2192 and LMC10Δ*hns*/pSS2192, Rns protein was present following growth at both 20°C and 37°C. Similar levels of Rns were expressed by the wild type and *hns* mutant ETEC strains. Temperature was found to have a limited effect on the production of Rns protein, as densitometric analysis revealed that the amount of Rns present in samples grown at 37°C was approximately 1.7-fold greater than those in samples grown at 20°C. Therefore, it appears that the levels of Rns protein present in the cell are not strongly thermoregulated.

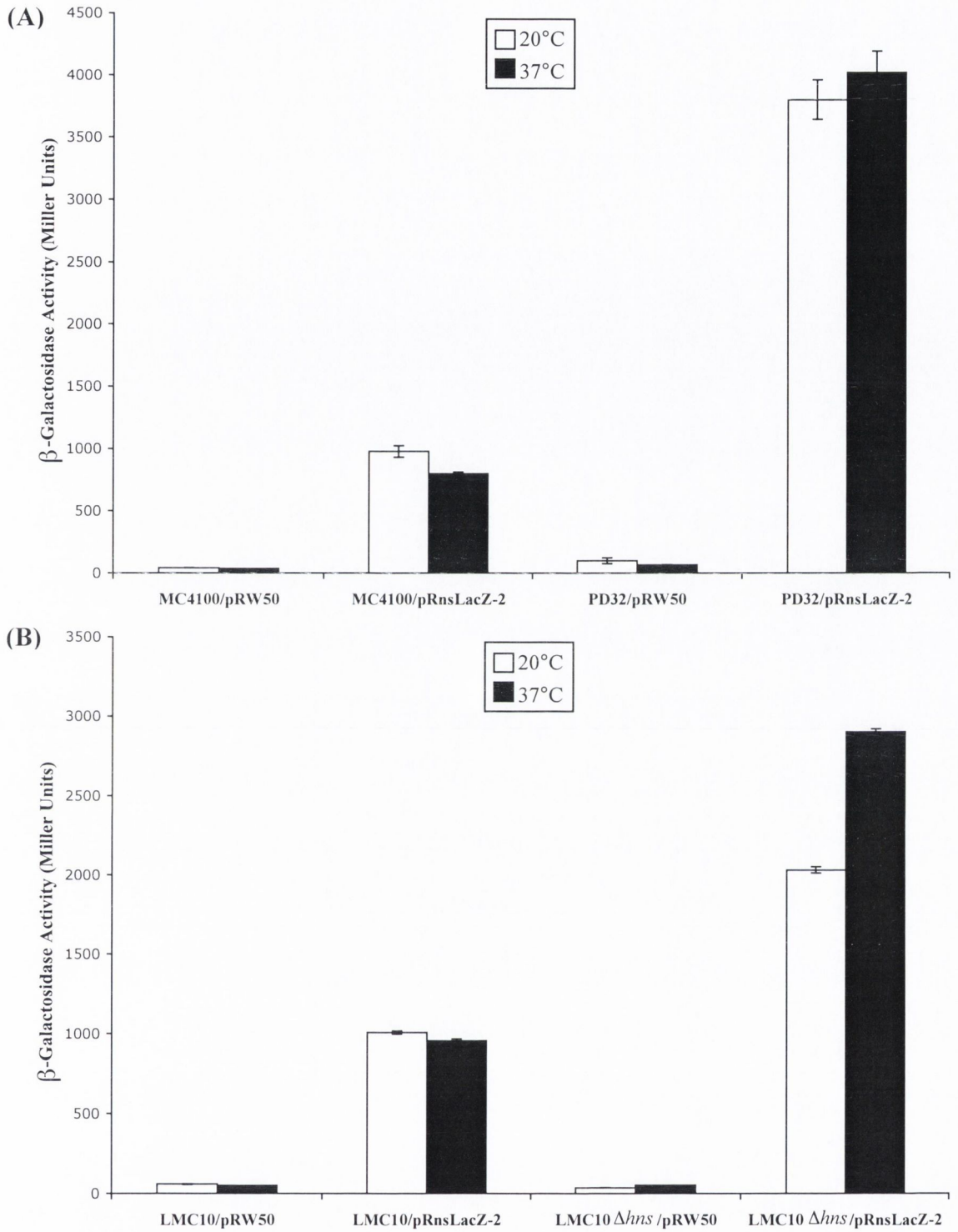

Fig. 6.6. Effect of temperature on *rns* promoter activity. The β-galactosidase activities of the indicated bacterial cultures (see horizontal axes) were measured following growth at 20°C or 37°C. Data represents averages of duplicate assays on duplicate cultures. Error bars indicate standard deviation values. Measurements were performed independently at least twice; a representative data set is shown.

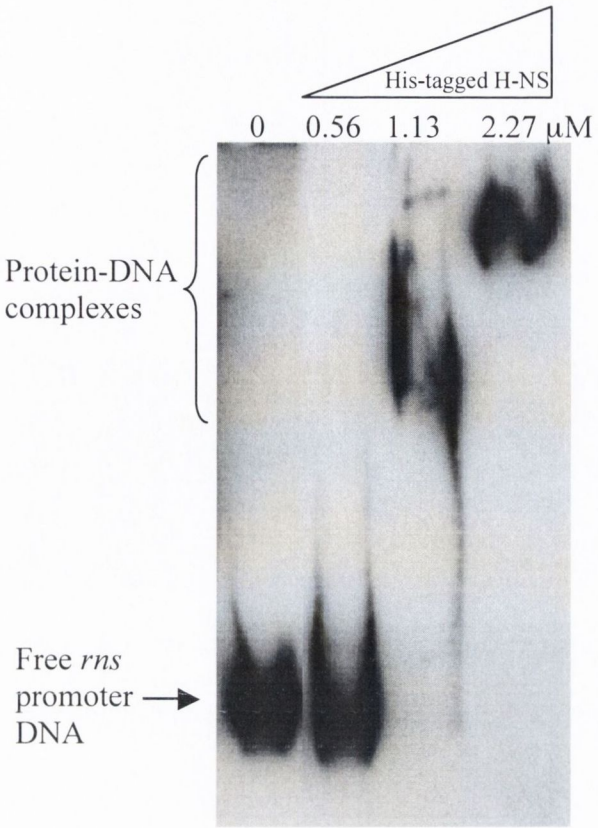

Fig. 6.7. Interaction of His-tagged H-NS with *rns* promoter DNA. EMSAs were carried out with approximately 20 pg of a biotinylated *rns* promoter DNA probe and a range of concentrations of His-tagged H-NS as indicated above each lane. The unbound DNA fragments and protein-DNA complexes are indicated on the left.

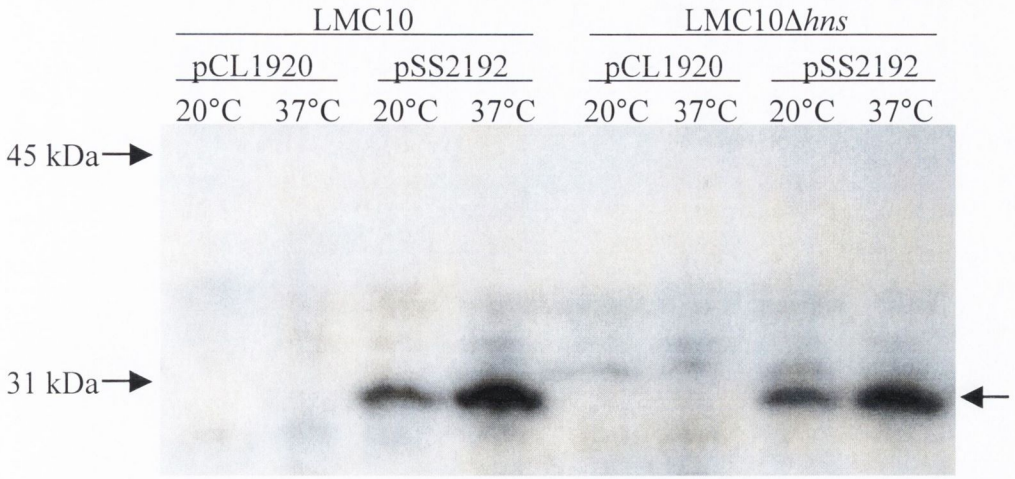

Fig. 6.8. Effect of temperature on Rns protein levels. Western immunoblot analysis (performed with anti-Rns antiserum) of whole cell lysates of ETEC strain LMC10 and LMC10Δ*hns* harbouring either pCL1920 (empty vector control) or pSS2192 (*rns*-containing plasmid) that had been grown at 20°C or 37°C as indicated above each lane. The positions of molecular mass markers are indicated on the left. An arrow on the right indicates the position of Rns.

6.3 Discussion

A change in temperature from that typically found in the outside environment (~20°C) to that found in a human host (37°C) is a key signal that induces the expression of many fimbriae in *E. coli*, including the CS1 fimbriae of ETEC. Analysis of this thermoregulatory response could lead to an improved understanding of the ETEC mechanism of pathogenesis. Therefore the aims of the study described in this chapter were to investigate the level at which temperature control of CS1 fimbrial production may occur, and to determine the involvement of H-NS in such thermal control. The influence of temperature and H-NS on CS1 expression, *coo* transcription, *rns* transcription and the cellular levels of Rns protein were assessed.

In the presence of Rns, production of the CooA major CS1 fimbrial subunit was detected following growth of wild type ETEC at 37°C but not after growth at 20°C. In contrast, CooA was detectable when the Rns-expressing *hns* mutant strain was grown at 20°C. To establish if the effects of temperature and H-NS on CooA production were exerted at the transcriptional level, *coo* promoter activity was measured in wild type and *hns* mutant backgrounds at 20°C and 37°C. The levels of Rns-dependent expression from the *coo* promoter following growth at 37°C were significantly greater than those observed after growth at 20°C. Therefore, thermoregulation of *coo* transcription was found to be occurring. The transcription of several other *E. coli* fimbrial genes is thermally modulated. Examples include the genes *fasA*, *papA* and *cfaB,* which encode subunit proteins of 987P, P and CFA/I fimbriae respectively (103, 153, 199). In addition to being controlled by temperature, *coo* transcription is regulated by the H-NS protein. While H-NS-mediated repression of the *coo* promoter at 37°C had been observed previously (see Chapter 5), H-NS was now found to also down-regulate *coo* transcription at 20°C. Thus, at 20°C in an *hns* mutant background, *coo* transcription was derepressed and a limited amount of CooA was produced. Thermoregulation was not abolished in the absence of H-NS however, as even in *hns* mutant strains increased levels of *coo* transcription and CooA production were observed following growth at 37°C, with respect to the levels detected after growth at 20°C. These results suggest that unlike the reported H-NS dependent thermoregulation of *pap* and *cfaB* transcription (199, 425), H-NS is not required for the thermal control of *coo* transcription. Therefore it appears that the thermoregulatory response is mediated by a factor other than H-NS.

To determine if the temperature-dependent activity of the *coo* promoter was an indirect effect caused by thermoregulation of *rns* expression, the effect of temperature on *rns* transcription was assessed. RT-PCR analysis demonstrated that the *rns* gene is transcribed at both 20°C and 37°C. In contrast, transcription of the homologous gene *cfaD*, which encodes an AraC-like activator of CFA/I fimbrial expression, is turned off at 20°C and turned on at 37°C (199). Thus, despite the close relationship between these ETEC genes, they are regulated differently by temperature. This may be due to the differences between the *rns and cfaD* promoters, especially upstream of the –35 hexamer (134). The reporter plasmid pRnsLacZ-2 was utilised to compare the levels of *rns* promoter activity at 20°C and 37°C. Measurements of β-galactosidase activity revealed that, in wild type *E. coli* strains, *rns* transcription is largely insensitive to temperature. Similar levels of *rns* promoter activity were detected following growth at both temperatures. This is in agreement with a study by Froehlich *et al.*, which also found that growth temperature had little effect on *rns* transcription (134). Thermoregulation occurs at the level of transcription of some genes encoding AraC family activators of fimbrial expression, such as *cfaD*, *perA* and *toxT* (199, 250, 284). It appears that regulation of *rns* more closely resembles that of *fasH* (also designated *fapR*), which encodes the Rns-related activator of 987P fimbrial genes, as *fasH* expression is also temperature-independent (103). Temperature was only found to affect *rns* transcription in the ETEC *hns* mutant background, where *rns* promoter activity at 20°C was 1.4-fold less than that at 37°C. It is possible that in the absence of H-NS, *rns* expression may be influenced by ETEC-specific regulatory elements.

It has previously been suggested that *rns* transcription is repressed by a negative regulator, speculated to be H-NS, that interacts with a region of DNA upstream of the *rns* promoter (134). The data presented in this chapter confirms that H-NS represses *rns* transcription, as expression from the *rns* promoter in an *hns* mutant was found to be significantly greater than in a wild type strain. It was also verified that this repression is likely due to H-NS interacting directly with the *rns* promoter region, as EMSAs revealed that H-NS binds to a *rns* promoter DNA probe. Therefore H-NS negatively regulates the production of CS1 fimbriae at two levels, by repressing the expression of the fimbrial genes themselves and the expression of the activator of the fimbrial genes. As was demonstrated for the *coo* promoter, H-NS-mediated repression of *rns* promoter activity was found to occur at both 20°C and 37°C. H-NS also represses the

transcription of several other genes encoding AraC family proteins, including *ureR, toxT* and *virF* (58, 284, 316). Unlike the regulation of *rns* however, H-NS-mediated repression of *virF* expression is temperature dependent (316).

Western immunoblot analysis was performed to monitor the influence of H-NS and temperature on the level of Rns protein in the cell. Rns levels appeared to be unaffected by H-NS and were not found to be strongly thermally regulated. The amount of Rns protein detected following growth of ETEC strains at 37°C was ~1.7-fold more than that observed after growth at 20°C, possibly due to an increase in translation efficiency at the higher temperature. It is possible that the slightly elevated levels of Rns protein present at 37°C make a contribution towards the increase in *coo* transcription at this temperature. However, it is unlikely that the variation in Rns levels at 20°C and 37°C is the primary cause of the temperature dependence of CS1 fimbrial production.

Therefore, of the various steps involved in the expression of CS1 fimbriae that were investigated in this study, it appeared that the most thermally sensitive process was *coo* transcription. The increase in Rns-dependent *coo* transcription as a response to a rise in temperature may be caused by a number of different mechanisms. The transcriptional activator of the *coo* promoter, Rns, was demonstrated to be present at both 20°C and 37°C. Despite this, the level of *coo* transcription that occurs at 20°C is significantly less than that at 37°C. It is possible that the conformation, and thus the activity of the Rns protein may be responsive to temperature changes. As a result Rns may then exhibit an enhanced ability to bind to, and activate, the *coo* promoter at 37°C. However examples of bacterial transcriptional regulators with intrinsic temperature-sensing capacity, such as the TlpA repressor of *S.* Typhimurium and the RheA repressor of *Streptomyces albus*, are rare (189, 364). Therefore, alternative potential explanations for the temperature regulation of Rns-dependent *coo* transcription must be considered.

It is possible that the DNA topology of the *coo* promoter region may be thermally sensitive. Temperature-induced changes in DNA bending and/or supercoiling could then influence the level of transcriptional activity from the *coo* promoter. At 20°C the promoter may adopt a conformation that is not conducive for efficient transcriptional activation by Rns, thus resulting in a low level of *coo* expression even though Rns is present. As the temperature increases to 37°C, the local DNA topology of the *coo* promoter may become altered such that Rns-dependent activation of transcription is facilitated, possibly due to improved interactions between Rns and RNA polymerase.

Changes in growth temperature are known to alter the supercoiling of DNA (149). Thermally mediated changes in the level of DNA supercoiling can, in turn, influence the rate of transcription from a promoter (90). Thermoregulation of gene expression due to such modifications of DNA superhelicity has been suggested to occur previously. The thermal activation of the *yop* virulence genes of *Yersinia enterocolitica* is reported to correlate with temperature-induced changes in DNA supercoiling (335). In addition, the lack of *virB* transcription at 30°C in *S. flexneri* is proposed to result from a reduction in negative DNA superhelicity in the vicinity of the *virB* promoter at this temperature (403). Induction of the *virB* promoter requires a temperature-dependent alteration of this local DNA superhelicity, which facilitates transcription activation by VirF (403).

Bent DNA frequently occurs in or around promoter regions and has been implicated in transcriptional regulation (288, 299). The *coo* promoter DNA sequence contains numerous short runs of A and T nucleotides, which are known to introduce localised intrinsic DNA bending (298). Furthermore, as described in Chapter 5, the *coo* promoter region was found by *in silico* analysis to contain predicted areas of DNA curvature. Thus, the *coo* promoter may contain an intrinsic DNA bend that inhibits transcription at 20°C but is responsive to changes in temperature. DNA bends have been demonstrated to collapse or "melt" at 37°C and this temperature-dependent conformational transition is postulated to play a role in the thermoregulation of virulence gene expression in *Y. enterocolitica* and other enteric pathogens (336). Similarly, the thermal activation of *virF* expression in *S. flexneri* appears to involve a loss of intrinsic curvature in the *virF* promoter region when the temperature rises above 32°C (315). It is proposed that an increase in temperature causes a transition from a promoter structure in which *virF* transcription is repressed by H-NS, to a more relaxed conformation which facilitates dissociation of H-NS from the promoter and activation of transcription (315). Therefore it is possible that, as in these examples, the DNA structure of the *coo* promoter acts as a thermosensor and its structural alterations regulate transcriptional activity.

The results presented in this chapter suggest that the temperature regulation of CS1 fimbrial expression is exerted mainly at the level of *coo* transcription, although the possibility remains that temperature effects may also operate at the post-transcriptional level. While it is hypothesised that the temperature sensing component is the Rns protein or the DNA topology of the *coo* promoter, further analysis is necessary to verify the molecular mechanism by which thermoregulation occurs.

Chapter 7 General Discussion

The Rns transcriptional regulator of ETEC was discovered almost 20 years ago when it was identified as being required for CS1 and CS2 fimbrial expression (49). Rns has subsequently been found to regulate the expression of other proteins including itself and the inner membrane protein NlpA (35, 134). Rns is one of the better characterised AraC-like transcriptional regulators, relative to many other members of this family of proteins. Amongst other findings, studies performed by the groups of Scott and Munson have identified several genes in the Rns regulon and the sites at which Rns binds to regulate these genes (35, 268, 269, 303). However, little is known about the structure of Rns, the critical regions of the protein, or the roles of such regions in Rns activity. Therefore, one of the major aims of this study was to gain an improved understanding of the structure-function relationship of Rns. Extensive mutagenesis was performed on the *rns* gene to generate Rns derivatives with mutations along the length of the protein. Scanning linker mutagenesis and random mutagenesis were employed to isolate several forms of mutants including two *rns* genes with 15 bp insertions downstream of the transcription start site, six Rns truncates and one point mutant. However, the majority of Rns mutants generated in this study contained pentapeptide insertions in either the N- or C-terminal half of the protein. The effect of these mutations on Rns activity was assessed to determine the areas of the protein that are vital to its function and the areas which can tolerate alteration and thus are less critical.

As is the case for many virulence regulators of the AraC family of proteins, the function of the N-terminal domain of Rns is unknown. It was recently suggested that this area of Rns is not involved in effector binding or dimerisation, the usual roles played by the N-terminal domains of AraC-like proteins (19). In a result similar to that found for the related regulator PerA (311), most of the N-terminal pentapeptide insertions generated in this work (10 out of 14) were found not to have a significant adverse effect on Rns activity. Therefore it appears that the N-terminal half of the protein is highly tolerant of alteration and the numerous disrupted areas are not critical for Rns function. However, the ability of a small number of Rns derivatives with N-terminal mutations to activate the *coo* promoter was found to be severely compromised. This implies that Rns activity is dependent on the regions affected by these mutations and thus some portions of the N-terminal domain are essential. This supports the findings of Basturea *et al.* who report that despite their unknown role, the N-terminal amino acids of Rns are crucial, as in their absence the Rns protein cannot bind to DNA or regulate its target promoters (19). These authors postulate that the essential N-

terminal residues participate in an intramolecular interaction, and thus facilitate DNA binding by the C-terminal domain of Rns (19). This is not unprecedented, as the N- and C-terminal domains of the XylS protein have also been proposed to interact with each other (201). However, it is worthwhile considering possible alternative roles for the N-terminal domain of Rns. Although there is currently no evidence that Rns responds to a chemical signal such as an effector ligand (19), the possibility remains that the N-terminal region of Rns may act as an environmental sensing domain, as was found for ToxT (317), and may respond to a physical signal such as temperature. During the analysis performed in this study on the thermoregulation of CS1 expression, it was demonstrated that both *rns* transcription and Rns protein levels were not strongly thermally regulated, yet Rns-dependent transcription from the *coo* promoter was sensitive to temperature changes. It is possible that the N-terminal domain of Rns undergoes a temperature-dependent conformational transition that causes the protein to become active as a positive transcriptional regulator at 37°C. Temperature sensing by regulators is rare but has been found to occur for a small number, including the TlpA and RheA proteins (189, 364). Future experiments to evaluate this proposed role for the N-terminal domain of Rns could involve determining whether different temperatures alter the circular dichroic profile of the protein, or the ability of Rns and its N-terminal mutant derivatives to bind to a thermally insensitive DNA probe. Another possible role for the N-terminus of Rns is that, as demonstrated for ToxT (317), it contains regions required for transcriptional activation. Work conducted in this study on the regulation of *coo* transcription by Rns and H-NS revealed that Rns does not solely act to overcome the effects of H-NS, but can also up-regulate *coo* expression in the absence of this repressor. Therefore it is possible that Rns activates the *coo* promoter by facilitating the binding of RNA polymerase there, and the N-terminal domain of Rns is involved in forming direct contacts with RNA polymerase subunit(s). This possibility could be assessed by establishing if Rns-dependent transcription requires the α-CTD of RNA polymerase or is affected by combinations of mutations in the σ^{70} subunit of RNA polymerase and/or the N-terminal domain of Rns. The precise role of this region of Rns remains an important unresolved issue.

One Rns mutant with an N-terminal pentapeptide insertion, RnsC102, was found to have retained a reduced ability to activate the *coo* and *rns* promoters and to have a decreased affinity for DNA binding. Analysis of the Rns amino acid sequence revealed that the insertion after residue C102 had occurred within a region of the protein that is

predicted to be disordered and that the insertion had eliminated this predicted disorder. Examination of the effects of further mutagenesis and deletion of this potentially disordered area, the NACRS amino acid sequence, led to the discovery that to be fully functional, the Rns protein appears to require a disordered region between its N- and C-terminal domains. It is postulated that this area of Rns forms, or is part of, a flexible interdomain linker that is necessary to correctly orientate the N- and C-terminal regions of the protein to enable them to participate in their functional roles. The NACRS sequence of Rns has some features in common with the flexible linkers of AraC, RhaR and RhaS (120, 212). It is possible that the activity of the Rns-related regulators, and other AraC-like proteins, may also depend on the presence of such a central disordered/flexible connecting sequence.

In contrast to the N-terminal mutations, all of the C-terminal pentapeptide insertions generated in this study had a severe negative effect on the activity of Rns at the *coo* promoter. This illustrates that the C-terminal domain of Rns is intolerant of disruption and the areas affected by the insertions; the two predicted HTH motifs, a long α-helix between these motifs and the final α-helix of the protein, are vital for Rns activity. It is postulated that these areas are involved in DNA binding and the lack of activity displayed by the mutants is due to their inability to bind to the *coo* promoter. This was confirmed for the RnsQ227 mutant, which contains an insertion immediately upstream of the predicted second HTH motif of Rns. This mutant was subjected to EMSA analysis and found not to bind to *coo*, or *rns*, promoter DNA probes. The essential C-terminal areas of Rns identified in this study may participate directly in DNA binding, i.e. they may form specific interactions with nucleotides in the binding sites at Rns-regulated promoters. This is most likely the case for the two predicted HTH motifs. Therefore, it appears that Rns uses both HTHs to bind DNA as has been proposed or demonstrated for the AraC family members ToxT (57), VirF (309) and PerA (311). The other C-terminal α-helices of Rns may not be directly involved in DNA binding, but are proposed to be required to correctly orientate or stabilise the HTHs to ensure these motifs are in the configuration necessary to interact with DNA.

In addition to its role in DNA binding, it is possible that the C-terminal domain of Rns is required for dimerisation of the protein. Three different techniques employed in this study indicated that the Rns protein is capable of dimerising. The results of a LexA-based gene fusion system and gel filtration chromatography demonstrated that Rns dimerises *in vivo* and *in vitro* respectively. Analysis of Rns dimerisation using the

technique of protein cross-linking was initially hindered by the apparent aggregation of the Rns protein. This phenomenon may have also affected the cross-linking results of Basturea *et al.* who report that they were unable to detect dimerisation of MBP-Rns (19). In this work, however, development of a urea-based denaturation method that reduces protein aggregation, in addition to SDS-PAGE and Western immunoblot analysis of cross-linked MBP-Rns, enabled the routine detection of putative dimeric forms of the protein. Therefore, in contrast to the findings of Basturea *et al.*, it is proposed that Rns is a dimer. The location of the dimerisation domain of Rns is currently unknown, although it has been reported that it does not reside in the N-terminal domain (19). The COILS programme predicted, with high probability, a coiled coil structure in the C-terminal domain of Rns between residues 195 and 222. A crucial α-helix, flanked by the HTH motifs of Rns, is predicted to occur within these residues. Coiled coil domains are commonly involved in protein oligomerisation (240). The AraC protein itself dimerises via an antiparallel coiled coil (378). Therefore, it is possible that the area of Rns located between its predicted HTHs contains the dimerisation domain of the protein. Such a location for the dimerisation domain of an AraC family member would be unusual, and an important future priority is further mutagenesis of Rns and cross-linking analysis to evaluate whether this region is indeed essential for dimerisation.

In addition to mutagenesis of the Rns protein, Rns binding sites I and II at the *coo* promoter were also subjected to mutagenesis during this study. This was done to gain an improved understanding of the interaction between Rns and its binding sites and of Rns-dependent activation of the *coo* promoter. The results of EMSAs conducted with mutation containing *coo* promoter probes and MBP-Rns revealed that the conserved TAT sequence present in Rns binding sites is critical for the interaction of Rns with its target DNA. In addition, it was found that although Rns can bind to either of the *coo* promoter sites individually, co-operative binding of Rns to the two sites may occur. Consistent with the reduced binding of Rns to mutated *coo* promoter DNA, it was demonstrated that mutation of either site reduced Rns-dependent transcription from the *coo* promoter. Mutation of site II had a greater effect than mutation of site I, however. Site II overlaps the -35 hexamer of the *coo* promoter. It is possible that whilst bound at this site Rns recruits RNA polymerase to the *coo* promoter by forming contacts with the subunits of the enzyme, and therefore site II plays a greater role in *coo* transcription activation than site I.

Analysis of reporter plasmids containing the mutated *coo* promoters, and *coo* promoter fragments missing upstream DNA sequences that were generated during reporter plasmid construction, revealed that these promoter variants had increased levels of basal transcriptional activity. This was unexpected and led to further consideration of the regulation of *coo* transcription. Production of CS1 fimbriae is energetically expensive and these surface appendages play a crucial role in ETEC pathogenesis (409). The *coo* promoter directs the expression of CS1 fimbriae. Thus it is possible that this promoter is subject to a complex regulation system, which involves a number of factors in addition to Rns that act to ensure fimbrial expression only occurs under appropriate conditions. Loss of sequences upstream of the *coo* promoter and/or the alteration of the region due to mutagenesis of the Rns binding sites may have influenced components of this putative delicately balanced and complex regulation system and thus led to the elevated basal levels of *coo* expression. It is possible that the *coo* promoter variants exhibit increased Rns-independent transcription because they have lost a repressive structure usually formed by *coo* promoter DNA and/or they are improved promoters for which RNA polymerase has a greater affinity than it does for the wild type promoter. Alternatively, the *coo* promoter alterations may have facilitated the binding of an activator other than Rns or they may have disrupted the activity of a repressor of *coo* transcription. H-NS is known to repress *coo* transcription (270), however the results of experiments performed in this study suggest that H-NS-mediated repression of the *coo* promoter has not been diminished by the loss of upstream sequences of *coo* DNA or the mutation of Rns binding sites I and/or II. It is possible that a repressor other than H-NS down regulates *coo* transcription and its activity was disrupted by the alterations to the *coo* promoter. It will be useful to perform transposon mutagenesis in an attempt to identify additional regulators of *coo* expression.

H-NS-dependent repression of *coo* transcription had previously been reported to mostly occur downstream of the transcription start site (270). However, the work performed in this study demonstrated that H-NS also exerts a repressive effect on a *coo* promoter fragment largely composed of DNA located upstream of the transcription start site. This H-NS-mediated down regulation of *coo* transcription was found to occur at 37°C and at 20°C. Therefore in the absence of H-NS, at 20°C *coo* transcription was derepressed and a limited amount of CooA was produced, in contrast to the lack of CS1 expression at 20°C in the wild type situation. The temperature dependence of *coo* transcription and CooA expression was not lost in an *hns* mutant however, unlike in

other thermally regulated systems such as the expression of P fimbriae (152, 425). Therefore, H-NS does not appear to be the factor responsible for the thermoregulation of CS1 fimbrial expression. This report was the first to demonstrate that H-NS acts directly at the *coo* locus. The results of EMSAs performed with purified His-tagged H-NS revealed that the protein binds to *coo* promoter DNA. Therefore it is possible that H-NS represses *coo* transcription by occluding the binding of RNA polymerase or forming a repressive nucleoprotein complex at the *coo* promoter. In addition, it was found that H-NS also represses *rns* transcription, and again purified H-NS was demonstrated to bind directly to *rns* promoter DNA. Thus it was revealed that H-NS exerts a negative effect on CS1 expression at the level of both *rns* transcription and *coo* transcription. This is consistent with the tight regulation of CS1 fimbrial production.

Analysis of the interplay between Rns and H-NS at the *coo* promoter revealed that it is unlikely that the two proteins have overlapping binding sites at this locus as it appears that, at least to some extent, Rns and H-NS can bind simultaneously to *coo* promoter DNA. Some competition occurs between the two regulators however, as it was revealed that H-NS could largely displace pre-bound Rns from the *coo* promoter DNA probe. In contrast, under the conditions of the competition EMSAs performed in this study, Rns could not displace pre-bound H-NS from the *coo* promoter. This may be due to the fact that the EMSAs were conducted *in vitro* using a linear DNA probe and although they are a useful experiment, they are not a perfect recreation of the natural situation. It is likely that several factors combine *in vivo* to regulate the activity of the *coo* promoter. A possible model based on current results, but not involving potential additional regulators, is that at 20°C H-NS is bound to the *coo* promoter from where it directly represses transcription. Rns is also present at this temperature and may bind to some extent to the *coo* promoter but *coo* transcription does not occur. This may be because the Rns protein is not in the correct conformation to activate transcription and/or it may be due to the *coo* promoter adopting a conformation that is not conducive for transcriptional activation. It is possible that the transition from 20°C to 37°C is required for the Rns protein to undergo a conformational change, thus becoming active, and/or the local DNA topology of the *coo* promoter to alter such that Rns is facilitated in activating transcription and perhaps also displacing H-NS. In order to evaluate whether the *coo* promoter is sensitive to changes in DNA topology, its level of activity in a gyrase mutant background or in the presence of gyrase inhibitors such as novobiocin could be assessed.

Temperature was found to play a role in the regulation of *coo* transcription. Indeed, of the processes leading to CS1 expression analysed in this study, *coo* transcription was the most thermally sensitive step. The possibility remains that further thermoregulation occurs subsequent to *coo* transcription. For example the *coo* transcript may act as an RNA thermometer and thus may undergo temperature-dependent translation, as occurs in the translation of the LcrF protein of *Y. pestis* (180). It will be useful to assess the temperature sensitivity of the post *coo* transcriptional events involved in CS1 production to build a complete picture of the thermoregulation of these fimbriae.

In conclusion, this work has led to an improved understanding of the regulation of CS1 fimbrial expression and of the Rns protein of ETEC. The recent discovery that Rns regulates the expression of NlpA, a protein thought to play a role in the biogenesis of LT-containing outer membrane vesicles (35), revealed that Rns influences two crucial aspects of ETEC virulence: fimbrial production and toxin release. As has been proposed for the related protein AggR in EAEC (170), Rns may be a key regulator of virulence in ETEC. Therefore, an improved characterisation of Rns is valuable not only to the study of AraC proteins, but also to the understanding of the pathogenesis of ETEC. It appears that Rns is a dimer, with each monomer consisting of two domains proposed to be connected via a flexible linker. The function of the N-terminal domain of Rns is unknown but it may have a role in environmental sensing and/or transcription activation. The C-terminal domain of Rns was demonstrated to be involved in DNA binding and may also play a role in dimerisation (Fig. 7.1 (A)). Production of CS1 fimbriae is subject to complex regulation. Expression of the Rns transcriptional activator of the fimbrial genes, and the fimbrial genes themselves was revealed to be repressed by H-NS. As well as regulation by H-NS and Rns, *coo* transcription was also found to be modulated by temperature and, potentially, DNA topology and additional regulators (Fig. 7.1 (B)).

Important future avenues of research in this area could include microarray and ChIP-on-chip studies to further investigate the Rns regulon and establish if this protein influences other virulence functions in ETEC. The solution of the structure of Rns would be invaluable, however, the insolubility of the protein and, as discovered in this work, its apparent tendency to aggregate may mean that this is unlikely to be available in the near future. In the absence of a 3D structure, the mutational analysis performed in this study has provided information on the structure-function relationship of Rns and serves as a basis for future research on this protein and other AraC family members.

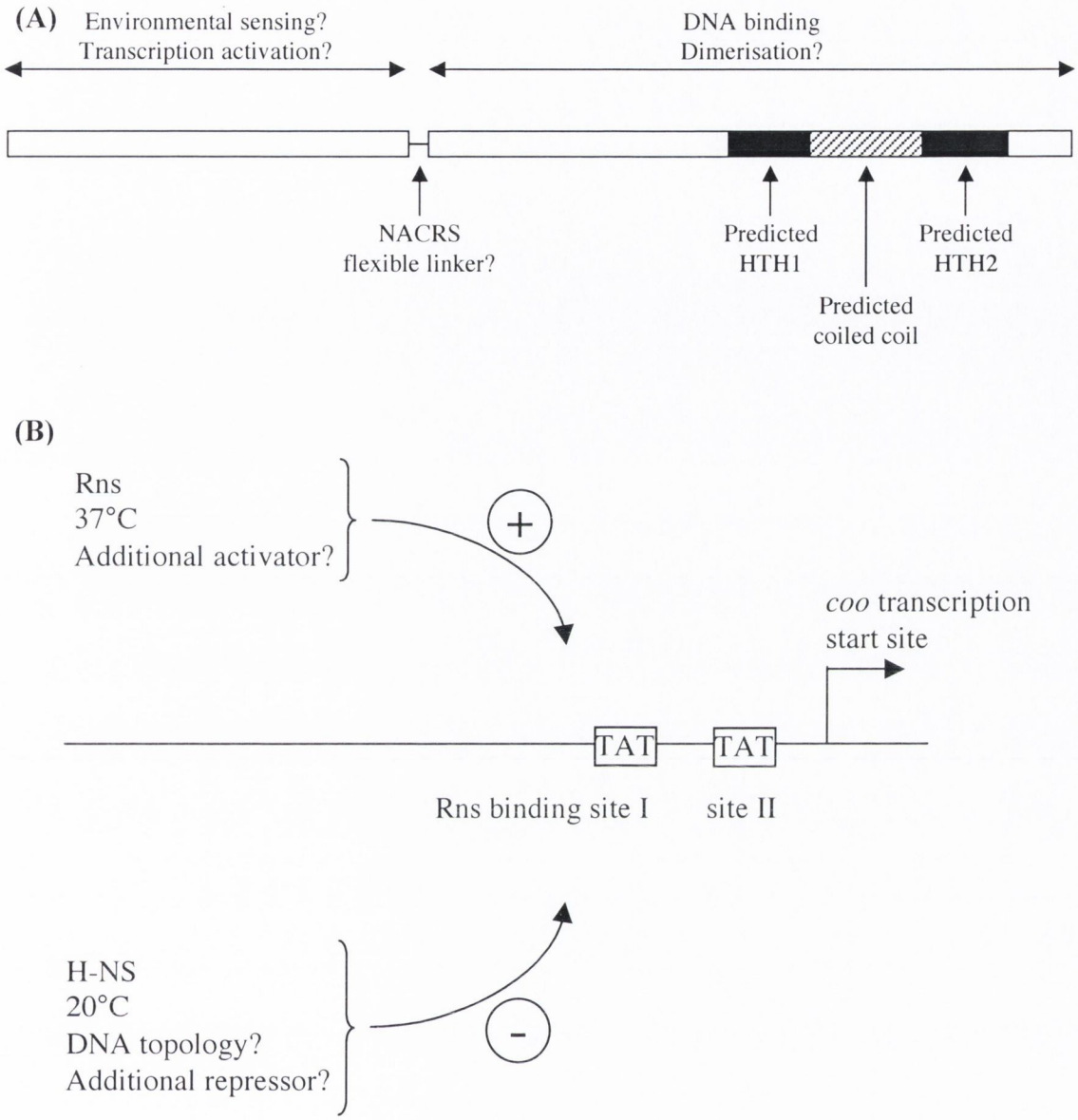

Fig. 7.1. Potential domain organisation of Rns and regulation of the *coo* promoter. **(A)** A schematic diagram of the potential two-domain structure of Rns is shown. The putative flexible linker region is represented as a line, while the N- and C-terminal domains are represented as rectangles above which their potential roles are indicated. Within the C-terminal domain, areas predicted (by *in silico* analysis) to form helix-turn-helix motifs or a coiled coil domain are shown as black or striped boxes respectively. **(B)** A simplified representation of the *coo* promoter region is shown. The transcription start site is indicated as a bent arrow and within Rns binding sites I and II the TAT sequences found to be critical for Rns binding are highlighted. Putative and confirmed factors involved in the regulation of *coo* transcription are listed on the left. Those that activate *coo* transcription are above the promoter and are indicated by a plus sign. Those that repress *coo* transcription are below the promoter and are indicated by a minus sign.

Bibliography

1. **Abraham, J. M., C. S. Freitag, J. R. Clements, and B. I. Eisenstein.** 1985. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. Proc Natl Acad Sci USA **82:**5724-5727.
2. **Adams, H., W. Teertstra, J. Demmers, R. Boesten, and J. Tommassen.** 2003. Interactions between phage-shock proteins in *Escherichia coli*. J Bacteriol **185:**1174-1180.
3. **Adler, B., C. Sasakawa, T. Tobe, S. Makino, K. Komatsu, and M. Yoshikawa.** 1989. A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. Mol Microbiol **3:**627-635.
4. **Afflerbach, H., O. Schroder, and R. Wagner.** 1999. Conformational change of the upstream DNA mediated by H-NS and Fis regulate *E. coli rrnB* P1 promoter activity. J Mol Biol **286:**339-353.
5. **Allen, P. M., I. Roberts, G. J. Boulnois, J. R. Saunders, and C. A. Hart.** 1987. Contribution of capsular polysaccharide and surface properties to virulence of *Escherichia coli* K1. Infect Immun **55:**2662-2668.
6. **Amann, E., B. Ochs, and K. J. Abel.** 1988. Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene **69:**301-315.
7. **Amit, R., A. B. Oppenheim, and J. Stavans.** 2003. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. Biophys J **84:**2467-2473.
8. **Anantha, R. P., A. L. McVeigh, L. H. Lee, M. K. Agnew, F. J. Cassels, D. A. Scott, T. S. Whittam, and S. J. Savarino.** 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.
9. **Anton, B. P., and E. A. Raleigh.** 2004. Transposon-mediated linker insertion scanning mutagenesis of the *Escherichia coli* McrA endonuclease. J Bacteriol **186:**5699-5707.
10. **Aravind, L., V. Anantharaman, S. Balaji, M. Mohan Babu, and L. M. Iyer.** 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev **29:**231-262.
11. **Atlung, T., and H. Ingmer.** 1997. H-NS: a modulator of environmentally regulated gene expression. Mol Microbiol **24:**7-17.
12. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1990. Current Protocols in Molecular Biology, vol. 1 Section 2.1A. John Wiley and Sons, New York.
13. **Babu, M. M., and S. A. Teichmann.** 2003. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. Nucleic Acids Res **31:**1234-1244.
14. **Badea, L., S. Doughty, L. Nicholls, J. Sloan, R. M. Robins-Browne, and E. L. Hartland.** 2003. Contribution of Efa1/LifA to the adherence of enteropathogenic *Escherichia coli* to epithelial cells. Microb Pathog **34:**205-215.
15. **Badger, J. L., C. A. Wass, S. J. Weissman, and K. S. Kim.** 2000. Application of signature-tagged mutagenesis for identification of *Escherichia coli* K1 genes that contribute to invasion of human brain microvascular endothelial cells. Infect Immun **68:**5056-5061.

16. **Barbosa, T. M., and S. B. Levy.** 2000. Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. J Bacteriol **182:**3467-3474.
17. **Barnard, A., A. Wolfe, and S. Busby.** 2004. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. Curr Opin Microbiol **7:**102-108.
18. **Barnhart, M. M., and M. R. Chapman.** 2006. Curli biogenesis and function. Annu Rev Microbiol **60:**131-147.
19. **Basturea, G. N., M. D. Bodero, M. E. Moreno, and G. P. Munson.** 2008. Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding. J Bacteriol **190:**2279-2285.
20. **Bäumler, A. J., R. M. Tsolis, and F. Heffron.** 1996. The *lpf* fimbrial operon mediates adhesion of *Salmonella typhimurium* to murine Peyer's patches. Proc Natl Acad Sci USA **93:**279-283.
21. **Behrens, M., J. Sheikh, and J. P. Nataro.** 2002. Regulation of the overlapping *pic/set* locus in *Shigella flexneri* and enteroaggregative *Escherichia coli*. Infect Immun **70:**2915-2925.
22. **Beloin, C., P. Deighan, M. Doyle, and C. J. Dorman.** 2003. *Shigella flexneri* 2a strain 2457T expresses three members of the H-NS-like protein family: characterisation of the Sfh protein. Mol Gen Genomics **270:**66-77.
23. **Benjamin, P., M. Federman, and C. A. Wanke.** 1995. Characterization of an invasive phenotype associated with enteroaggregative *Escherichia coli*. Infect Immun **63:**3417-3421.
24. **Benjelloun-Touimi, Z., P. J. Sansonetti, and C. Parsot.** 1995. SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion. Mol Microbiol **17:**123-135.
25. **Benz, I., and M. A. Schmidt.** 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. Infect Immun **57:**1506-1511.
26. **Bergthorsson, U., and H. Ochman.** 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol Biol Evol **15:**6-16.
27. **Bernet-Camard, M. F., M. H. Coconnier, S. Hudault, and A. L. Servin.** 1996. Pathogenicity of the diffusely adhering strain *Escherichia coli* C1845: F1845 adhesin-decay accelerating factor interaction, brush border microvillus injury, and actin disassembly in cultured human intestinal epithelial cells. Infect Immun **64:**1918-1928.
28. **Bernier, C., P. Gounon, and C. Le Bouguénec.** 2002. Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia coli* as a sensitive probe for detecting the AAF-encoding operon family. Infect Immun **70:**4302-4311.
29. **Bieber, D., S. W. Ramer, C. Y. Wu, W. J. Murray, T. Tobe, R. Fernandez, and G. K. Schoolnik.** 1998. Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. Science **280:**2114-2118.
30. **Biery, M. C., F. J. Stewart, A. E. Stellwagen, E. A. Raleigh, and N. L. Craig.** 2000. A simple *in vitro* Tn7-based transposition system with low target site selectivity for genome and gene analysis. Nucleic Acids Res **28:**1067-1077.
31. **Bilge, S. S., C. R. Clausen, W. Lau, and S. L. Moseley.** 1989. Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated *Escherichia coli* to HEp-2 cells. J Bacteriol **171:**4281-4289.

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

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

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

35. **Bodero, M. D., M. C. Pilonieta, and G. P. Munson.** 2007. Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic *Escherichia coli*. J Bacteriol **189:**1627-1632.

36. **Boisen, N., C. Struve, F. Scheutz, K. A. Krogfelt, and J. P. Nataro.** 2008. New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. Infect Immun **76:**3281-3292.

37. **Borukhov, S., and E. Nudler.** 2003. RNA polymerase holoenzyme: structure, function and biological implications. Curr Opin Microbiol **6:**93-100.

38. **Bouffartigues, E., M. Buckle, C. Badaut, A. Travers, and S. Rimsky.** 2007. H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. Nat Struct Mol Biol **14:**441-448.

39. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:**248-254.

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

41. **Browning, D. F., and S. J. W. Busby.** 2004. The regulation of bacterial transcription initiation Nat Rev Microbiol **2:**1-9.

42. **Brunelle, A., and R. Schleif.** 1989. Determining residue-base interactions between AraC protein and *araI* DNA. J Mol Biol **209:**607-622.

43. **Brzuszkiewicz, E., H. Brüggemann, H. Liesegang, M. Emmerth, T. Olschläger, G. Nagy, K. Albermann, C. Wagner, C. Buchrieser, L. Emody, G. Gottschalk, J. Hacker, and U. Dobrindt.** 2006. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc Natl Acad Sci USA **103:**12879-12884.

44. **Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti, and C. Parsot.** 2000. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. Mol Microbiol **38:**760-771.

45. **Bullitt, E., and L. Makowski.** 1995. Structural polymorphism of bacterial adhesion pili. Nature **373:**164-167.

46. **Bustos, S. A., and R. F. Schleif.** 1993. Functional domains of the AraC protein. Proc Natl Acad Sci USA **90:**5638-5642.

47. **Cao, Y., B. Hallet, D. J. Sherratt, and F. Hayes.** 1997. Structure-function correlations in the XerD site-specific recombinase revealed by pentapeptide scanning mutagenesis. J Mol Biol **274:**39-53.

48. **Caramel, A., and K. Schnetz.** 1998. Lac and λ repressors relieve silencing of the *Escherichia coli bgl* promoter. Activation by alteration of a repressing nucleoprotein complex J Mol Biol **284:**875-883.
49. **Caron, J., L. M. Coffield, and J. R. Scott.** 1989. A plasmid-encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. Proc Natl Acad Sci USA **86:**963-967.
50. **Caron, J., D. R. Maneval, J. B. Kaper, and J. R. Scott.** 1990. Association of *rns* homologs with colonization factor antigens in clinical *Escherichia coli* isolates. Infect Immun **58:**3442-3444.
51. **Caron, J., and J. R. Scott.** 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.
52. **Casadaban, M. J.** 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.
53. **Cass, L. G., and G. Wilcox.** 1986. Mutations in the *araC* regulatory gene of *Escherichia coli* B/r that affect repressor and activator functions of AraC protein. J Bacteriol **166:**892-900.
54. **Ceschini, S., G. Lupidi, M. Coletta, C. L. Pon, E. Fioretti, and M. Angeletti.** 2000. Multimeric self-assembly equilibria involving the histone-like protein H-NS. A thermodynamic study. J Biol Chem **275:**729-734.
55. **Chang, A. C., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J Bacteriol **134:**1141-1156.
56. **Chaveroche, M. K., J. M. Ghigo, and C. d'Enfert.** 2000. A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. Nucleic Acids Res **28:**E97.
57. **Childers, B. M., G. G. Weber, M. G. Prouty, M. M. Castaneda, F. Peng, and K. E. Klose.** 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.
58. **Coker, C., O. O. Bakare, and H. L. T. Mobley.** 2000. H-NS is a repressor of the *Proteus mirabilis* urease transcriptional activator gene *ureR*. J Bacteriol **182:**2649-2653.
59. **Collins, C. M., and S. E. D'Orazio.** 1993. Bacterial ureases: structure, regulation of expression and role in pathogenesis. Mol Microbiol **9:**907-913.
60. **Cookson, S. T., and J. P. Nataro.** 1996. Characterization of HEp-2 cell projection formation induced by diffusely adherent *Escherichia coli*. Microb Pathog **21:**421-434.
61. **Cowan, J. M., M. L. Urbanowski, M. Talmi, and G. V. Stauffer.** 1993. Regulation of the *Salmonella typhimurium metF* gene by the MetR protein. J Bacteriol **175:**5862-5866.
62. **Crowther, L. J., R. P. Anantha, and M. S. Donnenberg.** 2004. The inner membrane subassembly of the enteropathogenic *Escherichia coli* bundle-forming pilus machine. Mol Microbiol **52:**67-79.
63. **Crowther, L. J., A. Yamagata, L. Craig, J. A. Tainer, and M. S. Donnenberg.** 2005. The ATPase activity of BfpD is greatly enhanced by zinc and allosteric interactions with other Bfp proteins. J Biol Chem **280:**224839-24848.

64. **Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay, and G. J. Barton.** 1998. JPred: a consensus secondary structure prediction server. Bioinformatics **14:**892-893.
65. **Currie, M. G., K. F. Fok, J. Kato, R. J. Moore, F. K. Hamra, K. L. Duffin, and C. E. Smith.** 1992. Guanylin: an endogenous activator of intestinal guanylate cyclase. Proc Natl Acad Sci USA **89:**947-951.
66. **Czeczulin, J. R., S. Balepur, S. Hicks, A. Phillips, R. Hall, M. H. Kothary, F. Navarro-Garcia, and J. P. Nataro.** 1997. Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. Infect Immun **65:**4135-4145.
67. **D'Orazio, S. E., and C. M. Collins.** 1993. The plasmid-encoded urease gene cluster of the family *Enterobacteriaceae* is positively regulated by UreR, a member of the AraC family of transcriptional activators. J Bacteriol **175:**3459-3467.
68. **Dagberg, B., and B. E. Uhlin.** 1992. Regulation of virulence-associated plasmid genes in enteroinvasive *Escherichia coli*. J Bacteriol **174:**7606-7612.
69. **Daines, D. A., and R. P. Silver.** 2000. Evidence for multimerization of Neu proteins involved in polysialic acid synthesis in *Escherichia coli* K1 using improved LexA-based vectors. J Bacteriol **182:**5267-5270.
70. **Dame, R. T., C. Wyman, and N. Goosen.** 2000. H-NS meddiated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res **28:**3504-3510.
71. **Dame, R. T., C. Wyman, and N. Goosen.** 2001. Structural basis for preferential binding of H-NS to curved DNA. Biochimie **83:**231-234.
72. **Dame, R. T., C. Wyman, R. Wurm, R. Wagner, and N. Goosen.** 2002. Structural basis for H-NS-mediated trapping of RNA Polymerase in the open initiation complex at the *rrnB* P1. J Biol Chem **277:**2146-2150.
73. **Daniel, A., A. Singh, L. J. Crowther, P. J. Fernandes, W. Schreiber, and M. S. Donnenberg.** 2006. Interaction and localization studies of enteropathogenic *Escherichia coli* type IV bundle-forming pilus outer membrane components. Microbiology **152:**2405-2420.
74. **Darwin, K. H., and V. L. Miller.** 2001. Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*. EMBO J **20:**1850-1862.
75. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA **97:**6640-6645.
76. **Davis, J. M., S. B. Rasmussen, and A. D. O'Brien.** 2005. Cytotoxic necrotizing factor type 1 production by uropathogenic *Escherichia coli* modulates polymorphonuclear leukocyte function. Infect Immun **73:**5301-5310.
77. **Debroy, C., J. Yealy, R. A. Wilson, M. K. Bhan, and R. Kumar.** 1995. Antibodies raised against the outer membrane protein interrupt adherence of enteroaggregative *Escherichia coli*. Infect Immun **63:**2873-2879.
78. **Defez, R., and M. De Felice.** 1981. Cryptic operon for beta-glucoside metabolism in *Escherichia coli* K12: genetic evidence for a regulatory protein. Genetics **97:**11-25.
79. **Deighan, P., C. Beloin, and C. J. Dorman.** 2003. Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of *Shigella flexneri* 2a strain 2457T. Mol Microbiol **48:**1401-1416.

80. **Dersch, P., K. Schmidt, and E. Bremer.** 1993. Synthesis of the *Escherichia coli* K-12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. Mol Microbiol **8:**875-889.
81. **DiRita, V. J., C. Parsot, G. Jander, and J. J. Mekalanos.** 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. Proc Natl Acad Sci USA **88:**5403-5407.
82. **Dmitrova, M., G. Younes-Cauet, P. Oertel-Buchheit, D. Porte, M. Schnarr, and M. Granger-Schnarr.** 1998. A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in *Escherichia coli*. Mol Gen Genet **257:**205-212.
83. **Dobrindt, U.** 2005. (Patho-)Genomics of *Escherichia coli*. Int J Med Microbiol **295:**357-371.
84. **Dodson, K. W., F. Jacob-Dubuisson, R. T. Striker, and S. J. Hultgren.** 1993. Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. Proc Natl Acad Sci USA **90:**3670-3674.
85. **Dole, S., V. Nagarajavel, and K. Schnetz.** 2004. The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli bgl* operon downstream of the promoter. Mol Microbiol **52:**589-600.
86. **Domínguez-Cuevas, P., P. Marín, S. Busby, J. L. Ramos, and S. Marqués.** 2008. Roles of effectors in XylS-dependent transcription activation: intramolecular domain derepression and DNA binding. J Bacteriol **190:**3118-3128.
87. **Domínguez-Cuevas, P., P. Marín, S. Marqués, and J. L. Ramos.** 2008. XylS-Pm promoter interactions through two helix-turn-helix motifs: identifying XylS residues important for DNA binding and activation. J Mol Biol **375:**59-69.
88. **Donnenberg, M. S., J. A. Girón, J. P. Nataro, and J. B. Kaper.** 1992. A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. Mol Microbiol **6:**3427-3437.
89. **Donnenberg, M. S., S. Tzipori, M. L. McKee, A. D. O'Brien, J. Alroy, and J. B. Kaper.** 1993. The role of the *eae* gene of enterohemorrhagic *Escherichia coli* in intimate attachment *in vitro* and in a porcine model. J Clin Invest **92:**1418-1424.
90. **Dorman, C. J.** 1991. DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. Infect Immun **59:**745-749.
91. **Dorman, C. J.** 2007. H-NS, the genome sentinel. Nat Rev Microbiol **5:**157-161.
92. **Dorman, C. J.** 2004. H-NS: a universal regulator for a dynamic genome. Nat Rev Microbiol **2:**391-400.
93. **Dorman, C. J., and P. Deighan.** 2003. Regulation of gene expression by histone-like proteins in bacteria. Curr Opin Genet Dev **13:**105-107.
94. **Dorman, C. J., and C. F. Higgins.** 1987. Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. J Bacteriol **169:**3840-3843.
95. **Dorman, C. J., and M. E. Porter.** 1998. The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. Mol Microbiol **29:**677-684.
96. **Dove, S. L., S. G. Smith, and C. J. Dorman.** 1997. Control of *Escherichia coli* type 1 fimbrial gene expression in stationary phase: a negative role for RpoS. Mol Gen Genet **254:**13-20.

97. **Dower, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res **16:**6127-6145.
98. **Dudley, E. G., N. R. Thomson, J. Parkhill, N. P. Morin, and J. P. Nataro.** 2006. Proteomic and microarray characterization of the AggR regulon identifies a *pheU* pathogenicity island in enteroaggregative *Escherichia coli.* Mol Microbiol **61:**1267-1282.
99. **Dunker, A. K., C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic.** 2002. Intrinsic disorder and protein function. Biochemistry **41:**6573-6582.
100. **Dunker, A. K., E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger, and J. E. Villafranca.** 1998. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput **3:**473-484.
101. **Durand, J. M., B. Dagberg, B. E. Uhlin, and G. R. Björk.** 2000. Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. Mol Microbiol **35:**924-935.
102. **Dyson, H. J., and P. E. Wright.** 2002. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol **12:**54-60.
103. **Edwards, R. A., and D. M. Schifferli.** 1997. Differential regulation of *fasA* and *fasH* expression of *Escherichia coli* 987P fimbriae by environmental cues. Mol Microbiol **25:**797-809.
104. **Egan, S. M.** 2002. Growing repertoire of AraC/XylS activators. J Bacteriol **184:**5529-5532.
105. **Egan, S. M., and R. F. Schleif.** 1993. A regulatory cascade in the induction of *rhaBAD.* J Mol Biol **234:**87-98.
106. **Eichenbaum, Z., and Z. Livneh.** 1998. UV light induces IS*10* transposition in *Escherichia coli.* Genetics **149:**1173-1181.
107. **Elias, W. P., Jr., J. R. Czeczulin, I. R. Henderson, L. R. Trabulsi, and J. P. Nataro.** 1999. Organization of biogenesis genes for aggregative adherence fimbria II defines a virulence gene cluster in enteroaggregative *Escherichia coli.* J Bacteriol **181:**1779-1785.
108. **Elias, W. P., A. P. Uber, S. K. Tomita, L. R. Trabulsi, and T. A. Gomes.** 2002. Combinations of putative virulence markers in typical and variant enteroaggregative *Escherichia coli* strains from children with and without diarrhoea. Epidemiol Infect **129:**49-55.
109. **Elliott, E. J., R. M. Robins-Browne, E. V. O'Loughlin, V. Bennett-Wood, J. Bourke, P. Henning, G. G. Hogg, J. Knight, H. Powell, and D. Redmond.** 2001. Nationwide study of haemolytic uraemic syndrome: clinical, microbiological, and epidemiological features. Arch Dis Child **85:**125-131.
110. **Ellison, D. W., and V. L. Miller.** 2006. H-NS represses *inv* transcription in *Yersinia enterocolitica* through competition with RovA and interaction with YmoA. J Bacteriol **188:**5101-5112.
111. **Elsinghorst, E. A., and D. J. Kopecko.** 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli.* Infect Immun **60:**2409-2417.
112. **Elsinghorst, E. A., and J. A. Weitz.** 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.

113. **Emory, S. A., and J. G. Belasco.** 1990. The *ompA* 5' untranslated RNA segment functions in *Escherichia coli* as a growth-rate-regulated mRNA stabilizer whose activity is unrelated to translational efficiency. J Bacteriol **172:**4472-4481.
114. **Engler, M. J., and C. C. Richardson.** 1982. DNA Ligases, p. 3-30. *In* P. D. Boyer (ed.), The Enzymes, vol. 15. Academic Press, San Diego, CA.
115. **Englesberg, E., C. Squires, and F. Meronk, Jr.** 1969. The L-arabinose operon in *Escherichia coli* B-r: a genetic demonstration of two functional states of the product of a regulator gene. Proc Natl Acad Sci USA **62:**1100-1107.
116. **Epstein, E. A., and M. R. Chapman.** 2008. Polymerizing the fibre between bacteria and host cells: the biogenesis of functional amyloid fibres. Cell Microbiol **10:**1413-1420.
117. **Eslava, C., F. Navarro-García, J. R. Czeczulin, I. R. Henderson, A. Cravioto, and J. P. Nataro.** 1998. Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. Infect Immun **66:**3155-3163.
118. **Estrem, S. T., T. Gaal, W. Ross, and R. L. Gourse.** 1998. Identification of an UP element consensus sequence for bacterial promoters. Proc Natl Acad Sci USA **95:**9761-9766.
119. **Eustance, R. J., S. A. Bustos, and R. F. Schleif.** 1994. Reaching out. Locating and lengthening the interdomain linker in AraC protein. J Mol Biol **242:**330-338.
120. **Eustance, R. J., and R. F. Schleif.** 1996. The linker region of AraC protein. J Bacteriol **178:**7025-7030.
121. **Evans, D. G., and D. J. Evans, Jr.** 1978. New surface-associated heat-labile colonization factor antigen (CFA/II) produced by enterotoxigenic *Escherichia coli* of serogroups O6 and O8. Infect Immun **21:**638-647.
122. **Evans, D. G., D. J. Evans, Jr., and W. Tjoa.** 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.
123. **Fagan, R. P., and S. G. Smith.** 2007. The Hek outer membrane protein of *Escherichia coli* is an auto-aggregating adhesin and invasin. FEMS Microbiol Lett **269:**248-255.
124. **Fagan, R. P., M. A. Lambert, and S. G. Smith.** 2008. The Hek outer membrane protein of *Escherichia coli* strain RS218 binds to proteoglycan and utilizes a single extracellular loop for adherence, invasion, and autoaggregation. Infect Immun **76:**1135-1142.
125. **Falbo, V., T. Pace, L. Picci, E. Pizzi, and A. Caprioli.** 1993. Isolation and nucleotide sequence of the gene encoding cytotoxic necrotizing factor 1 of *Escherichia coli*. Infect Immun **61:**4909-4914.
126. **Falconi, M., B. Colonna, G. Prosseda, G. Micheli, and C. O. Gualerzi.** 1998. Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. EMBO J **17:**7033-7043.
127. **Falconi, M., M. T. Gualtieri, A. La Teana, M. A. Losso, and C. L. Pon.** 1988. Proteins from the prokaryotic nucleoid: primary and quaternary structure of the 15-kDa *Escherichia coli* DNA binding protein H-NS. Mol Microbiol **2:**323-329.

128. **Falconi, M., G. Prosseda, M. Giangrossi, E. Beghetto, and B. Colonna.** 2001. Involvement of FIS in the H-NS-mediated regulation of *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. Mol Microbiol **42:**439-452.
129. **Fleckenstein, J. M., J. T. Holland, and D. L. Hasty.** 2002. Interaction of an outer membrane protein of enterotoxigenic *Escherichia coli* with cell surface heparan sulfate proteoglycans. Infect Immun **70:**1530-1537.
130. **Fleckenstein, J. M., D. J. Kopecko, R. L. Warren, and E. A. Elsinghorst.** 1996. Molecular characterization of the *tia* invasion locus from enterotoxigenic *Escherichia coli*. Infect Immun **64:**2256-2265.
131. **Forsman, K., M. Göransson, and B. E. Uhlin.** 1989. Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. EMBO J **8:**1271-1277.
132. **Fowler, D. M., A. V. Koulov, W. E. Balch, and J. W. Kelly.** 2007. Functional amyloid--from bacteria to humans. Trends Biochem Sci **32:**217-224.
133. **Francklyn, C. S., and N. Lee.** 1988. AraC proteins with altered DNA sequence specificity which activate a mutant promoter in *Escherichia coli*. J Biol Chem **263:**4400-4407.
134. **Froehlich, B., L. Husmann, J. Caron, and J. R. Scott.** 1994. Regulation of *rns*, a positive regulatory factor for pili of enterotoxigenic *Escherichia coli*. J Bacteriol **176:**5385-5392.
135. **Froehlich, B., J. Parkhill, M. Sanders, M. A. Quail, and J. R. Scott.** 2005. The pCoo plasmid of enterotoxigenic *Escherichia coli* is a mosaic cointegrate. J Bacteriol **187:**6509-6516.
136. **Fujiyama, R., J. Nishi, N. Imuta, K. Tokuda, K. Manago, and Y. Kawano.** 2008. The *shf* gene of a *Shigella flexneri* homologue on the virulent plasmid pAA2 of enteroaggregative *Escherichia coli* 042 is required for firm biofilm formation. Curr Microbiol **56:**474-480.
137. **Fukuta, S., J. L. Magnani, E. M. Twiddy, R. K. Holmes, and V. Ginsburg.** 1988. Comparison of the carbohydrate-binding specificities of cholera toxin and *Escherichia coli* heat-labile enterotoxins LTh-I, LT-IIa, and LT-IIb. Infect Immun **56:**1748-1753.
138. **Gaastra, W., and A. M. Svennerholm.** 1996. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). Trends Microbiol **4:**444-452.
139. **Gallegos, M. T., R. Schleif, A. Bairoch, K. Hofmann, and J. L. Ramos.** 1997. Arac/XylS family of transcriptional regulators. Microbiol Mol Biol Rev **61:**393-410.
140. **Gally, D. L., J. A. Bogan, B. I. Eisenstein, and I. C. Blomfield.** 1993. Environmental regulation of the *fim* switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: effects of temperature and media. J Bacteriol **175:**6186-6193.
141. **Gendlina, I., D. M. Gutman, V. Thomas, and C. M. Collins.** 2002. Urea-dependent signal transduction by the virulence regulator UreR. J Biol Chem **277:**37349-37358.
142. **Ghosh, M., and R. F. Schleif.** 2001. Biophysical evidence of arm-domain interactions in AraC. Anal Biochem **295:**107-112.
143. **Gillette, W. K., R. G. Martin, and J. L. Rosner.** 2000. Probing the *Escherichia coli* transcriptional activator MarA using alanine-scanning mutagenesis: residues important for DNA binding and activation. J Mol Biol **299:**1245-1255.

144. **Girón, J. A., A. S. Ho, and G. K. Schoolnik.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710-713.
145. **Girón, J. A., A. S. Ho, and G. K. Schoolnik.** 1993. Characterization of fimbriae produced by enteropathogenic *Escherichia coli*. J Bacteriol **175:**7391-7403.
146. **Girón, J. A., T. Jones, F. Millán-Velasco, E. Castro-Muñoz, L. Zárate, J. Fry, G. Frankel, S. L. Moseley, B. Baudry, and J. B. Kaper.** 1991. Diffuse-adhering *Escherichia coli* (DAEC) as a putative cause of diarrhea in Mayan children in Mexico. J Infect Dis **163:**507-513.
147. **Girón, J. A., J. G. Xu, C. R. González, D. Hone, J. B. Kaper, and M. M. Levine.** 1995. Simultaneous expression of CFA/I and CS3 colonization factor antigens of enterotoxigenic *Escherichia coli* by Δ*aroC*, Δ*aroD Salmonella typhi* vaccine strain CVD 908. Vaccine **13:**939-946.
148. **Glinkowska, M., J. Majka, W. Messer, and G. Wegrzyn.** 2003. The mechanism of regulation of bacteriophage λ p_R promoter activity by *Escherichia coli* DnaA protein. J Biol Chem **278:**22250-22256.
149. **Goldstein, E., and K. Drlica.** 1984. Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. Proc Natl Acad Sci USA **81:**4046-4050.
150. **Gomez-Duarte, O. G., and J. B. Kaper.** 1995. A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. Infect Immun **63:**1767-1776.
151. **Gómez-Duarte, O. G., A. Ruiz-Tagle, D. C. Gómez, G. I. Viboud, K. G. Jarvis, J. B. Kaper, and J. A. Girón.** 1999. Identification of *lngA*, the structural gene of longus type IV pilus of enterotoxigenic *Escherichia coli*. Microbiology **145:**1809-1816.
152. **Göransson, M., B. Sondén, P. Nilsson, B. Dagberg, K. Forsman, K. Emanuelsson, and B. E. Uhlin.** 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. Nature **344:**682-685.
153. **Göransson, M., and B. E. Uhlin.** 1984. Environmental temperature regulates transcription of a virulence pili operon in *E. coli*. EMBO J **3:**2885-2888.
154. **Grainger, D. C., D. Hurd, M. D. Goldberg, and S. J. Busby.** 2006. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. Nucleic Acids Res **34:**4642-4652.
155. **Grainger, D. C., C. L. Webster, T. A. Belyaeva, E. I. Hyde, and S. J. Busby.** 2004. Transcription activation at the *Escherichia coli melAB* promoter: interactions of MelR with its DNA target site and with domain 4 of the RNA polymerase sigma subunit. Mol Microbiol **51:**1297-1309.
156. **Grange, P. A., A. K. Erickson, T. J. Anderson, and D. H. Francis.** 1998. Characterization of the carbohydrate moiety of intestinal mucin-type sialoglycoprotein receptors for the K88ac fimbrial adhesin of *Escherichia coli*. Infect Immun **66:**1613-1621.
157. **Griffith, I. P.** 1972. The effect of cross-links on the mobility of proteins in dodecyl sulphate-polyacrylamide gels. Biochem J **126:**553-560.
158. **Griffith, K. L., S. M. Becker, and R. E. Wolf, Jr.** 2005. Characterization of TetD as a transcriptional activator of a subset of genes of the *Escherichia coli* SoxS/MarA/Rob regulon. Mol Microbiol **56:**1103-1117.
159. **Griffith, K. L., and R. E. Wolf Jr.** 2002. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: identifying

amino acids important for DNA binding and transcription activation. J Mol Biol **322:**237-257.

160. **Guagliardi, A., L. Cerchia, and M. Rossi.** 1995. Prevention of *in vitro* protein thermal aggregation by the *Sulfolobus solfataricus* chaperonin. J Biol Chem **270:**28126-28132.
161. **Gunther, N. W., V. Lockatell, D. E. Johnson, and H. L. Mobley.** 2001. *In vivo* dynamics of type 1 fimbria regulation in uropathogenic *Escherichia coli* during experimental urinary tract infection. Infect Immun **69:**2838-2846.
162. **Gunther, N. W., J. A. Snyder, V. Lockatell, I. Blomfield, D. E. Johnson, and H. L. Mobley.** 2002. Assessment of virulence of uropathogenic *Escherichia coli* type 1 fimbrial mutants in which the invertible element is phase-locked on or off. Infect Immun **70:**3344-3354.
163. **Guyer, D. M., S. Radulovic, F. E. Jones, and H. L. Mobley.** 2002. Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. Infect Immun **70:**4539-4546.
164. **Ha, U. H., J. Kim, H. Badrane, J. Jia, H. V. Baker, D. Wu, and S. Jin.** 2004. An *in vivo* inducible gene of *Pseudomonas aeruginosa* encodes an anti-ExsA to suppress the type III secretion system. Mol Microbiol **54:**307-320.
165. **Haagmans, W., and M. van der Woude.** 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.
166. **Hahn, S., T. Dunn, and R. Schleif.** 1984. Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. J Mol Biol **180:**61-72.
167. **Hallet, B., D. J. Sherratt, and F. Hayes.** 1997. Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. Nucleic Acids Res **25:**1866-1867.
168. **Hammar, M., A. Arnqvist, Z. Bian, A. Olsén, and S. Normark.** 1995. Expression of two *csg* operons is required for production of fibronectin- and congo red-binding curli polymers in *Escherichia coli* K-12. Mol Microbiol **18:**661-670.
169. **Hammar, M., Z. Bian, and S. Normark.** 1996. Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*. Proc Natl Acad Sci USA **93:**6562-6566.
170. **Harrington, S. M., E. G. Dudley, and J. P. Nataro.** 2006. Pathogenesis of enteroaggregative *Escherichia coli* infection. FEMS Microbiol Lett **254:**12-18.
171. **Hassanain, H. H., W. Dai, and S. L. Gupta.** 1993. Enhanced gel mobility shift assay for DNA-binding factors. Anal Biochem **213:**162-167.
172. **Hautefort, I., M. J. Proenca, and J. C. Hinton.** 2003. Single-copy green fluorescent protein gene fusions allow accurate measurement of *Salmonella* gene expression *in vitro* and during infection of mammalian cells. Appl Environ Microbiol **69:**7480-7491.
173. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res **8:**11-22.
174. **Hayes, F.** 2003. Transposon-based strategies for microbial functional genomics and proteomics. Annu Rev Genet **37:**3-29.

175. **Hayes, F., and B. Hallet.** 2000. Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. Trends Microbiol **8:**571-577.
176. **Hendrickson, W., and R. Schleif.** 1985. A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. Proc Natl Acad Sci USA **82:**3129-3133.
177. **Hernday, A., M. Krabbe, B. Braaten, and D. Low.** 2002. Self-perpetuating epigenetic pili switches in bacteria. Proc Natl Acad Sci USA **99:**16470-16476.
178. **Hernday, A. D., B. A. Braaten, G. Broitman-Maduro, P. Engelberts, and D. A. Low.** 2004. Regulation of the pap epigenetic switch by CpxAR: phosphorylated CpxR inhibits transition to the phase ON state by competition with Lrp. Mol Cell **16:**537-547.
179. **Hernday, A. D., B. A. Braaten, and D. A. Low.** 2003. The mechanism by which DNA adenine methylase and PapI activate the *pap* epigenetic switch. Mol Cell **12:**947-957.
180. **Hoe, N. P., and J. D. Goguen.** 1993. Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. J Bacteriol **175:**7901-7909.
181. **Hoffman, J. A., J. L. Badger, Y. Zhang, S. H. Huang, and K. S. Kim.** 2000. *Escherichia coli* K1 *aslA* contributes to invasion of brain microvascular endothelial cells *in vitro* and *in vivo*. Infect Immun **68:**5062-5067.
182. **Holcroft, C. C., and S. M. Egan.** 2000. Interdependence of activation at *rhaSR* by cyclic AMP receptor protein, the RNA polymerase alpha subunit C-terminal domain, and *rhaR*. J Bacteriol **182:**6774-6782.
183. **Holcroft, C. C., and S. M. Egan.** 2000. Roles of cyclic AMP receptor protein and the carboxyl-terminal domain of the alpha subunit in transcription activation of the *Escherichia coli rhaBAD* operon. J Bacteriol **182:**3529-3535.
184. **Hu, J. C., E. K. O'Shea, P. S. Kim, and R. T. Sauer.** 1990. Sequence requirements for coiled-coils: analysis with lamda repressor-GCN4 leucine zipper fusions. Science **250:**1400-1403.
185. **Huang, S. H., Y. H. Chen, Q. Fu, M. Stins, Y. Wang, C. Wass, and K. S. Kim.** 1999. Identification and characterization of an *Escherichia coli* invasion gene locus, ibeB, required for penetration of brain microvascular endothelial cells. Infect Immun **67:**2103-2109.
186. **Huang, S. H., C. Wass, Q. Fu, N. V. Prasadarao, M. Stins, and K. S. Kim.** 1995. *Escherichia coli* invasion of brain microvascular endothelial cells *in vitro* and *in vivo*: molecular cloning and characterization of invasion gene *ibe10*. Infect Immun **63:**4470-4475.
187. **Hulbert, R. R., and R. K. Taylor.** 2002. Mechanism of ToxT-dependent transcriptional activation at the *Vibrio cholerae tcpA* promoter. J Bacteriol **184:**5533-5544.
188. **Huott, P. A., W. Liu, J. A. McRoberts, R. A. Giannella, and K. Dharmsathaphorn.** 1988. Mechanism of action of *Escherichia coli* heat stable enterotoxin in a human colonic cell line. J Clin Invest **82:**514-523.
189. **Hurme, R., K. D. Berndt, S. J. Normark, and M. Rhen.** 1997. A proteinaceous gene regulatory thermometer in *Salmonella*. Cell **90:**55-64.
190. **Hurme, R., and M. Rhen.** 1998. Temperature sensing in bacterial gene regulation - what it all boils down to. Mol Microbiol **30:**1-6.
191. **Hurt, J. K., S. Olgen, and G. A. Garcia.** 2007. Site-specific modification of *Shigella flexneri virF* mRNA by tRNA-guanine transglycosylase *in vitro*. Nucleic Acids Res **35:**4905-4913.

192. **Ibarra, J. A., E. Pérez-Rueda, L. Segovia, and J. L. Puente.** 2007. The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. Genetica **133:**65-76.
193. **Ibarra, J. A., M. I. Villalba, and J. L. Puente.** 2003. Identification of the DNA binding sites of PerA, the transcriptional activator of the *bfp* and *per* operons in enteropathogenic *Escherichia coli*. J Bacteriol **185:**2835-2847.
194. **Inouye, S., A. Nakazawa, and T. Nakazawa.** 1986. Nucleotide sequence of the regulatory gene *xylS* on the *Pseudomonas putida* TOL plasmid and identification of the protein product. Gene **44:**235-242.
195. **Jarvis, K. G., J. A. Girón, A. E. Jerse, T. K. McDaniel, M. S. Donnenberg, and J. B. Kaper.** 1995. Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. Proc Natl Acad Sci USA **92:**7996-8000.
196. **Johnson, J. R.** 1991. Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev **4:**80-128.
197. **Johnson, J. R., M. A. Kuskowski, A. Gajewski, S. Soto, J. P. Horcajada, M. T. Jimenez de Anta, and J. Vila.** 2005. Extended virulence genotypes and phylogenetic background of *Escherichia coli* isolates from patients with cystitis, pyelonephritis, or prostatitis. J Infect Dis **191:**46-50.
198. **Johnson, J. R., and A. L. Stell.** 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. J Infect Dis **181:**261-272.
199. **Jordi, B. J., B. Dagberg, L. A. de Haan, A. M. Hamers, B. A. van der Zeijst, W. Gaastra, and B. E. Uhlin.** 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*. EMBO J **11:**2627-2632.
200. **Jordi, B. J., G. A. Willshaw, B. A. van der Zeijst, and W. Gaastra.** 1992. The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic *Escherichia coli*. DNA Seq **2:**257-263.
201. **Kaldalu, N., U. Toots, V. de Lorenzo, and M. Ustav.** 2000. Functional domains of the TOL plasmid transcription factor XylS. J Bacteriol **182:**1118-1126.
202. **Kaper, J. B., J. P. Nataro, and H. L. Mobley.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123-140.
203. **Kapust, R. B., and D. S. Waugh.** 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci **8:**1668-1674.
204. **Keane, W. F., R. Welch, G. Gekker, and P. K. Peterson.** 1987. Mechanism of *Escherichia coli* alpha-hemolysin-induced injury to isolated renal tubular cells. Am J Pathol **126:**350-357.
205. **Kellermann, O. K., and T. Ferenci.** 1982. Maltose-binding protein from *Escherichia coli*. Methods Enzymol **90 Pt E:**459-463.
206. **Kenny, B., R. DeVinney, M. Stein, D. J. Reinscheid, E. A. Frey, and B. B. Finlay.** 1997. Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. Cell **91:**511-520.
207. **Khan, N. A., Y. Wang, K. J. Kim, J. W. Chung, C. A. Wass, and K. S. Kim.** 2002. Cytotoxic necrotizing factor-1 contributes to *Escherichia coli* K1 invasion of the central nervous system. J Biol Chem **277:**15607-15612.
208. **Kim, K. S.** 2003. Pathogenesis of bacterial meningitis: from bacteraemia to neuronal injury. Nat Rev Neurosci **4:**376-385.

209. **Klemm, P.** 1986. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. EMBO J **5:**1389-1393.
210. **Kohn, W. D., C. T. Mant, and R. S. Hodges.** 1997. Alpha-helical protein assembly motifs. J Biol Chem **272:**2583-2586.
211. **Kolin, A., V. Balasubramaniam, J. M. Skredenske, J. R. Wickstrum, and S. M. Egan.** 2008. Differences in the mechanism of the allosteric l-rhamnose responses of the AraC/XylS family transcription activators RhaS and RhaR. Mol Microbiol **68:**448-461.
212. **Kolin, A., V. Jevtic, L. Swint-Kruse, and S. M. Egan.** 2007. Linker regions of the RhaS and RhaR proteins. J Bacteriol **189:**269-271.
213. **Konkel, M. E., and K. Tilly.** 2000. Temperature-regulated expression of bacterial virulence genes. Microbes Infect **2:**157-166.
214. **Korhonen, T. K., M. V. Valtonen, J. Parkkinen, V. Vaisanen-Rhen, J. Finne, F. Orskov, I. Orskov, S. B. Svenson, and P. H. Makela.** 1985. Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. Infect Immun **48:**486-491.
215. **Korhonen, T. K., R. Virkola, and H. Holthöfer.** 1986. Localization of binding sites for purified *Escherichia coli* P fimbriae in the human kidney. Infect Immun **54:**328-332.
216. **Kozmik, Z., P. Urbanek, and V. Paces.** 1990. Albumin improves formation and detection of some specific protein-DNA complexes in the mobility shift assay. Nucleic Acids Res **18:**2198.
217. **Krukonis, E. S., R. R. Yu, and V. J. Dirita.** 2000. The *Vibrio cholerae* ToxR/TcpP/ToxT virulence cascade: distinct roles for two membrane-localized transcriptional activators on a single promoter. Mol Microbiol **38:**67-84.
218. **Kuehn, M. J., J. Heuser, S. Normark, and S. J. Hultgren.** 1992. P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. Nature **356:**252-255.
219. **Kuehn, M. J., S. Normark, and S. J. Hultgren.** 1991. Immunoglobulin-like PapD chaperone caps and uncaps interactive surfaces of nascently translocated pilus subunits. Proc Natl Acad Sci USA **88:**10586-10590.
220. **Kuipers, O. P.** 1996. Random mutagenesis by using mixtures of dNTP and dITP in PCR. Methods Mol Biol **57:**351-356.
221. **Kwon, H. J., M. H. Bennik, B. Demple, and T. Ellenberger.** 2000. Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. Nat Struct Biol **7:**424-430.
222. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.
223. **Lan, R., M. C. Alles, K. Donohoe, M. B. Martinez, and P. R. Reeves.** 2004. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun **72:**5080-5088.
224. **Langermann, S., S. Palaszynski, M. Barnhart, G. Auguste, J. S. Pinkner, J. Burlein, P. Barren, S. Koenig, S. Leath, C. H. Jones, and S. J. Hultgren.** 1997. Prevention of mucosal *Escherichia coli* infection by FimH-adhesin-based systemic vaccination. Science **276:**607-611.
225. **Lawson, C. L., D. Swigon, K. S. Murakami, S. A. Darst, H. M. Berman, and R. H. Ebright.** 2004. Catabolite activator protein: DNA binding and transcription activation. Curr Opin Struct Biol **14:**10-20.

226. **Lerner, C. G., and M. Inouye.** 1990. Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability. Nucleic Acids Res **18:**4631.
227. **Li, H., L. Qian, Z. Chen, D. Thibault, G. Liu, T. Liu, and D. G. Thanassi.** 2004. The outer membrane usher forms a twin-pore secretion complex. J Mol Biol **344:**1397-1407.
228. **Lindenthal, C., and E. A. Elsinghorst.** 2001. Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestine epithelial cells. Infect Immun **69:**52-57.
229. **Lindenthal, C., and E. A. Elsinghorst.** 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. Infect Immun **67:**4084-4091.
230. **Linding, R., R. B. Russell, V. Neduva, and T. J. Gibson.** 2003. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res **31:**3701-3708.
231. **Liu, W., Y. Qi, and F. M. Hulett.** 1998. Sites internal to the coding regions of *phoA* and *pstS* bind PhoP and are required for full promoter activity. Mol Microbiol **28:**119-130.
232. **Lobell, R. B., and R. F. Schleif.** 1990. DNA looping and unlooping by AraC protein. Science **250:**528-532.
233. **Lodge, J., J. Fear, S. Busby, P. Gunasekaran, and N. R. Kamini.** 1992. Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. FEMS Microbiol Lett **74:**271-276.
234. **Louise, C. B., and T. G. Obrig.** 1991. Shiga toxin-associated hemolytic-uremic syndrome: combined cytotoxic effects of Shiga toxin, interleukin-1 beta, and tumor necrosis factor alpha on human vascular endothelial cells *in vitro*. Infect Immun **59:**4173-4179.
235. **Low, A. S., F. Dziva, A. G. Torres, J. L. Martinez, T. Rosser, S. Naylor, K. Spears, N. Holden, A. Mahajan, J. Findlay, J. Sales, D. G. Smith, J. C. Low, M. P. Stevens, and D. L. Gally.** 2006. Cloning, expression, and characterization of fimbrial operon F9 from enterohemorrhagic *Escherichia coli* O157:H7. Infect Immun **74:**2233-2244.
236. **Low, A. S., N. Holden, T. Rosser, A. J. Roe, C. Constantinidou, J. L. Hobman, D. G. Smith, J. C. Low, and D. L. Gally.** 2006. Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. Environ Microbiol **8:**1033-1047.
237. **Lucchini, S., G. Rowley, M. D. Goldberg, D. Hurd, M. Harrison, and J. C. Hinton.** 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog **2:**e81.
238. **Lucht, J. M., P. Dersch, B. Kempf, and E. Bremer.** 1994. Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled *proU* operon of *Escherichia coli*. J Biol Chem **269:**6578-6586.
239. **Ludwig, A., C. von Rhein, S. Bauer, C. Hüttinger, and W. Goebel.** 2004. Molecular analysis of cytolysin A (ClyA) in pathogenic *Escherichia coli* strains. J Bacteriol **186:**5311-5320.
240. **Lupas, A.** 1996. Coiled coils: new structures and new functions. Trends Biochem Sci **21:**375-382.
241. **Lupas, A., M. Van Dyke, and J. Stock.** 1991. Predicting coiled coils from protein sequences. Science **252:**1162-1164.

242. **Mandel, M., and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J Mol Biol **53:**159-162.
243. **Manzanera, M., S. Marques, and J. L. Ramos.** 2000. Mutational analysis of the highly conserved C-terminal residues of the XylS protein, a member of the AraC family of transcriptional regulators. FEBS Lett **476:**312-317.
244. **Marron, M. B., and C. J. Smyth.** 1995. Molecular analysis of the *cso* operon of enterotoxigenic *Escherichia coli* reveals that CsoA is the adhesin of CS1 fimbriae and that the accessory genes are interchangeable with those of the cfa operon. Microbiology **141:**2849-2859.
245. **Martin, R. G., E. S. Bartlett, J. L. Rosner, and M. E. Wall.** 2008. Activation of the *Escherichia coli marA/soxS/rob* regulon in response to transcriptional activator concentration. J Mol Biol **380:**278-284.
246. **Martin, R. G., W. K. Gillette, S. Rhee, and J. L. Rosner.** 1999. Structural requirements for marbox function in transcriptional activation of *mar/sox/rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. Mol Microbiol **34:**431-441.
247. **Martin, R. G., W. K. Gillette, and J. L. Rosner.** 2000. Promoter discrimination by the related transcriptional activators MarA and SoxS: differential regulation by differential binding. Mol Microbiol **35:**623-634.
248. **Martin, R. G., and J. L. Rosner.** 2001. The AraC transcriptional activators. Curr Opin Microbiol **4:**132-137.
249. **Martínez-Antonio, A., and J. Collado-Vides.** 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol **6:**482-489.
250. **Martínez-Laguna, Y., E. Calva, and J. L. Puente.** 1999. Autoactivation and environmental regulation of *bfpT* expression, the gene coding for the transcriptional activator of *bfpA* in enteropathogenic *Escherichia coli*. Mol Microbiol **33:**153-166.
251. **McClain, M. S., I. C. Blomfield, and B. I. Eisenstein.** 1991. Roles of *fimB* and *fimE* in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. J Bacteriol **173:**5308-5314.
252. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc Natl Acad Sci USA **92:**1664-1668.
253. **Mekalanos, J. J.** 1992. Environmental signals controlling expression of virulence determinants in bacteria. J Bacteriol **174:**1-7.
254. **Mellies, J. L., S. J. Elliott, V. Sperandio, M. S. Donnenberg, and J. B. Kaper.** 1999. The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). Mol Microbiol **33:**296-306.
255. **Ménard, L. P., and J. D. Dubreuil.** 2002. Enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 (EAST1): a new toxin with an old twist. Crit Rev Microbiol **28:**43-60.
256. **Ménard, R., M. C. Prévost, P. Gounon, P. Sansonetti, and C. Dehio.** 1996. The secreted Ipa complex of *Shigella flexneri* promotes entry into mammalian cells. Proc Natl Acad Sci USA **93:**1254-1258.
257. **Mermod, N., J. L. Ramos, A. Bairoch, and K. N. Timmis.** 1987. The *xylS* gene positive regulator of TOL plasmid pWWO: identification, sequence

analysis and overproduction leading to constitutive expression of *meta* cleavage operon. Mol Gen Genet **207:**349-354.

258. **Miliotis, M. D., H. J. Koornhof, and J. I. Phillips.** 1989. Invasive potential of noncytotoxic enteropathogenic *Escherichia coli* in an *in vitro* Henle 407 cell model. Infect Immun **57:**1928-1935.
259. **Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.
260. **Mol, O., and B. Oudega.** 1996. Molecular and structural aspects of fimbriae biosynthesis and assembly in *Escherichia coli*. FEMS Microbiol Rev **19:**25-52.
261. **Monje-Casas, F., J. Jurado, M. J. Prieto-Alamo, A. Holmgren, and C. Pueyo.** 2001. Expression analysis of the *nrdHIEF* operon from *Escherichia coli*. Conditions that trigger the transcript level *in vivo*. J Biol Chem **276:**18031-18037.
262. **Moon, H. W., S. C. Whipp, R. A. Argenzio, M. M. Levine, and R. A. Giannella.** 1983. Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. Infect Immun **41:**1340-1351.
263. **Moreira, C. G., S. M. Carneiro, J. P. Nataro, L. R. Trabulsi, and W. P. Elias.** 2003. Role of type I fimbriae in the aggregative adhesion pattern of enteroaggregative *Escherichia coli*. FEMS Microbiol Lett **226:**79-85.
264. **Mounier, J., T. Vasselon, R. Hellio, M. Lesourd, and P. J. Sansonetti.** 1992. *Shigella flexneri* enters human colonic Caco-2 epithelial cells through the basolateral pole. Infect Immun **60:**237-248.
265. **Mullany, P., A. M. Field, M. M. McConnell, S. M. Scotland, H. R. Smith, and B. Rowe.** 1983. Expression of plasmids coding for colonization factor antigen II (CFA/II) and enterotoxin production in *Escherichia coli*. J Gen Microbiol **129:**3591-3601.
266. **Munson, G. P., L. G. Holcomb, H. L. Alexander, and J. R. Scott.** 2002. *In vitro* identification of Rns-regulated genes. J Bacteriol **184:**1196-1199.
267. **Munson, G. P., L. G. Holcomb, and J. R. Scott.** 2001. Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. Infect Immun **69:**186-193.
268. **Munson, G. P., and J. R. Scott.** 1999. Binding site recognition by Rns, a virulence regulator in the AraC family. J Bacteriol **181:**2110-2117.
269. **Munson, G. P., and J. R. Scott.** 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.
270. **Murphree, D., B. Froehlich, and J. R. Scott.** 1997. Transcriptional control of genes encoding CS1 pili: negative regulation by a silencer and positive regulation by Rns. J Bacteriol **179:**5736-5743.
271. **Narberhaus, F., T. Waldminghaus, and S. Chowdhury.** 2006. RNA thermometers. FEMS Microbiol Rev **30:**3-16.
272. **Nataro, J. P.** 2005. Enteroaggregative *Escherichia coli* pathogenesis. Curr Opin Gastroenterol **21:**4-8.
273. **Nataro, J. P., Y. Deng, D. R. Maneval, A. L. German, W. C. Martin, and M. M. Levine.** 1992. Aggregative adherence fimbriae I of enteroaggregative *Escherichia coli* mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. Infect Immun **60:**2297-2304.
274. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11:**142-201.

275. **Nataro, J. P., J. B. Kaper, R. Robins-Browne, V. Prado, P. Vial, and M. M. Levine.** 1987. Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells. Pediatr Infect Dis J **6:**829-831.

276. **Nataro, J. P., I. C. Scaletsky, J. B. Kaper, M. M. Levine, and L. R. Trabulsi.** 1985. Plasmid-mediated factors conferring diffuse and localized adherence of enteropathogenic *Escherichia coli*. Infect Immun **48:**378-383.

277. **Nataro, J. P., J. Seriwatana, A. Fasano, D. R. Maneval, L. D. Guers, F. Noriega, F. Dubovsky, M. M. Levine, and J. G. Morris, Jr.** 1995. Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains. Infect Immun **63:**4721-4728.

278. **Nataro, J. P., D. Yikang, D. Yingkang, and K. Walker.** 1994. AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. J Bacteriol **176:**4691-4699.

279. **Navarre, W. W., M. McClelland, S. J. Libby, and F. C. Fang.** 2007. Silencing of xenogeneic DNA by H-NS - facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. Genes Dev **21:**1456-1471.

280. **Niland, P., R. Huhne, and B. Muller-Hill.** 1996. How AraC interacts specifically with its target DNAs. J Mol Biol **264:**667-674.

281. **Nishi, J., J. Sheikh, K. Mizuguchi, B. Luisi, V. Burland, A. Boutin, D. J. Rose, F. R. Blattner, and J. P. Nataro.** 2003. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. J Biol Chem **278:**45680-45689.

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

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

284. **Nye, M. B., J. D. Pfau, K. Skorupski, and R. K. Taylor.** 2000. *Vibrio cholerae* H-NS silences virulence gene expression at multiple steps in the ToxR regulatory cascade. J Bacteriol **182:**4295-4303.

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

286. **O'Halloran, T. V., B. Frantz, M. K. Shin, D. M. Ralston, and J. G. Wright.** 1989. The MerR heavy metal receptor mediates positive activation in a topologically novel transcription complex. Cell **56:**119-129.

287. **Oelschlaeger, T. A., T. J. Barrett, and D. J. Kopecko.** 1994. Some structures and processes of human epithelial cells involved in uptake of enterohemorrhagic *Escherichia coli* 0157:H7. Infect Immun **62:**5142-5150.

288. **Ohyama, T.** 2001. Intrinsic DNA bends: an organizer of local chromatin structure for transcription. Bioessays **23:**708-715.

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

290. **Ono, S., M. D. Goldberg, T. Olsson, D. Esposito, J. C. D. Hinton, and J. E. Ladbury.** 2005. H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. Biochem J **391:**203-213.
291. **Oshima, T., S. Ishikawa, K. Kurokawa, H. Aiba, and N. Ogasawara.** 2006. *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. DNA Res **13:**141-153.
292. **Parkkinen, J., T. K. Korhonen, A. Pere, J. Hacker, and S. Soinila.** 1988. Binding sites in the rat brain for *Escherichia coli* S fimbriae associated with neonatal meningitis. J Clin Invest **81:**860-865.
293. **Parsot, C.** 2005. *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. FEMS Microbiol Lett **252:**11-18.
294. **Patel, S. K., J. Dotson, K. P. Allen, and J. M. Fleckenstein.** 2004. Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. Infect Immun **72:**1786-1794.
295. **Peiffer, I., M. F. Bernet-Camard, M. Rousset, and A. L. Servin.** 2001. Impairments in enzyme activity and biosynthesis of brush border-associated hydrolases in human intestinal Caco-2/TC7 cells infected by members of the Afa/Dr family of diffusely adhering *Escherichia coli*. Cell Microbiol **3:**341-357.
296. **Peiffer, I., A. L. Servin, and M. F. Bernet-Camard.** 1998. Piracy of decay-accelerating factor (CD55) signal transduction by the diffusely adhering strain *Escherichia coli* C1845 promotes cytoskeletal F-actin rearrangements in cultured human intestinal INT407 cells. Infect Immun **66:**4036-4042.
297. **Perez-Casal, J., J. S. Swartley, and J. R. Scott.** 1990. Gene encoding the major subunit of CS1 pili of human enterotoxigenic *Escherichia coli*. Infect Immun **58:**3594-3600.
298. **Pérez-Martín, J., and V. de Lorenzo.** 1997. Clues and consequences of DNA bending in transcription. Annu Rev Microbiol **51:**593-628.
299. **Pérez-Martín, J., F. Rojo, and V. de Lorenzo.** 1994. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol Rev **58:**268-290.
300. **Pérez-Rueda, E., and J. Collado-Vides.** 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res **28:**1838-1847.
301. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Pósfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529-533.
302. **Pichel, M., N. Binsztein, and G. Viboud.** 2000. CS22, a novel human enterotoxigenic *Escherichia coli* adhesin, is related to CS15. Infect Immun **68:**3280-3285.
303. **Pilonieta, M. C., M. D. Bodero, and G. P. Munson.** 2007. CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic *Escherichia coli*. J Bacteriol **189:**5060-5067.
304. **Plano, G. V.** 2004. Modulation of AraC family member activity by protein ligands. Mol Microbiol **54:**287-290.

305. **Pomposiello, P. J., M. H. Bennik, and B. Demple.** 2001. Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. J Bacteriol **183:**3890-3902.
306. **Poore, C. A., C. Coker, J. D. Dattlebaum, and H. L. T. Mobley.** 2001. Identification of the domains of UreR, an AraC-like transcriptional regulator of the urease gene cluster in *Proteus mirabilis*. J Bacteriol **183:**4526-4535.
307. **Poore, C. A., and H. L. T. Mobley.** 2003. Differential regulation of the *Proteus mirabilis* urease gene cluster by UreR and H-NS. Microbiology **149:**3383-3394.
308. **Porter, M. E., and C. J. Dorman.** 1997. Differential regulation of the plasmid-encoded genes in the *Shigella flexneri* virulence regulon. Mol Gen Genet **256:**93-103.
309. **Porter, M. E., and C. J. Dorman.** 2002. *In vivo* DNA-binding and oligomerization properties of the *Shigella flexneri* AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis. J Bacteriol **184:**531-539.
310. **Porter, M. E., and C. J. Dorman.** 1997. Positive regulation of *Shigella flexneri* virulence genes by integration host factor. J Bacteriol **179:**6537-6550.
311. **Porter, M. E., P. Mitchell, A. J. Roe, A. Free, D. G. E. Smith, and D. L. Gally.** 2004. Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. Mol Microbiol **54:**1117-1133.
312. **Porter, M. E., S. G. Smith, and C. J. Dorman.** 1998. Two highly related regulatory proteins, *Shigella flexneri* VirF and enterotoxigenic *Escherichia coli* Rns, have common and distinct regulatory properties. FEMS Microbiol Lett **162:**303-309.
313. **Prasadarao, N. V., C. A. Wass, and K. S. Kim.** 1996. Endothelial cell GlcNAc beta 1-4GlcNAc epitopes for outer membrane protein A enhance traversal of *Escherichia coli* across the blood-brain barrier. Infect Immun **64:**154-160.
314. **Prasadarao, N. V., C. A. Wass, J. N. Weiser, M. F. Stins, S. H. Huang, and K. S. Kim.** 1996. Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. Infect Immun **64:**146-153.
315. **Prosseda, G., M. Falconi, M. Giangrossi, C. O. Gualerzi, G. Micheli, and B. Colonna.** 2004. The *virF* promoter in *Shigella*: more than just a curved DNA stretch. Mol Microbiol **51:**523-537.
316. **Prosseda, G., P. A. Fradiani, M. Di Lorenzo, M. Falconi, G. Micheli, M. Casalino, M. Nicoletti, and B. Colonna.** 1998. A role for H-NS in the regulation of the*virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. Res Microbiol **149:**15-25.
317. **Prouty, M. G., C. R. Osorio, and K. E. Klose.** 2005. Characterisation of functional domains of the *Vibrio cholerae* virulence regulator ToxT. Mol Microbiol **58:**1143-1156.
318. **Puente, J. L., D. Bieber, S. W. Ramer, W. Murray, and G. K. Schoolnik.** 1996. The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals Mol Microbiol **20:**87-100.
319. **Qadri, F., A. M. Svennerholm, A. S. Faruque, and R. B. Sack.** 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev **18:**465-483.
320. **Ramos, J. L., A. Stolz, R. W., and K. N. Timmis.** 1986. Altered effector specificities in regulators of gene expression: TOL plasmid *xylS* mutants and

their use to engineer expansion of the range of aromatics degraded by bacteria. Proc Natl Acad Sci USA **83:**8467-8471.

321. **Rao, L., W. Ross, J. A. Appleman, T. Gaal, S. Leirmo, P. J. Schlax, M. T. Record Jr., and R. L. Gourse.** 1994. Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J Mol Biol **235:**1421-1435.
322. **Reece, K. S., and G. J. Phillips.** 1995. New plasmids carrying antibiotic-resistance cassettes. Gene **165:**141-142.
323. **Reed, W. L., and R. F. Schleif.** 1999. Hemiplegic mutations in AraC protein. J Mol Biol **294:**417-425.
324. **Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:**64-67.
325. **Remaut, H., C. Tang, N. S. Henderson, J. S. Pinkner, T. Wang, S. J. Hultgren, D. G. Thanassi, G. Waksman, and H. Li.** 2008. Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. Cell **133:**640-652.
326. **Rendón, M. A., Z. Saldaña, A. L. Erdem, V. Monteiro-Neto, A. Vázquez, J. B. Kaper, J. L. Puente, and J. A. Girón.** 2007. Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization. Proc Natl Acad Sci USA **104:**10637-10642.
327. **Rhee, S., R. G. Martin, J. L. Rosner, and D. R. Davies.** 1998. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. Proc Natl Acad Sci USA **95:**10413-10418.
328. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med **308:**681-685.
329. **Rimsky, S.** 2004. Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. Curr Opin Microbiol **7:**109-114.
330. **Rimsky, S., F. Zuber, M. Buckle, and H. Buc.** 2001. A molecular mechanism for the repression of transcription by the H-NS protein. Mol Microbiol **42:**1311-1323.
331. **Robbins, J. B., G. H. McCracken, Jr., E. C. Gotschlich, F. Orskov, I. Orskov, and L. A. Hanson.** 1974. *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. N Engl J Med **290:**1216-1220.
332. **Robinson, L. S., E. M. Ashman, S. J. Hultgren, and M. R. Chapman.** 2006. Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. Mol Microbiol **59:**870-881.
333. **Rodgers, M. E., N. D. Holder, S. Dirla, and R. Schleif.** 2008. Functional modes of the regulatory arm of AraC. Proteins:Epub ahead of print.
334. **Roesch, P. L., and I. C. Blomfield.** 1998. Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the *fim* switch to stimulate site-specific recombination in *Escherichia coli*. Mol Microbiol **27:**751-761.
335. **Rohde, J. R., J. M. Fox, and S. A. Minnich.** 1994. Thermoregulation in *Yersinia enterocolitica* is coincident with changes in DNA supercoiling. Mol Microbiol **12:**187-199.
336. **Rohde, J. R., X. S. Luan, H. Rohde, J. M. Fox, and S. A. Minnich.** 1999. The *Yersinia enterocolitica* pYV virulence plasmid contains multiple intrinsic DNA bends which melt at 37°C J Bacteriol **181:**4198-4204.

337. **Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse.** 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science **262:**1407-1413.
338. **Rost, B., G. Yachdav, and J. Liu.** 2004. The PredictProtein server. Nucleic Acids Res **32:**W321-W326.
339. **Ruiz, R., S. Marques, and J. L. Ramos.** 2003. Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL *meta*-cleavage pathway, are involved in dimerization. J Bacteriol **185:**3036-3041.
340. **Ruiz, R., J. L. Ramos, and S. M. Egan.** 2001. Interactions of the XylS regulators with the C-terminal domain of the RNA polymerase alpha subunit influence the expression level from the cognate Pm promoter. FEBS Lett **491:**207-211.
341. **Sagne, C., M. F. Isambert, J. P. Henry, and B. Gasnier.** 1996. SDS-resistant aggregation of membrane proteins: application to the purification of the vesicular monoamine transporter. Biochem J **316:**825-831.
342. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Muilis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487-491.
343. **Sakellaris, H., D. P. Balding, and J. R. Scott.** 1996. Assembly proteins of CS1 pili of enterotoxigenic *Escherichia coli*. Mol Microbiol **21:**529-541.
344. **Sakellaris, H., G. P. Munson, and J. R. Scott.** 1999. A conserved residue in the tip proteins of CS1 and CFA/I pili of enterotoxigenic *Escherichia coli* that is essential for adherence. Proc Natl Acad Sci USA **96:**12828-12832.
345. **Sakellaris, H., V. R. Penumalli, and J. R. Scott.** 1999. The level of expression of the minor pilin subunit, CooD, determines the number of CS1 pili assembled on the cell surface of *Escherichia coli*. J Bacteriol **181:**1694-1697.
346. **Sakellaris, H., and J. R. Scott.** 1998. New tools in an old trade: CS1 pilus morphogenesis. Mol Microbiol **30:**681-687.
347. **Sansonetti, P. J., and A. Phalipon.** 1999. M cells as ports of entry for enteroinvasive pathogens: mechanisms of interaction, consequences for the disease process. Semin Immunol **11:**193-203.
348. **Savarino, S. J., A. Fasano, J. Watson, B. M. Martin, M. M. Levine, S. Guandalini, and P. Guerry.** 1993. Enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 represents another subfamily of *E. coli* heat-stable toxin. Proc Natl Acad Sci USA **90:**3093-3097.
349. **Savelkoul, P. H., G. A. Willshaw, M. M. McConnell, H. R. Smith, A. M. Hamers, B. A. van der Zeijst, and W. Gaastra.** 1990. Expression of CFA/I fimbriae is positively regulated. Microb Pathog **8:**91-99.
350. **Saviola, B., R. Seabold, and R. F. Schleif.** 1998. Arm-domain interactions in AraC. J Mol Biol **278:**539-548.
351. **Scaletsky, I. C., S. H. Fabbricotti, R. L. Carvalho, C. R. Nunes, H. S. Maranhão, M. B. Morais, and U. Fagundes-Neto.** 2002. Diffusely adherent *Escherichia coli* as a cause of acute diarrhea in young children in Northeast Brazil: a case-control study. J Clin Microbiol **40:**645-648.
352. **Scaletsky, I. C., M. L. Silva, and L. R. Trabulsi.** 1984. Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. Infect Immun **45:**534-536.

353. **Schechter, L. M., S. M. Damrauer, and C. A. Lee.** 1999. Two AraC/XylS family members can independently counteract the effect of repressing sequences upstream of the *hilA* promoter. Mol Microbiol **32:**629-642.
354. **Schleif, R.** 2003. AraC protein: a love-hate relationship. Bioessays **25:**274-282.
355. **Schleif, R.** 2000. Regulation of the L-arabinose operon of *Escherichia coli*. Trends Genet **16:**559-565.
356. **Schmidt, H., H. Karch, and L. Beutin.** 1994. The large-sized plasmids of enterohemorrhagic *Escherichia coli* O157 strains encode hemolysins which are presumably members of the *E. coli* alpha-hemolysin family. FEMS Microbiol Lett **117:**189-196.
357. **Schmidt, S. A., D. Bieber, S. W. Ramer, J. Hwang, C. Y. Wu, and G. Schoolnik.** 2001. Structure-function analysis of BfpB, a secretin-like protein encoded by the bundle-forming-pilus operon of enteropathogenic *Escherichia coli*. J Bacteriol **183:**4848-4859.
358. **Schneiders, T., T. M. Barbosa, L. M. McMurry, and S. B. Levy.** 2004. The *Escherichia coli* transcriptional regulator MarA directly represses transcription of *purA* and *hdeA*. J Biol Chem **279:**9037-9042.
359. **Schnetz, K., and J. C. Wang.** 1996. Silencing of the *Escherichia coli bgl* promoter: effects of template supercoiling and cell extracts on promoter activity *in vitro*. Nucleic Acids Res **24:**2422-2428.
360. **Schuhmacher, D. A., and K. E. Klose.** 1999. Environmental signals modulate ToxT-dependent virulence factor expression in *Vibrio cholerae*. J Bacteriol **181:**1508-1514.
361. **Schulz, S., C. K. Green, P. S. Yuen, and D. L. Garbers.** 1990. Guanylyl cyclase is a heat-stable enterotoxin receptor. Cell **63:**941-948.
362. **Seabold, R. R., and R. F. Schleif.** 1998. Apo-AraC actively seeks to loop. J Mol Biol **278:**529-538.
363. **Sears, C. L., and J. B. Kaper.** 1996. Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion. Microbiol Rev **60:**167-215.
364. **Servant, P., C. Grandvalet, and P. Mazodier.** 2000. The RheA repressor is the thermosensor of the HSP18 heat shock response in *Streptomyces albus*. Proc Natl Acad Sci USA **97:**3538-3543.
365. **Servin, A. L.** 2005. Pathogenesis of Afa/Dr diffusely adhering *Escherichia coli*. Clin Microbiol Rev **18:**264-292.
366. **Shearwin, K. E., I. B. Dodd, and J. B. Egan.** 2002. The helix-turn-helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences. J Biol Chem **277:**3186-3194.
367. **Sheikh, J., J. R. Czeczulin, S. Harrington, S. Hicks, I. R. Henderson, C. Le Bouguénec, P. Gounon, A. Phillips, and J. P. Nataro.** 2002. A novel dispersin protein in enteroaggregative *Escherichia coli*. J Clin Invest **110:**1329-1337.
368. **Sherlock, O., M. A. Schembri, A. Reisner, and P. Klemm.** 2004. Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. J Bacteriol **186:**8058-8065.
369. **Sherlock, O., R. M. Vejborg, and P. Klemm.** 2005. The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. Infect Immun **73:**1954-1963.
370. **Silber, K. R., and R. T. Sauer.** 1994. Deletion of the *prc* (*tsp*) gene provides evidence for additional tail-specific proteolytic activity in *Escherichia coli* K-12. Mol Gen Genet **242:**237-240.

371. **Skerra, A.** 1994. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. Gene **151:**131-135.
372. **Smith, H. R., G. A. Willshaw, and B. Rowe.** 1982. Mapping of a plasmid, coding for colonization factor antigen I and heat-stable enterotoxin production, isolated from an enterotoxigenic strain of *Escherichia coli*. J Bacteriol **149:**264-275.
373. **Smith, S. G.** 1995. Regulation of CS1 fimbrial expression in enterotoxigenic *Escherichia coli*. PhD Thesis. Trinity College Dublin.
374. **Smyth, C. J.** 1982. Two mannose-resistant haemagglutinins on enterotoxigenic *Escherichia coli* of serotype O6:K15:H16 or H-isolated from travellers' and infantile diarrhoea. J Gen Microbiol **128:**2081-2096.
375. **Smyth, C. J., M. B. Marron, J. M. Twohig, and S. G. Smith.** 1996. Fimbrial adhesins: similarities and variations in structure and biogenesis. FEMS Immunol Med Microbiol **16:**127-139.
376. **Snyder, J. A., B. J. Haugen, C. V. Lockatell, N. Maroncle, E. C. Hagan, D. E. Johnson, R. A. Welch, and H. L. Mobley.** 2005. Coordinate expression of fimbriae in uropathogenic *Escherichia coli*. Infect Immun **73:**7588-7596.
377. **Sohel, I., J. L. Puente, S. W. Ramer, D. Bieber, C. Y. Wu, and G. K. Schoolnik.** 1996. Enteropathogenic *Escherichia coli*: identification of a gene cluster coding for bundle-forming pilus morphogenesis. J Bacteriol **178:**2613-2628.
378. **Soisson, S. M., B. MacDougall-Shackleton, R. Schleif, and C. Wolberger.** 1997. Structural basis for ligand-regulated oligomerization of AraC. Science **276:**421-425.
379. **Sondén, B., and B. E. Uhlin.** 1996. Coordinated and differential expression of histone-like proteins in *Escherichia coli*: regulation and function of the H-NS analog StpA. EMBO J **15:**4970-4980.
380. **Soto, G. E., and S. J. Hultgren.** 1999. Bacterial adhesins: common themes and variations in architecture and assembly. J Bacteriol **181:**1059-1071.
381. **Soulié, S., L. Denoroy, J. P. Le Caer, N. Hamasaki, J. D. Groves, and M. le Maire.** 1998. Treatment with crystalline ultra-pure urea reduces the aggregation of integral membrane proteins without inhibiting N-terminal sequencing. J Biochem **124:**417-420.
382. **Soulié, S., J. V. Moller, P. Falson, and M. le Maire.** 1996. Urea reduces the aggregation of membrane proteins on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Anal Biochem **236:**363-364.
383. **Spangler, B. D.** 1992. Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin. Microbiol Rev **56:**622-647.
384. **Spee, J. H., W. M. de Vos, and O. P. Kuipers.** 1993. Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. Nucleic Acids Res **21:**777-778.
385. **Spencer, J., H. Chart, H. R. Smith, and B. Rowe.** 1998. Expression of membrane-associated proteins by strains of enteroaggregative *Escherichia coli*. FEMS Microbiol Lett **161:**325-330.
386. **Spurio, R., M. Falconi, A. Brandi, C. L. Pon, and C. O. Gualerzi.** 1997. The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. EMBO J **16:**1795-1805.

387. **Spurlino, J. C., G.-Y. Lu, and F. A. Quiocho.** 1991. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem **266:**5202-5219.
388. **Stamm, W. E., and S. R. Norrby.** 2001. Urinary tract infections: disease panorama and challenges. J Infect Dis **183:**S1-S4.
389. **Starks, A. M., B. J. Froehlich, T. N. Jones, and J. R. Scott.** 2006. Assembly of CS1 pili: the role of specific residues of the major pilin, CooA. J Bacteriol **188:**231-239.
390. **Stella, S., M. Falconi, M. Lammi, C. O. Gualerzi, and C. L. Pon.** 2006. Environmental control of the *in vivo* oligomerization of nucleoid protein H-NS. J Mol Biol **355:**169-174.
391. **Stellwagen, A. E., and N. L. Craig.** 1997. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. Genetics **145:**573-585.
392. **Stins, M. F., N. V. Prasadarao, L. Ibric, C. A. Wass, P. Luckett, and K. S. Kim.** 1994. Binding characteristics of S fimbriated *Escherichia coli* to isolated brain microvascular endothelial cells. Am J Pathol **145:**1228-1236.
393. **Stone, K. D., H. Z. Zhang, L. K. Carlson, and M. S. Donnenberg.** 1996. A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for the biogenesis of a type IV pilus. Mol Microbiol **20:**325-337.
394. **Studier, F. W.** 2005. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif **41:**207-234.
395. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol **185:**60-89.
396. **Taniguchi, T., Y. Akeda, A. Haba, Y. Yasuda, K. Yamamoto, T. Honda, and K. Tochikubo.** 2001. Gene cluster for assembly of pilus colonization factor antigen III of enterotoxigenic *Escherichia coli*. Infect Immun **69:**5864-5873.
397. **Tatsuno, I., M. Horie, H. Abe, T. Miki, K. Makino, H. Shinagawa, H. Taguchi, S. Kamiya, T. Hayashi, and C. Sasakawa.** 2001. *toxB* gene on pO157 of enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. Infect Immun **69:**6660-6669.
398. **Teng, C. H., M. Cai, S. Shin, Y. Xie, K. J. Kim, N. A. Khan, F. Di Cello, and K. S. Kim.** 2005. *Escherichia coli* K1 RS218 interacts with human brain microvascular endothelial cells via type 1 fimbria bacteria in the fimbriated state. Infect Immun **73:**2923-2931.
399. **Thomas, L. V., M. M. McConnell, B. Rowe, and A. M. Field.** 1985. The possession of three novel coli surface antigens by enterotoxigenic *Escherichia coli* strains positive for the putative colonization factor PCF8775. J Gen Microbiol **131:**2319-2326.
400. **Thomas, V. J., and C. M. Collins.** 1999. Identification of UreR binding sites in the *Enterobacteriaceae* plasmid-encoded and *Proteus mirabilis* urease gene operons. Mol Microbiol **31:**1417-1428.
401. **Tobe, T., G. K. Schoolnik, I. Sohel, V. H. Bustamante, and J. L. Puente.** 1996. Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. Mol Microbiol **21:**963-975.
402. **Tobe, T., M. Yoshikawa, T. Mizuno, and C. Sasakawa.** 1993. Transcriptional control of the invasion regulatory gene *virB* of *Shigella flexneri*: activation by VirF and repression by H-NS. J Bacteriol **175:**6142-6149.

403. **Tobe, T., M. Yoshikawa, and C. Sasakawa.** 1995. Thermoregulation of *virB* transcription in *Shigella flexneri* by sensing of changes in local DNA superhelicity. J Bacteriol **177:**1094-1097.
404. **Tobin, J. F., and R. F. Schleif.** 1990. Transcription from the *rha* operon p_{sr} promoter. J Mol Biol **211:**1-4.
405. **Toma, C., N. Higa, S. Iyoda, M. Rivas, and M. Iwanaga.** 2006. The long polar fimbriae genes identified in Shiga toxin-producing *Escherichia coli* are present in other diarrheagenic *E. coli* and in the standard *E. coli* collection of reference (ECOR) strains. Res Microbiol **157**.
406. **Torres, A. G., J. A. Giron, N. T. Perna, V. Burland, F. R. Blattner, F. Avelino-Flores, and J. B. Kaper.** 2002. Identification and characterization of *lpfABCC'DE*, a fimbrial operon of enterohemorrhagic *Escherichia coli* O157:H7. Infect Immun **70:**5416-5427.
407. **Tozzi, A. E., A. Caprioli, F. Minelli, A. Gianviti, L. De Petris, A. Edefonti, G. Montini, A. Ferretti, T. De Palo, M. Gaido, and G. Rizzoni.** 2003. Shiga toxin-producing *Escherichia coli* infections associated with hemolytic uremic syndrome, Italy, 1988-2000. Emerg Infect Dis **9:**106-108.
408. **Tupper, A. E., T. A. Owen-Hughes, D. W. Ussery, D. S. Santos, D. J. P. Ferguson, J. M. Sidebotham, J. C. D. Hinton, and C. F. Higgins.** 1994. The chromatin-associated protein H-NS alters DNA topology *in vitro*. EMBO J **13:**258-268.
409. **Turner, S. M., A. Scott-Tucker, L. M. Cooper, and I. R. Henderson.** 2006. Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*. FEMS Microbiol Lett **263:**10-20.
410. **Usein, C. R., M. Damian, D. Tatu-Chitoiu, C. Capusa, R. Fagaras, D. Tudorache, M. Nica, and C. Le Bouguénec.** 2001. Prevalence of virulence genes in *Escherichia coli* strains isolated from Romanian adult urinary tract infection cases. J Cell Mol Med **5:**303-310.
411. **van der Woude, M., B. Braaten, and D. Low.** 1996. Epigenetic phase variation of the *pap* operon in *Escherichia coli*. Trends Microbiol **4:**5-9.
412. **van der Woude, M. W.** 2006. Re-examining the role and random nature of phase variation. FEMS Microbiol Lett **254:**190-197.
413. **van der Woude, M. W., and A. J. Bäumler.** 2004. Phase and antigenic variation in bacteria. Clin Microbiol Rev **17:**581-611.
414. **van der Woude, M. W., and D. A. Low.** 1994. Leucine-responsive regulatory protein and deoxyadenosine methylase control the phase variation and expression of the *sfa* and *daa* pili operons in *Escherichia coli*. Mol Microbiol **11:**605-618.
415. **Vía, P., J. Badía, L. Baldomà, N. Obradors, and J. Aguilar.** 1996. Transcriptional regulation of the *Escherichia coli rhaT* gene. Microbiology **142:**1833-1840.
416. **Vicente, M., K. F. Chater, and V. de Lorenzo.** 1999. Bacterial transcription factors involved in global regulation. Mol Microbiol **33:**8-17.
417. **Viswanathan, P., T. Ueki, S. Inouye, and L. Kroos.** 2007. Combinatorial regulation of genes essential for *Myxococcus xanthus* development involves a response regulator and a LysR-type regulator. Proc Natl Acad Sci USA **104:**7969-7974.
418. **Voegele, K., H. Sakellaris, and J. R. Scott.** 1997. CooB plays a chaperone-like role for the proteins involved in formation of CS1 pili of enterotoxigenic *Escherichia coli*. Proc Natl Acad Sci USA **94:**13257-13261.

419. **Vullo, A., O. Bortolami, G. Pollastri, and S. C. E. Tosatto.** 2006. Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. Nucleic Acids Res **34:**W164-W168.

420. **Wadood, A., M. Dohmoto, S. Sugiura, and K. Yamaguchi.** 1997. Characterization of copy number mutants of plasmid pSC101. J Gen Appl Microbiol **43:**309-316.

421. **Wang, Y., Z. G. Wen, and K. S. Kim.** 2004. Role of S fimbriae in *Escherichia coli* K1 binding to brain microvascular endothelial cells *in vitro* and penetration into the central nervous system *in vivo*. Microb Pathog **37:**287-293.

422. **Welch, R. A., V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci USA **99:**17020-17024.

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

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

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

426. **White-Ziegler, C. A., A. M. Black, S. H. Eliades, S. Young, and K. Porter.** 2002. The N-acetyltransferase RimJ responds to environmental stimuli to repress *pap* fimbrial transcription in *Escherichia coli*. J Bacteriol **184:**4334-4342.

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

428. **Wickstrum, J. R., and S. M. Egan.** 2004. Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR. J Bacteriol **186:**6277-6285.

429. **Wickstrum, J. R., J. M. Skredenske, A. Kolin, D. J. Jin, J. Fang, and S. M. Egan.** 2007. Transcription activation by the DNA-binding domain of the AraC family protein RhaS in the absence of its effector-binding domain. J Bacteriol **189:**4984-4993.

430. **Wilcox, G., and P. Meuris.** 1976. Stabilization and size of AraC protein. Mol Gen Genet **145:**97-100.

431. **Williams, R. M., S. Rimsky, and H. Buc.** 1996. Probing the structure, function, and interactions of the *Escherichia coli* H-NS and StpA proteins by using dominant negative derivatives. J Bacteriol **178:**4335-4343.

432. **Withey, J. H., and V. J. DiRita.** 2006. The toxbox: specific DNA sequence requirements for activation of *Vibrio cholerae* virulence genes by ToxT. Mol Microbiol **59:**1779-1789.

433. **Wolf, M. K.** 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.
434. **Wright, P. E., and H. J. Dyson.** 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol **293:**321-331.
435. **Wu, M., and R. Schleif.** 2001. Mapping arm-DNA-binding domain interactions in AraC. J Mol Biol **307:**1001-1009.
436. **Xia, G., D. Manen, Y. Yu, and L. Caro.** 1993. *In vivo* and *in vitro* studies of a copy number mutation of the RepA replication protein of plasmid pSC101. J Bacteriol **175:**4165-4175.
437. **Xia, Y., D. Gally, K. Forsman-Semb, and B. E. Uhlin.** 2000. Regulatory cross-talk between adhesin operons in *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein. EMBO J **19:**1450-1457.
438. **Xicohtencatl-Cortes, J., V. Monteiro-Neto, M. A. Ledesma, D. M. Jordan, O. Francetic, J. B. Kaper, J. L. Puente, and J. A. Girón.** 2007. Intestinal adherence associated with type IV pili of enterohemorrhagic *Escherichia coli* O157:H7. J Clin Invest **117:**3519-3529.
439. **Yamada, H., S. Muramatsu, and T. Mizuno.** 1990. An *Escherichia coli* protein that preferentially binds to sharply curved DNA. J Biochem **108:**420-425.
440. **Yamamoto, S.** 2007. Molecular epidemiology of uropathogenic *Escherichia coli*. J Infect Chemother **13:**68-73.
441. **Yamamoto, T., and P. Echeverria.** 1996. Detection of the enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 gene sequences in enterotoxigenic *E. coli* strains pathogenic for humans. Infect Immun **64:**1441-1445.
442. **Yang, J., D. L. Baldi, M. Tauschek, R. A. Strugnell, and R. M. Robins-Browne.** 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.
443. **Yu, R. R., and V. J. DiRita.** 1999. Analysis of an autoregulatory loop controlling ToxT, cholera toxin, and toxin-coregulated pilus production in *Vibrio cholerae*. J Bacteriol **181:**2584-2592.
444. **Yu, R. R., and V. J. DiRita.** 2002. Regulation of gene expression in *Vibrio cholerae* by ToxT involves both antirepression and RNA polymerase stimulation. Mol Microbiol **43:**119-134.
445. **Zhang, A., and M. Belfort.** 1992. Nucleotide sequence of a newly identified *Escherichia coli* gene, *stpA,* encoding an H-NS-like protein. Nucleic Acids Res **20:**6735.
446. **Zhang, H. Z., S. Lory, and M. S. Donnenberg.** 1994. A plasmid-encoded prepilin peptidase gene from enteropathogenic *Escherichia coli*. J Bacteriol **176:**6885-6891.
447. **Zhang, X., T. Reeder, and R. Schleif.** 1996. Transcription activation parameters at *ara* p_{BAD}. J Mol Biol **258:**14-24.
448. **Zhang, X., and R. Schleif.** 1998. Catabolite gene activator protein mutations affecting activity of the *araBAD* promoter. J Bacteriol **180:**195-200.
449. **Zychlinsky, A., M. C. Prevost, and P. J. Sansonetti.** 1992. *Shigella flexneri* induces apoptosis in infected macrophages. Nature **358:**167-169.